Linear Algebra and Differential Equations

MATH 54

Second Custom Edition for University of California, Berkeley

From Linear Algebra and Its Applications , 6/e, David C. Lay / Steven R. Lay / Judi J. McDonald. Copyright © 2021 by Pearson Education.

From Linear Algebra and Its Applications , 5/e, David C. Lay / Steven R. Lay / Judi J. McDonald. Copyright © 2016 by Pearson Education.

From Fundamentals of Differential Equations , 9/e, R. Kent Nagle / Edward B. Saff / Arthur David Snider. Copyright © 2018 by Pearson Education.

| PEARSON COLLECTIONS |

PEARSON

Cover Art: Courtesy of Photodisc/Getty Images

Pearson Learning Solutions, 501 Boylston Street, Suite 900, Boston, MA 02116

A Pearson Education Company
www.pearsoned.com

ISBN 10: 0137114036

ISBN 13: 9780137114030

Printed in the USA

Table of Contents

Preface

The response of students and teachers to the first five editions of *Linear Algebra and Its Applications* has been most gratifying. This *Sixth Edition* provides substantial support both for teaching and for using technology in the course. As before, the text provides a modern elementary introduction to linear algebra and a broad selection of interesting classical and leading-edge applications. The material is accessible to students with the maturity that should come from successful completion of two semesters of college-level mathematics, usually calculus.

The main goal of the text is to help students master the basic concepts and skills they will use later in their careers. The topics here follow the recommendations of the original Linear Algebra Curriculum Study Group (LACSG), which were based on a careful investigation of the real needs of the students and a consensus among professionals in many disciplines that use linear algebra. Ideas being discussed by the second Linear Algebra Curriculum Study Group (LACSG 2.0) have also been included. We hope this course will be one of the most useful and interesting mathematics classes taken by undergraduates.

What's New in This Edition

The *Sixth Edition* has exciting new material, examples, and online resources. After talking with high-tech industry researchers and colleagues in applied areas, we added new topics, vignettes, and applications with the intention of highlighting for students and faculty the linear algebraic foundational material for machine learning, artificial intelligence, data science, and digital signal processing.

Content Changes

- Since matrix multiplication is a highly useful skill, we added new examples in Chapter 2 to show how matrix multiplication is used to identify patterns and scrub data. Corresponding exercises have been created to allow students to explore using matrix multiplication in various ways.
- In our conversations with colleagues in industry and electrical engineering, we heard repeatedly how important understanding abstract vector spaces is to their work. After reading the reviewers' comments for Chapter 4, we reorganized the chapter, condensing some of the material on column, row, and null spaces; moving Markov chains to the end of Chapter 5; and creating a new section on signal processing. We view signals

as an infinite dimensional vector space and illustrate the usefulness of linear transformations to filter out unwanted "vectors" (a.k.a. noise), analyze data, and enhance signals.

- By moving Markov chains to the end of Chapter 5, we can now discuss the steady state vector as an eigenvector. We also reorganized some of the summary material on determinants and change of basis to be more specific to the way they are used in this chapter.
- In Chapter 6, we present pattern recognition as an application of orthogonality, and the section on linear models now illustrates how machine learning relates to curve fitting.
- Chapter 9 on optimization was previously available only as an online file. It has now been moved into the regular textbook where it is more readily available to faculty and students. After an opening section on finding optimal strategies to two-person zero-sum games, the rest of the chapter presents an introduction to linear programming—from two-dimensional problems that can be solved geometrically to higher dimensional problems that are solved using the Simplex Method.

Other Changes

- In the high-tech industry, where most computations are done on computers, judging the validity of information and computations is an important step in preparing and analyzing data. In this edition, students are encouraged to learn to analyze their own computations to see if they are consistent with the data at hand and the questions being asked. For this reason, we have added "Reasonable Answers" advice and exercises to guide students.
- We have added a list of projects to the end of each chapter (available online at `bit.ly/30IM8gT` and in MyLab Math). Some of these projects were previously available online and have a wide range of themes from using linear transformations to create art to exploring additional ideas in mathematics. They can be used for group work or to enhance the learning of individual students.
- Free-response writing exercises have been added to MyLab Math, allowing faculty to ask more sophisticated questions online and create a paperless class without losing the richness of discussing how concepts relate to each other and introductory proof writing.
- The electronic interactive textbook has been changed from Wolfram CDF to Wolfram Cloud format. This allows faculty and students to interact with figures and examples on a wider variety of electronic, devices, without the need to install the CDF Player.
- PowerPoint lecture slides have been updated to cover all sections of the text and cover them more thoroughly.

Distinctive Features

Early Introduction of Key Concepts

Many fundamental ideas of linear algebra are introduced within the first seven lectures, in the concrete setting of $\mathbb{R}^n$, and then gradually examined from different points of view. Later generalizations of these concepts appear as natural extensions of familiar ideas, visualized through the geometric intuition developed in Chapter 1. A major achievement of this text is that the level of difficulty is fairly even throughout the course.

A Modern View of Matrix Multiplication

Good notation is crucial, and the text reflects the way scientists and engineers actually use linear algebra in practice. The definitions and proofs focus on the columns of a matrix rather than on the matrix entries. A central theme is to view a matrix–vector product $A\mathbf{x}$ as a linear combination of the columns of A. This modern approach simplifies many arguments, and it ties vector space ideas into the study of linear systems.

Linear Transformations

Linear transformations form a "thread" that is woven into the fabric of the text. Their use enhances the geometric flavor of the text. In Chapter 1, for instance, linear transformations provide a dynamic and graphical view of matrix–vector multiplication.

Eigenvalues and Dynamical Systems

Eigenvalues appear fairly early in the text, in Chapters 5 and 7. Because this material is spread over several weeks, students have more time than usual to absorb and review these critical concepts. Eigenvalues are motivated by and applied to discrete and continuous dynamical systems, which appear in Sections 1.10, 4.8, and 5.9, and in five sections of Chapter 5. Some courses reach Chapter 5 after about five weeks by covering Sections 2.8 and 2.9 instead of Chapter 4. These two optional sections present all the vector space concepts from Chapter 4 needed for Chapter 5.

Orthogonality and Least-Squares Problems

These topics receive a more comprehensive treatment than is commonly found in beginning texts. The original Linear Algebra Curriculum Study Group has emphasized the need for a substantial unit on orthogonality and least-squares problems, because orthogonality plays such an important role in computer calculations and numerical linear algebra and because inconsistent linear systems arise so often in practical work.

Pedagogical Features

Applications

A broad selection of applications illustrates the power of linear algebra to explain fundamental principles and simplify calculations in engineering, computer science, mathematics, physics, biology, economics, and statistics. Some applications appear in separate sections; others are treated in examples and exercises. In addition, each chapter opens with an introductory vignette that sets the stage for some application of linear algebra and provides a motivation for developing the mathematics that follows.

A Strong Geometric Emphasis

Every major concept in the course is given a geometric interpretation, because many students learn better when they can visualize an idea. There are substantially more drawings here than usual, and some of the figures have never before appeared in a linear algebra text. Interactive versions of these figures, and more, appear in the electronic version of the textbook.

Examples

This text devotes a larger proportion of its expository material to examples than do most linear algebra texts. There are more examples than an instructor would ordinarily present in class. But because the examples are written carefully, with lots of detail, students can read them on their own.

Theorems and Proofs

Important results are stated as theorems. Other useful facts are displayed in tinted boxes, for easy reference. Most of the theorems have formal proofs, written with the beginner student in mind. In a few cases, the essential calculations of a proof are exhibited in a carefully chosen example. Some routine verifications are saved for exercises, when they will benefit students.

Practice Problems

A few carefully selected Practice Problems appear just before each exercise set. Complete solutions follow the exercise set. These problems either focus on potential trouble spots in the exercise set or provide a "warm-up" for the exercises, and the solutions often contain helpful hints or warnings about the homework.

Exercises

The abundant supply of exercises ranges from routine computations to conceptual questions that require more thought. A good number of innovative questions pinpoint conceptual difficulties that we have found on student papers over the years. Each exercise set is carefully arranged in the same general order as the text; homework assignments are readily available when only part of a section is discussed. A notable feature of the exercises is their numerical simplicity. Problems "unfold" quickly, so students spend little time on numerical calculations. The exercises concentrate on teaching understanding rather than mechanical calculations. The exercises in the *Sixth Edition* maintain the integrity of the exercises from previous editions, while providing fresh problems for students and instructors.

Exercises marked with the symbol **T** are designed to be worked with the aid of a "matrix program" (a computer program, such as MATLAB, Maple, Mathematica, MathCad, or Derive, or a programmable calculator with matrix capabilities, such as those manufactured by Texas Instruments).

True/False Questions

To encourage students to read all of the text and to think critically, we have developed over 300 simple true/false questions that appear throughout the text, just after the computational problems. They can be answered directly from the text, and they prepare students for the conceptual problems that follow. Students appreciate these questions-after they get used to the importance of reading the text carefully. Based on class testing and discussions with students, we decided not to put the answers in the text. (The *Study Guide* tells the students where to find the answers to the odd-numbered questions.) An additional 150 true/false questions (mostly at the ends of chapters) test understanding of the material. The text does provide simple T/F answers to most of these supplementary exercises, but it omits the justifications for the answers (which usually require some thought).

Writing Exercises

An ability to write coherent mathematical statements in English is essential for all students of linear algebra, not just those who may go to graduate school in mathematics. The text includes many exercises for which a written justification is part of the answer. Conceptual exercises that require a short proof usually contain hints that help a student get started. For all odd-numbered writing exercises, either a solution is included at the back of the text or a hint is provided and the solution is given in the *Study Guide,* described below.

Projects

A list of projects (available online at `bit.ly/30IM8gT`) have been identified at the end of each chapter. They can be used by individual students or in groups. These projects provide the opportunity for students to explore fundamental concepts and applications in more detail. Two of the projects even encourage students to engage their creative side and use linear transformations to build artwork.

Reasonable Answers

Many of our students will enter a workforce where important decisions are being made based on answers provided by computers and other machines. The Reasonable Answers boxes and exercises help students develop an awareness of the need to analyze their answers for correctness and accuracy.

Computational Topics

The text stresses the impact of the computer on both the development and practice of linear algebra in science and engineering. Frequent Numerical Notes draw attention to issues in computing and distinguish between theoretical concepts, such as matrix inversion, and computer implementations, such as LU factorizations.

Acknowledgments

David Lay was grateful to many people who helped him over the years with various aspects of this book. He was particularly grateful to Israel Gohberg and Robert Ellis for more than fifteen years of research collaboration, which greatly shaped his view of linear algebra. And he was privileged to be a member of the Linear Algebra Curriculum Study Group along with David Carlson, Charles Johnson, and Duane Porter. Their creative ideas about teaching linear algebra have influenced this text in significant ways. He often spoke fondly of three good friends who guided the development of the book nearly from the beginning—giving wise counsel and encouragement—Greg Tobin, publisher; Laurie Rosatone, former editor; and William Hoffman, former editor.

Judi and Steven have been privileged to work on recent editions of Professor David Lay's linear algebra book. In making this revision, we have attempted to maintain the basic approach and the clarity of style that has made earlier editions popular with students and faculty. We thank Eric Schulz for sharing his considerable technological and pedagogical expertise in the creation of the electronic textbook. His help and encouragement were essential in bringing the figures and examples to life in the Wolfram Cloud version of this textbook.

Mathew Hudelson has been a valuable colleague in preparing the *Sixth Edition;* he is always willing to brainstorm about concepts or ideas and test out new writing and exercises. He contributed the idea for new vignette for Chapter 3 and the accompanying project. He has helped with new exercises throughout the text. Harley Weston has provided Judi with many years of good conversations about how, why, and who we appeal to when we present mathematical material in different ways. Katerina Tsatsomeros' artistic side has been a definite asset when we needed artwork to transform (the fish and the sheep), improved writing in the new introductory vignettes, or information from the perspective of college-age students.

We appreciate the encouragement and shared expertise from Nella Ludlow, Thomas Fischer, Amy Johnston, Cassandra Seubert, and Mike Manzano. They provided information about important applications of linear algebra and ideas for new examples and exercises. In particular, the new vignettes and material in Chapters 4 and 6 were inspired by conversations with these individuals.

We are energized by Sepideh Stewart and the other new Linear Algebra Curriculum Study Group (LACSG 2.0) members: Sheldon Axler, Rob Beezer, Eugene Boman, Minerva Catral, Guershon Harel, David Strong, and Megan Wawro. Initial meetings of this group have provided valuable guidance in revising the *Sixth Edition.*

We sincerely thank the following reviewers for their careful analyses and constructive suggestions:

Maila C. Brucal-Hallare, *Norfolk State University*
Kristen Campbell, *Elgin Community College*
Charles Conrad, *Volunteer State Community College*
R. Darrell Finney, *Wilkes Community College*
Xiaofeng Gu, *University of West Georgia*
Jeong Mi-Yoon, *University of Houston–Downtown*
Michael T. Muzheve, *Texas A&M U.–Kingsville*
Iason Rusodimos, *Perimeter C. at Georgia State U.*
Rebecca Swanson, *Colorado School of Mines*
Casey Wynn, *Kenyon College*
Taoye Zhang, *Penn State U.–Worthington Scranton*
Steven Burrow, *Central Texas College*
J. S. Chahal, *Brigham Young University*
Kevin Farrell, *Lyndon State College*
Chris Fuller, *Cumberland University*
Jeffrey Jauregui, *Union College*
Christopher Murphy, *Guilford Tech. C.C.*
Charles I. Odion, *Houston Community College*
Desmond Stephens, *Florida Ag. and Mech. U.*
Jiyuan Tao, *Loyola University–Maryland*
Amy Yielding, *Eastern Oregon University*
Houlong Zhuang, *Arizona State University*

We appreciate the proofreading and suggestions provided by John Samons and Jennifer Blue. Their careful eye has helped to minimize errors in this edition.

We thank Kristina Evans, Phil Oslin, and Jean Choe for their work in setting up and maintaining the online homework to accompany the text in MyLab Math, and for continuing to work with us to improve it. The reviews of the online homework done by Joan Saniuk, Robert Pierce, Doron Lubinsky and Adriana Corinaldesi were greatly appreciated. We also thank the faculty at University of California Santa Barbara, University of Alberta, Washington State University and the Georgia Institute of Technology for their feedback on the MyLab Math course. Joe Vetere has provided much appreciated technical help with the *Study Guide* and *Instructor's Solutions Manual*.

We thank Jeff Weidenaar, our content manager, for his continued careful, well-thought-out advice. Project Manager Ron Hampton has been a tremendous help guiding us through the production process. We are also grateful to Stacey Sveum and Rosemary Morton, our marketers, and Jon Krebs, our editorial associate, who have also contributed to the success of this edition.

Steven R. Lay and Judi J. McDonald

1 Linear Equations in Linear Algebra

Introductory Example

LINEAR MODELS IN ECONOMICS AND ENGINEERING

It was late summer in 1949. Harvard Professor Wassily Leontief was carefully feeding the last of his punched cards into the university's Mark II computer. The cards contained information about the U.S. economy and represented a summary of more than 250,000 pieces of information produced by the U.S. Bureau of Labor Statistics after two years of intensive work. Leontief had divided the U.S. economy into 500 "sectors," such as the coal industry, the automotive industry, communications, and so on. For each sector, he had written a linear equation that described how the sector distributed its output to the other sectors of the economy. Because the Mark II, one of the largest computers of its day, could not handle the resulting system of 500 equations in 500 unknowns, Leontief had distilled the problem into a system of 42 equations in 42 unknowns.

Programming the Mark II computer for Leontief's 42 equations had required several months of effort, and he was anxious to see how long the computer would take to solve the problem. The Mark II hummed and blinked for 56 hours before finally producing a solution. We will discuss the nature of this solution in Sections 1.6 and 2.6.

Leontief, who was awarded the 1973 Nobel Prize in Economic Science, opened the door to a new era in mathematical modeling in economics. His efforts at Harvard in 1949 marked one of the first significant uses of computers to analyze what was then a large-scale mathematical model. Since that time, researchers in many other fields have employed computers to analyze mathematical models. Because of the massive amounts of data involved, the models are usually *linear*; that is, they are described by *systems of linear equations*.

The importance of linear algebra for applications has risen in direct proportion to the increase in computing power, with each new generation of hardware and software triggering a demand for even greater capabilities. Computer science is thus intricately linked with linear algebra through the explosive growth of parallel processing and large-scale computations.

Scientists and engineers now work on problems far more complex than even dreamed possible a few decades ago. Today, linear algebra has more potential value for students in many scientific and business fields than any other undergraduate mathematics subject! The material in this text provides the foundation for further work in many interesting areas. Here are a few possibilities; others will be described later.

- *Oil exploration.* When a ship searches for offshore oil deposits, its computers solve thousands of separate systems of linear equations *every day*. The seismic data for the equations are obtained from underwater shock waves created by explosions from air guns. The waves bounce off subsurface rocks and are measured by geophones attached to mile-long cables behind the ship.

- *Linear programming.* Many important management decisions today are made on the basis of linear programming models that use hundreds of variables. The airline industry, for instance, employs linear programs that schedule flight crews, monitor the locations of aircraft, or plan the varied schedules of support services such as maintenance and terminal operations.
- *Electrical networks.* Engineers use simulation software to design electrical circuits and microchips involving millions of transistors. Such software relies on linear algebra techniques and systems of linear equations.
- *Artificial intelligence.* Linear algebra plays a key role in everything from scrubbing data to facial recognition.
- *Signals and signal processing.* From a digital photograph to the daily price of a stock, important information is recorded as a signal and processed using linear transformations.
- *Machine learning.* Machines (specifically computers) use linear algebra to learn about anything from online shopping preferences to speech recognition.

Systems of linear equations lie at the heart of linear algebra, and this chapter uses them to introduce some of the central concepts of linear algebra in a simple and concrete setting. Sections 1.1 and 1.2 present a systematic method for solving systems of linear equations. This algorithm will be used for computations throughout the text. Sections 1.3 and 1.4 show how a system of linear equations is equivalent to a *vector equation* and to a *matrix equation*. This equivalence will reduce problems involving linear combinations of vectors to questions about systems of linear equations. The fundamental concepts of spanning, linear independence, and linear transformations, studied in the second half of the chapter, will play an essential role throughout the text as we explore the beauty and power of linear algebra.

1.1 Systems of Linear Equations

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

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

where b and the **coefficients** $a_1, \dots, a_n$ are real or complex numbers, usually known in advance. The subscript n may be any positive integer. In textbook examples and exercises, n is normally between 2 and 5. In real-life problems, n might be 50 or 5000, or even larger.

The equations

$$4x_1 - 5x_2 + 2 = x_1 \quad \text{and} \quad x_2 = 2\left(\sqrt{6} - x_1\right) + x_3$$

are both linear because they can be rearranged algebraically as in equation (1):

$$3x_1 - 5x_2 = -2 \quad \text{and} \quad 2x_1 + x_2 - x_3 = 2\sqrt{6}$$

The equations

$$4x_1 - 5x_2 = x_1x_2 \quad \text{and} \quad x_2 = 2\sqrt{x_1} - 6$$

are not linear because of the presence of x_1x_2 in the first equation and $\sqrt{x_1}$ in the second.

A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables—say, $x_1, \dots, x_n$. An example is

$$\begin{aligned} 2x_1 - x_2 + 1.5x_3 &= 8 \\ x_1 - 4x_3 &= -7 \end{aligned} \tag{2}$$

A **solution** of the system is a list $(s_1, s_2, \dots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \dots, s_n$ are substituted for $x_1, \dots, x_n$, respectively. For instance, $(5, 6.5, 3)$ is a solution of system (2) because, when these values are substituted in (2) for x_1, x_2, x_3, respectively, the equations simplify to $8 = 8$ and $-7 = -7$.

The set of all possible solutions is called the **solution set** of the linear system. Two linear systems are called **equivalent** if they have the same solution set. That is, each solution of the first system is a solution of the second system, and each solution of the second system is a solution of the first.

Finding the solution set of a system of two linear equations in two variables is easy because it amounts to finding the intersection of two lines. A typical problem is

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

The graphs of these equations are lines, which we denote by ℓ_1 and ℓ_2. A pair of numbers (x_1, x_2) satisfies *both* equations in the system if and only if the point (x_1, x_2) lies on both ℓ_1 and ℓ_2. In the system above, the solution is the single point $(3, 2)$, as you can easily verify. See Figure 1.

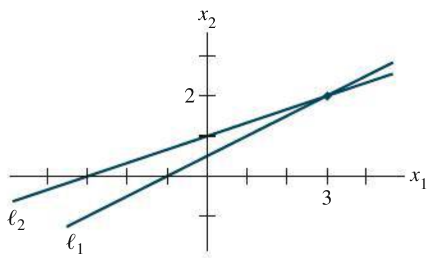


FIGURE 1 Exactly one solution.

Of course, two lines need not intersect in a single point—they could be parallel, or they could coincide and hence "intersect" at every point on the line. Figure 2 shows the graphs that correspond to the following systems:

$$\text{(a)}\quad \begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 3 \end{aligned} \qquad\qquad \text{(b)}\quad \begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 1 \end{aligned}$$

Figures 1 and 2 illustrate the following general fact about linear systems, to be verified in Section 1.2.

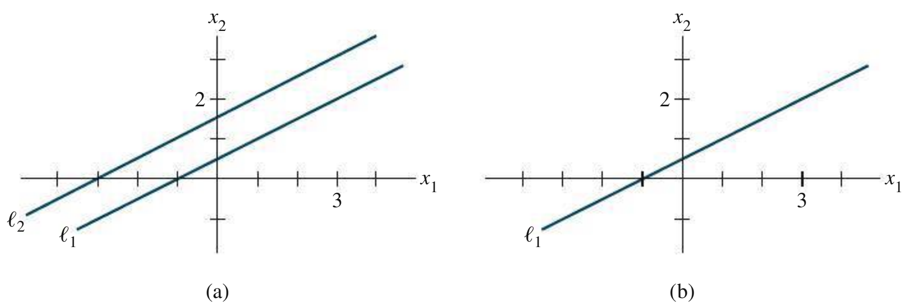


FIGURE 2 (a) No solution. (b) Infinitely many solutions.

A system of linear equations has

1. no solution, or
2. exactly one solution, or
3. infinitely many solutions.

A system of linear equations is said to be **consistent** if it has either one solution or infinitely many solutions; a system is **inconsistent** if it has no solution.

Matrix Notation

The essential information of a linear system can be recorded compactly in a rectangular array called a **matrix**. Given the system

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 - 5x_3 &= 10 \end{aligned} \tag{3}$$

with the coefficients of each variable aligned in columns, the matrix

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$$

is called the **coefficient matrix** (or **matrix of coefficients**) of the system (3), and the matrix

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

is called the **augmented matrix** of the system. (The second row here contains a zero because the second equation could be written as $0 \cdot x_1 + 2x_2 - 8x_3 = 8$.) An augmented matrix of a system consists of the coefficient matrix with an added column containing the constants from the respective right sides of the equations.

The **size** of a matrix tells how many rows and columns it has. The augmented matrix (4) above has 3 rows and 4 columns and is called a 3×4 (read "3 by 4") matrix. If m and n are positive integers, an $\boldsymbol{m \times n}$ **matrix** is a rectangular array of numbers with m rows and n columns. (The number of rows always comes first.) Matrix notation will simplify the calculations in the examples that follow.

Solving a Linear System

This section and the next describe an algorithm, or a systematic procedure, for solving linear systems. The basic strategy is *to replace one system with an equivalent system (that is one with the same solution set) that is easier to solve.*

Roughly speaking, use the x_1 term in the first equation of a system to eliminate the x_1 terms in the other equations. Then use the x_2 term in the second equation to eliminate the x_2 terms in the other equations, and so on, until you finally obtain a very simple equivalent system of equations.

Three basic operations are used to simplify a linear system: Replace one equation by the sum of itself and a multiple of another equation, interchange two equations, and multiply all the terms in an equation by a nonzero constant. After the first example, you will see why these three operations do not change the solution set of the system.

EXAMPLE 1 Solve system (3).

SOLUTION The elimination procedure is shown here with and without matrix notation, and the results are placed side by side for comparison:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 - 5x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

Keep x_1 in the first equation and eliminate it from the other equations. To do so, add -5 times equation 1 to equation 3. After some practice, this type of calculation is usually performed mentally:

$$\begin{array}{r} -5 \cdot [\text{equation 1}] \\ +\,[\text{equation 3}] \\ \hline [\text{new equation 3}] \end{array} \qquad \begin{array}{r} -5x_1 + 10x_2 - 5x_3 = 0 \\ 5x_1 - 5x_3 = 10 \\ \hline 10x_2 - 10x_3 = 10 \end{array}$$

The result of this calculation is written in place of the original third equation:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 10x_2 - 10x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

Now, multiply equation 2 by $\frac{1}{2}$ in order to obtain 1 as the coefficient for x_2. (This calculation will simplify the arithmetic in the next step.)

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ 10x_2 - 10x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

Use the x_2 in equation 2 to eliminate the $10x_2$ in equation 3. The "mental" computation is

$$\begin{array}{r} -10 \cdot [\text{equation 2}] \\ +\,[\text{equation 3}] \\ \hline [\text{new equation 3}] \end{array} \qquad \begin{array}{r} -10x_2 + 40x_3 = -40 \\ 10x_2 - 10x_3 = 10 \\ \hline 30x_3 = -30 \end{array}$$

The result of this calculation is written in place of the previous third equation (row):

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ 30x_3 &= -30 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 30 & -30 \end{bmatrix}$$

Now, multiply equation 3 by $\frac{1}{30}$ in order to obtain 1 as the coefficient for x_3. (This calculation will simplify the arithmetic in the next step.)

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

The new system has a *triangular* form (the intuitive term *triangular* will be replaced by a precise term in the next section):

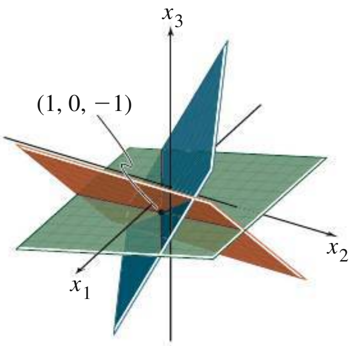
$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Eventually, you want to eliminate the $-2x_2$ term from equation 1, but it is more efficient to use the x_3 in equation 3 first, to eliminate the $-4x_3$ and $+x_3$ terms in equations 2 and 1. The two "mental" calculations are

$$\begin{array}{rr} 4\cdot[\text{equation 3}] & 4x_3 = -4 \\ +\,[\text{equation 2}] & x_2 - 4x_3 = 4 \\ \hline [\text{new equation 2}] & x_2 = 0 \end{array} \qquad \begin{array}{rr} -1\cdot[\text{equation 3}] & -x_3 = 1 \\ +\,[\text{equation 1}] & x_1 - 2x_2 + x_3 = 0 \\ \hline [\text{new equation 1}] & x_1 - 2x_2 = 1 \end{array}$$

It is convenient to combine the results of these two operations:

$$\begin{aligned} x_1 - 2x_2 &= 1 \\ x_2 &= 0 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$



Each of the original equations determines a plane in three-dimensional space. The point $(1, 0, -1)$ lies in all three planes.

Now, having cleaned out the column above the x_3 in equation 3, move back to the x_2 in equation 2 and use it to eliminate the $-2x_2$ above it. Because of the previous work with x_3, there is now no arithmetic involving x_3 terms. Add 2 times equation 2 to equation 1 and obtain the system:

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

The work is essentially done. It shows that the only solution of the original system is $(1, 0, -1)$. However, since there are so many calculations involved, it is a good practice to check the work. To verify that $(1, 0, -1)$ *is* a solution, substitute these values into the left side of the original system, and compute:

$$\begin{aligned} 1(1) - 2(0) + 1(-1) &= 1 - 0 - 1 = 0 \\ 2(0) - 8(-1) &= \phantom{1 -{}} 0 + 8 = 8 \\ 5(1) - 5(-1) &= 5 + 5 = 10 \end{aligned}$$

The results agree with the right side of the original system, so $(1, 0, -1)$ is a solution of the system. ■

Example 1 illustrates how operations on equations in a linear system correspond to operations on the appropriate rows of the augmented matrix. The three basic operations listed earlier correspond to the following operations on the augmented matrix.

ELEMENTARY ROW OPERATIONS

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.[1]
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a nonzero constant.

[1] A common paraphrase of row replacement is "Add to one row a multiple of another row."

Row operations can be applied to any matrix, not merely to one that arises as the augmented matrix of a linear system. Two matrices are called **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into the other.

It is important to note that row operations are *reversible*. If two rows are interchanged, they can be returned to their original positions by another interchange. If a row is scaled by a nonzero constant c, then multiplying the new row by $1/c$ produces the original row. Finally, consider a replacement operation involving two rows—say, rows 1 and 2—and suppose that c times row 1 is added to row 2 to produce a new row 2. To "reverse" this operation, add $-c$ times row 1 to (new) row 2 and obtain the original row 2. See Exercises 39–42 at the end of this section.

At the moment, we are interested in row operations on the augmented matrix of a system of linear equations. Suppose a system is changed to a new one via row operations. By considering each type of row operation, you can see that any solution of the original system remains a solution of the new system. Conversely, since the original system can be produced via row operations on the new system, each solution of the new system is also a solution of the original system. This discussion justifies the following statement.

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Though Example 1 is lengthy, you will find that after some practice, the calculations go quickly. Row operations in the text and exercises will usually be extremely easy to perform, allowing you to focus on the underlying concepts. Still, you must learn to perform row operations accurately because they will be used throughout the text.

The rest of this section shows how to use row operations to determine the size of a solution set, without completely solving the linear system.

Existence and Uniqueness Questions

Section 1.2 will show why a solution set for a linear system contains either no solutions, one solution, or infinitely many solutions. Answers to the following two questions will determine the nature of the solution set for a linear system.

To determine which possibility is true for a particular system, we ask two questions.

TWO FUNDAMENTAL QUESTIONS ABOUT A LINEAR SYSTEM

1. Is the system consistent; that is, does at least one solution *exist*?
2. If a solution exists, is it the *only* one; that is, is the solution *unique*?

These two questions will appear throughout the text, in many different guises. This section and the next will show how to answer these questions via row operations on the augmented matrix.

EXAMPLE 2 Determine if the following system is consistent:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 \phantom{{}-2x_2} - 5x_3 &= 10 \end{aligned}$$

SOLUTION This is the system from Example 1. Suppose that we have performed the row operations necessary to obtain the triangular form

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

At this point, we know x_3. Were we to substitute the value of x_3 into equation 2, we could compute x_2 and hence could determine x_1 from equation 1. So a solution exists; the system is consistent. (In fact, x_2 is uniquely determined by equation 2 since x_3 has only one possible value, and x_1 is therefore uniquely determined by equation 1. So the solution is unique.) ■

EXAMPLE 3 Determine if the following system is consistent:

$$\begin{aligned} x_2 - 4x_3 &= 8 \\ 2x_1 - 3x_2 + 2x_3 &= 1 \\ 4x_1 - 8x_2 + 12x_3 &= 1 \end{aligned} \tag{5}$$

SOLUTION The augmented matrix is

$$\begin{bmatrix} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 4 & -8 & 12 & 1 \end{bmatrix}$$

To obtain an x_1 in the first equation, interchange rows 1 and 2:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 4 & -8 & 12 & 1 \end{bmatrix}$$

To eliminate the $4x_1$ term in the third equation, add -2 times row 1 to row 3:

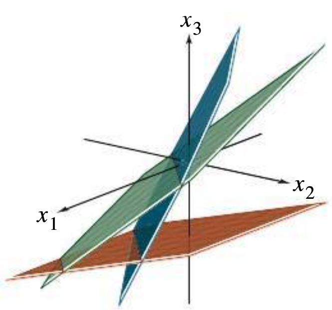
$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & -2 & 8 & -1 \end{bmatrix} \tag{6}$$

Next, use the x_2 term in the second equation to eliminate the $-2x_2$ term from the third equation. Add 2 times row 2 to row 3:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 15 \end{bmatrix} \tag{7}$$

The augmented matrix is now in triangular form. To interpret it correctly, go back to equation notation:

$$\begin{aligned} 2x_1 - 3x_2 + 2x_3 &= 1 \\ x_2 - 4x_3 &= 8 \\ 0 &= 15 \end{aligned} \tag{8}$$



The system is inconsistent because there is no point that lies on all three planes.

The equation $0 = 15$ is a short form of $0x_1 + 0x_2 + 0x_3 = 15$. This system in triangular form obviously has a built-in contradiction. There are no values of x_1, x_2, x_3 that satisfy (8) because the equation $0 = 15$ is never true. Since (8) and (5) have the same solution set, the original system is inconsistent (it has no solution). ■

Pay close attention to the augmented matrix in (7). Its last row is typical of an inconsistent system in triangular form.

Reasonable Answers

Once you have one or more solutions to a system of equations, remember to check your answer by substituting the solution you found back into the original equation. For example, if you found $(2, 1, -1)$ was a solution to the system of equations

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

you could substitute your solution into the original equations to get

$$\begin{aligned} 2 - 2(1) + (-1) &= -1 \neq 2 \\ 2 \qquad - 2(-1) &= 4 \neq -2 \\ 1 + (-1) &= 0 \neq 3 \end{aligned}$$

It is now clear that there must have been an error in your original calculations. If upon rechecking your arithmetic, you find the answer $(2, 1, 2)$, you can see that

$$\begin{aligned} 2 - 2(1) + (2) &= 2 = 2 \\ 2 \qquad - 2(2) &= -2 = -2 \\ 1 + 2 &= 3 = 3 \end{aligned}$$

and you can now be confident you have a correct solution to the given system of equations.

Numerical Note

In real-world problems, systems of linear equations are solved by a computer. For a square coefficient matrix, computer programs nearly always use the elimination algorithm given here and in Section 1.2, modified slightly for improved accuracy.

The vast majority of linear algebra problems in business and industry are solved with programs that use *floating point arithmetic*. Numbers are represented as decimals $\pm .d_1 \cdots d_p \times 10^r$, where r is an integer and the number p of digits to the right of the decimal point is usually between 8 and 16. Arithmetic with such numbers typically is inexact, because the result must be rounded (or truncated) to the number of digits stored. "Roundoff error" is also introduced when a number such as $1/3$ is entered into the computer, since its decimal representation must be approximated by a finite number of digits. Fortunately, inaccuracies in floating point arithmetic seldom cause problems. The numerical notes in this book will occasionally warn of issues that you may need to consider later in your career.

Practice Problems

Throughout the text, practice problems should be attempted before working the exercises. Solutions appear after each exercise set.

1. State in words the next elementary row operation that should be performed on the system in order to solve it. [More than one answer is possible in (a).]

Practice Problems (Continued)

a. $$\begin{aligned} x_1 + 4x_2 - 2x_3 + 8x_4 &= 12 \\ x_2 - 7x_3 + 2x_4 &= -4 \\ 5x_3 - x_4 &= 7 \\ x_3 + 3x_4 &= -5 \end{aligned}$$

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

2. The augmented matrix of a linear system has been transformed by row operations into the form below. Determine if the system is consistent.

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

3. Is $(3, 4, -2)$ a solution of the following system?

$$\begin{aligned} 5x_1 - x_2 + 2x_3 &= 7 \\ -2x_1 + 6x_2 + 9x_3 &= 0 \\ -7x_1 + 5x_2 - 3x_3 &= -7 \end{aligned}$$

4. For what values of h and k is the following system consistent?

$$\begin{aligned} 2x_1 - x_2 &= h \\ -6x_1 + 3x_2 &= k \end{aligned}$$

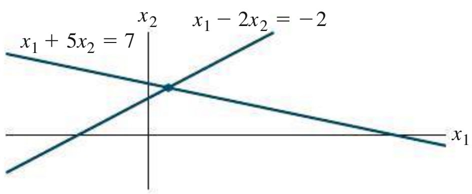
1.1 Exercises

Solve each system in Exercises 1–4 by using elementary row operations on the equations or on the augmented matrix. Follow the systematic elimination procedure described in this section.

1. $$\begin{aligned} x_1 + 5x_2 &= 7 \\ -2x_1 - 7x_2 &= -5 \end{aligned}$$

2. $$\begin{aligned} 2x_1 + 4x_2 &= -4 \\ 5x_1 + 7x_2 &= 11 \end{aligned}$$

3. Find the point (x_1, x_2) that lies on the line $x_1 + 5x_2 = 7$ and on the line $x_1 - 2x_2 = -2$. See the figure.



4. Find the point of intersection of the lines $x_1 - 5x_2 = 1$ and $3x_1 - 7x_2 = 5$.

Consider each matrix in Exercises 5 and 6 as the augmented matrix of a linear system. State in words the next two elementary row operations that should be performed in the process of solving the system.

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

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

In Exercises 7–10, the augmented matrix of a linear system has been reduced by row operations to the form shown. In each case, continue the appropriate row operations and describe the solution set of the original system.

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

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

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

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

Solve the systems in Exercises 11–14.

11. $$\begin{aligned} x_2 + 4x_3 &= -4 \\ x_1 + 3x_2 + 3x_3 &= -2 \\ 3x_1 + 7x_2 + 5x_3 &= 6 \end{aligned}$$

12. $$\begin{aligned} x_1 - 3x_2 + 4x_3 &= -4 \\ 3x_1 - 7x_2 + 7x_3 &= -8 \\ -4x_1 + 6x_2 + 2x_3 &= 4 \end{aligned}$$

13. $$\begin{aligned} x_1 \qquad - 3x_3 &= 8 \\ 2x_1 + 2x_2 + 9x_3 &= 7 \\ x_2 + 5x_3 &= -2 \end{aligned}$$

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

15. Verify that the solution you found to Exercise 11 is correct by substituting the values you obtained back into the original equations.

16. Verify that the solution you found to Exercise 12 is correct by substituting the values you obtained back into the original equations.

17. Verify that the solution you found to Exercise 13 is correct by substituting the values you obtained back into the original equations.

18. Verify that the solution you found to Exercise 14 is correct by substituting the values you obtained back into the original equations.

Determine if the systems in Exercises 19 and 20 are consistent. Do not completely solve the systems.

19. $$\begin{aligned} x_1 \qquad + 3x_3 \qquad &= 2 \\ x_2 \qquad - 3x_4 &= 3 \\ -2x_2 + 3x_3 + 2x_4 &= 1 \\ 3x_1 \qquad\qquad + 7x_4 &= -5 \end{aligned}$$

20. $$\begin{aligned} x_1 \qquad\qquad - 2x_4 &= -3 \\ 2x_2 + 2x_3 \qquad &= 0 \\ x_3 + 3x_4 &= 1 \\ -2x_1 + 3x_2 + 2x_3 + x_4 &= 5 \end{aligned}$$

21. Do the three lines $x_1 - 4x_2 = 1$, $2x_1 - x_2 = -3$, and $-x_1 - 3x_2 = 4$ have a common point of intersection? Explain.

22. Do the three planes $x_1 + 2x_2 + x_3 = 4$, $x_2 - x_3 = 1$, and $x_1 + 3x_2 = 0$ have at least one common point of intersection? Explain.

In Exercises 23–26, determine the value(s) of h such that the matrix is the augmented matrix of a consistent linear system.

23. $\begin{bmatrix} 1 & h & 4 \\ 3 & 6 & 8 \end{bmatrix}$

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

25. $\begin{bmatrix} 1 & 3 & -2 \\ -4 & h & 8 \end{bmatrix}$

26. $\begin{bmatrix} 2 & -3 & h \\ -6 & 9 & 5 \end{bmatrix}$

In Exercises 27–34, key statements from this section are either quoted directly, restated slightly (but still true), or altered in some way that makes them false in some cases. Mark each statement True or False, and *justify* your answer. (If true, give the approximate location where a similar statement appears, or refer to a definition or theorem. If false, give the location of a statement that has been quoted or used incorrectly, or cite an example that shows the statement is not true in all cases.) Similar true/false questions will appear in many sections of the text and will be flagged with a **(T/F)** at the beginning of the question.

27. **(T/F)** Every elementary row operation is reversible.

28. **(T/F)** Elementary row operations on an augmented matrix never change the solution set of the associated linear system.

29. **(T/F)** A 5×6 matrix has six rows.

30. **(T/F)** Two matrices are row equivalent if they have the same number of rows.

31. **(T/F)** The solution set of a linear system involving variables $x_1, \dots, x_n$ is a list of numbers $(s_1, \dots, s_n)$ that makes each equation in the system a true statement when the values $s_1, \dots, s_n$ are substituted for $x_1, \dots, x_n$, respectively.

32. **(T/F)** An inconsistent system has more than one solution.

33. **(T/F)** Two fundamental questions about a linear system involve existence and uniqueness.

34. **(T/F)** Two linear systems are equivalent if they have the same solution set.

35. Find an equation involving g, h, and k that makes this augmented matrix correspond to a consistent system:

$$\begin{bmatrix} 1 & -4 & 7 & g \\ 0 & 3 & -5 & h \\ -2 & 5 & -9 & k \end{bmatrix}$$

36. Construct three different augmented matrices for linear systems whose solution set is $x_1 = -2$, $x_2 = 1$, $x_3 = 0$.

37. Suppose the system below is consistent for all possible values of f and g. What can you say about the coefficients c and d? Justify your answer.

$$\begin{aligned} x_1 + 3x_2 &= f \\ cx_1 + dx_2 &= g \end{aligned}$$

38. Suppose a, b, c, and d are constants such that a is not zero and the system below is consistent for all possible values of f and g. What can you say about the numbers a, b, c, and d? Justify your answer.

$$\begin{aligned} ax_1 + bx_2 &= f \\ cx_1 + dx_2 &= g \end{aligned}$$

In Exercises 39–42, find the elementary row operation that transforms the first matrix into the second, and then find the reverse row operation that transforms the second matrix into the first.

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

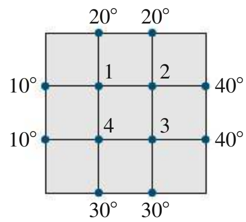
40. $\begin{bmatrix} 1 & 3 & -4 \\ 0 & -2 & 6 \\ 0 & -5 & 9 \end{bmatrix}, \begin{bmatrix} 1 & 3 & -4 \\ 0 & 1 & -3 \\ 0 & -5 & 9 \end{bmatrix}$

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

42. $\begin{bmatrix} 1 & 2 & -5 & 0 \\ 0 & 1 & -3 & -2 \\ 0 & -3 & 9 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 2 & -5 & 0 \\ 0 & 1 & -3 & -2 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

An important concern in the study of heat transfer is to determine the steady-state temperature distribution of a thin plate when the temperature around the boundary is known. Assume the plate shown in the figure represents a cross section of a metal beam, with negligible heat flow in the direction perpendicular to the plate. Let $T_1, \dots, T_4$ denote the temperatures at the four interior nodes of the mesh in the figure. The temperature at a node is approximately equal to the average of the four nearest nodes—to the left, above, to the right, and below.[2] For instance,

$$T_1 = (10 + 20 + T_2 + T_4)/4, \quad \text{or} \quad 4T_1 - T_2 - T_4 = 30$$



43. Write a system of four equations whose solution gives estimates for the temperatures $T_1, \dots, T_4$.

44. Solve the system of equations from Exercise 43. [*Hint:* To speed up the calculations, interchange rows 1 and 4 before starting "replace" operations.]

[2] See Frank M. White, *Heat and Mass Transfer* (Reading, MA: Addison-Wesley Publishing, 1991), pp. 145–149.

Solutions to Practice Problems

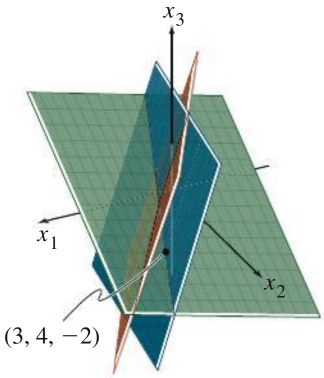
1. a. For "hand computation," the best choice is to interchange equations 3 and 4. Another possibility is to multiply equation 3 by $1/5$. Or, replace equation 4 by its sum with $-1/5$ times row 3. (In any case, do not use the x_2 in equation 2 to eliminate the $4x_2$ in equation 1. Wait until a triangular form has been reached and the x_3 terms and x_4 terms have been eliminated from the first two equations.)

b. The system is in triangular form. Further simplification begins with the x_4 in the fourth equation. Use the x_4 to eliminate all x_4 terms above it. The appropriate step now is to add 2 times equation 4 to equation 1. (After that, move to equation 3, multiply it by $1/2$, and then use the equation to eliminate the x_3 terms above it.)

2. The system corresponding to the augmented matrix is

$$\begin{aligned} x_1 + 5x_2 + 2x_3 &= -6 \\ 4x_2 - 7x_3 &= 2 \\ 5x_3 &= 0 \end{aligned}$$

The third equation makes $x_3 = 0$, which is certainly an allowable value for x_3. After eliminating the x_3 terms in equations 1 and 2, you could go on to solve for unique values for x_2 and x_1. Hence a solution exists, and it is unique. Contrast this situation with that in Example 3.



Since $(3, 4, -2)$ satisfies the first two equations, it is on the line of the intersection of the first two planes. Since $(3, 4, -2)$ does not satisfy all three equations, it does not lie on all three planes.

3. It is easy to check if a specific list of numbers is a solution. Set $x_1 = 3$, $x_2 = 4$, and $x_3 = -2$, and find that

$$\begin{aligned} 5(3) - (4) + 2(-2) &= 15 - 4 - 4 = 7 \\ -2(3) + 6(4) + 9(-2) &= -6 + 24 - 18 = 0 \\ -7(3) + 5(4) - 3(-2) &= -21 + 20 + 6 = 5 \end{aligned}$$

Although the first two equations are satisfied, the third is not, so $(3, 4, -2)$ is not a solution of the system. Notice the use of parentheses when making the substitutions. They are strongly recommended as a guard against arithmetic errors.

4. When the second equation is replaced by its sum with 3 times the first equation, the system becomes

$$\begin{aligned} 2x_1 - x_2 &= h \\ 0 &= k + 3h \end{aligned}$$

If $k + 3h$ is nonzero, the system has no solution. The system is consistent for any values of h and k that make $k + 3h = 0$.

1.2 Row Reduction and Echelon Forms

This section refines the method of Section 1.1 into a row reduction algorithm that will enable us to analyze any system of linear equations.[1] By using only the first part of the algorithm, we will be able to answer the fundamental existence and uniqueness questions posed in Section 1.1.

The algorithm applies to any matrix, whether or not the matrix is viewed as an augmented matrix for a linear system. So the first part of this section concerns an arbitrary rectangular matrix and begins by introducing two important classes of matrices that include the "triangular" matrices of Section 1.1. In the definitions that follow, a *nonzero* row or column in a matrix means a row or column that contains at least one nonzero entry; a **leading entry** of a row refers to the leftmost nonzero entry (in a nonzero row).

DEFINITION

A rectangular matrix is in **echelon form** (or **row echelon form**) if it has the following three properties:

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

If a matrix in echelon form satisfies the following additional conditions, then it is in **reduced echelon form** (or **reduced row echelon form**):

4. The leading entry in each nonzero row is 1.
5. Each leading 1 is the only nonzero entry in its column.

[1] The algorithm here is a variant of what is commonly called *Gaussian elimination*. A similar elimination method for linear systems was used by Chinese mathematicians in about 250 B.C. The process was unknown in Western culture until the nineteenth century, when a famous German mathematician, Carl Friedrich Gauss, discovered it. A German engineer, Wilhelm Jordan, popularized the algorithm in an 1888 text on geodesy.

An **echelon matrix** (respectively, **reduced echelon matrix**) is one that is in echelon form (respectively, reduced echelon form). Property 2 says that the leading entries form an *echelon* ("steplike") pattern that moves down and to the right through the matrix. Property 3 is a simple consequence of property 2, but we include it for emphasis.

The "triangular" matrices of Section 1.1, such as

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 5/2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

are in echelon form. In fact, the second matrix is in reduced echelon form. Here are additional examples.

EXAMPLE 1 The following matrices are in echelon form. The leading entries ($\blacksquare$) may have any nonzero value; the starred entries ($*$) may have any value (including zero).

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

The following matrices are in reduced echelon form because the leading entries are 1's, and there are 0's below *and above* each leading 1.

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

Any nonzero matrix may be **row reduced** (that is, transformed by elementary row operations) into more than one matrix in echelon form, using different sequences of row operations. However, the reduced echelon form one obtains from a matrix is unique. The following theorem is proved in Appendix A at the end of the text.

THEOREM 1 **Uniqueness of the Reduced Echelon Form**

Each matrix is row equivalent to one and only one reduced echelon matrix.

If a matrix A is row equivalent to an echelon matrix U, we call U **an echelon form** (or row echelon form) **of** A; if U is in reduced echelon form, we call U **the reduced echelon form of** A. [Most matrix programs and calculators with matrix capabilities use the abbreviation RREF for reduced (row) echelon form. Some use REF for (row) echelon form.]

Pivot Positions

When row operations on a matrix produce an echelon form, further row operations to obtain the reduced echelon form do not change the positions of the leading entries. Since the reduced echelon form is unique, *the leading entries are always in the same positions* *in any echelon form obtained from a given matrix.* These leading entries correspond to leading 1's in the reduced echelon form.

DEFINITION

A **pivot position** in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column of A that contains a pivot position.

In Example 1, the squares (■) identify the pivot positions. Many fundamental concepts in the first four chapters will be connected in one way or another with pivot positions in a matrix.

EXAMPLE 2 Row reduce the matrix A below to echelon form, and locate the pivot columns of A.

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

SOLUTION Use the same basic strategy as in Section 1.1. The top of the leftmost nonzero column is the first pivot position. A nonzero entry, or *pivot*, must be placed in this position. A good choice is to interchange rows 1 and 4 (because the mental computations in the next step will not involve fractions).

$$\begin{array}{c} \text{Pivot} \\ \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \\ \text{Pivot column} \end{array}$$

Create zeros below the pivot, 1, by adding multiples of the first row to the rows below, and obtain matrix (1) below. The pivot position in the second row must be as far left as possible—namely in the second column. Choose the 2 in this position as the next pivot.

$$\begin{array}{c} \text{Pivot} \\ \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \\ \text{Next pivot column} \end{array} \tag{1}$$

Add $-5/2$ times row 2 to row 3, and add $3/2$ times row 2 to row 4.

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

The matrix in (2) is different from any encountered in Section 1.1. There is no way to create a leading entry in column 3! (We can't use row 1 or 2 because doing so would destroy the echelon arrangement of the leading entries already produced.) However, if we interchange rows 3 and 4, we can produce a leading entry in column 4.

Pivot

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{General form: } \begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Pivot columns

The matrix is in echelon form and thus reveals that columns 1, 2, and 4 of A are pivot columns.

Pivot positions

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

Pivot columns ■

A **pivot**, as illustrated in Example 2, is a nonzero number in a pivot position that is used as needed to create zeros via row operations. The pivots in Example 2 were 1, 2, and -5. Notice that these numbers are not the same as the actual elements of A in the highlighted pivot positions shown in (3).

With Example 2 as a guide, we are ready to describe an efficient procedure for transforming a matrix into an echelon or reduced echelon matrix. Careful study and mastery of this procedure now will pay rich dividends later in the course.

The Row Reduction Algorithm

The algorithm that follows consists of four steps, and it produces a matrix in echelon form. A fifth step produces a matrix in reduced echelon form. We illustrate the algorithm by an example.

EXAMPLE 3 Apply elementary row operations to transform the following matrix first into echelon form and then into reduced echelon form:

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

SOLUTION

Step 1

Begin with the leftmost nonzero column. This is a pivot column. The pivot position is at the top.

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

Pivot column

> **Step 2**
>
> Select a nonzero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

Interchange rows 1 and 3. (We could have interchanged rows 1 and 2 instead.)

Pivot

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

> **Step 3**
>
> Use row replacement operations to create zeros in all positions below the pivot.

As a preliminary step, we could divide the top row by the pivot, 3. But with two 3's in column 1, it is just as easy to add -1 times row 1 to row 2.

Pivot

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

> **Step 4**
>
> Cover (or ignore) the row containing the pivot position and cover all rows, if any, above it. Apply steps 1–3 to the submatrix that remains. Repeat the process until there are no more nonzero rows to modify.

With row 1 covered, step 1 shows that column 2 is the next pivot column; for step 2, select as a pivot the "top" entry in that column.

Pivot

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

New pivot column

For step 3, we could insert an optional step of dividing the "top" row of the submatrix by the pivot, 2. Instead, we add $-3/2$ times the "top" row to the row below. This produces

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

When we cover the row containing the second pivot position for step 4, we are left with a new submatrix having only one row:

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

Pivot

Steps 1–3 require no work for this submatrix, and we have reached an echelon form of the full matrix. If we want the reduced echelon form, we perform one more step.

Step 5

Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. If a pivot is not 1, make it 1 by a scaling operation.

The rightmost pivot is in row 3. Create zeros above it, adding suitable multiples of row 3 to rows 2 and 1.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 2 & -4 & 4 & 0 & -14 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{matrix} \leftarrow \text{Row 1} + (-6)\cdot\text{row 3} \\ \leftarrow \text{Row 2} + (-2)\cdot\text{row 3} \\ \end{matrix}$$

The next pivot is in row 2. Scale this row, dividing by the pivot.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \leftarrow \text{Row scaled by } \tfrac{1}{2}$$

Create a zero in column 2 by adding 9 times row 2 to row 1.

$$\begin{bmatrix} 3 & 0 & -6 & 9 & 0 & -72 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{matrix} \leftarrow \text{Row 1} + (9)\cdot\text{row 2} \\ \\ \end{matrix}$$

Finally, scale row 1, dividing by the pivot, 3.

$$\begin{bmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{matrix} \leftarrow \text{Row scaled by } \tfrac{1}{3} \\ \\ \end{matrix}$$

This is the reduced echelon form of the original matrix. ■

The combination of steps 1–4 is called the **forward phase** of the row reduction algorithm. Step 5, which produces the unique reduced echelon form, is called the **backward phase**.

Numerical Note

In step 2 on page 17, a computer program usually selects as a pivot the entry in a column having the largest absolute value. This strategy, called **partial pivoting**, is used because it reduces roundoff errors in the calculations.

Solutions of Linear Systems

The row reduction algorithm leads directly to an explicit description of the solution set of a linear system when the algorithm is applied to the augmented matrix of the system.

Suppose, for example, that the augmented matrix of a linear system has been changed into the equivalent *reduced* echelon form

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

There are three variables because the augmented matrix has four columns. The associated system of equations is

$$\begin{aligned} x_1 \qquad - 5x_3 &= 1 \\ x_2 + \quad x_3 &= 4 \\ 0 &= 0 \end{aligned} \tag{4}$$

The variables x_1 and x_2 corresponding to pivot columns in the matrix are called **basic variables**.[2] The other variable, x_3, is called a **free variable**.

Whenever a system is consistent, as in (4), the solution set can be described explicitly by solving the *reduced* system of equations for the basic variables in terms of the free variables. This operation is possible because the reduced echelon form places each basic variable in one and only one equation. In (4), solve the first equation for x_1 and the second for x_2. (Ignore the third equation; it offers no restriction on the variables.)

$$\begin{cases} x_1 = 1 + 5x_3 \\ x_2 = 4 - x_3 \\ x_3 \text{ is free} \end{cases} \tag{5}$$

The statement "x_3 is free" means that you are free to choose any value for x_3. Once that is done, the formulas in (5) determine the values for x_1 and x_2. For instance, when $x_3 = 0$, the solution is $(1, 4, 0)$; when $x_3 = 1$, the solution is $(6, 3, 1)$. *Each different choice of x_3 determines a (different) solution of the system, and every solution of the system is determined by a choice of x_3.*

EXAMPLE 4 Find the general solution of the linear system whose augmented matrix has been reduced to

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

SOLUTION The matrix is in echelon form, but we want the reduced echelon form before solving for the basic variables. The row reduction is completed next. The symbol $\sim$ before a matrix indicates that the matrix is row equivalent to the preceding matrix.

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 2 & -8 & 0 & 10 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

[2] Some texts use the term *leading variables* because they correspond to the columns containing leading entries.

There are five variables because the augmented matrix has six columns. The associated system now is

$$\begin{aligned} x_1 + 6x_2 + 3x_4 &= 0 \\ x_3 - 4x_4 &= 5 \\ x_5 &= 7 \end{aligned} \tag{6}$$

The pivot columns of the matrix are 1, 3, and 5, so the basic variables are x_1, x_3, and x_5. The remaining variables, x_2 and x_4, must be free. Solve for the basic variables to obtain the general solution:

$$\begin{cases} x_1 = -6x_2 - 3x_4 \\ x_2 \text{ is free} \\ x_3 = 5 + 4x_4 \\ x_4 \text{ is free} \\ x_5 = 7 \end{cases} \tag{7}$$

Note that the value of x_5 is already fixed by the third equation in system (6). ■

Parametric Descriptions of Solution Sets

The descriptions in (5) and (7) are *parametric descriptions* of solution sets in which the free variables act as parameters. *Solving a system* amounts to finding a parametric description of the solution set or determining that the solution set is empty.

Whenever a system is consistent and has free variables, the solution set has many parametric descriptions. For instance, in system (4), we may add 5 times equation 2 to equation 1 and obtain the equivalent system

$$\begin{aligned} x_1 + 5x_2 &= 21 \\ x_2 + x_3 &= 4 \end{aligned}$$

We could treat x_2 as a parameter and solve for x_1 and x_3 in terms of x_2, and we would have an accurate description of the solution set. However, to be consistent, we make the (arbitrary) convention of always using the free variables as the parameters for describing a solution set. (The answer section at the end of the text also reflects this convention.)

Whenever a system is inconsistent, the solution set is empty, even when the system has free variables. In this case, the solution set has *no* parametric representation.

Back-Substitution

Consider the following system, whose augmented matrix is in echelon form but is *not* in reduced echelon form:

$$\begin{aligned} x_1 - 7x_2 + 2x_3 - 5x_4 + 8x_5 &= 10 \\ x_2 - 3x_3 + 3x_4 + x_5 &= -5 \\ x_4 - x_5 &= 4 \end{aligned}$$

A computer program would solve this system by back-substitution, rather than by computing the reduced echelon form. That is, the program would solve equation 3 for x_4 in terms of x_5 and substitute the expression for x_4 into equation 2, solve equation 2 for x_2, and then substitute the expressions for x_2 and x_4 into equation 1 and solve for x_1.

Our matrix format for the backward phase of row reduction, which produces the reduced echelon form, has the same number of arithmetic operations as back-substitution. But the discipline of the matrix format substantially reduces the likelihood of errors during hand computations. The best strategy is to use only the *reduced* echelon form to solve a system! The *Study Guide* that accompanies this text offers several helpful suggestions for performing row operations accurately and rapidly.

Numerical Note

In general, the forward phase of row reduction takes much longer than the backward phase. An algorithm for solving a system is usually measured in flops (or floating point operations). A **flop** is one arithmetic operation $(+, -, *, /)$ on two real floating point numbers.[3] For an $n \times (n+1)$ matrix, the reduction to echelon form can take $2n^3/3 + n^2/2 - 7n/6$ flops (which is approximately $2n^3/3$ flops when n is moderately large—say, $n \geq 30$). In contrast, further reduction to reduced echelon form needs at most n^2 flops.

Existence and Uniqueness Questions

Although a nonreduced echelon form is a poor tool for solving a system, this form is just the right device for answering two fundamental questions posed in Section 1.1.

EXAMPLE 5 Determine the existence and uniqueness of the solutions to the system

$$\begin{aligned} 3x_2 - 6x_3 + 6x_4 + 4x_5 &= -5 \\ 3x_1 - 7x_2 + 8x_3 - 5x_4 + 8x_5 &= 9 \\ 3x_1 - 9x_2 + 12x_3 - 9x_4 + 6x_5 &= 15 \end{aligned}$$

SOLUTION The augmented matrix of this system was row reduced in Example 3 to

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \tag{8}$$

The basic variables are x_1, x_2, and x_5; the free variables are x_3 and x_4. There is no equation such as $0 = 1$ that would indicate an inconsistent system, so we could use back-substitution to find a solution. But the *existence* of a solution is already clear in (8). Also, the solution is *not unique* because there are free variables. Each different choice of x_3 and x_4 determines a different solution. Thus the system has infinitely many solutions. ■

When a system is in echelon form and contains no equation of the form $0 = b$, with b nonzero, every nonzero equation contains a basic variable with a nonzero coefficient. Either the basic variables are completely determined (with no free variables) or at least one of the basic variables may be expressed in terms of one or more free variables. In the former case, there is a unique solution; in the latter case, there are infinitely many solutions (one for each choice of values for the free variables).

These remarks justify the following theorem.

[3] Traditionally, a *flop* was only a multiplication or division because addition and subtraction took much less time and could be ignored. The definition of *flop* given here is preferred now, as a result of advances in computer architecture. See Golub and Van Loan, *Matrix Computations*, 2nd ed. (Baltimore: The Johns Hopkins Press, 1989), pp. 19–20.

THEOREM 2

Existence and Uniqueness Theorem

A linear system is consistent if and only if the rightmost column of the augmented matrix is *not* a pivot column—that is, if and only if an echelon form of the augmented matrix has *no* row of the form

$$[\,0 \quad \cdots \quad 0 \quad b\,] \qquad \text{with } b \text{ nonzero}$$

If a linear system is consistent, then the solution set contains either (i) a unique solution, when there are no free variables, or (ii) infinitely many solutions, when there is at least one free variable.

The following procedure outlines how to find and describe all solutions of a linear system.

USING ROW REDUCTION TO SOLVE A LINEAR SYSTEM

1. Write the augmented matrix of the system.
2. Use the row reduction algorithm to obtain an equivalent augmented matrix in echelon form. Decide whether the system is consistent. If there is no solution, stop; otherwise, go to the next step.
3. Continue row reduction to obtain the reduced echelon form.
4. Write the system of equations corresponding to the matrix obtained in step 3.
5. Rewrite each nonzero equation from step 4 so that its one basic variable is expressed in terms of any free variables appearing in the equation.

Reasonable Answers

Remember that each augmented matrix corresponds to a system of equations. If you row reduce the augmented matrix $\begin{bmatrix} 1 & -2 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 0 & 1 & 1 & 3 \end{bmatrix}$ to get the matrix $\begin{bmatrix} 1 & 0 & 3 & 8 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, the solution set is

$$\begin{cases} x_1 = 8 - 3x_3 \\ x_2 = 3 - x_3 \\ x_3 \text{ is free} \end{cases}$$

The system of equations corresponding to the original augmented matrix is

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

You can now check whether your solution is correct by substituting it into the original equations. Notice that you can just leave the free variables in the solution.

$$\begin{aligned}
(8-3x_3) \;-\; 2(3-x_3) \;+\; (x_3) &= 8-3x_3-6+2x_3+\;x_3 = 2\\
(8-3x_3) \;-\; (3-x_3) \;+\; 2(x_3) &= 8-3x_3-3+\;x_3+2x_3 = 5\\
(3-x_3) \;+\; (x_3) &= 3-\;x_3+\;x_3 = 3
\end{aligned}$$

You can now be confident you have a correct solution to the system of equations represented by the augmented matrix.

Practice Problems

1. Find the general solution of the linear system whose augmented matrix is

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

2. Find the general solution of the system

$$\begin{aligned}
x_1 - 2x_2 - \;x_3 + 3x_4 &= 0\\
-2x_1 + 4x_2 + 5x_3 - 5x_4 &= 3\\
3x_1 - 6x_2 - 6x_3 + 8x_4 &= 2
\end{aligned}$$

3. Suppose a 4×7 coefficient matrix for a system of equations has 4 pivots. Is the system consistent? If the system is consistent, how many solutions are there?

1.2 Exercises

In Exercises 1 and 2, determine which matrices are in reduced echelon form and which others are only in echelon form.

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

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

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

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

Row reduce the matrices in Exercises 3 and 4 to reduced echelon form. Circle the pivot positions in the final matrix and in the original matrix, and list the pivot columns.

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

5. Describe the possible echelon forms of a nonzero 2×2 matrix. Use the symbols ■, $*$, and 0, as in the first part of Example 1.

6. Repeat Exercise 5 for a nonzero 3×2 matrix.

Find the general solutions of the systems whose augmented matrices are given in Exercises 7–14.

7. $\begin{bmatrix} 1 & 3 & 4 & 7 \\ 3 & 9 & 7 & 6 \end{bmatrix}$ **8.** $\begin{bmatrix} 1 & 4 & 0 & 7 \\ 2 & 7 & 0 & 11 \end{bmatrix}$

9. $\begin{bmatrix} 0 & 1 & -6 & 5 \\ 1 & -2 & 7 & -4 \end{bmatrix}$ **10.** $\begin{bmatrix} 1 & -2 & -1 & 3 \\ 3 & -6 & -2 & 2 \end{bmatrix}$

11. $\begin{bmatrix} 3 & -4 & 2 & 0 \\ -9 & 12 & -6 & 0 \\ -6 & 8 & -4 & 0 \end{bmatrix}$ **12.** $\begin{bmatrix} 1 & -7 & 0 & 6 & 5 \\ 0 & 0 & 1 & -2 & -3 \\ -1 & 7 & -4 & 2 & 7 \end{bmatrix}$

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

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

You may find it helpful to review the information in the Reasonable Answers box from this section before answering Exercises 15–18.

15. Write down the equations corresponding to the augmented matrix in Exercise 9 and verify your answer to Exercise 9 is correct by substituting the solutions you obtained back into the original equations.

16. Write down the equations corresponding to the augmented matrix in Exercise 10 and verify your answer to Exercise 10 is correct by substituting the solutions you obtained back into the original equations.

17. Write down the equations corresponding to the augmented matrix in Exercise 11 and verify your answer to Exercise 11 is correct by substituting the solutions you obtained back into the original equations.

18. Write down the equations corresponding to the augmented matrix in Exercise 12 and verify your answer to Exercise 12 is correct by substituting the solutions you obtained back into the original equations.

Exercises 19 and 20 use the notation of Example 1 for matrices in echelon form. Suppose each matrix represents the augmented matrix for a system of linear equations. In each case, determine if the system is consistent. If the system is consistent, determine if the solution is unique.

19. a. $\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & 0 \end{bmatrix}$

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

20. a. $\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & 0 \end{bmatrix}$

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

In Exercises 21 and 22, determine the value(s) of h such that the matrix is the augmented matrix of a consistent linear system.

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

22. $\begin{bmatrix} 1 & -3 & -2 \\ 5 & h & -7 \end{bmatrix}$

In Exercises 23 and 24, choose h and k such that the system has (a) no solution, (b) a unique solution, and (c) many solutions. Give separate answers for each part.

23. $\begin{aligned} x_1 + hx_2 &= 2 \\ 4x_1 + 8x_2 &= k \end{aligned}$

24. $\begin{aligned} x_1 + 3x_2 &= 2 \\ 3x_1 + hx_2 &= k \end{aligned}$

In Exercises 25–34, mark each statement True or False **(T/F)**. Justify each answer.[4]

25. **(T/F)** In some cases, a matrix may be row reduced to more than one matrix in reduced echelon form, using different sequences of row operations.

26. **(T/F)** The echelon form of a matrix is unique.

27. **(T/F)** The row reduction algorithm applies only to augmented matrices for a linear system.

28. **(T/F)** The pivot positions in a matrix depend on whether row interchanges are used in the row reduction process.

29. **(T/F)** A basic variable in a linear system is a variable that corresponds to a pivot column in the coefficient matrix.

30. **(T/F)** Reducing a matrix to echelon form is called the *forward phase* of the row reduction process.

31. **(T/F)** Finding a parametric description of the solution set of a linear system is the same as *solving* the system.

32. **(T/F)** Whenever a system has free variables, the solution set contains a unique solution.

33. **(T/F)** If one row in an echelon form of an augmented matrix is $[0 \quad 0 \quad 0 \quad 0 \quad 5]$, then the associated linear system is inconsistent.

34. **(T/F)** A general solution of a system is an explicit description of all solutions of the system.

35. Suppose a 3×5 *coefficient* matrix for a system has three pivot columns. Is the system consistent? Why or why not?

36. Suppose a system of linear equations has a 3×5 *augmented* matrix whose fifth column is a pivot column. Is the system consistent? Why (or why not)?

37. Suppose the coefficient matrix of a system of linear equations has a pivot position in every row. Explain why the system is consistent.

38. Suppose the coefficient matrix of a linear system of three equations in three variables has a pivot in each column. Explain why the system has a unique solution.

39. Restate the last sentence in Theorem 2 using the concept of pivot columns: "If a linear system is consistent, then the solution is unique if and only if ________."

40. What would you have to know about the pivot columns in an augmented matrix in order to know that the linear system is consistent and has a unique solution?

41. A system of linear equations with fewer equations than unknowns is sometimes called an *underdetermined system.*

[4] True/false questions of this type will appear in many sections. Methods for justifying your answers were described before the True or False exercises in Section 1.1.

Suppose that such a system happens to be consistent. Explain why there must be an infinite number of solutions.

42. Give an example of an inconsistent underdetermined system of two equations in three unknowns.

43. A system of linear equations with more equations than unknowns is sometimes called an *overdetermined system*. Can such a system be consistent? Illustrate your answer with a specific system of three equations in two unknowns.

44. Suppose an $n \times (n+1)$ matrix is row reduced to reduced echelon form. Approximately what fraction of the total number of operations (flops) is involved in the backward phase of the reduction when $n = 30$? when $n = 300$?

Suppose experimental data are represented by a set of points in the plane. An **interpolating polynomial** for the data is a polynomial whose graph passes through every point. In scientific work, such a polynomial can be used, for example, to estimate values between the known data points. Another use is to create curves for graphical images on a computer screen. One method for finding an interpolating polynomial is to solve a system of linear equations.

45. Find the interpolating polynomial $p(t) = a_0 + a_1 t + a_2 t^2$ for the data $(1, 12)$, $(2, 15)$, $(3, 16)$. That is, find a_0, a_1, and a_2 such that

$$\begin{aligned} a_0 + a_1(1) + a_2(1)^2 &= 12 \\ a_0 + a_1(2) + a_2(2)^2 &= 15 \\ a_0 + a_1(3) + a_2(3)^2 &= 16 \end{aligned}$$

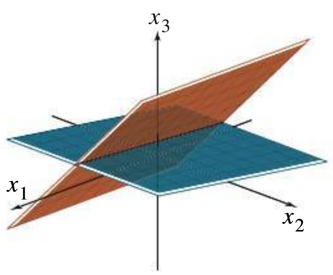
T **46.** In a wind tunnel experiment, the force on a projectile due to air resistance was measured at different velocities:

Velocity (100 ft/sec)	0	2	4	6	8	10
Force (100 lb)	0	2.90	14.8	39.6	74.3	119

Find an interpolating polynomial for these data and estimate the force on the projectile when the projectile is traveling at 750 ft/sec. Use $p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5$. What happens if you try to use a polynomial of degree less than 5? (Try a cubic polynomial, for instance.)[5]

[5] Exercises marked with the symbol T are designed to be worked with the aid of a " Matrix program" (a computer program, such as MATLAB, Maple, Mathematica, MathCad, Octave, or Derive, or a programmable calculator with matrix capabilities, such as those manufactured by Texas Instruments or Hewlett-Packard).

Solutions to Practice Problems



The general solution of the system of equations is the line of intersection of the two planes.

1. The reduced echelon form of the augmented matrix and the corresponding system are

$$\begin{bmatrix} 1 & 0 & -8 & -3 \\ 0 & 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{aligned} x_1 \qquad - 8x_3 &= -3 \\ x_2 - \ \ x_3 &= -1 \end{aligned}$$

The basic variables are x_1 and x_2, and the general solution is

$$\begin{cases} x_1 = -3 + 8x_3 \\ x_2 = -1 + x_3 \\ x_3 \text{ is free} \end{cases}$$

Note: It is essential that the general solution describe each variable, with any parameters clearly identified. The following statement does *not* describe the solution:

$$\begin{cases} x_1 = -3 + 8x_3 \\ x_2 = -1 + x_3 \\ x_3 = 1 + x_2 \quad \text{Incorrect solution} \end{cases}$$

This description implies that x_2 and x_3 are *both* free, which certainly is not the case.

2. Row reduce the system's augmented matrix:

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

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

Solutions to Practice Problems (Continued)

This echelon matrix shows that the system is *inconsistent*, because its rightmost column is a pivot column; the third row corresponds to the equation $0 = 5$. There is no need to perform any more row operations. Note that the presence of the free variables in this problem is irrelevant because the system is inconsistent.

3. Since the coefficient matrix has four pivots, there is a pivot in every row of the coefficient matrix. This means that when the coefficient matrix is row reduced, it will *not* have a row of zeros, thus the corresponding row reduced augmented matrix can never have a row of the form $[0 \;\; 0 \;\; \cdots \;\; 0 \;\; b]$, where b is a nonzero number. By Theorem 2, the system is consistent. Moreover, since there are seven columns in the coefficient matrix and only four pivot columns, there will be three free variables resulting in infinitely many solutions.

1.3 Vector Equations

Important properties of linear systems can be described with the concept and notation of vectors. This section connects equations involving vectors to ordinary systems of equations. The term *vector* appears in a variety of mathematical and physical contexts, which we will discuss in Chapter 4, "Vector Spaces." Until then, *vector* will mean an *ordered list of numbers*. This simple idea enables us to get to interesting and important applications as quickly as possible.

Vectors in $\mathbb{R}^2$

A matrix with only one column is called a **column vector** or simply a **vector**. Examples of vectors with two entries are

$$\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} .2 \\ .3 \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

where w_1 and w_2 are any real numbers. The set of all vectors with two entries is denoted by $\mathbb{R}^2$ (read "r-two"). The $\mathbb{R}$ stands for the real numbers that appear as entries in the vectors, and the exponent 2 indicates that each vector contains two entries.[1]

Two vectors in $\mathbb{R}^2$ are **equal** if and only if their corresponding entries are equal. Thus $\begin{bmatrix} 4 \\ 7 \end{bmatrix}$ and $\begin{bmatrix} 7 \\ 4 \end{bmatrix}$ are *not* equal, because vectors in $\mathbb{R}^2$ are *ordered pairs* of real numbers.

Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$, their **sum** is the vector $\mathbf{u} + \mathbf{v}$ obtained by adding corresponding entries of $\mathbf{u}$ and $\mathbf{v}$. For example,

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

Given a vector $\mathbf{u}$ and a real number c, the **scalar multiple** of $\mathbf{u}$ by c is the vector $c\mathbf{u}$ obtained by multiplying each entry in $\mathbf{u}$ by c. For instance,

$$\text{if} \quad \mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad \text{and} \quad c = 5, \qquad \text{then} \quad c\mathbf{u} = 5\begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 15 \\ -5 \end{bmatrix}$$

[1] Most of the text concerns vectors and matrices that have only real entries. However, all definitions and theorems in Chapters 1–5, and in most of the rest of the text, remain valid if the entries are complex numbers. Complex vectors and matrices arise naturally, for example, in electrical engineering and physics.

The number c in $c\mathbf{u}$ is called a **scalar**; it is written in lightface type to distinguish it from the boldface vector $\mathbf{u}$.

The operations of scalar multiplication and vector addition can be combined, as in the following example.

EXAMPLE 1 Given $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$, find $4\mathbf{u}$, $(-3)\mathbf{v}$, and $4\mathbf{u} + (-3)\mathbf{v}$.

SOLUTION

$$4\mathbf{u} = \begin{bmatrix} 4 \\ -8 \end{bmatrix}, \qquad (-3)\mathbf{v} = \begin{bmatrix} -6 \\ 15 \end{bmatrix}$$

and

$$4\mathbf{u} + (-3)\mathbf{v} = \begin{bmatrix} 4 \\ -8 \end{bmatrix} + \begin{bmatrix} -6 \\ 15 \end{bmatrix} = \begin{bmatrix} -2 \\ 7 \end{bmatrix}$$ ■

Sometimes, for convenience (and also to save space), this text may write a column vector such as $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ in the form $(3, -1)$. In this case, the parentheses and the comma distinguish the vector $(3, -1)$ from the 1×2 row matrix $\begin{bmatrix} 3 & -1 \end{bmatrix}$, written with brackets and no comma. Thus

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

because the matrices have different shapes, even though they have the same entries.

Geometric Descriptions of $\mathbb{R}^2$

Consider a rectangular coordinate system in the plane. Because each point in the plane is determined by an ordered pair of numbers, *we can identify a geometric point* (a, b) *with the column vector* $\begin{bmatrix} a \\ b \end{bmatrix}$. So we may regard $\mathbb{R}^2$ as the set of all points in the plane. See Figure 1.

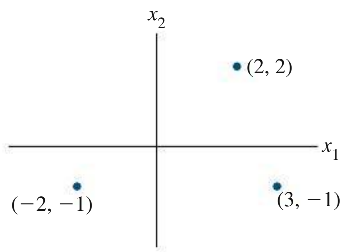


FIGURE 1 Vectors as points.

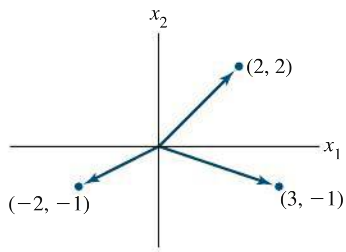


FIGURE 2 Vectors with arrows.

The geometric visualization of a vector such as $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ is often aided by including an arrow (directed line segment) from the origin $(0, 0)$ to the point $(3, -1)$, as in Figure 2. In this case, the individual points along the arrow itself have no special significance.[2]

The sum of two vectors has a useful geometric representation. The following rule can be verified by analytic geometry.

[2] In physics, arrows can represent forces and usually are free to move about in space. This interpretation of vectors will be discussed in Section 4.1.

Parallelogram Rule for Addition

If **u** and **v** in $\mathbb{R}^2$ are represented as points in the plane, then $\mathbf{u} + \mathbf{v}$ corresponds to the fourth vertex of the parallelogram whose other vertices are **u**, **0**, and **v**. See Figure 3.

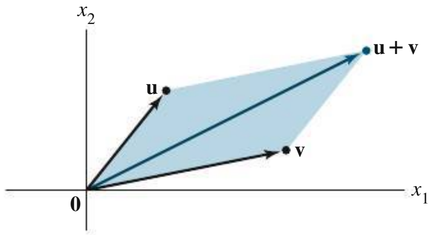


FIGURE 3 The parallelogram rule.

EXAMPLE 2 The vectors $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ are displayed in Figure 4. ■

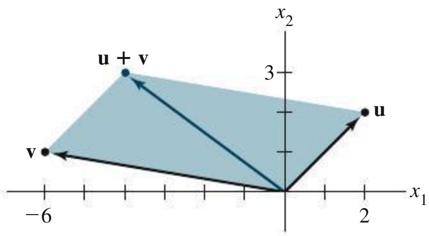
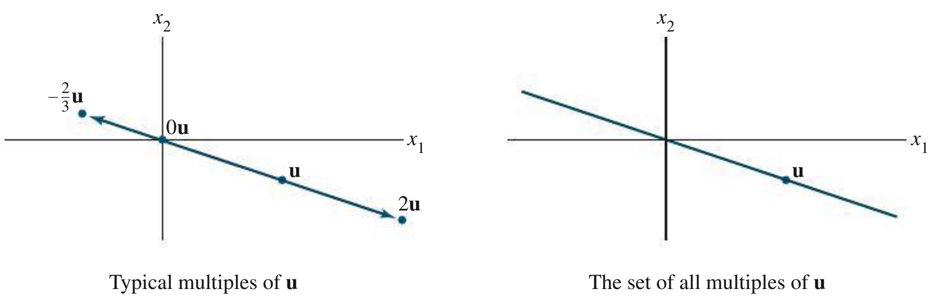


FIGURE 4

The next example illustrates the fact that the set of all scalar multiples of one fixed nonzero vector is a line through the origin, $(0, 0)$.

EXAMPLE 3 Let $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$. Display the vectors **u**, 2**u**, and $-\frac{2}{3}\mathbf{u}$ on a graph.

SOLUTION See Figure 5, where **u**, $2\mathbf{u} = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$, and $-\frac{2}{3}\mathbf{u} = \begin{bmatrix} -2 \\ 2/3 \end{bmatrix}$ are displayed. The arrow for 2**u** is twice as long as the arrow for **u**, and the arrows point in the same direction. The arrow for $-\frac{2}{3}\mathbf{u}$ is two-thirds the length of the arrow for **u**, and the arrows point in opposite directions. In general, the length of the arrow for $c\mathbf{u}$ is $|c|$ times the length of the arrow for **u**. [Recall that the length of the line segment from $(0, 0)$ to (a, b) is $\sqrt{a^2 + b^2}$.



■

FIGURE 5

Vectors in $\mathbb{R}^3$

Vectors in $\mathbb{R}^3$ are 3×1 column matrices with three entries. They are represented geometrically by points in a three-dimensional coordinate space, with arrows from the origin sometimes included for visual clarity. The vectors $\mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ and $2\mathbf{a}$ are displayed in Figure 6.

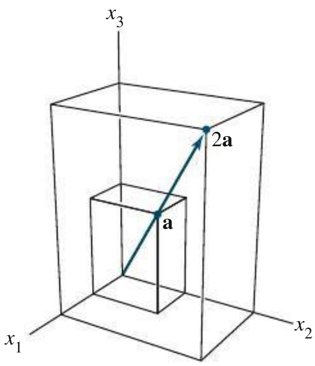


FIGURE 6
Scalar multiples.

Vectors in $\mathbb{R}^n$

If n is a positive integer, $\mathbb{R}^n$ (read "r-n") denotes the collection of all lists (or *ordered n-tuples*) of n real numbers, usually written as $n \times 1$ column matrices, such as

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

The vector whose entries are all zero is called the **zero vector** and is denoted by $\mathbf{0}$. (The number of entries in $\mathbf{0}$ will be clear from the context.)

Equality of vectors in $\mathbb{R}^n$ and the operations of scalar multiplication and vector addition in $\mathbb{R}^n$ are defined entry by entry just as in $\mathbb{R}^2$. These operations on vectors have the following properties, which can be verified directly from the corresponding properties for real numbers. See Practice Problem 1 and Exercises 41 and 42 at the end of this section.

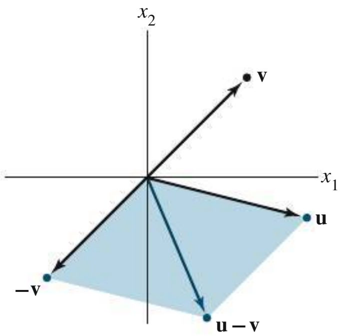


FIGURE 7
Vector subtraction.

Algebraic Properties of $\mathbb{R}^n$

For all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$ and all scalars c and d:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$,
where $-\mathbf{u}$ denotes $(-1)\mathbf{u}$

(v) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

For simplicity of notation, a vector such as $\mathbf{u} + (-1)\mathbf{v}$ is often written as $\mathbf{u} - \mathbf{v}$. Figure 7 shows $\mathbf{u} - \mathbf{v}$ as the sum of $\mathbf{u}$ and $-\mathbf{v}$.

Linear Combinations

Given vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ in $\mathbb{R}^n$ and given scalars $c_1, c_2, \dots, c_p$, the vector $\mathbf{y}$ defined by

$$\mathbf{y} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$$

is called a **linear combination** of $\mathbf{v}_1, \dots, \mathbf{v}_p$ with **weights** $c_1, \dots, c_p$. Algebraic Property (ii) above permits us to omit parentheses when forming such a linear combination. The weights in a linear combination can be any real numbers, including zero. For example, some linear combinations of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are

$$\sqrt{3}\,\mathbf{v}_1 + \mathbf{v}_2, \quad \tfrac{1}{2}\mathbf{v}_1 \ (= \tfrac{1}{2}\mathbf{v}_1 + 0\mathbf{v}_2), \quad \text{and} \quad \mathbf{0} \ (= 0\mathbf{v}_1 + 0\mathbf{v}_2)$$

EXAMPLE 4 Figure 8 identifies selected linear combinations of $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. (Note that sets of parallel grid lines are drawn through integer multiples of $\mathbf{v}_1$ and $\mathbf{v}_2$.) Estimate the linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$ that generate the vectors $\mathbf{u}$ and $\mathbf{w}$.

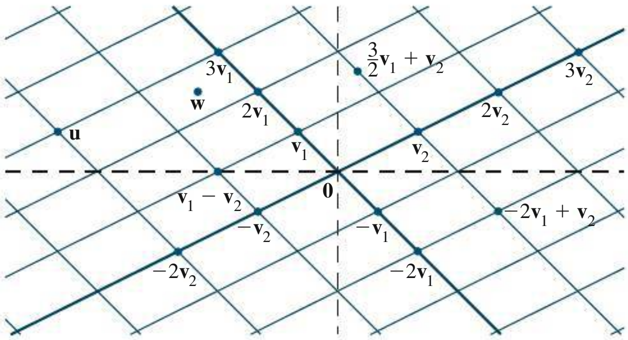


FIGURE 8 Linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.

SOLUTION The parallelogram rule shows that $\mathbf{u}$ is the sum of $3\mathbf{v}_1$ and $-2\mathbf{v}_2$; that is,

$$\mathbf{u} = 3\mathbf{v}_1 - 2\mathbf{v}_2$$

This expression for $\mathbf{u}$ can be interpreted as instructions for traveling from the origin to $\mathbf{u}$ along two straight paths. First, travel 3 units in the $\mathbf{v}_1$ direction to $3\mathbf{v}_1$, and then travel -2 units in the $\mathbf{v}_2$ direction (parallel to the line through $\mathbf{v}_2$ and $\mathbf{0}$). Next, although the vector $\mathbf{w}$ is not on a grid line, $\mathbf{w}$ appears to be about halfway between two pairs of grid lines, at the vertex of a parallelogram determined by $(5/2)\mathbf{v}_1$ and $(-1/2)\mathbf{v}_2$. (See Figure 9.) Thus a reasonable estimate for $\mathbf{w}$ is

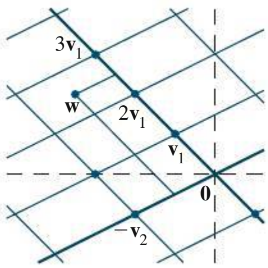


FIGURE 9

$$\mathbf{w} = \tfrac{5}{2}\mathbf{v}_1 - \tfrac{1}{2}\mathbf{v}_2$$ ■

The next example connects a problem about linear combinations to the fundamental existence question studied in Sections 1.1 and 1.2.

EXAMPLE 5 Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$. Determine whether $\mathbf{b}$ can be generated (or written) as a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. That is, determine whether weights x_1 and x_2 exist such that

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b} \tag{1}$$

If vector equation (1) has a solution, find it.

SOLUTION Use the definitions of scalar multiplication and vector addition to rewrite the vector equation

$$x_1 \underset{\substack{\uparrow \\ \mathbf{a}_1}}{\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}} + x_2 \underset{\substack{\uparrow \\ \mathbf{a}_2}}{\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}} = \underset{\substack{\uparrow \\ \mathbf{b}}}{\begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}}$$

which is the same as

$$\begin{bmatrix} x_1 \\ -2x_1 \\ -5x_1 \end{bmatrix} + \begin{bmatrix} 2x_2 \\ 5x_2 \\ 6x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

and

$$\begin{bmatrix} x_1 + 2x_2 \\ -2x_1 + 5x_2 \\ -5x_1 + 6x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix} \tag{2}$$

The vectors on the left and right sides of (2) are equal if and only if their corresponding entries are both equal. That is, x_1 and x_2 make the vector equation (1) true if and only if x_1 and x_2 satisfy the system

$$\begin{aligned} x_1 + 2x_2 &= 7 \\ -2x_1 + 5x_2 &= 4 \\ -5x_1 + 6x_2 &= -3 \end{aligned} \tag{3}$$

To solve this system, row reduce the augmented matrix of the system as follows:[3]

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 9 & 18 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

The solution of (3) is $x_1 = 3$ and $x_2 = 2$. Hence $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$, with weights $x_1 = 3$ and $x_2 = 2$. That is,

$$3\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix} \qquad \blacksquare$$

Observe in Example 5 that the original vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{b}$ are the columns of the augmented matrix that we row reduced:

$$\begin{array}{c} \begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \\ \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{b} \end{array} \end{array}$$

For brevity, write this matrix in a way that identifies its columns—namely

$$[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{b}\,] \tag{4}$$

It is clear how to write this augmented matrix immediately from vector equation (1), without going through the intermediate steps of Example 5. Take the vectors in the order in which they appear in (1) and put them into the columns of a matrix as in (4).

The discussion above is easily modified to establish the following fundamental fact.

[3] The symbol $\sim$ between matrices denotes row equivalence (Section 1.2).

A vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

has the same solution set as the linear system whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix} \tag{5}$$

In particular, $\mathbf{b}$ can be generated by a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ if and only if there exists a solution to the linear system corresponding to the matrix (5).

One of the key ideas in linear algebra is to study the set of all vectors that can be generated or written as a linear combination of a fixed set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of vectors.

DEFINITION

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are in $\mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is denoted by Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ and is called the **subset of $\mathbb{R}^n$ spanned** (or **generated**) **by $\mathbf{v}_1, \ldots, \mathbf{v}_p$**. That is, Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is the collection of all vectors that can be written in the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$$

with $c_1, \ldots, c_p$ scalars.

Asking whether a vector $\mathbf{b}$ is in Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ amounts to asking whether the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{b}$$

has a solution, or, equivalently, asking whether the linear system with augmented matrix $[\,\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_p \quad \mathbf{b}\,]$ has a solution.

Note that Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ contains every scalar multiple of $\mathbf{v}_1$ (for example), since $c\mathbf{v}_1 = c\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p$. In particular, the zero vector must be in Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

A Geometric Description of Span $\{\mathbf{v}\}$ and Span $\{\mathbf{u}, \mathbf{v}\}$

Let $\mathbf{v}$ be a nonzero vector in $\mathbb{R}^3$. Then Span $\{\mathbf{v}\}$ is the set of all scalar multiples of $\mathbf{v}$, which is the set of points on the line in $\mathbb{R}^3$ through $\mathbf{v}$ and $\mathbf{0}$. See Figure 10.

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in $\mathbb{R}^3$, with $\mathbf{v}$ not a multiple of $\mathbf{u}$, then Span $\{\mathbf{u}, \mathbf{v}\}$ is the plane in $\mathbb{R}^3$ that contains $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{0}$. In particular, Span $\{\mathbf{u}, \mathbf{v}\}$ contains the line in $\mathbb{R}^3$ through $\mathbf{u}$ and $\mathbf{0}$ and the line through $\mathbf{v}$ and $\mathbf{0}$. See Figure 11.

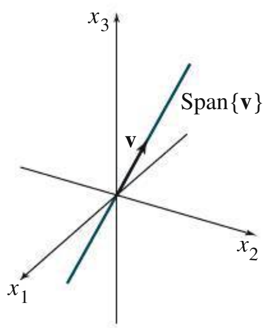


FIGURE 10 Span $\{\mathbf{v}\}$ as a line through the origin.

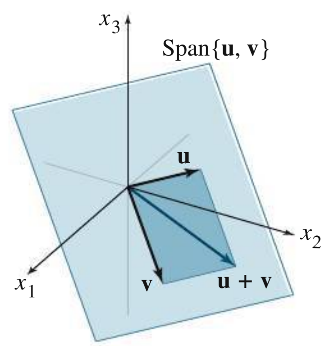


FIGURE 11 Span $\{\mathbf{u}, \mathbf{v}\}$ as a plane through the origin.

EXAMPLE 6 Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix}$. Then Span $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a plane through the origin in $\mathbb{R}^3$. Is $\mathbf{b}$ in that plane?

SOLUTION Does the equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ have a solution? To answer this, row reduce the augmented matrix $[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{b}\,]$:

$$\begin{bmatrix} 1 & 5 & -3 \\ -2 & -13 & 8 \\ 3 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & -18 & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & 0 & -2 \end{bmatrix}$$

The third equation is $0 = -2$, which shows that the system has no solution. The vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ has no solution, and so $\mathbf{b}$ is *not* in Span $\{\mathbf{a}_1, \mathbf{a}_2\}$. ■

Linear Combinations in Applications

The final example shows how scalar multiples and linear combinations can arise when a quantity such as "cost" is broken down into several categories. The basic principle for the example concerns the cost of producing several units of an item when the cost per unit is known:

$$\begin{Bmatrix} \text{number} \\ \text{of units} \end{Bmatrix} \cdot \begin{Bmatrix} \text{cost} \\ \text{per unit} \end{Bmatrix} = \begin{Bmatrix} \text{total} \\ \text{cost} \end{Bmatrix}$$

EXAMPLE 7 A company manufactures two products. For \$1.00 worth of product B, the company spends \$.45 on materials, \$.25 on labor, and \$.15 on overhead. For \$1.00 worth of product C, the company spends \$.40 on materials, \$.30 on labor, and \$.15 on overhead. Let

$$\mathbf{b} = \begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} .40 \\ .30 \\ .15 \end{bmatrix}$$

Then $\mathbf{b}$ and $\mathbf{c}$ represent the "costs per dollar of income" for the two products.

a. What economic interpretation can be given to the vector 100$\mathbf{b}$?

b. Suppose the company wishes to manufacture x_1 dollars worth of product B and x_2 dollars worth of product C. Give a vector that describes the various costs the company will have (for materials, labor, and overhead).

SOLUTION

a. Compute

$$100\mathbf{b} = 100\begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} = \begin{bmatrix} 45 \\ 25 \\ 15 \end{bmatrix}$$

The vector 100$\mathbf{b}$ lists the various costs for producing \$100 worth of product B—namely \$45 for materials, \$25 for labor, and \$15 for overhead.

b. The costs of manufacturing x_1 dollars worth of B are given by the vector $x_1\mathbf{b}$, and the costs of manufacturing x_2 dollars worth of C are given by $x_2\mathbf{c}$. Hence the total costs for both products are given by the vector $x_1\mathbf{b} + x_2\mathbf{c}$. ■

Practice Problems

1. Prove that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for any $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$.
2. For what value(s) of h will $\mathbf{y}$ be in Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

3. Let $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{u}$, and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Suppose the vectors $\mathbf{u}$ and $\mathbf{v}$ are in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. Show that $\mathbf{u} + \mathbf{v}$ is also in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. [*Hint:* The solution requires the use of the definition of the span of a set of vectors. It is useful to review this definition before starting this exercise.]

1.3 Exercises

In Exercises 1 and 2, compute $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - 2\mathbf{v}$.

1. $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -3 \\ 3 \end{bmatrix}$

2. $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

In Exercises 3 and 4, display the following vectors using arrows on an xy-graph: $\mathbf{u}$, $\mathbf{v}$, $-\mathbf{v}$, $-2\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, $\mathbf{u} - \mathbf{v}$, and $\mathbf{u} - 2\mathbf{v}$. Notice that $\mathbf{u} - \mathbf{v}$ is the vertex of a parallelogram whose other vertices are $\mathbf{u}$, $\mathbf{0}$, and $-\mathbf{v}$.

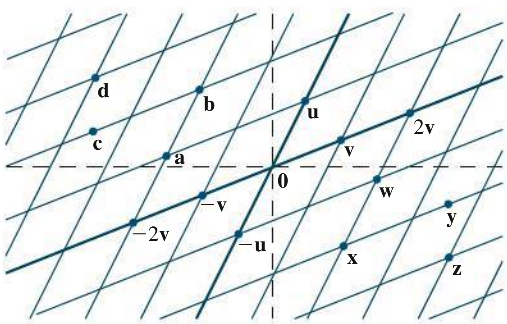
3. $\mathbf{u}$ and $\mathbf{v}$ as in Exercise 1
4. $\mathbf{u}$ and $\mathbf{v}$ as in Exercise 2

In Exercises 5 and 6, write a system of equations that is equivalent to the given vector equation.

5. $x_1 \begin{bmatrix} 6 \\ -1 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ -5 \end{bmatrix}$

6. $x_1 \begin{bmatrix} -2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 8 \\ 5 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Use the accompanying figure to write each vector listed in Exercises 7 and 8 as a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Is every vector in $\mathbb{R}^2$ a linear combination of $\mathbf{u}$ and $\mathbf{v}$?



7. Vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$
8. Vectors $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$

In Exercises 9 and 10, write a vector equation that is equivalent to the given system of equations.

9.
$$\begin{aligned} x_2 + 5x_3 &= 0 \\ 4x_1 + 6x_2 - x_3 &= 0 \\ -x_1 + 3x_2 - 8x_3 &= 0 \end{aligned}$$

10.
$$\begin{aligned} 4x_1 + x_2 + 3x_3 &= 9 \\ x_1 - 7x_2 - 2x_3 &= 2 \\ 8x_1 + 6x_2 - 5x_3 &= 15 \end{aligned}$$

In Exercises 11 and 12, determine if $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

11. $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 5 \\ -6 \\ 8 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}$

12. $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 5 \\ 5 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -5 \\ 11 \\ -7 \end{bmatrix}$

In Exercises 13 and 14, determine if $\mathbf{b}$ is a linear combination of the vectors formed from the columns of the matrix A.

13. $A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 5 \\ -2 & 8 & -4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ -7 \\ -3 \end{bmatrix}$

14. $A = \begin{bmatrix} 1 & -2 & -6 \\ 0 & 3 & 7 \\ 1 & -2 & 5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 11 \\ -5 \\ 9 \end{bmatrix}$

In Exercises 15 and 16, list five vectors in Span $\{\mathbf{v}_1, \mathbf{v}_2\}$. For each vector, show the weights on $\mathbf{v}_1$ and $\mathbf{v}_2$ used to generate the vector and list the three entries of the vector. Do not make a sketch.

15. $\mathbf{v}_1 = \begin{bmatrix} 7 \\ 1 \\ -6 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}$

16. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix}$

17. Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ h \end{bmatrix}$. For what value(s) of h is $\mathbf{b}$ in the plane spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$?

18. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 8 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} h \\ -5 \\ -3 \end{bmatrix}$. For what value(s) of h is $\mathbf{y}$ in the plane generated by $\mathbf{v}_1$ and $\mathbf{v}_2$?

19. Give a geometric description of Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the vectors $\mathbf{v}_1 = \begin{bmatrix} 8 \\ 2 \\ -6 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 12 \\ 3 \\ -9 \end{bmatrix}$.

20. Give a geometric description of Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the vectors in Exercise 16.

21. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Show that $\begin{bmatrix} h \\ k \end{bmatrix}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$ for all h and k.

22. Construct a 3×3 matrix A, with nonzero entries, and a vector $\mathbf{b}$ in $\mathbb{R}^3$ such that $\mathbf{b}$ is *not* in the set spanned by the columns of A.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** Another notation for the vector $\begin{bmatrix} -4 \\ 3 \end{bmatrix}$ is $[-4 \;\; 3]$.

24. **(T/F)** Any list of five real numbers is a vector in $\mathbb{R}^5$.

25. **(T/F)** The points in the plane corresponding to $\begin{bmatrix} -2 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} -5 \\ 2 \end{bmatrix}$ lie on a line through the origin.

26. **(T/F)** The vector $\mathbf{u}$ results when a vector $\mathbf{u} - \mathbf{v}$ is added to the vector $\mathbf{v}$.

27. **(T/F)** An example of a linear combination of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is the vector $\frac{1}{2}\mathbf{v}_1$.

28. **(T/F)** The weights $c_1, \ldots, c_p$ in a linear combination $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ cannot all be zero.

29. **(T/F)** The solution set of the linear system whose augmented matrix is $[\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{b} \,]$ is the same as the solution set of the equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}$.

30. **(T/F)** When $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, Span $\{\mathbf{u}, \mathbf{v}\}$ contains the line through $\mathbf{u}$ and the origin.

31. **(T/F)** The set Span $\{\mathbf{u}, \mathbf{v}\}$ is always visualized as a plane through the origin.

32. **(T/F)** Asking whether the linear system corresponding to an augmented matrix $[\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{b} \,]$ has a solution amounts to asking whether $\mathbf{b}$ is in Span $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

33. Let $A = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 3 & -2 \\ -2 & 6 & 3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ -4 \end{bmatrix}$. Denote the columns of A by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, and let $W = \text{Span}\,\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

a. Is $\mathbf{b}$ in $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$? How many vectors are in $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$?

b. Is $\mathbf{b}$ in W? How many vectors are in W?

c. Show that $\mathbf{a}_1$ is in W. [*Hint:* Row operations are unnecessary.]

34. Let $A = \begin{bmatrix} 2 & 0 & 6 \\ -1 & 8 & 5 \\ 1 & -2 & 1 \end{bmatrix}$, let $\mathbf{b} = \begin{bmatrix} 10 \\ 3 \\ 3 \end{bmatrix}$, and let W be the set of all linear combinations of the columns of A.

a. Is $\mathbf{b}$ in W?

b. Show that the third column of A is in W.

35. A mining company has two mines. One day's operation at mine 1 produces ore that contains 20 metric tons of copper and 550 kilograms of silver, while one day's operation at mine 2 produces ore that contains 30 metric tons of copper and 500 kilograms of silver. Let $\mathbf{v}_1 = \begin{bmatrix} 20 \\ 550 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 30 \\ 500 \end{bmatrix}$. Then $\mathbf{v}_1$ and $\mathbf{v}_2$ represent the "output per day" of mine 1 and mine 2, respectively.

a. What physical interpretation can be given to the vector $5\mathbf{v}_1$?

b. Suppose the company operates mine 1 for x_1 days and mine 2 for x_2 days. Write a vector equation whose solution gives the number of days each mine should operate in order to produce 150 tons of copper and 2825 kilograms of silver. Do not solve the equation.

T c. Solve the equation in (b).

36. A steam plant burns two types of coal: anthracite (A) and bituminous (B). For each ton of A burned, the plant produces 27.6 million Btu of heat, 3100 grams (g) of sulfur dioxide, and 250 g of particulate matter (solid-particle pollutants). For each ton of B burned, the plant produces 30.2 million Btu, 6400 g of sulfur dioxide, and 360 g of particulate matter.

a. How much heat does the steam plant produce when it burns x_1 tons of A and x_2 tons of B?

b. Suppose the output of the steam plant is described by a vector that lists the amounts of heat, sulfur dioxide, and particulate matter. Express this output as a linear combination of two vectors, assuming that the plant burns x_1 tons of A and x_2 tons of B.

T c. Over a certain time period, the steam plant produced 162 million Btu of heat, 23,610 g of sulfur dioxide, and 1623 g of particulate matter. Determine how many tons of each type of coal the steam plant must have burned. Include a vector equation as part of your solution.

37. Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be points in $\mathbb{R}^3$ and suppose that for $j = 1, \dots, k$ an object with mass m_j is located at point $\mathbf{v}_j$. Physicists call such objects *point masses*. The total mass of the system of point masses is

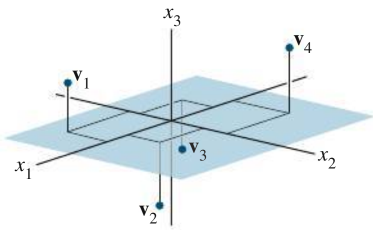
$$m = m_1 + \cdots + m_k$$

The *center of mass* (or *center of gravity*) of the system is

$$\overline{\mathbf{v}} = \frac{1}{m}[m_1\mathbf{v}_1 + \cdots + m_k\mathbf{v}_k]$$

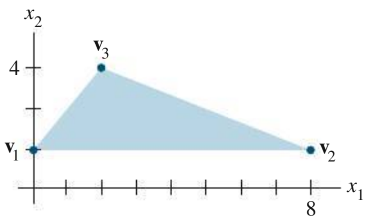
Compute the center of gravity of the system consisting of the following point masses (see the figure):

Point	Mass
$\mathbf{v}_1 = (5, -4, 3)$	2 g
$\mathbf{v}_2 = (4, 3, -2)$	5 g
$\mathbf{v}_3 = (-4, -3, -1)$	2 g
$\mathbf{v}_4 = (-9, 8, 6)$	1 g



38. Let $\mathbf{v}$ be the center of mass of a system of point masses located at $\mathbf{v}_1, \dots, \mathbf{v}_k$ as in Exercise 37. Is $\mathbf{v}$ in Span $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$? Explain.

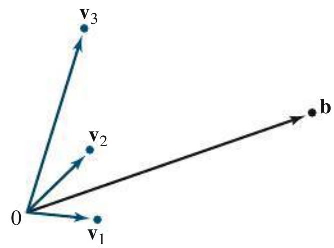
39. A thin triangular plate of uniform density and thickness has vertices at $\mathbf{v}_1 = (0, 1)$, $\mathbf{v}_2 = (8, 1)$, and $\mathbf{v}_3 = (2, 4)$, as in the figure below, and the mass of the plate is 3 g.



a. Find the (x, y)-coordinates of the center of mass of the plate. This "balance point" of the plate coincides with the center of mass of a system consisting of three 1-gram point masses located at the vertices of the plate.

b. Determine how to distribute an additional mass of 6 g at the three vertices of the plate to move the balance point of the plate to $(2, 2)$. [*Hint:* Let w_1, w_2, and w_3 denote the masses added at the three vertices, so that $w_1 + w_2 + w_3 = 6$.]

40. Consider the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{b}$ in $\mathbb{R}^2$, shown in the figure. Does the equation $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}$ have a solution? Is the solution unique? Use the figure to explain your answers.



41. Use the vectors $\mathbf{u} = (u_1, \dots, u_n)$, $\mathbf{v} = (v_1, \dots, v_n)$, and $\mathbf{w} = (w_1, \dots, w_n)$ to verify the following algebraic properties of $\mathbb{R}^n$.

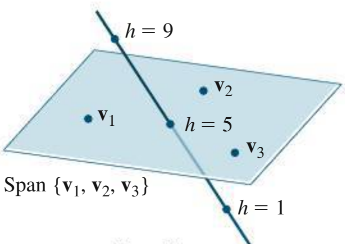
a. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

b. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for each scalar c

42. Use the vector $\mathbf{u} = (u_1, \dots, u_n)$ to verify the following algebraic properties of $\mathbb{R}^n$.

a. $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0}$

b. $c(d\mathbf{u}) = (cd)\mathbf{u}$ for all scalars c and d



The points $\begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$ lie on a line that intersects the plane when $h = 5$.

Solutions to Practice Problems

1. Take arbitrary vectors $\mathbf{u} = (u_1, \dots, u_n)$ and $\mathbf{v} = (v_1, \dots, v_n)$ in $\mathbb{R}^n$, and compute

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1 + v_1, \dots, u_n + v_n) && \text{Definition of vector addition} \\ &= (v_1 + u_1, \dots, v_n + u_n) && \text{Commutativity of addition in } \mathbb{R} \\ &= \mathbf{v} + \mathbf{u} && \text{Definition of vector addition} \end{aligned}$$

2. The vector $\mathbf{y}$ belongs to Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if there exist scalars x_1, x_2, x_3 such that

$$x_1\begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} + x_2\begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix} + x_3\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

This vector equation is equivalent to a system of three linear equations in three unknowns. If you row reduce the augmented matrix for this system, you find that

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

The system is consistent if and only if there is no pivot in the fourth column. That is, $h - 5$ must be 0. So $\mathbf{y}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if $h = 5$.

Remember: The presence of a free variable in a system does not guarantee that the system is consistent.

3. Since the vectors $\mathbf{u}$ and $\mathbf{v}$ are in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, there exist scalars c_1, c_2, c_3 and d_1, d_2, d_3 such that

$$\mathbf{u} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 \quad \text{and} \quad \mathbf{v} = d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + d_3\mathbf{w}_3.$$

Notice

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 + d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + d_3\mathbf{w}_3 \\ &= (c_1 + d_1)\mathbf{w}_1 + (c_2 + d_2)\mathbf{w}_2 + (c_3 + d_3)\mathbf{w}_3 \end{aligned}$$

Since $c_1 + d_1, c_2 + d_2$, and $c_3 + d_3$ are also scalars, the vector $\mathbf{u} + \mathbf{v}$ is in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

1.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$

A fundamental idea in linear algebra is to view a linear combination of vectors as the product of a matrix and a vector. The following definition permits us to rephrase some of the concepts of Section 1.3 in new ways.

DEFINITION

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and if $\mathbf{x}$ is in $\mathbb{R}^n$, then the **product of A and $\mathbf{x}$**, denoted by $A\mathbf{x}$, is **the linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights**; that is,

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

Note that $A\mathbf{x}$ is defined only if the number of columns of A equals the number of entries in $\mathbf{x}$.

EXAMPLE 1

a. $$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 2 \\ -5 \end{bmatrix} + 7\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \begin{bmatrix} 6 \\ -15 \end{bmatrix} + \begin{bmatrix} -7 \\ 21 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

b. $$\begin{bmatrix} 2 & -3 \\ 8 & 0 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = 4\begin{bmatrix} 2 \\ 8 \\ -5 \end{bmatrix} + 7\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ -20 \end{bmatrix} + \begin{bmatrix} -21 \\ 0 \\ 14 \end{bmatrix} = \begin{bmatrix} -13 \\ 32 \\ -6 \end{bmatrix}$$ ■

EXAMPLE 2 For $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in $\mathbb{R}^m$, write the linear combination $3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3$ as a matrix times a vector.

SOLUTION Place $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ into the columns of a matrix A and place the weights 3, -5, and 7 into a vector $\mathbf{x}$. That is,

$$3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix} = A\mathbf{x}$$

■

Section 1.3 showed how to write a system of linear equations as a vector equation involving a linear combination of vectors. For example, the system

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

is equivalent to

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -5 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{2}$$

As in Example 2, the linear combination on the left side is a matrix times a vector, so that (2) becomes

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{3}$$

Equation (3) has the form $A\mathbf{x} = \mathbf{b}$. Such an equation is called a **matrix equation**, to distinguish it from a vector equation such as is shown in (2).

Notice how the matrix in (3) is just the matrix of coefficients of the system (1). Similar calculations show that any system of linear equations, or any vector equation such as (2), can be written as an equivalent matrix equation in the form $A\mathbf{x} = \mathbf{b}$. This simple observation will be used repeatedly throughout the text.

Here is the formal result.

THEOREM 3

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and if $\mathbf{b}$ is in $\mathbb{R}^m$, the matrix equation

$$A\mathbf{x} = \mathbf{b} \tag{4}$$

has the same solution set as the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b} \tag{5}$$

which, in turn, has the same solution set as the system of linear equations whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix} \tag{6}$$

Theorem 3 provides a powerful tool for gaining insight into problems in linear algebra, because a system of linear equations may now be viewed in three different but equivalent ways: as a matrix equation, as a vector equation, or as a system of linear equations. Whenever you construct a mathematical model of a problem in real life, you are free to choose whichever viewpoint is most natural. Then you may switch from one formulation of a problem to another whenever it is convenient. In any case, the matrix equation (4), the vector equation (5), and the system of equations are all solved in the same way—by row reducing the augmented matrix (6). Other methods of solution will be discussed later.

Existence of Solutions

The definition of $A\mathbf{x}$ leads directly to the following useful fact.

> The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Section 1.3 considered the existence question, "Is $\mathbf{b}$ in Span $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$?" Equivalently, "Is $A\mathbf{x} = \mathbf{b}$ consistent?" A harder existence problem is to determine whether the equation $A\mathbf{x} = \mathbf{b}$ is consistent *for all* possible $\mathbf{b}$.

EXAMPLE 3 Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent for all possible b_1, b_2, b_3?

SOLUTION Row reduce the augmented matrix for $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 7 & 5 & b_3 + 3b_1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) \end{bmatrix}$$

The third entry in column 4 equals $b_1 - \frac{1}{2}b_2 + b_3$. The equation $A\mathbf{x} = \mathbf{b}$ is *not* consistent for every $\mathbf{b}$ because some choices of $\mathbf{b}$ can make $b_1 - \frac{1}{2}b_2 + b_3$ nonzero. ■

The reduced matrix in Example 3 provides a description of all $\mathbf{b}$ for which the equation $A\mathbf{x} = \mathbf{b}$ *is* consistent: The entries in $\mathbf{b}$ must satisfy

$$b_1 - \tfrac{1}{2}b_2 + b_3 = 0$$

This is the equation of a plane through the origin in $\mathbb{R}^3$. The plane is the set of all linear combinations of the three columns of A. See Figure 1.

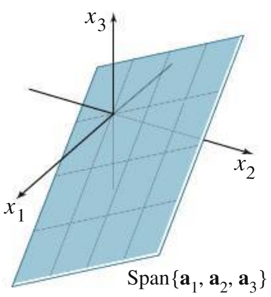


FIGURE 1
The columns of $A = [\,\mathbf{a}_1 \ \ \mathbf{a}_2 \ \ \mathbf{a}_3\,]$ span a plane through $\mathbf{0}$.

The equation $A\mathbf{x} = \mathbf{b}$ in Example 3 fails to be consistent for all $\mathbf{b}$ because the echelon form of A has a row of zeros. If A had a pivot in all three rows, we would not care about the calculations in the augmented column because in this case an echelon form of the augmented matrix could not have a row such as $[\,0 \ \ 0 \ \ 0 \ \ 1\,]$.

In the next theorem, the sentence "The columns of A span $\mathbb{R}^m$" means that *every* $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A. In general, a set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^m$ **spans** (or **generates**) $\mathbb{R}^m$ if every vector in $\mathbb{R}^m$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_p$—that is, if Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\} = \mathbb{R}^m$.

THEOREM 4

Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true statements or they are all false.

a. For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.
b. Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.
c. The columns of A span $\mathbb{R}^m$.
d. A has a pivot position in every row.

Theorem 4 is one of the most useful theorems in this chapter. Statements (a), (b), and (c) are equivalent because of the definition of $A\mathbf{x}$ and what it means for a set of vectors to span $\mathbb{R}^m$. The discussion after Example 3 suggests why (a) and (d) are equivalent; a proof is given at the end of the section. The exercises will provide examples of how Theorem 4 is used.

Warning: Theorem 4 is about a *coefficient matrix*, not an augmented matrix. If an augmented matrix $[\,A \quad \mathbf{b}\,]$ has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ may or may not be consistent.

Computation of $A\mathbf{x}$

The calculations in Example 1 were based on the definition of the product of a matrix A and a vector $\mathbf{x}$. The following simple example will lead to a more efficient method for calculating the entries in $A\mathbf{x}$ when working problems by hand.

EXAMPLE 4 Compute $A\mathbf{x}$, where $A = \begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

SOLUTION From the definition,

$$\begin{aligned}
\begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= x_1\begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} + x_2\begin{bmatrix} 3 \\ 5 \\ -2 \end{bmatrix} + x_3\begin{bmatrix} 4 \\ -3 \\ 8 \end{bmatrix} \\
&= \begin{bmatrix} 2x_1 \\ -x_1 \\ 6x_1 \end{bmatrix} + \begin{bmatrix} 3x_2 \\ 5x_2 \\ -2x_2 \end{bmatrix} + \begin{bmatrix} 4x_3 \\ -3x_3 \\ 8x_3 \end{bmatrix} \qquad (7) \\
&= \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ -x_1 + 5x_2 - 3x_3 \\ 6x_1 - 2x_2 + 8x_3 \end{bmatrix}
\end{aligned}$$

The first entry in the product $A\mathbf{x}$ is a sum of products (sometimes called a *dot product*), using the first row of A and the entries in $\mathbf{x}$. That is,

$$\begin{bmatrix} 2 & 3 & 4 \\ & & \\ & & \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ \\ \\ \end{bmatrix}$$

This matrix shows how to compute the first entry in $A\mathbf{x}$ directly, without writing down all the calculations shown in (7). Similarly, the second entry in $A\mathbf{x}$ can be calculated at once by multiplying the entries in the second row of A by the corresponding entries in $\mathbf{x}$ and then summing the resulting products:

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

Likewise, the third entry in $A\mathbf{x}$ can be calculated from the third row of A and the entries in $\mathbf{x}$. ■

Row–Vector Rule for Computing $A\mathbf{x}$

If the product $A\mathbf{x}$ is defined, then the ith entry in $A\mathbf{x}$ is the sum of the products of corresponding entries from row i of A and from the vector $\mathbf{x}$.

EXAMPLE 5

a. $\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 1\cdot 4 + 2\cdot 3 + (-1)\cdot 7 \\ 0\cdot 4 + (-5)\cdot 3 + 3\cdot 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$

b. $\begin{bmatrix} 2 & -3 \\ 8 & 0 \\ -5 & 2 \end{bmatrix}\begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} 2\cdot 4 + (-3)\cdot 7 \\ 8\cdot 4 + 0\cdot 7 \\ (-5)\cdot 4 + 2\cdot 7 \end{bmatrix} = \begin{bmatrix} -13 \\ 32 \\ -6 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r \\ s \\ t \end{bmatrix} = \begin{bmatrix} 1\cdot r + 0\cdot s + 0\cdot t \\ 0\cdot r + 1\cdot s + 0\cdot t \\ 0\cdot r + 0\cdot s + 1\cdot t \end{bmatrix} = \begin{bmatrix} r \\ s \\ t \end{bmatrix}$ ■

By definition, the matrix in Example 5(c) with 1's on the diagonal and 0's elsewhere is called an **identity matrix** and is denoted by I. The calculation in part (c) shows that $I\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^3$. There is an analogous $n \times n$ identity matrix, sometimes written as I_n. As in part (c), $I_n\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^n$.

Properties of the Matrix–Vector Product $A\mathbf{x}$

The facts in the next theorem are important and will be used throughout the text. The proof relies on the definition of $A\mathbf{x}$ and the algebraic properties of $\mathbb{R}^n$.

THEOREM 5

If A is an $m \times n$ matrix, $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, and c is a scalar, then:

a. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$;

b. $A(c\mathbf{u}) = c(A\mathbf{u})$.

PROOF For simplicity, take $n = 3$, $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$, and $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^3$. (The proof of the general case is similar.) For $i = 1, 2, 3$, let u_i and v_i be the ith entries in $\mathbf{u}$ and $\mathbf{v}$, respectively. To prove statement (a), compute $A(\mathbf{u} + \mathbf{v})$ as a linear combination of the columns of A using the entries in $\mathbf{u} + \mathbf{v}$ as weights.

$$\begin{aligned} A(\mathbf{u} + \mathbf{v}) &= [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix} \\ &= \underbrace{(u_1 + v_1)}_{\text{Entries in } \mathbf{u}+\mathbf{v}}\mathbf{a}_1 + (u_2 + v_2)\mathbf{a}_2 + (u_3 + v_3)\underbrace{\mathbf{a}_3}_{\text{Columns of } A} \\ &= (u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3) + (v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3\mathbf{a}_3) \\ &= A\mathbf{u} + A\mathbf{v} \end{aligned}$$

To prove statement (b), compute $A(c\mathbf{u})$ as a linear combination of the columns of A using the entries in $c\mathbf{u}$ as weights.

$$\begin{aligned} A(c\mathbf{u}) &= [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]\begin{bmatrix} cu_1 \\ cu_2 \\ cu_3 \end{bmatrix} = (cu_1)\mathbf{a}_1 + (cu_2)\mathbf{a}_2 + (cu_3)\mathbf{a}_3 \\ &= c(u_1\mathbf{a}_1) + c(u_2\mathbf{a}_2) + c(u_3\mathbf{a}_3) \\ &= c(u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3) \\ &= c(A\mathbf{u}) \end{aligned}$$ ■

Numerical Note

To optimize a computer algorithm to compute $A\mathbf{x}$, the sequence of calculations should involve data stored in contiguous memory locations. The most widely used professional algorithms for matrix computations are written in Fortran, a language that stores a matrix as a set of columns. Such algorithms compute $A\mathbf{x}$ as a linear combination of the columns of A. In contrast, if a program is written in the popular language C, which stores matrices by rows, $A\mathbf{x}$ should be computed via the alternative rule that uses the rows of A.

PROOF OF THEOREM 4 As was pointed out after Theorem 4, statements (a), (b), and (c) are logically equivalent. So, it suffices to show (for an arbitrary matrix A) that (a) and (d) are either both true or both false. This will tie all four statements together.

Let U be an echelon form of A. Given $\mathbf{b}$ in $\mathbb{R}^m$, we can row reduce the augmented matrix $[\,A \quad \mathbf{b}\,]$ to an augmented matrix $[\,U \quad \mathbf{d}\,]$ for some $\mathbf{d}$ in $\mathbb{R}^m$:

$$[\,A \quad \mathbf{b}\,] \sim \cdots \sim [\,U \quad \mathbf{d}\,]$$

If statement (d) is true, then each row of U contains a pivot position and there can be no pivot in the augmented column. So $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b}$, and (a) is true. If (d) is false, the last row of U is all zeros. Let $\mathbf{d}$ be any vector with a 1 in its last entry. Then $[\,U \quad \mathbf{d}\,]$ represents an *inconsistent* system. Since row operations are reversible, $[\,U \quad \mathbf{d}\,]$ can be transformed into the form $[\,A \quad \mathbf{b}\,]$. The new system $A\mathbf{x} = \mathbf{b}$ is also inconsistent, and (a) is false. ■

Practice Problems

1. Let $A = \begin{bmatrix} 1 & 5 & -2 & 0 \\ -3 & 1 & 9 & -5 \\ 4 & -8 & -1 & 7 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ -4 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -7 \\ 9 \\ 0 \end{bmatrix}$. It can be shown that $\mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$. Use this fact to exhibit $\mathbf{b}$ as a specific linear combination of the columns of A.
2. Let $A = \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$. Verify Theorem 5(a) in this case by computing $A(\mathbf{u} + \mathbf{v})$ and $A\mathbf{u} + A\mathbf{v}$.
3. Construct a 3×3 matrix A and vectors $\mathbf{b}$ and $\mathbf{c}$ in $\mathbb{R}^3$ so that $A\mathbf{x} = \mathbf{b}$ has a solution, but $A\mathbf{x} = \mathbf{c}$ does not.

1.4 Exercises

Compute the products in Exercises 1–4 using (a) the definition, as in Example 1, and (b) the row–vector rule for computing $A\mathbf{x}$. If a product is undefined, explain why.

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

2. $\begin{bmatrix} 2 \\ 6 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

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

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

In Exercises 5–8, use the definition of $A\mathbf{x}$ to write the matrix equation as a vector equation, or vice versa.

5. $\begin{bmatrix} 5 & 1 & -8 & 4 \\ -2 & -7 & 3 & -5 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -8 \\ 16 \end{bmatrix}$

6. $\begin{bmatrix} 7 & -3 \\ 2 & 1 \\ 9 & -6 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ 12 \\ -4 \end{bmatrix}$

7. $x_1\begin{bmatrix} 4 \\ -1 \\ 7 \\ -4 \end{bmatrix} + x_2\begin{bmatrix} -5 \\ 3 \\ -5 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} 7 \\ -8 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ -8 \\ 0 \\ -7 \end{bmatrix}$

8. $z_1\begin{bmatrix} 4 \\ -2 \end{bmatrix} + z_2\begin{bmatrix} -4 \\ 5 \end{bmatrix} + z_3\begin{bmatrix} -5 \\ 4 \end{bmatrix} + z_4\begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 13 \end{bmatrix}$

In Exercises 9 and 10, write the system first as a vector equation and then as a matrix equation.

9. $\begin{aligned} 3x_1 + x_2 - 5x_3 &= 9 \\ x_2 + 4x_3 &= 0 \end{aligned}$

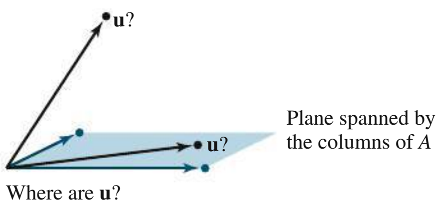
10. $\begin{aligned} 8x_1 - x_2 &= 4 \\ 5x_1 + 4x_2 &= 1 \\ x_1 - 3x_2 &= 2 \end{aligned}$

Given A and $\mathbf{b}$ in Exercises 11 and 12, write the augmented matrix for the linear system that corresponds to the matrix equation $A\mathbf{x} = \mathbf{b}$. Then solve the system and write the solution as a vector.

11. $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 5 \\ -2 & -4 & -3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -2 \\ 2 \\ 9 \end{bmatrix}$

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

13. Let $\mathbf{u} = \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}$ and $A = \begin{bmatrix} 3 & -5 \\ -2 & 6 \\ 1 & 1 \end{bmatrix}$. Is $\mathbf{u}$ in the plane in $\mathbb{R}^3$ spanned by the columns of A? (See the figure.) Why or why not?



14. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -3 \\ 2 \end{bmatrix}$ and $A = \begin{bmatrix} 5 & 8 & 7 \\ 0 & 1 & -1 \\ 1 & 3 & 0 \end{bmatrix}$. Is $\mathbf{u}$ in the subset of $\mathbb{R}^3$ spanned by the columns of A? Why or why not?

15. Let $A = \begin{bmatrix} 2 & -1 \\ -6 & 3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. Show that the equation $A\mathbf{x} = \mathbf{b}$ does not have a solution for all possible $\mathbf{b}$, and describe the set of all $\mathbf{b}$ for which $A\mathbf{x} = \mathbf{b}$ *does* have a solution.

16. Repeat Exercise 15: $A = \begin{bmatrix} 1 & -3 & -4 \\ -3 & 2 & 6 \\ 5 & -1 & -8 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$.

Exercises 17–20 refer to the matrices A and B below. Make appropriate calculations that justify your answers and mention an appropriate theorem.

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

17. How many rows of A contain a pivot position? Does the equation $A\mathbf{x} = \mathbf{b}$ have a solution for each $\mathbf{b}$ in $\mathbb{R}^4$?

18. Do the columns of B span $\mathbb{R}^4$? Does the equation $B\mathbf{x} = \mathbf{y}$ have a solution for each $\mathbf{y}$ in $\mathbb{R}^4$?

19. Can each vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix A above? Do the columns of A span $\mathbb{R}^4$?

20. Can every vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix B above? Do the columns of B span $\mathbb{R}^3$?

21. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$.

Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^4$? Why or why not?

22. Let $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ -3 \\ 8 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}$.

Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^3$? Why or why not?

In Exercises 23–34, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** The equation $A\mathbf{x} = \mathbf{b}$ is referred to as a *vector equation*.

24. **(T/F)** Every matrix equation $A\mathbf{x} = \mathbf{b}$ corresponds to a vector equation with the same solution set.

25. **(T/F)** If the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent, then $\mathbf{b}$ is not in the set spanned by the columns of A.

26. **(T/F)** A vector $\mathbf{b}$ is a linear combination of the columns of a matrix A if and only if the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution.

27. **(T/F)** The equation $A\mathbf{x} = \mathbf{b}$ is consistent if the augmented matrix $[\ A \ \ \mathbf{b}\ ]$ has a pivot position in every row.

28. **(T/F)** If A is an $m \times n$ matrix whose columns do not span $\mathbb{R}^m$, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for some $\mathbf{b}$ in $\mathbb{R}^m$.

29. **(T/F)** The first entry in the product $A\mathbf{x}$ is a sum of products.

30. **(T/F)** Any linear combination vectors can always be written in the form $A\mathbf{x}$ for a suitable matrix A and vector $\mathbf{x}$.

31. **(T/F)** If the columns of an $m \times n$ matrix A span $\mathbb{R}^m$, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent for each $\mathbf{b}$ in $\mathbb{R}^m$.

32. **(T/F)** The solution set of a linear system whose augmented matrix is $[\ \mathbf{a}_1 \ \ \mathbf{a}_2 \ \ \mathbf{a}_3 \ \ \mathbf{b}\ ]$ is the same as the solution set of $A\mathbf{x} = \mathbf{b}$, if $A = [\ \mathbf{a}_1 \ \ \mathbf{a}_2 \ \ \mathbf{a}_3\ ]$.

33. **(T/F)** If A is an $m \times n$ matrix and if the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for some $\mathbf{b}$ in $\mathbb{R}^m$, then A cannot have a pivot position in every row.

34. **(T/F)** If the augmented matrix $[\,A \quad \mathbf{b}\,]$ has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent.

35. Note that $\begin{bmatrix} 4 & -3 & 1 \\ 5 & -2 & 5 \\ -6 & 2 & -3 \end{bmatrix}\begin{bmatrix} -3 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -7 \\ -3 \\ 10 \end{bmatrix}$. Use this fact (and no row operations) to find scalars c_1, c_2, c_3 such that $\begin{bmatrix} -7 \\ -3 \\ 10 \end{bmatrix} = c_1\begin{bmatrix} 4 \\ 5 \\ -6 \end{bmatrix} + c_2\begin{bmatrix} -3 \\ -2 \\ 2 \end{bmatrix} + c_3\begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}$.

36. Let $\mathbf{u} = \begin{bmatrix} 7 \\ 2 \\ 5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}$.

It can be shown that $3\mathbf{u} - 5\mathbf{v} - \mathbf{w} = \mathbf{0}$. Use this fact (and no row operations) to find x_1 and x_2 that satisfy the equation $\begin{bmatrix} 7 & 3 \\ 2 & 1 \\ 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}$.

37. Let $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, and $\mathbf{v}$ represent vectors in $\mathbb{R}^5$, and let x_1, x_2, and x_3 denote scalars. Write the following vector equation as a matrix equation. Identify any symbols you choose to use.

$$x_1\mathbf{q}_1 + x_2\mathbf{q}_2 + x_3\mathbf{q}_3 = \mathbf{v}$$

38. Rewrite the (numerical) matrix equation below in symbolic form as a vector equation, using symbols $\mathbf{v}_1, \mathbf{v}_2, \ldots$ for the vectors and $c_1, c_2, \ldots$ for scalars. Define what each symbol represents, using the data given in the matrix equation.

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

39. Construct a 3×3 matrix, not in echelon form, whose columns span $\mathbb{R}^3$. Show that the matrix you construct has the desired property.

40. Construct a 3×3 matrix, not in echelon form, whose columns do *not* span $\mathbb{R}^3$. Show that the matrix you construct has the desired property.

STUDY GUIDE offers additional resources for mastering the concept of span.

41. Let A be a 3×2 matrix. Explain why the equation $A\mathbf{x} = \mathbf{b}$ cannot be consistent for all $\mathbf{b}$ in $\mathbb{R}^3$. Generalize your argument to the case of an arbitrary A with more rows than columns.

42. Could a set of three vectors in $\mathbb{R}^4$ span all of $\mathbb{R}^4$? Explain. What about n vectors in $\mathbb{R}^m$ when n is less than m?

43. Suppose A is a 4×3 matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^4$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. What can you say about the reduced echelon form of A? Justify your answer.

44. Suppose A is a 3×3 matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^3$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. Explain why the columns of A must span $\mathbb{R}^3$.

45. Let A be a 3×4 matrix, let $\mathbf{y}_1$ and $\mathbf{y}_2$ be vectors in $\mathbb{R}^3$, and let $\mathbf{w} = \mathbf{y}_1 + \mathbf{y}_2$. Suppose $\mathbf{y}_1 = A\mathbf{x}_1$ and $\mathbf{y}_2 = A\mathbf{x}_2$ for some vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^4$. What fact allows you to conclude that the system $A\mathbf{x} = \mathbf{w}$ is consistent? (*Note:* $\mathbf{x}_1$ and $\mathbf{x}_2$ denote vectors, not scalar entries in vectors.)

46. Let A be a 5×3 matrix, let $\mathbf{y}$ be a vector in $\mathbb{R}^3$, and let $\mathbf{z}$ be a vector in $\mathbb{R}^5$. Suppose $A\mathbf{y} = \mathbf{z}$. What fact allows you to conclude that the system $A\mathbf{x} = 4\mathbf{z}$ is consistent?

T In Exercises 47–50, determine if the columns of the matrix span $\mathbb{R}^4$.

47. $\begin{bmatrix} 7 & 2 & -5 & 8 \\ -5 & -3 & 4 & -9 \\ 6 & 10 & -2 & 7 \\ -7 & 9 & 2 & 15 \end{bmatrix}$

48. $\begin{bmatrix} 5 & -7 & -4 & 9 \\ 6 & -8 & -7 & 5 \\ 4 & -4 & -9 & -9 \\ -9 & 11 & 16 & 7 \end{bmatrix}$

49. $\begin{bmatrix} 12 & -7 & 11 & -9 & 5 \\ -9 & 4 & -8 & 7 & -3 \\ -6 & 11 & -7 & 3 & -9 \\ 4 & -6 & 10 & -5 & 12 \end{bmatrix}$

50. $\begin{bmatrix} 8 & 11 & -6 & -7 & 13 \\ -7 & -8 & 5 & 6 & -9 \\ 11 & 7 & -7 & -9 & -6 \\ -3 & 4 & 1 & 8 & 7 \end{bmatrix}$

T 51. Find a column of the matrix in Exercise 49 that can be deleted and yet have the remaining matrix columns still span $\mathbb{R}^4$.

T 52. Find a column of the matrix in Exercise 50 that can be deleted and yet have the remaining matrix columns still span $\mathbb{R}^4$. Can you delete more than one column?

Solutions to Practice Problems

1. The matrix equation

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

is equivalent to the vector equation

$$3\begin{bmatrix}1\\-3\\4\end{bmatrix}-2\begin{bmatrix}5\\1\\-8\end{bmatrix}+0\begin{bmatrix}-2\\9\\-1\end{bmatrix}-4\begin{bmatrix}0\\-5\\7\end{bmatrix}=\begin{bmatrix}-7\\9\\0\end{bmatrix},$$

which expresses **b** as a linear combination of the columns of A.

2. $\mathbf{u}+\mathbf{v}=\begin{bmatrix}4\\-1\end{bmatrix}+\begin{bmatrix}-3\\5\end{bmatrix}=\begin{bmatrix}1\\4\end{bmatrix}$

$$A(\mathbf{u}+\mathbf{v})=\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}=\begin{bmatrix}2+20\\3+4\end{bmatrix}=\begin{bmatrix}22\\7\end{bmatrix}$$

$$\begin{aligned}A\mathbf{u}+A\mathbf{v}&=\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}4\\-1\end{bmatrix}+\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}-3\\5\end{bmatrix}\\&=\begin{bmatrix}3\\11\end{bmatrix}+\begin{bmatrix}19\\-4\end{bmatrix}=\begin{bmatrix}22\\7\end{bmatrix}\end{aligned}$$

Remark: There are, in fact, infinitely many correct solutions to Practice Problem 3. When creating matrices to satisfy specified criteria, it is often useful to create matrices that are straightforward, such as those already in reduced echelon form. Here is one possible solution:

3. Let

$$A=\begin{bmatrix}1&0&1\\0&1&1\\0&0&0\end{bmatrix},\ \mathbf{b}=\begin{bmatrix}3\\2\\0\end{bmatrix},\ \text{and } \mathbf{c}=\begin{bmatrix}3\\2\\1\end{bmatrix}.$$

Notice the reduced echelon form of the augmented matrix corresponding to $A\mathbf{x}=\mathbf{b}$ is

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

which corresponds to a consistent system, and hence $A\mathbf{x}=\mathbf{b}$ has solutions. The reduced echelon form of the augmented matrix corresponding to $A\mathbf{x}=\mathbf{c}$ is

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

which corresponds to an inconsistent system, and hence $A\mathbf{x}=\mathbf{c}$ does not have any solutions.

1.5 Solution Sets of Linear Systems

Solution sets of linear systems are important objects of study in linear algebra. They will appear later in several different contexts. This section uses vector notation to give explicit and geometric descriptions of such solution sets.

Homogeneous Linear Systems

A system of linear equations is said to be **homogeneous** if it can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ *always* has at least one solution, namely $\mathbf{x} = \mathbf{0}$ (the zero vector in $\mathbb{R}^n$). This zero solution is usually called the **trivial solution**. For a given equation $A\mathbf{x} = \mathbf{0}$, the important question is whether there exists a **nontrivial solution**, that is, a nonzero vector $\mathbf{x}$ that satisfies $A\mathbf{x} = \mathbf{0}$. The Existence and Uniqueness Theorem in Section 1.2 (Theorem 2) leads immediately to the following fact.

> The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.

EXAMPLE 1 Determine if the following homogeneous system has a nontrivial solution. Then describe the solution set.

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0 \\ 6x_1 + x_2 - 8x_3 &= 0 \end{aligned}$$

SOLUTION Let A be the matrix of coefficients of the system and row reduce the augmented matrix $[\,A \quad \mathbf{0}\,]$ to echelon form:

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

Since x_3 is a free variable, $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions (one for each nonzero choice of x_3). To describe the solution set, continue the row reduction of $[\,A \quad \mathbf{0}\,]$ to *reduced* echelon form:

$$\begin{bmatrix} 1 & 0 & -\frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{aligned} x_1 \qquad - \tfrac{4}{3}x_3 &= 0 \\ x_2 \qquad\qquad &= 0 \\ 0 &= 0 \end{aligned}$$

Solve for the basic variables x_1 and x_2 and obtain $x_1 = \frac{4}{3}x_3$, $x_2 = 0$, with x_3 free. As a vector, the general solution of $A\mathbf{x} = \mathbf{0}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} = x_3\mathbf{v}, \quad \text{where } \mathbf{v} = \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}$$

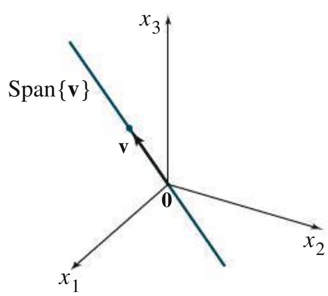


FIGURE 1

Here x_3 is factored out of the expression for the general solution vector. This shows that every solution of $A\mathbf{x} = \mathbf{0}$ in this case is a scalar multiple of $\mathbf{v}$. The trivial solution is obtained by choosing $x_3 = 0$. Geometrically, the solution set is a line through $\mathbf{0}$ in $\mathbb{R}^3$. See Figure 1. ■

Notice that a nontrivial solution $\mathbf{x}$ can have some zero entries so long as not all of its entries are zero.

EXAMPLE 2 A single linear equation can be treated as a very simple system of equations. Describe all solutions of the homogeneous "system"

$$10x_1 - 3x_2 - 2x_3 = 0 \tag{1}$$

SOLUTION There is no need for matrix notation. Solve for the basic variable x_1 in terms of the free variables. The general solution is $x_1 = .3x_2 + .2x_3$, with x_2 and x_3 free. As a vector, the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .3x_2 + .2x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .3x_2 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} .2x_3 \\ 0 \\ x_3 \end{bmatrix}$$

$$= x_2 \underset{\substack{\uparrow \\ \mathbf{u}}}{\begin{bmatrix} .3 \\ 1 \\ 0 \end{bmatrix}} + x_3 \underset{\substack{\uparrow \\ \mathbf{v}}}{\begin{bmatrix} .2 \\ 0 \\ 1 \end{bmatrix}} \quad \text{(with } x_2, x_3 \text{ free)} \tag{2}$$

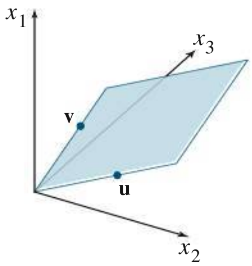


FIGURE 2

This calculation shows that every solution of (1) is a linear combination of the vectors **u** and **v**, shown in (2). That is, the solution set is Span $\{\mathbf{u}, \mathbf{v}\}$. Since neither **u** nor **v** is a scalar multiple of the other, the solution set is a plane through the origin. See Figure 2. ■

Examples 1 and 2, along with the exercises, illustrate the fact that the solution set of a homogeneous equation $A\mathbf{x} = \mathbf{0}$ can always be expressed explicitly as Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ for suitable vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$. If the only solution is the zero vector, then the solution set is Span $\{\mathbf{0}\}$. If the equation $A\mathbf{x} = \mathbf{0}$ has only one free variable, the solution set is a line through the origin, as in Figure 1. A plane through the origin, as in Figure 2, provides a good mental image for the solution set of $A\mathbf{x} = \mathbf{0}$ when there are two or more free variables. Note, however, that a similar figure can be used to visualize Span $\{\mathbf{u}, \mathbf{v}\}$ even when **u** and **v** do not arise as solutions of $A\mathbf{x} = \mathbf{0}$. See Figure 11 in Section 1.3.

Parametric Vector Form

The original equation (1) for the plane in Example 2 is an *implicit* description of the plane. Solving this equation amounts to finding an *explicit* description of the plane as the set spanned by **u** and **v**. Equation (2) is called a **parametric vector equation** of the plane. Sometimes such an equation is written as

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v} \quad (s, t \text{ in } \mathbb{R})$$

to emphasize that the parameters vary over all real numbers. In Example 1, the equation $\mathbf{x} = x_3\mathbf{v}$ (with x_3 free), or $\mathbf{x} = t\mathbf{v}$ (with t in $\mathbb{R}$), is a parametric vector equation of a line. Whenever a solution set is described explicitly with vectors as in Examples 1 and 2, we say that the solution is in **parametric vector form**.

Solutions of Nonhomogeneous Systems

When a nonhomogeneous linear system has many solutions, the general solution can be written in parametric vector form as one vector plus an arbitrary linear combination of vectors that satisfy the corresponding homogeneous system.

EXAMPLE 3 Describe all solutions of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}$$

SOLUTION Here A is the matrix of coefficients from Example 1. Row operations on $[\,A \quad \mathbf{b}\,]$ produce

$$\begin{bmatrix} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 \qquad - \tfrac{4}{3}x_3 &= -1 \\ x_2 \qquad\qquad &= 2 \\ 0 &= 0 \end{aligned}$$

Thus $x_1 = -1 + \frac{4}{3}x_3$, $x_2 = 2$, and x_3 is free. As a vector, the general solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 + \frac{4}{3}x_3 \\ 2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = \underset{\substack{\uparrow \\ \mathbf{p}}}{\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}} + x_3 \underset{\substack{\uparrow \\ \mathbf{v}}}{\begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}}$$

The equation $\mathbf{x} = \mathbf{p} + x_3\mathbf{v}$, or, writing t as a general parameter,

$$\mathbf{x} = \mathbf{p} + t\mathbf{v} \quad (t \text{ in } \mathbb{R}) \tag{3}$$

describes the solution set of $A\mathbf{x} = \mathbf{b}$ in parametric vector form. Recall from Example 1 that the solution set of $A\mathbf{x} = \mathbf{0}$ has the parametric vector equation

$$\mathbf{x} = t\mathbf{v} \quad (t \text{ in } \mathbb{R}) \tag{4}$$

[with the same $\mathbf{v}$ that appears in (3)]. Thus the solutions of $A\mathbf{x} = \mathbf{b}$ are obtained by adding the vector $\mathbf{p}$ to the solutions of $A\mathbf{x} = \mathbf{0}$. The vector $\mathbf{p}$ itself is just one particular solution of $A\mathbf{x} = \mathbf{b}$ [corresponding to $t = 0$ in (3)]. ■

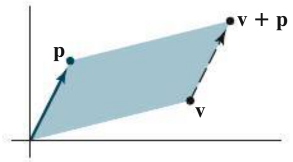


FIGURE 3
Adding $\mathbf{p}$ to $\mathbf{v}$ translates $\mathbf{v}$ to $\mathbf{v} + \mathbf{p}$.

To describe the solution set of $A\mathbf{x} = \mathbf{b}$ geometrically, we can think of vector addition as a *translation*. Given $\mathbf{v}$ and $\mathbf{p}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, the effect of adding $\mathbf{p}$ to $\mathbf{v}$ is to *move* $\mathbf{v}$ in a direction parallel to the line through $\mathbf{p}$ and $\mathbf{0}$. We say that $\mathbf{v}$ is **translated by p** to $\mathbf{v} + \mathbf{p}$. See Figure 3. If each point on a line L in $\mathbb{R}^2$ or $\mathbb{R}^3$ is translated by a vector $\mathbf{p}$, the result is a line parallel to L. See Figure 4.

Suppose L is the line through $\mathbf{0}$ and $\mathbf{v}$, described by equation (4). Adding $\mathbf{p}$ to each point on L produces the translated line described by equation (3). Note that $\mathbf{p}$ is on the line in equation (3). We call (3) **the equation of the line through p parallel to v**. Thus *the solution set of* $A\mathbf{x} = \mathbf{b}$ *is a line through* $\mathbf{p}$ *parallel to the solution set of* $A\mathbf{x} = \mathbf{0}$. Figure 5 illustrates this case.

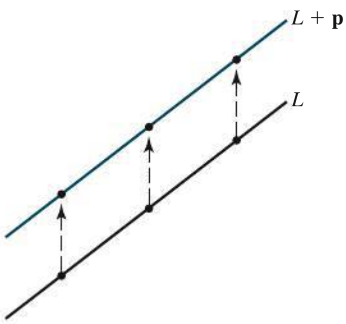


FIGURE 4
Translated line.

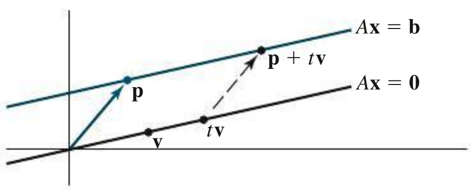


FIGURE 5 Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

The relation between the solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$ shown in Figure 5 generalizes to any *consistent* equation $A\mathbf{x} = \mathbf{b}$, although the solution set will be larger than a line when there are several free variables. The following theorem gives the precise statement. See Exercise 37 at the end of this section for a proof.

THEOREM 6

Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some given $\mathbf{b}$, and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Theorem 6 says that if $A\mathbf{x} = \mathbf{b}$ has a solution, then the solution set is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$, using any particular solution $\mathbf{p}$ of $A\mathbf{x} = \mathbf{b}$ for the translation. Figure 6 illustrates the case in which there are two free variables. Even when $n > 3$, our mental image of the solution set of a consistent system $A\mathbf{x} = \mathbf{b}$ (with $\mathbf{b} \neq \mathbf{0}$) is either a single nonzero point or a line or plane not passing through the origin.

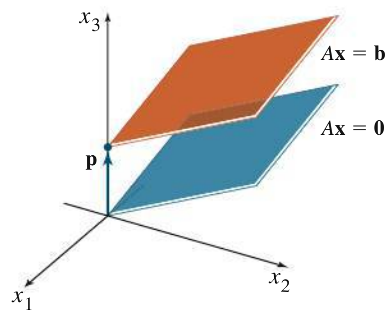


FIGURE 6 Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

Warning: Theorem 6 and Figure 6 apply only to an equation $A\mathbf{x} = \mathbf{b}$ that has at least one nonzero solution $\mathbf{p}$. When $A\mathbf{x} = \mathbf{b}$ has no solution, the solution set is empty.

The following algorithm outlines the calculations shown in Examples 1, 2, and 3.

WRITING A SOLUTION SET (OF A CONSISTENT SYSTEM) IN PARAMETRIC VECTOR FORM

1. Row reduce the augmented matrix to reduced echelon form.
2. Express each basic variable in terms of any free variables appearing in an equation.
3. Write a typical solution $\mathbf{x}$ as a vector whose entries depend on the free variables, if any.
4. Decompose $\mathbf{x}$ into a linear combination of vectors (with numeric entries) using the free variables as parameters.

Reasonable Answers

To verify that the solutions you found are indeed solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$, simply multiply the matrix by each vector in your solution and check that the result is the zero vector. For example, if $A = \begin{bmatrix} 1 & -2 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 0 & 1 & 1 & 3 \end{bmatrix}$, and you found the homogeneous solutions to

Reasonable Answers (Continued)

be $x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, check $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$ and $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$. Then $A\left(x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}\right)$ $=x_3A\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4A\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, which is equal to $x_3\begin{bmatrix}0\\0\\0\end{bmatrix}+x_4\begin{bmatrix}0\\0\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$, as desired.

If you are solving $A\mathbf{x}=\mathbf{b}$, then you can again verify that you have correct solutions by multiplying the matrix by each vector in your solutions. The product of A with the first vector (the one that is *not* part of the solution to the homogeneous equation) should be $\mathbf{b}$. The product of A with the remaining vectors (the ones that are part of the solution to the homogeneous equation) should of course be $\mathbf{0}$.

For example, to verify that $\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$ are solutions to $A\mathbf{x}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, check $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}2\\1\\1\\2\end{bmatrix}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, and use the calculations from above. Notice $A\left(\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}\right)$ $=A\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3A\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4A\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, which is equal to $\begin{bmatrix}5\\13\\8\end{bmatrix}+x_3\begin{bmatrix}0\\0\\0\end{bmatrix}$ $+x_4\begin{bmatrix}0\\0\\0\end{bmatrix}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, as desired.

Practice Problems

1. Each of the following equations determines a plane in $\mathbb{R}^3$. Do the two planes intersect? If so, describe their intersection.

$$\begin{aligned} x_1 + 4x_2 - 5x_3 &= 0 \\ 2x_1 - x_2 + 8x_3 &= 9 \end{aligned}$$

2. Write the general solution of $10x_1 - 3x_2 - 2x_3 = 7$ in parametric vector form, and relate the solution set to the one found in Example 2.
3. Prove the first part of Theorem 6: Suppose that $\mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$, so that $A\mathbf{p} = \mathbf{b}$. Let $\mathbf{v}_h$ be any solution to the homogeneous equation $A\mathbf{x} = \mathbf{0}$, and let $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$. Show that $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$.

1.5 Exercises

In Exercises 1–4, determine if the system has a nontrivial solution. Try to use as few row operations as possible.

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

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

3. $$\begin{aligned} -3x_1 + 5x_2 - 7x_3 &= 0 \\ -6x_1 + 7x_2 + x_3 &= 0 \end{aligned}$$

4. $$\begin{aligned} -5x_1 + 7x_2 + 9x_3 &= 0 \\ x_1 - 2x_2 + 6x_3 &= 0 \end{aligned}$$

In Exercises 5 and 6, follow the method of Examples 1 and 2 to write the solution set of the given homogeneous system in parametric vector form.

5. $$\begin{aligned} x_1 + 3x_2 + x_3 &= 0 \\ -4x_1 - 9x_2 + 2x_3 &= 0 \\ -3x_2 - 6x_3 &= 0 \end{aligned}$$

6. $$\begin{aligned} x_1 + 3x_2 - 5x_3 &= 0 \\ x_1 + 4x_2 - 8x_3 &= 0 \\ -3x_1 - 7x_2 + 9x_3 &= 0 \end{aligned}$$

In Exercises 7–12, describe all solutions of $A\mathbf{x} = \mathbf{0}$ in parametric vector form, where A is row equivalent to the given matrix.

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

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

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

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

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

12. $\begin{bmatrix} 1 & 5 & 2 & -6 & 9 & 0 \\ 0 & 0 & 1 & -7 & 4 & -8 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

You may find it helpful to review the information in the Reasonable Answers box from this section before answering Exercises 13–16.

13. Verify that the solutions you found to Exercise 9 are indeed homogeneous solutions.
14. Verify that the solutions you found to Exercise 10 are indeed homogeneous solutions.
15. Verify that the solutions you found to Exercise 11 are indeed homogeneous solutions.
16. Verify that the solutions you found to Exercise 12 are indeed homogeneous solutions.
17. Suppose the solution set of a certain system of linear equations can be described as $x_1 = 5 + 4x_3$, $x_2 = -2 - 7x_3$, with x_3 free. Use vectors to describe this set as a line in $\mathbb{R}^3$.
18. Suppose the solution set of a certain system of linear equations can be described as $x_1 = 3x_4$, $x_2 = 8 + x_4$, $x_3 = 2 - 5x_4$, with x_4 free. Use vectors to describe this set as a line in $\mathbb{R}^4$.
19. Follow the method of Example 3 to describe the solutions of the following system in parametric vector form. Also, give a geometric description of the solution set and compare it to that in Exercise 5.

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

20. As in Exercise 19, describe the solutions of the following system in parametric vector form, and provide a geometric comparison with the solution set in Exercise 6.

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

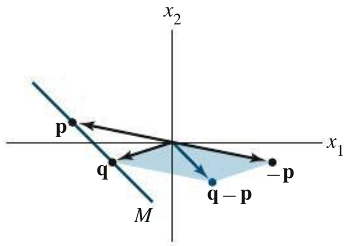
21. Describe and compare the solution sets of $x_1 + 9x_2 - 4x_3 = 0$ and $x_1 + 9x_2 - 4x_3 = -2$.
22. Describe and compare the solution sets of $x_1 - 3x_2 + 5x_3 = 0$ and $x_1 - 3x_2 + 5x_3 = 4$.

In Exercises 23 and 24, find the parametric equation of the line through **a** parallel to **b**.

23. $\mathbf{a} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$ **24.** $\mathbf{a} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -7 \\ 8 \end{bmatrix}$

In Exercises 25 and 26, find a parametric equation of the line M through **p** and **q**. [*Hint:* M is parallel to the vector $\mathbf{q} - \mathbf{p}$. See the figure below.]

25. $\mathbf{p} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}, \mathbf{q} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ **26.** $\mathbf{p} = \begin{bmatrix} -6 \\ 3 \end{bmatrix}, \mathbf{q} = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$



The line through **p** and **q**.

In Exercises 27–36, mark each statement True or False **(T/F)**. Justify each answer.

27. **(T/F)** A homogeneous equation is always consistent.

28. **(T/F)** If **x** is a nontrivial solution of $A\mathbf{x} = \mathbf{0}$, then every entry in **x** is nonzero.

29. **(T/F)** The equation $A\mathbf{x} = \mathbf{0}$ gives an explicit description of its solution set.

30. **(T/F)** The equation $\mathbf{x} = x_2\mathbf{u} + x_3\mathbf{v}$, with x_2 and x_3 free (and neither **u** nor **v** a multiple of the other), describes a plane through the origin.

31. **(T/F)** The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution if and only if the equation has at least one free variable.

32. **(T/F)** The equation $A\mathbf{x} = \mathbf{b}$ is homogeneous if the zero vector is a solution.

33. **(T/F)** The equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ describes a line through **v** parallel to **p**.

34. **(T/F)** The effect of adding **p** to a vector is to move the vector in a direction parallel to **p**.

35. **(T/F)** The solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the equation $A\mathbf{x} = \mathbf{0}$.

36. **(T/F)** The solution set of $A\mathbf{x} = \mathbf{b}$ is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$.

37. Prove the second part of Theorem 6: Let **w** be any solution of $A\mathbf{x} = \mathbf{b}$, and define $\mathbf{v}_h = \mathbf{w} - \mathbf{p}$. Show that $\mathbf{v}_h$ is a solution of $A\mathbf{x} = \mathbf{0}$. This shows that every solution of $A\mathbf{x} = \mathbf{b}$ has the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, with **p** a particular solution of $A\mathbf{x} = \mathbf{b}$ and $\mathbf{v}_h$ a solution of $A\mathbf{x} = \mathbf{0}$.

38. Suppose $A\mathbf{x} = \mathbf{b}$ has a solution. Explain why the solution is unique precisely when $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

39. Suppose A is the 3×3 *zero* matrix (with all zero entries). Describe the solution set of the equation $A\mathbf{x} = \mathbf{0}$.

40. If $\mathbf{b} \neq \mathbf{0}$, can the solution set of $A\mathbf{x} = \mathbf{b}$ be a plane through the origin? Explain.

In Exercises 41–44, (a) does the equation $A\mathbf{x} = \mathbf{0}$ have a nontrivial solution and (b) does the equation $A\mathbf{x} = \mathbf{b}$ have at least one solution for every possible **b**?

41. A is a 3×3 matrix with three pivot positions.

42. A is a 3×3 matrix with two pivot positions.

43. A is a 3×2 matrix with two pivot positions.

44. A is a 2×4 matrix with two pivot positions.

45. Given $A = \begin{bmatrix} -2 & -6 \\ 7 & 21 \\ -3 & -9 \end{bmatrix}$, find one nontrivial solution of $A\mathbf{x} = \mathbf{0}$ by inspection. [*Hint:* Think of the equation $A\mathbf{x} = \mathbf{0}$ written as a vector equation.]

46. Given $A = \begin{bmatrix} 4 & -6 \\ -8 & 12 \\ 6 & -9 \end{bmatrix}$, find one nontrivial solution of $A\mathbf{x} = \mathbf{0}$ by inspection.

47. Construct a 3×3 nonzero matrix A such that the vector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

48. Construct a 3×3 nonzero matrix A such that the vector $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

49. Construct a 2×2 matrix A such that the solution set of the equation $A\mathbf{x} = \mathbf{0}$ is the line in $\mathbb{R}^2$ through $(4, 1)$ and the origin. Then, find a vector **b** in $\mathbb{R}^2$ such that the solution set of $A\mathbf{x} = \mathbf{b}$ is *not* a line in $\mathbb{R}^2$ parallel to the solution set of $A\mathbf{x} = \mathbf{0}$. Why does this *not* contradict Theorem 6?

50. Suppose A is a 3×3 matrix and **y** is a vector in $\mathbb{R}^3$ such that the equation $A\mathbf{x} = \mathbf{y}$ does *not* have a solution. Does there exist a vector **z** in $\mathbb{R}^3$ such that the equation $A\mathbf{x} = \mathbf{z}$ has a unique solution? Discuss.

51. Let A be an $m \times n$ matrix and let **u** be a vector in $\mathbb{R}^n$ that satisfies the equation $A\mathbf{x} = \mathbf{0}$. Show that for any scalar c, the vector $c\mathbf{u}$ also satisfies $A\mathbf{x} = \mathbf{0}$. [That is, show that $A(c\mathbf{u}) = \mathbf{0}$.]

52. Let A be an $m \times n$ matrix, and let **u** and **v** be vectors in $\mathbb{R}^n$ with the property that $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. Explain why $A(\mathbf{u} + \mathbf{v})$ must be the zero vector. Then explain why $A(c\mathbf{u} + d\mathbf{v}) = \mathbf{0}$ for each pair of scalars c and d.

Solutions to Practice Problems

1. Row reduce the augmented matrix:

$$\begin{bmatrix} 1 & 4 & -5 & 0 \\ 2 & -1 & 8 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & -5 & 0 \\ 0 & -9 & 18 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & -2 & -1 \end{bmatrix}$$

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

Thus $x_1 = 4 - 3x_3$, $x_2 = -1 + 2x_3$, with x_3 free. The general solution in parametric vector form is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 - 3x_3 \\ -1 + 2x_3 \\ x_3 \end{bmatrix} = \underset{\overset{\uparrow}{\mathbf{p}}}{\begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}} + x_3 \underset{\overset{\uparrow}{\mathbf{v}}}{\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}}$$

The intersection of the two planes is the line through $\mathbf{p}$ in the direction of $\mathbf{v}$.

2. The augmented matrix $\begin{bmatrix} 10 & -3 & -2 & 7 \end{bmatrix}$ is row equivalent to $\begin{bmatrix} 1 & -.3 & -.2 & .7 \end{bmatrix}$, and the general solution is $x_1 = .7 + .3x_2 + .2x_3$, with x_2 and x_3 free. That is,

$$\begin{aligned} \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .7 + .3x_2 + .2x_3 \\ x_2 \\ x_3 \end{bmatrix} &= \begin{bmatrix} .7 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} .3 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} .2 \\ 0 \\ 1 \end{bmatrix} \\ &= \mathbf{p} + x_2\mathbf{u} + x_3\mathbf{v} \end{aligned}$$

The solution set of the nonhomogeneous equation $A\mathbf{x} = \mathbf{b}$ is the translated plane $\mathbf{p} + \text{Span}\{\mathbf{u}, \mathbf{v}\}$, which passes through $\mathbf{p}$ and is parallel to the solution set of the homogeneous equation in Example 2.

3. Using Theorem 5 from Section 1.4, notice

$$A(\mathbf{p} + \mathbf{v}_h) = A\mathbf{p} + A\mathbf{v}_h = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

hence $\mathbf{p} + \mathbf{v}_h$ is a solution to $A\mathbf{x} = \mathbf{b}$.

1.6 Applications of Linear Systems

You might expect that a real-life problem involving linear algebra would have only one solution, or perhaps no solution. The purpose of this section is to show how linear systems with many solutions can arise naturally. The applications here come from economics, chemistry, and network flow.

A Homogeneous System in Economics

The system of 500 equations in 500 variables, mentioned in this chapter's introduction, is now known as a Leontief "input–output" (or "production") model.[1] Section 2.6 will examine this model in more detail, when more theory and better notation are available. For now, we look at a simpler "exchange model," also due to Leontief.

[1] See Wassily W. Leontief, "Input–Output Economics," *Scientific American*, October 1951, pp. 15–21.

Suppose a nation's economy is divided into many sectors, such as various manufacturing, communication, entertainment, and service industries. Suppose that for each sector we know its total output for one year and we know exactly how this output is divided or "exchanged" among the other sectors of the economy. Let the total dollar value of a sector's output be called the **price** of that output. Leontief proved the following result.

> There exist *equilibrium prices* that can be assigned to the total outputs of the various sectors in such a way that the income of each sector exactly balances its expenses.

The following example shows how to find the equilibrium prices.

EXAMPLE 1 Suppose an economy consists of the Coal, Electric (power), and Steel sectors, and the output of each sector is distributed among the various sectors as shown in Table 1, where the entries in a column represent the fractional parts of a sector's total output.

The second column of Table 1, for instance, says that the total output of the Electric sector is divided as follows: 40% to Coal, 50% to Steel, and the remaining 10% to Electric. (Electric treats this 10% as an expense it incurs in order to operate its business.) Since all output must be taken into account, the decimal fractions in each column must sum to 1.

Denote the prices (in dollar values) of the total annual outputs of the Coal, Electric, and Steel sectors by p_C, p_E, and p_S, respectively. If possible, find equilibrium prices that make each sector's income match its expenditures.

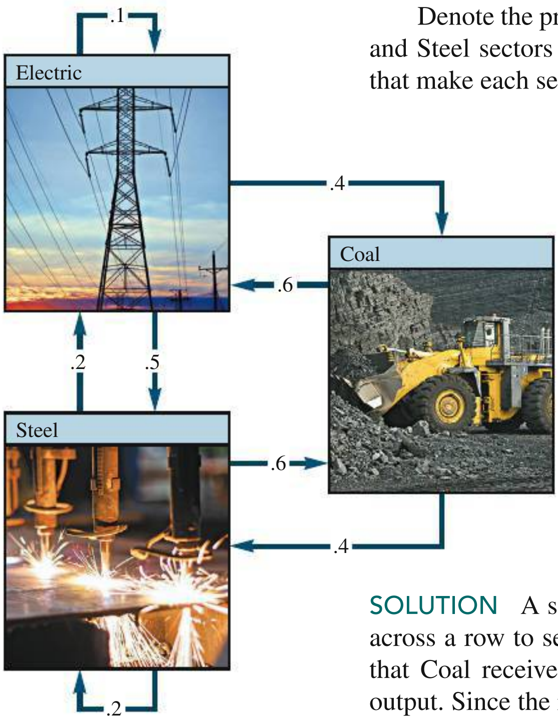


TABLE 1 A Simple Economy

Distribution of Output from			
Coal	**Electric**	**Steel**	**Purchased by**
.0	.4	.6	Coal
.6	.1	.2	Electric
.4	.5	.2	Steel

SOLUTION A sector looks down a column to see where its output goes, and it looks across a row to see what it needs as inputs. For instance, the first row of Table 1 says that Coal receives (and pays for) 40% of the Electric output and 60% of the Steel output. Since the respective values of the total outputs are p_E and p_S, Coal must spend $.4p_E$ dollars for its share of Electric's output and $.6p_S$ for its share of Steel's output. Thus Coal's total expenses are $.4p_E + .6p_S$. To make Coal's income, p_C, equal to its expenses, we want

$$p_C = .4p_E + .6p_S \qquad (1)$$

The second row of the exchange table shows that the Electric sector spends $.6p_C$ for coal, $.1p_E$ for electricity, and $.2p_S$ for steel. Hence the income/expense requirement for Electric is

$$p_E = .6p_C + .1p_E + .2p_S \qquad (2)$$

Finally, the third row of the exchange table leads to the final requirement:

$$p_S = .4p_C + .5p_E + .2p_S \tag{3}$$

To solve the system of equations (1), (2), and (3), move all the unknowns to the left sides of the equations and combine like terms. [For instance, on the left side of (2), write $p_E - .1p_E$ as $.9p_E$.]

$$\begin{aligned} p_C - .4p_E - .6p_S &= 0 \\ -.6p_C + .9p_E - .2p_S &= 0 \\ -.4p_C - .5p_E + .8p_S &= 0 \end{aligned}$$

Row reduction is next. For simplicity here, decimals are rounded to two places.

$$\begin{bmatrix} 1 & -.4 & -.6 & 0 \\ -.6 & .9 & -.2 & 0 \\ -.4 & -.5 & .8 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -.4 & -.6 & 0 \\ 0 & .66 & -.56 & 0 \\ 0 & -.66 & .56 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -.4 & -.6 & 0 \\ 0 & .66 & -.56 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & -.4 & -.6 & 0 \\ 0 & 1 & -.85 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -.94 & 0 \\ 0 & 1 & -.85 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution is $p_C = .94p_S$, $p_E = .85p_S$, and p_S is free. The equilibrium price vector for the economy has the form

$$\mathbf{p} = \begin{bmatrix} p_C \\ p_E \\ p_S \end{bmatrix} = \begin{bmatrix} .94p_S \\ .85p_S \\ p_S \end{bmatrix} = p_S \begin{bmatrix} .94 \\ .85 \\ 1 \end{bmatrix}$$

Any (nonnegative) choice for p_S results in a choice of equilibrium prices. For instance, if we take p_S to be 100 (or \$100 million), then $p_C = 94$ and $p_E = 85$. The incomes and expenditures of each sector will be equal if the output of Coal is priced at \$94 million, that of Electric at \$85 million, and that of Steel at \$100 million. ■

Balancing Chemical Equations

Chemical equations describe the quantities of substances consumed and produced by chemical reactions. For instance, when propane gas burns, the propane (C_3H_8) combines with oxygen (O_2) to form carbon dioxide (CO_2) and water (H_2O), according to an equation of the form

$$(x_1)C_3H_8 + (x_2)O_2 \rightarrow (x_3)CO_2 + (x_4)H_2O \tag{4}$$

To "balance" this equation, a chemist must find whole numbers $x_1, \dots, x_4$ such that the total numbers of carbon (C), hydrogen (H), and oxygen (O) atoms on the left match the corresponding numbers of atoms on the right (because atoms are neither destroyed nor created in the reaction).

A systematic method for balancing chemical equations is to set up a vector equation that describes the numbers of atoms of each type present in a reaction. Since equation (4) involves three types of atoms (carbon, hydrogen, and oxygen), construct a vector in $\mathbb{R}^3$ for each reactant and product in (4) that lists the numbers of "atoms per molecule," as follows:

$$C_3H_8: \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix}, \quad O_2: \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \quad CO_2: \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad H_2O: \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \begin{array}{l} \leftarrow \text{Carbon} \\ \leftarrow \text{Hydrogen} \\ \leftarrow \text{Oxygen} \end{array}$$

To balance equation (4), the coefficients $x_1, \ldots, x_4$ must satisfy

$$x_1\begin{bmatrix}3\\8\\0\end{bmatrix}+x_2\begin{bmatrix}0\\0\\2\end{bmatrix}=x_3\begin{bmatrix}1\\0\\2\end{bmatrix}+x_4\begin{bmatrix}0\\2\\1\end{bmatrix}$$

To solve, move all the terms to the left (changing the signs in the third and fourth vectors):

$$x_1\begin{bmatrix}3\\8\\0\end{bmatrix}+x_2\begin{bmatrix}0\\0\\2\end{bmatrix}+x_3\begin{bmatrix}-1\\0\\-2\end{bmatrix}+x_4\begin{bmatrix}0\\-2\\-1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

Row reduction of the augmented matrix for this equation leads to the general solution

$$x_1 = \tfrac{1}{4}x_4,\; x_2 = \tfrac{5}{4}x_4,\; x_3 = \tfrac{3}{4}x_4,\; \text{ with } x_4 \text{ free}$$

Since the coefficients in a chemical equation must be integers, take $x_4 = 4$, in which case $x_1 = 1$, $x_2 = 5$, and $x_3 = 3$. The balanced equation is

$$C_3H_8 + 5O_2 \to 3CO_2 + 4H_2O$$

The equation would also be balanced if, for example, each coefficient were doubled. For most purposes, however, chemists prefer to use a balanced equation whose coefficients are the smallest possible whole numbers.

Network Flow

Systems of linear equations arise naturally when scientists, engineers, or economists study the flow of some quantity through a network. For instance, urban planners and traffic engineers monitor the pattern of traffic flow in a grid of city streets. Electrical engineers calculate current flow through electrical circuits. Economists analyze the distribution of products from manufacturers to consumers through a network of wholesalers and retailers. For many networks, the systems of equations involve hundreds or even thousands of variables and equations.

A *network* consists of a set of points called *junctions*, or *nodes*, with lines or arcs called *branches* connecting some or all of the junctions. The direction of flow in each branch is indicated, and the flow amount (or rate) is either shown or is denoted by a variable.

The basic assumption of network flow is that the total flow into the network equals the total flow out of the network and that the total flow into a junction equals the total flow out of the junction. For example, Figure 1 shows 30 units flowing into a junction through one branch, with x_1 and x_2 denoting the flows out of the junction through other branches. Since the flow is "conserved" at each junction, we must have $x_1 + x_2 = 30$. In a similar fashion, the flow at each junction is described by a linear equation. The problem of network analysis is to determine the flow in each branch when partial information (such as the flow into and out of the network) is known.

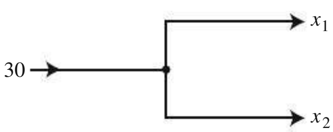


FIGURE 1
A junction or node.

EXAMPLE 2 The network in Figure 2 shows the traffic flow (in vehicles per hour) over several one-way streets in downtown Baltimore during a typical early afternoon. Determine the general flow pattern for the network.

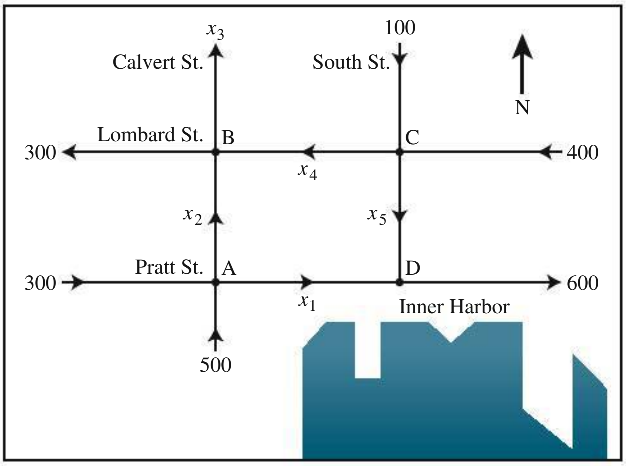


FIGURE 2 Baltimore streets.

SOLUTION Write equations that describe the flow, and then find the general solution of the system. Label the street intersections (junctions) and the unknown flows in the branches, as shown in Figure 2. At each intersection, set the flow in equal to the flow out.

Intersection	Flow in		Flow out
A	$300 + 500$	$=$	$x_1 + x_2$
B	$x_2 + x_4$	$=$	$300 + x_3$
C	$100 + 400$	$=$	$x_4 + x_5$
D	$x_1 + x_5$	$=$	600

Also, the total flow into the network $(500 + 300 + 100 + 400)$ equals the total flow out of the network $(300 + x_3 + 600)$, which simplifies to $x_3 = 400$. Combine this equation with a rearrangement of the first four equations to obtain the following system of equations:

$$\begin{aligned} x_1 + x_2 \qquad\qquad\qquad &= 800 \\ x_2 - x_3 + x_4 \qquad &= 300 \\ x_4 + x_5 &= 500 \\ x_1 \qquad\qquad\qquad + x_5 &= 600 \\ x_3 \qquad\qquad &= 400 \end{aligned}$$

Row reduction of the associated augmented matrix leads to

$$\begin{aligned} x_1 \qquad\qquad + x_5 &= 600 \\ x_2 \qquad - x_5 &= 200 \\ x_3 \qquad\quad &= 400 \\ x_4 + x_5 &= 500 \end{aligned}$$

The general flow pattern for the network is described by

$$\begin{cases} x_1 = 600 - x_5 \\ x_2 = 200 + x_5 \\ x_3 = 400 \\ x_4 = 500 - x_5 \\ x_5 \text{ is free} \end{cases}$$

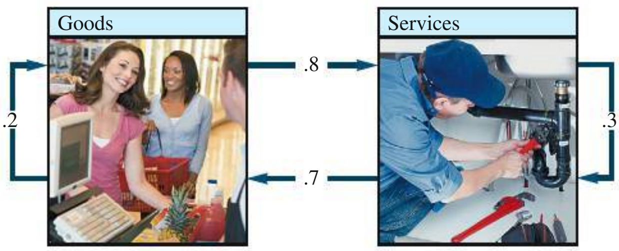
A negative flow in a network branch corresponds to flow in the direction opposite to that shown on the model. Since the streets in this problem are one way, none of the variables here can be negative. This fact leads to certain limitations on the possible values of the variables. For instance, $x_5 \leq 500$ because x_4 cannot be negative. Other constraints on the variables are considered in Practice Problem 2. ■

Practice Problems

1. Suppose an economy has three sectors: Agriculture, Mining, and Manufacturing. Agriculture sells 5% of its output to Mining and 30% to Manufacturing, and retains the rest. Mining sells 20% of its output to Agriculture and 70% to Manufacturing, and retains the rest. Manufacturing sells 20% of its output to Agriculture and 30% to Mining, and retains the rest. Determine the exchange table for this economy, where the columns describe how the output of each sector is exchanged among the three sectors.
2. Consider the network flow studied in Example 2. Determine the possible range of values of x_1 and x_2. [*Hint:* The example showed that $x_5 \leq 500$. What does this imply about x_1 and x_2? Also, use the fact that $x_5 \geq 0$.]

1.6 Exercises

1. Suppose an economy has only two sectors, Goods and Services. Each year, Goods sells 80% of its output to Services and keeps the rest, while Services sells 70% of its output to Goods and retains the rest. Find equilibrium prices for the annual outputs of the Goods and Services sectors that make each sector's income match its expenditures.



2. Find another set of equilibrium prices for the economy in Example 1. Suppose the same economy used Japanese yen instead of dollars to measure the value of the various sectors' outputs. Would this change the problem in any way? Discuss.

3. Consider an economy with three sectors, Chemicals & Metals, Fuels & Power, and Machinery. Chemicals sells 30% of its output to Fuels and 50% to Machinery and retains the rest. Fuels sells 80% of its output to Chemicals and 10% to Machinery and retains the rest. Machinery sells 40% to Chemicals and 40% to Fuels and retains the rest.

   a. Construct the exchange table for this economy.

   b. Develop a system of equations that leads to prices at which each sector's income matches its expenses. Then write the augmented matrix that can be row reduced to find these prices.

   T c. Find a set of equilibrium prices when the price for the Machinery output is 100 units.

4. Suppose an economy has four sectors, Agriculture (A), Energy (E), Manufacturing (M), and Transportation (T). Sector A sells 10% of its output to E and 25% to M and retains the rest. Sector E sells 30% of its output to A, 35% to M, and 25% to T and retains the rest. Sector M sells 30% of its output to A, 15% to E, and 40% to T and retains the rest. Sector T sells 20% of its output to A, 10% to E, and 30% to M and retains the rest.

   a. Construct the exchange table for this economy.

   T b. Find a set of equilibrium prices for the economy.

Balance the chemical equations in Exercises 5–10 using the vector equation approach discussed in this section.

5. Boron sulfide reacts violently with water to form boric acid and hydrogen sulfide gas (the smell of rotten eggs). The unbalanced equation is

   $B_2S_3 + H_2O \rightarrow H_3BO_3 + H_2S$

   [For each compound, construct a vector that lists the numbers of atoms of boron, sulfur, hydrogen, and oxygen.]

6. When solutions of sodium phosphate and barium nitrate are mixed, the result is barium phosphate (as a precipitate) and sodium nitrate. The unbalanced equation is

   $Na_3PO_4 + Ba(NO_3)_2 \rightarrow Ba_3(PO_4)_2 + NaNO_3$

   [For each compound, construct a vector that lists the numbers of atoms of sodium (Na), phosphorus, oxygen, barium, and nitrogen. For instance, barium nitrate corresponds to $(0, 0, 6, 1, 2)$.]

7. Alka-Seltzer contains sodium bicarbonate ($NaHCO_3$) and citric acid ($H_3C_6H_5O_7$). When a tablet is dissolved in water, the following reaction produces sodium citrate, water, and carbon dioxide (gas):

$$NaHCO_3 + H_3C_6H_5O_7 \rightarrow Na_3C_6H_5O_7 + H_2O + CO_2$$

8. The following reaction between potassium permanganate ($KMnO_4$) and manganese sulfate in water produces manganese dioxide, potassium sulfate, and sulfuric acid:

$$KMnO_4 + MnSO_4 + H_2O \rightarrow MnO_2 + K_2SO_4 + H_2SO_4$$

[For each compound, construct a vector that lists the numbers of atoms of potassium (K), manganese, oxygen, sulfur, and hydrogen.]

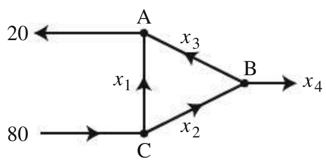
T 9. If possible, use exact arithmetic or rational format for calculations in balancing the following chemical reaction:

$$PbN_6 + CrMn_2O_8 \rightarrow Pb_3O_4 + Cr_2O_3 + MnO_2 + NO$$

T 10. The chemical reaction below can be used in some industrial processes, such as the production of arsene (AsH_3). Use exact arithmetic or rational format for calculations to balance this equation.

$$MnS + As_2Cr_{10}O_{35} + H_2SO_4$$
$$\rightarrow HMnO_4 + AsH_3 + CrS_3O_{12} + H_2O$$

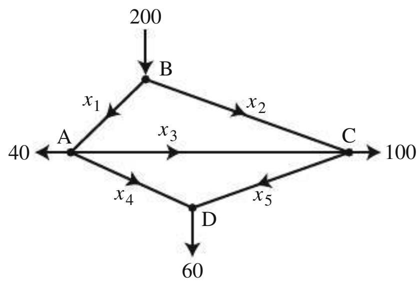
11. Find the general flow pattern of the network shown in the figure. Assuming that the flows are all nonnegative, what is the largest possible value for x_3?



12. a. Find the general traffic pattern in the freeway network shown in the figure. (Flow rates are in cars/minute.)

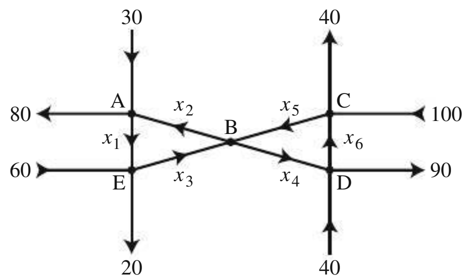
b. Describe the general traffic pattern when the road whose flow is x_4 is closed.

c. When $x_4 = 0$, what is the minimum value of x_1?

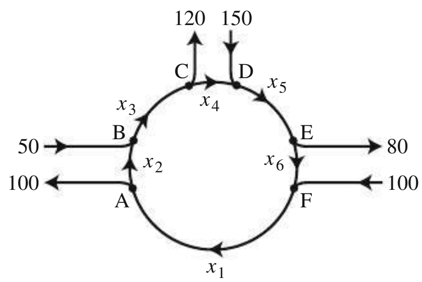


13. a. Find the general flow pattern in the network shown in the figure.

b. Assuming that the flow must be in the directions indicated, find the minimum flows in the branches denoted by x_2, x_3, x_4, and x_5.



14. Intersections in England are often constructed as one-way "roundabouts," such as the one shown in the figure. Assume that traffic must travel in the directions shown. Find the general solution of the network flow. Find the smallest possible value for x_6.



Solutions to Practice Problems

1. Write the percentages as decimals. Since all output must be taken into account, each column must sum to 1. This fact helps to fill in any missing entries.

Distribution of Output from			
Agriculture	**Mining**	**Manufacturing**	**Purchased by**
.65	.20	.20	Agriculture
.05	.10	.30	Mining
.30	.70	.50	Manufacturing

Solutions to Practice Problems (Continued)

2. Since $x_5 \leq 500$, the equations D and A for x_1 and x_2 imply that $x_1 \geq 100$ and $x_2 \leq 700$. The fact that $x_5 \geq 0$ implies that $x_1 \leq 600$ and $x_2 \geq 200$. So, $100 \leq x_1 \leq 600$, and $200 \leq x_2 \leq 700$.

1.7 Linear Independence

The homogeneous equations in Section 1.5 can be studied from a different perspective by writing them as vector equations. In this way, the focus shifts from the unknown solutions of $A\mathbf{x} = \mathbf{0}$ to the vectors that appear in the vector equations.

For instance, consider the equation

$$x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{1}$$

This equation has a trivial solution, of course, where $x_1 = x_2 = x_3 = 0$. As in Section 1.5, the main issue is whether the trivial solution is the *only one*.

DEFINITION

An indexed set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be **linearly independent** if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution. The set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is said to be **linearly dependent** if there exist weights $c_1, \dots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0} \tag{2}$$

Equation (2) is called a **linear dependence relation** among $\mathbf{v}_1, \dots, \mathbf{v}_p$ when the weights are not all zero. An indexed set is linearly dependent if and only if it is not linearly independent. For brevity, we may say that $\mathbf{v}_1, \dots, \mathbf{v}_p$ are linearly dependent when we mean that $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a linearly dependent set. We use analogous terminology for linearly independent sets.

EXAMPLE 1 Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

a. Determine if the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

b. If possible, find a linear dependence relation among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

SOLUTION

a. We must determine if there is a nontrivial solution of equation (1) above. Row operations on the associated augmented matrix show that

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

Clearly, x_1 and x_2 are basic variables, and x_3 is free. Each nonzero value of x_3 determines a nontrivial solution of (1). Hence $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly dependent (and not linearly independent).

b. To find a linear dependence relation among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, completely row reduce the augmented matrix and write the new system:

$$\begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{aligned} x_1 \qquad - 2x_3 &= 0 \\ x_2 + \ \ x_3 &= 0 \\ 0 &= 0 \end{aligned}$$

Thus $x_1 = 2x_3$, $x_2 = -x_3$, and x_3 is free. Choose any nonzero value for x_3—say, $x_3 = 5$. Then $x_1 = 10$ and $x_2 = -5$. Substitute these values into equation (1) and obtain

$$10\mathbf{v}_1 - 5\mathbf{v}_2 + 5\mathbf{v}_3 = \mathbf{0}$$

This is one (out of infinitely many) possible linear dependence relations among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. ■

Linear Independence of Matrix Columns

Suppose that we begin with a matrix $A = [\,\mathbf{a}_1 \ \ \cdots \ \ \mathbf{a}_n\,]$ instead of a set of vectors. The matrix equation $A\mathbf{x} = \mathbf{0}$ can be written as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

Each linear dependence relation among the columns of A corresponds to a nontrivial solution of $A\mathbf{x} = \mathbf{0}$. Thus we have the following important fact.

> The columns of a matrix A are linearly independent if and only if the equation $A\mathbf{x} = \mathbf{0}$ has *only* the trivial solution. (3)

EXAMPLE 2 Determine if the columns of the matrix $A = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 2 & -1 \\ 5 & 8 & 0 \end{bmatrix}$ are linearly independent.

SOLUTION To study $A\mathbf{x} = \mathbf{0}$, row reduce the augmented matrix:

$$\begin{bmatrix} 0 & 1 & 4 & 0 \\ 1 & 2 & -1 & 0 \\ 5 & 8 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & -2 & 5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 13 & 0 \end{bmatrix}$$

At this point, it is clear that there are three basic variables and no free variables. So the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, and the columns of A are linearly independent. ■

Sets of One or Two Vectors

A set containing only one vector—say, $\mathbf{v}$—is linearly independent if and only if $\mathbf{v}$ is not the zero vector. This is because the vector equation $x_1\mathbf{v} = \mathbf{0}$ has only the trivial solution when $\mathbf{v} \neq \mathbf{0}$. The zero vector is linearly dependent because $x_1\mathbf{0} = \mathbf{0}$ has many nontrivial solutions.

The next example will explain the nature of a linearly dependent set of two vectors.

EXAMPLE 3 Determine if the following sets of vectors are linearly independent.

a. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ b. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

SOLUTION

a. Notice that $\mathbf{v}_2$ is a multiple of $\mathbf{v}_1$, namely $\mathbf{v}_2 = 2\mathbf{v}_1$. Hence $-2\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$, which shows that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent.

b. The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are certainly *not* multiples of one another. Could they be linearly dependent? Suppose c and d satisfy

$$c\mathbf{v}_1 + d\mathbf{v}_2 = \mathbf{0}$$

If $c \neq 0$, then we can solve for $\mathbf{v}_1$ in terms of $\mathbf{v}_2$, namely $\mathbf{v}_1 = (-d/c)\mathbf{v}_2$. This result is impossible because $\mathbf{v}_1$ is *not* a multiple of $\mathbf{v}_2$. So c must be zero. Similarly, d must also be zero. Thus $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set. ■

The arguments in Example 3 show that you can always decide *by inspection* when a set of two vectors is linearly dependent. Row operations are unnecessary. Simply check whether at least one of the vectors is a scalar times the other. (The test applies only to sets of *two* vectors.)

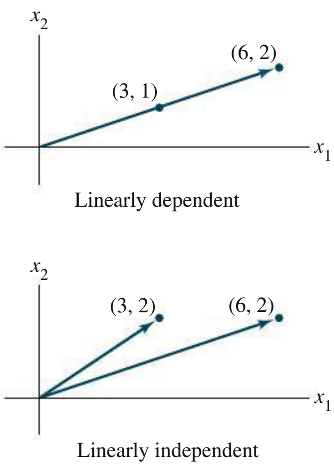


FIGURE 1

A set of two vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if at least one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other.

In geometric terms, two vectors are linearly dependent if and only if they lie on the same line through the origin. Figure 1 shows the vectors from Example 3.

Sets of Two or More Vectors

The proof of the next theorem is similar to the solution of Example 3. Details are given at the end of this section.

THEOREM 7

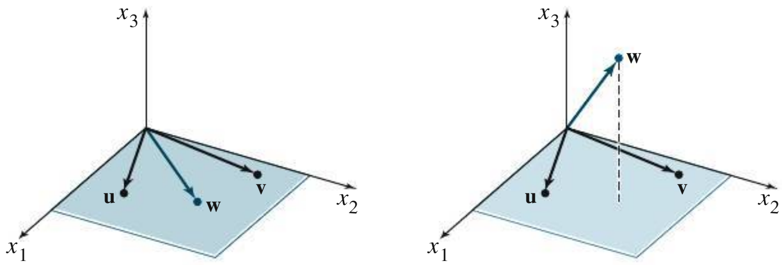
Characterization of Linearly Dependent Sets

An indexed set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ of two or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1 \neq \mathbf{0}$, then some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\mathbf{v}_1, \dots, \mathbf{v}_{j-1}$.

Warning: Theorem 7 does *not* say that *every* vector in a linearly dependent set is a linear combination of the preceding vectors. A vector in a linearly dependent set may fail to be a linear combination of the other vectors. See Practice Problem 1(c).

EXAMPLE 4 Let $\mathbf{u} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}$. Describe the set spanned by $\mathbf{u}$ and $\mathbf{v}$, and explain why a vector $\mathbf{w}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$ if and only if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent.

SOLUTION The vectors **u** and **v** are linearly independent because neither vector is a multiple of the other, and so they span a plane in $\mathbb{R}^3$. (See Section 1.3.) In fact, Span$\{\mathbf{u}, \mathbf{v}\}$ is the x_1x_2-plane (with $x_3 = 0$). If **w** is a linear combination of **u** and **v**, then $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent, by Theorem 7. Conversely, suppose that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent. By Theorem 7, some vector in $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a linear combination of the preceding vectors (since $\mathbf{u} \neq \mathbf{0}$). That vector must be **w**, since **v** is not a multiple of **u**. So **w** is in Span$\{\mathbf{u}, \mathbf{v}\}$. See Figure 2. ■



Linearly dependent, **w** in Span$\{\mathbf{u}, \mathbf{v}\}$

Linearly independent, **w** *not* in Span$\{\mathbf{u}, \mathbf{v}\}$

FIGURE 2 Linear dependence in $\mathbb{R}^3$.

Example 4 generalizes to any set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ in $\mathbb{R}^3$ with **u** and **v** linearly independent. The set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ will be linearly dependent if and only if **w** is in the plane spanned by **u** and **v**.

The next two theorems describe special cases in which the linear dependence of a set is automatic. Moreover, Theorem 8 will be a key result for work in later chapters.

THEOREM 8

If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.

$$\begin{array}{cc} & p \\ n & \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \end{array}$$

FIGURE 3
If $p > n$, the columns are linearly dependent.

PROOF Let $A = [\, \mathbf{v}_1 \ \cdots \ \mathbf{v}_p \,]$. Then A is $n \times p$, and the equation $A\mathbf{x} = \mathbf{0}$ corresponds to a system of n equations in p unknowns. If $p > n$, there are more variables than equations, so there must be a free variable. Hence $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, and the columns of A are linearly dependent. See Figure 3 for a matrix version of this theorem. ■

Warning: Theorem 8 says nothing about the case in which the number of vectors in the set does *not* exceed the number of entries in each vector.

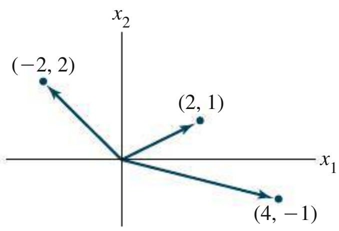


FIGURE 4
A linearly dependent set in $\mathbb{R}^2$.

EXAMPLE 5 The vectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \end{bmatrix}$ are linearly dependent by Theorem 8, because there are three vectors in the set and there are only two entries in each vector. Notice, however, that none of the vectors is a multiple of one of the other vectors. See Figure 4. ■

THEOREM 9

If a set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ contains the zero vector, then the set is linearly dependent.

PROOF By renumbering the vectors, we may suppose $\mathbf{v}_1 = \mathbf{0}$. Then the equation $1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p = \mathbf{0}$ shows that S is linearly dependent. ∎

EXAMPLE 6 Determine by inspection if the given set is linearly dependent.

a. $\begin{bmatrix} 1 \\ 7 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 9 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 8 \end{bmatrix}$ b. $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 8 \end{bmatrix}$ c. $\begin{bmatrix} -2 \\ 4 \\ 6 \\ 10 \end{bmatrix}, \begin{bmatrix} 3 \\ -6 \\ -9 \\ 15 \end{bmatrix}$

SOLUTION

a. The set contains four vectors, each of which has only three entries. So the set is linearly dependent by Theorem 8.

b. Theorem 8 does not apply here because the number of vectors does not exceed the number of entries in each vector. Since the zero vector is in the set, the set is linearly dependent by Theorem 9.

c. Compare the corresponding entries of the two vectors. The second vector seems to be $-3/2$ times the first vector. This relation holds for the first three pairs of entries, but fails for the fourth pair. Thus neither of the vectors is a multiple of the other, and hence they are linearly independent. ∎

In general, you should read a section thoroughly *several* times to absorb an important concept such as linear independence. The notes in the *Study Guide* for this section will help you learn to form mental images of key ideas in linear algebra. For instance, the following proof is worth reading carefully because it shows how the definition of linear independence can be *used*.

PROOF OF THEOREM 7 (Characterization of Linearly Dependent Sets)
If some $\mathbf{v}_j$ in S equals a linear combination of the other vectors, then $\mathbf{v}_j$ can be subtracted from both sides of the equation, producing a linear dependence relation with a nonzero weight (-1) on $\mathbf{v}_j$. [For instance, if $\mathbf{v}_1 = c_2\mathbf{v}_2 + c_3\mathbf{v}_3$, then $\mathbf{0} = (-1)\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + 0\mathbf{v}_4 + \cdots + 0\mathbf{v}_p$.] Thus S is linearly dependent.

Conversely, suppose S is linearly dependent. If $\mathbf{v}_1$ is zero, then it is a (trivial) linear combination of the other vectors in S. Otherwise, $\mathbf{v}_1 \neq \mathbf{0}$, and there exist weights $c_1, \ldots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

Let j be the largest subscript for which $c_j \neq 0$. If $j = 1$, then $c_1\mathbf{v}_1 = \mathbf{0}$, which is impossible because $\mathbf{v}_1 \neq \mathbf{0}$. So $j > 1$, and

$$c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j + 0\mathbf{v}_{j+1} + \cdots + 0\mathbf{v}_p = \mathbf{0}$$
$$c_j\mathbf{v}_j = -c_1\mathbf{v}_1 - \cdots - c_{j-1}\mathbf{v}_{j-1}$$
$$\mathbf{v}_j = \left(-\frac{c_1}{c_j}\right)\mathbf{v}_1 + \cdots + \left(-\frac{c_{j-1}}{c_j}\right)\mathbf{v}_{j-1}$$ ∎

Practice Problems

1. Let $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -6 \\ 1 \\ 7 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 0 \\ -5 \\ 2 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 3 \\ 7 \\ -5 \end{bmatrix}$.

   a. Are the sets $\{\mathbf{u}, \mathbf{v}\}$, $\{\mathbf{u}, \mathbf{w}\}$, $\{\mathbf{u}, \mathbf{z}\}$, $\{\mathbf{v}, \mathbf{w}\}$, $\{\mathbf{v}, \mathbf{z}\}$, and $\{\mathbf{w}, \mathbf{z}\}$ each linearly independent? Why or why not?

b. Does the answer to Part (a) imply that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ is linearly independent?

c. To determine if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ is linearly dependent, is it wise to check if, say, $\mathbf{w}$ is a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{z}$?

d. Is $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ linearly dependent?

2. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly dependent set of vectors in $\mathbb{R}^n$ and $\mathbf{v}_4$ is a vector in $\mathbb{R}^n$. Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is also a linearly dependent set.

1.7 Exercises

In Exercises 1–4, determine if the vectors are linearly independent. Justify each answer.

1. $\begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 2 \\ -6 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 6 \end{bmatrix}$

2. $\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -8 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \end{bmatrix}$

4. $\begin{bmatrix} -1 \\ 4 \end{bmatrix}, \begin{bmatrix} -2 \\ 8 \end{bmatrix}$

In Exercises 5–8, determine if the columns of the matrix form a linearly independent set. Justify each answer.

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

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

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

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

In Exercises 9 and 10, (a) for what values of h is $\mathbf{v}_3$ in Span $\{\mathbf{v}_1, \mathbf{v}_2\}$, and (b) for what values of h is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly *dependent*? Justify each answer.

9. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -3 \\ 10 \\ -6 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -7 \\ h \end{bmatrix}$

10. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -5 \\ -3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ 10 \\ 6 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -10 \\ h \end{bmatrix}$

In Exercises 11–14, find the value(s) of h for which the vectors are linearly *dependent*. Justify each answer.

11. $\begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ h \end{bmatrix}$

12. $\begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} -6 \\ 7 \\ -3 \end{bmatrix}, \begin{bmatrix} 8 \\ h \\ 4 \end{bmatrix}$

13. $\begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ -9 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ h \\ -9 \end{bmatrix}$

14. $\begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} -5 \\ 7 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ h \end{bmatrix}$

Determine by inspection whether the vectors in Exercises 15–20 are linearly *independent*. Justify each answer.

15. $\begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 7 \end{bmatrix}$

16. $\begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix}, \begin{bmatrix} 6 \\ -3 \\ 9 \end{bmatrix}$

17. $\begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -6 \\ 5 \\ 4 \end{bmatrix}$

18. $\begin{bmatrix} 4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 8 \\ 1 \end{bmatrix}$

19. $\begin{bmatrix} -8 \\ 12 \\ -4 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$

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

In Exercises 21–28, mark each statement True or False **(T/F)**. Justify each answer on the basis of a careful reading of the text.

21. **(T/F)** The columns of a matrix A are linearly independent if the equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution.

22. **(T/F)** Two vectors are linearly dependent if and only if they lie on a line through the origin.

23. **(T/F)** If S is a linearly dependent set, then each vector is a linear combination of the other vectors in S.

24. **(T/F)** If a set contains fewer vectors than there are entries in the vectors, then the set is linearly independent.

25. **(T/F)** The columns of any 4×5 matrix are linearly dependent.

26. **(T/F)** If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, and if $\mathbf{z}$ is in Span $\{\mathbf{x}, \mathbf{y}\}$, then $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly dependent.

27. **(T/F)** If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, and if $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly dependent, then $\mathbf{z}$ is in Span $\{\mathbf{x}, \mathbf{y}\}$.

28. **(T/F)** If a set in $\mathbb{R}^n$ is linearly dependent, then the set contains more vectors than there are entries in each vector.

In Exercises 29–32, describe the possible echelon forms of the matrix. Use the notation of Example 1 in Section 1.2.

29. A is a 3×3 matrix with linearly independent columns.

30. A is a 2×2 matrix with linearly dependent columns.

31. A is a 4×2 matrix, $A = [\mathbf{a}_1 \quad \mathbf{a}_2]$, and $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_1$.

32. A is a 4×3 matrix, $A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3]$, such that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is linearly independent and $\mathbf{a}_3$ is not in Span $\{\mathbf{a}_1, \mathbf{a}_2\}$.

33. How many pivot columns must a 7×5 matrix have if its columns are linearly independent? Why?

34. How many pivot columns must a 5×7 matrix have if its columns span $\mathbb{R}^5$? Why?

35. Construct 3×2 matrices A and B such that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution and $B\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

36. a. Fill in the blank in the following statement: "If A is an $m \times n$ matrix, then the columns of A are linearly independent if and only if A has _____ pivot columns."

b. Explain why the statement in (a) is true.

Exercises 37 and 38 should be solved *without performing row operations*. [*Hint:* Write $A\mathbf{x} = \mathbf{0}$ as a vector equation.]

37. Given $A = \begin{bmatrix} 2 & 3 & 5 \\ -5 & 1 & -4 \\ -3 & -1 & -4 \\ 1 & 0 & 1 \end{bmatrix}$, observe that the third column is the sum of the first two columns. Find a nontrivial solution of $A\mathbf{x} = \mathbf{0}$.

38. Given $A = \begin{bmatrix} 4 & 1 & 6 \\ -7 & 5 & 3 \\ 9 & -3 & 3 \end{bmatrix}$, observe that the first column plus twice the second column equals the third column. Find a nontrivial solution of $A\mathbf{x} = \mathbf{0}$.

Each statement in Exercises 39–44 is either true (in all cases) or false (for at least one example). If false, construct a specific example to show that the statement is not always true. Such an example is called a *counterexample* to the statement. If a statement is true, give a justification. (One specific example cannot explain why a statement is always true. You will have to do more work here than in Exercises 21–28.)

39. **(T/F-C)** If $\mathbf{v}_1, \dots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3 = 2\mathbf{v}_1 + \mathbf{v}_2$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly dependent.

40. **(T/F-C)** If $\mathbf{v}_1, \dots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3 = \mathbf{0}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly dependent.

41. **(T/F-C)** If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in $\mathbb{R}^4$ and $\mathbf{v}_2$ is not a scalar multiple of $\mathbf{v}_1$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

42. **(T/F-C)** If $\mathbf{v}_1, \dots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3$ is *not* a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent.

43. **(T/F-C)** If $\mathbf{v}_1, \dots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is also linearly dependent.

44. **(T/F-C)** If $\mathbf{v}_1, \dots, \mathbf{v}_4$ are linearly independent vectors in $\mathbb{R}^4$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is also linearly independent. [*Hint:* Think about $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 + 0 \cdot \mathbf{v}_4 = \mathbf{0}$.]

45. Suppose A is an $m \times n$ matrix with the property that for all $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ has at most one solution. Use the definition of linear independence to explain why the columns of A must be linearly independent.

46. Suppose an $m \times n$ matrix A has n pivot columns. Explain why for each $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ has at most one solution. [*Hint:* Explain why $A\mathbf{x} = \mathbf{b}$ cannot have infinitely many solutions.]

T In Exercises 47 and 48, use as many columns of A as possible to construct a matrix B with the property that the equation $B\mathbf{x} = \mathbf{0}$ has only the trivial solution. Solve $B\mathbf{x} = \mathbf{0}$ to verify your work.

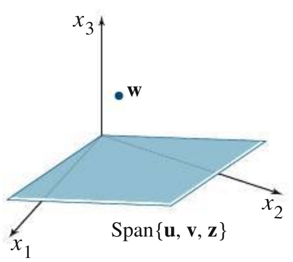
47. $A = \begin{bmatrix} 8 & -3 & 0 & -7 & 2 \\ -9 & 4 & 5 & 11 & -7 \\ 6 & -2 & 2 & -4 & 4 \\ 5 & -1 & 7 & 0 & 10 \end{bmatrix}$

48. $A = \begin{bmatrix} 12 & 10 & -6 & -3 & 7 & 10 \\ -7 & -6 & 4 & 7 & -9 & 5 \\ 9 & 9 & -9 & -5 & 5 & -1 \\ -4 & -3 & 1 & 6 & -8 & 9 \\ 8 & 7 & -5 & -9 & 11 & -8 \end{bmatrix}$

T **49.** With A and B as in Exercise 47 select a column $\mathbf{v}$ of A that was not used in the construction of B and determine if $\mathbf{v}$ is in the set spanned by the columns of B. (Describe your calculations.)

T **50.** Repeat Exercise 49 with the matrices A and B from Exercise 48. Then give an explanation for what you discover, assuming that B was constructed as specified.

STUDY GUIDE offers additional resources for mastering the concept of linear independence.



Solutions to Practice Problems

1. a. Yes. In each case, neither vector is a multiple of the other. Thus each set is linearly independent.

b. No. The observation in Part (a), by itself, says nothing about the linear independence of $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$.

c. No. When testing for linear independence, it is usually a poor idea to check if one selected vector is a linear combination of the others. It may happen that the selected vector is not a linear combination of the others and yet the whole set of vectors is linearly dependent. In this practice problem, $\mathbf{w}$ is not a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{z}$.

d. Yes, by Theorem 8. There are more vectors (four) than entries (three) in them.

2. Applying the definition of linearly dependent to $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ implies that there exist scalars c_1, c_2, and c_3, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}.$$

Adding $0\,\mathbf{v}_4 = \mathbf{0}$ to both sides of this equation results in

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + 0\,\mathbf{v}_4 = \mathbf{0}.$$

Since c_1, c_2, c_3 and 0 are not *all* zero, the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ satisfies the definition of a linearly dependent set.

1.8 Introduction to Linear Transformations

The difference between a matrix equation $A\mathbf{x} = \mathbf{b}$ and the associated vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ is merely a matter of notation. However, a matrix equation $A\mathbf{x} = \mathbf{b}$ can arise in linear algebra (and in applications such as computer graphics and signal processing) in a way that is not directly connected with linear combinations of vectors. This happens when we think of the matrix A as an object that "acts" on a vector $\mathbf{x}$ by multiplication to produce a new vector called $A\mathbf{x}$.

For instance, the equations

$$\underset{\substack{\uparrow\\ A}}{\begin{bmatrix}4&-3&1&3\\2&0&5&1\end{bmatrix}}\underset{\substack{\uparrow\\ \mathbf{x}}}{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}}=\underset{\substack{\uparrow\\ \mathbf{b}}}{\begin{bmatrix}5\\8\end{bmatrix}} \quad\text{and}\quad \underset{\substack{\uparrow\\ A}}{\begin{bmatrix}4&-3&1&3\\2&0&5&1\end{bmatrix}}\underset{\substack{\uparrow\\ \mathbf{u}}}{\begin{bmatrix}1\\4\\-1\\3\end{bmatrix}}=\underset{\substack{\uparrow\\ \mathbf{0}}}{\begin{bmatrix}0\\0\end{bmatrix}}$$

say that multiplication by A transforms $\mathbf{x}$ into $\mathbf{b}$ and transforms $\mathbf{u}$ into the zero vector. See Figure 1.

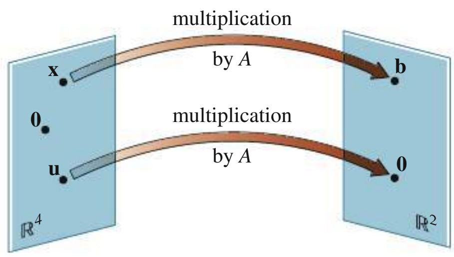


FIGURE 1 Transforming vectors via matrix multiplication.

From this new point of view, solving the equation $A\mathbf{x} = \mathbf{b}$ amounts to finding all vectors $\mathbf{x}$ in $\mathbb{R}^4$ that are transformed into the vector $\mathbf{b}$ in $\mathbb{R}^2$ under the "action" of multiplication by A.

The correspondence from $\mathbf{x}$ to $A\mathbf{x}$ is a *function* from one set of vectors to another. This concept generalizes the common notion of a function as a rule that transforms one real number into another.

A **transformation** (or **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T, and $\mathbb{R}^m$

is called the **codomain** of T. The notation $T : \mathbb{R}^n \to \mathbb{R}^m$ indicates that the domain of T is $\mathbb{R}^n$ and the codomain is $\mathbb{R}^m$. For $\mathbf{x}$ in $\mathbb{R}^n$, the vector $T(\mathbf{x})$ in $\mathbb{R}^m$ is called the **image** of $\mathbf{x}$ (under the action of T). The set of all images $T(\mathbf{x})$ is called the **range** of T. See Figure 2.

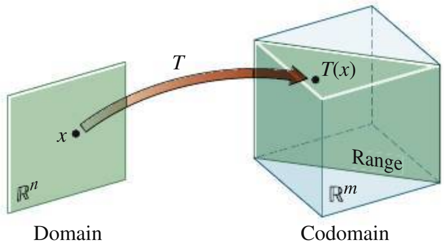
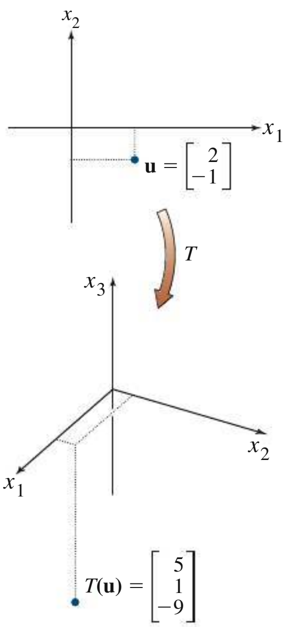


FIGURE 2 Domain, codomain, and range of $T : \mathbb{R}^n \to \mathbb{R}^m$.

The new terminology in this section is important because a dynamic view of matrix–vector multiplication is the key to understanding several ideas in linear algebra and to building mathematical models of physical systems that evolve over time. Such *dynamical systems* will be discussed in Sections 1.10, 4.8, and throughout Chapter 5.

Matrix Transformations

The rest of this section focuses on mappings associated with matrix multiplication. For each $\mathbf{x}$ in $\mathbb{R}^n$, $T(\mathbf{x})$ is computed as $A\mathbf{x}$, where A is an $m \times n$ matrix. For simplicity, we sometimes denote such a *matrix transformation* by $\mathbf{x} \mapsto A\mathbf{x}$. Observe that the domain of T is $\mathbb{R}^n$ when A has n columns and the codomain of T is $\mathbb{R}^m$ when each column of A has m entries. The range of T is the set of all linear combinations of the columns of A, because each image $T(\mathbf{x})$ is of the form $A\mathbf{x}$.



EXAMPLE 1 Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$, and define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$, so that

$$T(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{bmatrix}$$

a. Find $T(\mathbf{u})$, the image of $\mathbf{u}$ under the transformation T.
b. Find an $\mathbf{x}$ in $\mathbb{R}^2$ whose image under T is $\mathbf{b}$.
c. Is there more than one $\mathbf{x}$ whose image under T is $\mathbf{b}$?
d. Determine if $\mathbf{c}$ is in the range of the transformation T.

SOLUTION

a. Compute

$$T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$$

b. Solve $T(\mathbf{x}) = \mathbf{b}$ for $\mathbf{x}$. That is, solve $A\mathbf{x} = \mathbf{b}$, or

$$\begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix} \tag{1}$$

Using the method discussed in Section 1.4, row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & -.5 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1.5 \\ 0 & 1 & -.5 \\ 0 & 0 & 0 \end{bmatrix} \tag{2}$$

Hence $x_1 = 1.5$, $x_2 = -.5$, and $\mathbf{x} = \begin{bmatrix} 1.5 \\ -.5 \end{bmatrix}$. The image of this $\mathbf{x}$ under T is the given vector $\mathbf{b}$.

c. Any $\mathbf{x}$ whose image under T is $\mathbf{b}$ must satisfy equation (1). From (2), it is clear that equation (1) has a unique solution. So there is exactly one $\mathbf{x}$ whose image is $\mathbf{b}$.

d. The vector $\mathbf{c}$ is in the range of T if $\mathbf{c}$ is the image of some $\mathbf{x}$ in $\mathbb{R}^2$, that is, if $\mathbf{c} = T(\mathbf{x})$ for some $\mathbf{x}$. This is just another way of asking if the system $A\mathbf{x} = \mathbf{c}$ is consistent. To find the answer, row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 14 & -7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -35 \end{bmatrix}$$

The third equation, $0 = -35$, shows that the system is inconsistent. So $\mathbf{c}$ is *not* in the range of T. ■

The question in Example 1(c) is a *uniqueness* problem for a system of linear equations, translated here into the language of matrix transformations: Is $\mathbf{b}$ the image of a *unique* $\mathbf{x}$ in $\mathbb{R}^n$? Similarly, Example 1(d) is an *existence* problem: Does there *exist* an $\mathbf{x}$ whose image is $\mathbf{c}$?

The next two matrix transformations can be viewed geometrically. They reinforce the dynamic view of a matrix as something that transforms vectors into other vectors. Section 2.7 contains other interesting examples connected with computer graphics.

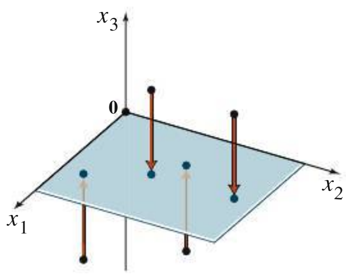


FIGURE 3
A projection transformation.

EXAMPLE 2 If $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ *projects* points in $\mathbb{R}^3$ onto the x_1x_2-plane because

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$$

See Figure 3. ■

sheep

sheared sheep

EXAMPLE 3 Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is called a **shear transformation**. It can be shown that if T acts on each point in the 2×2 square shown in Figure 4, then the set of images forms the sheared parallelogram. The key idea is to show that T maps line segments onto line segments (as shown in Exercise 35) and then to check that the corners of the square map onto the vertices of the parallelogram. For instance, the image of the point $\mathbf{u} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ is $T(\mathbf{u}) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, and the image of $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$. T deforms the square as if the top of the square were pushed to the right while the base is held fixed. Shear transformations appear in physics, geology, and crystallography. ■

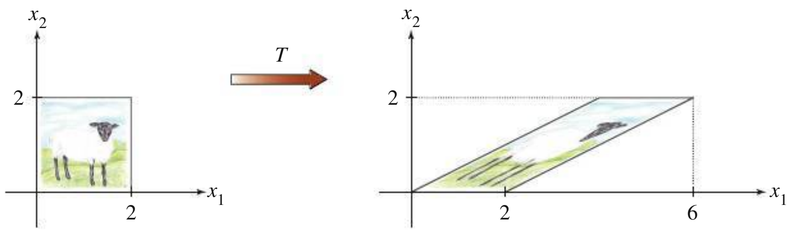


FIGURE 4 A shear transformation.

Linear Transformations

Theorem 5 in Section 1.4 shows that if A is $m \times n$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ has the properties

$$A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} \quad \text{and} \quad A(c\mathbf{u}) = cA\mathbf{u}$$

for all $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$ and all scalars c. These properties, written in function notation, identify the most important class of transformations in linear algebra.

DEFINITION

A transformation (or mapping) T is **linear** if

(i) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$, $\mathbf{v}$ in the domain of T;

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and all $\mathbf{u}$ in the domain of T.

Every matrix transformation is a linear transformation. Important examples of linear transformations that are not matrix transformations will be discussed in Chapters 4 and 5.

Linear transformations *preserve the operations of vector addition and scalar multiplication*. Property (i) says that the result $T(\mathbf{u}+\mathbf{v})$ of first adding $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ and then applying T is the same as first applying T to $\mathbf{u}$ and to $\mathbf{v}$ and then adding $T(\mathbf{u})$ and $T(\mathbf{v})$ in $\mathbb{R}^m$. These two properties lead easily to the following useful facts.

If T is a linear transformation, then

$$T(\mathbf{0}) = \mathbf{0} \tag{3}$$

and

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}) \tag{4}$$

for all vectors $\mathbf{u}$, $\mathbf{v}$ in the domain of T and all scalars c, d.

Property (3) follows from condition (ii) in the definition, because $T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}$. Property (4) requires both (i) and (ii):

$$T(c\mathbf{u} + d\mathbf{v}) = T(c\mathbf{u}) + T(d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$$

Observe that *if a transformation satisfies* (4) *for all* $\mathbf{u}$, $\mathbf{v}$ *and* c, d, *it must be linear*. (Set $c = d = 1$ for preservation of addition, and set $d = 0$ for preservation of scalar multiplication.) Repeated application of (4) produces a useful generalization:

$$T(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + \cdots + c_pT(\mathbf{v}_p) \tag{5}$$

In engineering and physics, (5) is referred to as a *superposition principle*. Think of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ as signals that go into a system and $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_p)$ as the responses of that system to the signals. The system satisfies the superposition principle if whenever an input is expressed as a linear combination of such signals, the system's response is *the same* linear combination of the responses to the individual signals. We will return to this idea in Chapter 4.

EXAMPLE 4 Given a scalar r, define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = r\mathbf{x}$. T is called a **contraction** when $0 \le r \le 1$ and a **dilation** when $r > 1$. Let $r = 3$, and show that T is a linear transformation.

SOLUTION Let $\mathbf{u}$, $\mathbf{v}$ be in $\mathbb{R}^2$ and let c, d be scalars. Then

$$\begin{aligned} T(c\mathbf{u} + d\mathbf{v}) &= 3(c\mathbf{u} + d\mathbf{v}) && \text{Definition of } T \\ &= 3c\mathbf{u} + 3d\mathbf{v} && \\ &= c(3\mathbf{u}) + d(3\mathbf{v}) && \Big\} \text{ Vector arithmetic} \\ &= cT(\mathbf{u}) + dT(\mathbf{v}) \end{aligned}$$

Thus T is a linear transformation because it satisfies (4). See Figure 5. ■

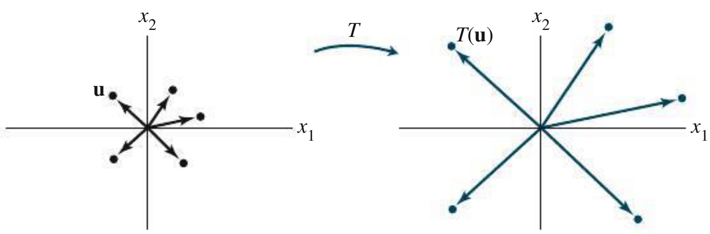


FIGURE 5 A dilation transformation.

EXAMPLE 5 Define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

Find the images under T of $\mathbf{u} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$.

SOLUTION

$$T(\mathbf{u}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}, \qquad T(\mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix},$$

$$T(\mathbf{u} + \mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix}$$

Note that $T(\mathbf{u} + \mathbf{v})$ is obviously equal to $T(\mathbf{u}) + T(\mathbf{v})$. It appears from Figure 6 that T rotates $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} + \mathbf{v}$ counterclockwise about the origin through 90°. In fact, T transforms the entire parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ into the one determined by $T(\mathbf{u})$ and $T(\mathbf{v})$. (See Exercise 36.) ■

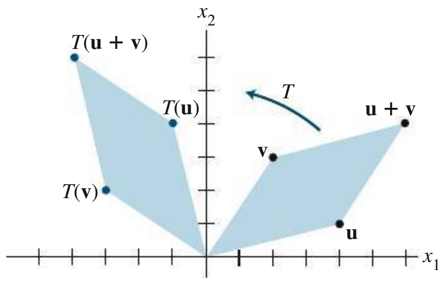


FIGURE 6 A rotation transformation.

The final example is not geometrical; instead, it shows how a linear mapping can transform one type of data into another.

EXAMPLE 6 A company manufactures two products, B and C. Using data from Example 7 in Section 1.3, we construct a "unit cost" matrix, $U = [\,\mathbf{b}\quad \mathbf{c}\,]$, whose columns describe the "costs per dollar of output" for the products:

$$\begin{array}{cc} & \text{Product} \\ & \begin{array}{cc} \text{B} & \text{C} \end{array} \\ U = & \begin{bmatrix} .45 & .40 \\ .25 & .30 \\ .15 & .15 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{Materials} \\ \text{Labor} \\ \text{Overhead} \end{array}$$

Let $\mathbf{x} = (x_1, x_2)$ be a "production" vector, corresponding to x_1 dollars of product B and x_2 dollars of product C, and define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$T(\mathbf{x}) = U\mathbf{x} = x_1 \begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} + x_2 \begin{bmatrix} .40 \\ .30 \\ .15 \end{bmatrix} = \begin{bmatrix} \text{Total cost of materials} \\ \text{Total cost of labor} \\ \text{Total cost of overhead} \end{bmatrix}$$

The mapping T transforms a list of production quantities (measured in dollars) into a list of total costs. The linearity of this mapping is reflected in two ways:

1. If production is increased by a factor of, say, 4, from $\mathbf{x}$ to $4\mathbf{x}$, then the costs will increase by the same factor, from $T(\mathbf{x})$ to $4T(\mathbf{x})$.
2. If $\mathbf{x}$ and $\mathbf{y}$ are production vectors, then the total cost vector associated with the combined production $\mathbf{x} + \mathbf{y}$ is precisely the sum of the cost vectors $T(\mathbf{x})$ and $T(\mathbf{y})$. ■

Practice Problems

1. Suppose $T : \mathbb{R}^5 \to \mathbb{R}^2$ and $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A and for each $\mathbf{x}$ in $\mathbb{R}^5$. How many rows and columns does A have?
2. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Give a geometric description of the transformation $\mathbf{x} \mapsto A\mathbf{x}$.
3. The line segment from $\mathbf{0}$ to a vector $\mathbf{u}$ is the set of points of the form $t\mathbf{u}$, where $0 \le t \le 1$. Show that a linear transformation T maps this segment into the segment between $\mathbf{0}$ and $T(\mathbf{u})$.

1.8 Exercises

1. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, and define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$. Find the images under T of $\mathbf{u} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$.

2. Let $A = \begin{bmatrix} .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$. Find $T(\mathbf{u})$ and $T(\mathbf{v})$.

In Exercises 3–6, with T defined by $T(\mathbf{x}) = A\mathbf{x}$, find a vector $\mathbf{x}$ whose image under T is $\mathbf{b}$, and determine whether $\mathbf{x}$ is unique.

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

4. $A = \begin{bmatrix} 1 & -3 & 2 \\ 0 & 1 & -4 \\ 3 & -5 & -9 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 6 \\ -7 \\ -9 \end{bmatrix}$

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

6. $A = \begin{bmatrix} 1 & -2 & 1 \\ 3 & -4 & 5 \\ 0 & 1 & 1 \\ -3 & 5 & -4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 9 \\ 3 \\ -6 \end{bmatrix}$

7. Let A be a 6×5 matrix. What must a and b be in order to define $T : \mathbb{R}^a \to \mathbb{R}^b$ by $T(\mathbf{x}) = A\mathbf{x}$?

8. How many rows and columns must a matrix A have in order to define a mapping from $\mathbb{R}^4$ into $\mathbb{R}^5$ by the rule $T(\mathbf{x}) = A\mathbf{x}$?

For Exercises 9 and 10, find all $\mathbf{x}$ in $\mathbb{R}^4$ that are mapped into the zero vector by the transformation $\mathbf{x} \mapsto A\mathbf{x}$ for the given matrix A.

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

10. $A = \begin{bmatrix} 1 & 3 & 9 & 2 \\ 1 & 0 & 3 & -4 \\ 0 & 1 & 2 & 3 \\ -2 & 3 & 0 & 5 \end{bmatrix}$

11. Let $\mathbf{b} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$, and let A be the matrix in Exercise 9. Is $\mathbf{b}$ in the range of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$? Why or why not?

12. Let $\mathbf{b} = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 4 \end{bmatrix}$, and let A be the matrix in Exercise 10. Is $\mathbf{b}$ in the range of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$? Why or why not?

In Exercises 13–16, use a rectangular coordinate system to plot $\mathbf{u} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, and their images under the given transformation T. (Make a separate and reasonably large sketch for each exercise.) Describe geometrically what T does to each vector $\mathbf{x}$ in $\mathbb{R}^2$.

13. $T(\mathbf{x}) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

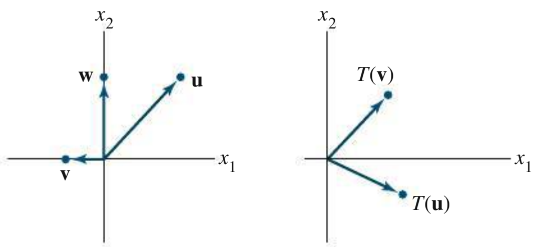
14. $T(\mathbf{x}) = \begin{bmatrix} .5 & 0 \\ 0 & .5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

15. $T(\mathbf{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

16. $T(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

17. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{u} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ into $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and maps $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ into $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$. Use the fact that T is linear to find the images under T of $3\mathbf{u}$, $2\mathbf{v}$, and $3\mathbf{u} + 2\mathbf{v}$.

18. The figure shows vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, along with the images $T(\mathbf{u})$ and $T(\mathbf{v})$ under the action of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$. Copy this figure carefully, and draw the image $T(\mathbf{w})$ as accurately as possible. [*Hint:* First, write $\mathbf{w}$ as a linear combination of $\mathbf{u}$ and $\mathbf{v}$.]



19. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{y}_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, and $\mathbf{y}_2 = \begin{bmatrix} -1 \\ 6 \end{bmatrix}$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{e}_1$ into $\mathbf{y}_1$ and maps $\mathbf{e}_2$ into $\mathbf{y}_2$. Find the images of $\begin{bmatrix} 5 \\ -3 \end{bmatrix}$ and $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

20. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} 7 \\ -3 \end{bmatrix}$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{x}$ into $x_1\mathbf{v}_1 + x_2\mathbf{v}_2$. Find a matrix A such that $T(\mathbf{x})$ is $A\mathbf{x}$ for each $\mathbf{x}$.

In Exercises 21–30, mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** A linear transformation is a special type of function.

22. **(T/F)** Every matrix transformation is a linear transformation.

23. **(T/F)** If A is a 3×5 matrix and T is a transformation defined by $T(\mathbf{x}) = A\mathbf{x}$, then the domain of T is $\mathbb{R}^3$.

24. **(T/F)** The codomain of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is the set of all linear combinations of the columns of A.

25. **(T/F)** If A is an $m \times n$ matrix, then the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is $\mathbb{R}^m$.

26. **(T/F)** If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and if $\mathbf{c}$ is in $\mathbb{R}^m$, then a uniqueness question is "Is $\mathbf{c}$ in the range of T?"

27. **(T/F)** Every linear transformation is a matrix transformation.

28. **(T/F)** A linear transformation preserves the operations of vector addition and scalar multiplication.

29. **(T/F)** A transformation T is linear if and only if $T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2)$ for all $\mathbf{v}_1$ and $\mathbf{v}_2$ in the domain of T and for all scalars c_1 and c_2.

30. **(T/F)** The superposition principle is a physical description of a linear transformation.

31. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that reflects each point through the x_1-axis. (See Practice Problem 2.) Make two sketches similar to Figure 6 that illustrate properties (i) and (ii) of a linear transformation.

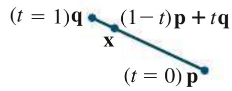
32. Suppose vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ span $\mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Suppose $T(\mathbf{v}_i) = \mathbf{0}$ for $i = 1, \ldots, p$. Show that T is the zero transformation. That is, show that if $\mathbf{x}$ is any vector in $\mathbb{R}^n$, then $T(\mathbf{x}) = \mathbf{0}$.

33. Given $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{p}$ in $\mathbb{R}^n$, the line through $\mathbf{p}$ in the direction of $\mathbf{v}$ has the parametric equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$. Show that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ maps this line onto another line or onto a single point (a *degenerate line*).

34. Let $\mathbf{u}$ and $\mathbf{v}$ be linearly independent vectors in $\mathbb{R}^3$, and let P be the plane through $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{0}$. The parametric equation of P is $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$ (with s, t in $\mathbb{R}$). Show that a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ maps P onto a plane through $\mathbf{0}$, or onto a line through $\mathbf{0}$, or onto just the origin in $\mathbb{R}^3$. What must be true about $T(\mathbf{u})$ and $T(\mathbf{v})$ in order for the image of the plane P to be a plane?

35. a. Show that the line through vectors $\mathbf{p}$ and $\mathbf{q}$ in $\mathbb{R}^n$ may be written in the parametric form $\mathbf{x} = (1-t)\mathbf{p} + t\mathbf{q}$. (Refer to the figure with Exercises 25 and 26 in Section 1.5.)

b. The line segment from $\mathbf{p}$ to $\mathbf{q}$ is the set of points of the form $(1-t)\mathbf{p} + t\mathbf{q}$ for $0 \le t \le 1$ (as shown in the figure below). Show that a linear transformation T maps this line segment onto a line segment or onto a single point.



36. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. It can be shown that the set P of all points in the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ has the form $a\mathbf{u} + b\mathbf{v}$, for $0 \le a \le 1, 0 \le b \le 1$. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Explain why the image of a point in P under the transformation T lies in the parallelogram determined by $T(\mathbf{u})$ and $T(\mathbf{v})$.

37. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = mx + b$.

a. Show that f is a linear transformation when $b = 0$.

b. Find a property of a linear transformation that is violated when $b \neq 0$.

c. Why is f called a linear function?

38. An *affine transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ has the form $T(x) = A\mathbf{x} + \mathbf{b}$, with A an $m \times n$ matrix and $\mathbf{b}$ in $\mathbb{R}^m$. Show that T is *not* a linear transformation when $\mathbf{b} \neq \mathbf{0}$. (Affine transformations are important in computer graphics.)

39. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be a linearly dependent set in $\mathbb{R}^n$. Explain why the set $\{T(\mathbf{v}_1), T(\mathbf{v}_2), T(\mathbf{v}_3)\}$ is linearly dependent.

In Exercises 40–44, column vectors are written as rows, such as $\mathbf{x} = (x_1, x_2)$, and $T(\mathbf{x})$ is written as $T(x_1, x_2)$.

40. Show that the transformation T defined by $T(x_1, x_2) = (4x_1 - 2x_2, 3|x_2|)$ is not linear.

41. Show that the transformation T defined by $T(x_1, x_2) = (2x_1 - 3x_2, x_1 + 4, 5x_2)$ is not linear.

42. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that if T maps two linearly independent vectors onto a linearly dependent set, then the equation $T(\mathbf{x}) = \mathbf{0}$ has a nontrivial solution. [*Hint:* Suppose $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are linearly independent and yet $T(\mathbf{u})$ and $T(\mathbf{v})$ are linearly dependent. Then $c_1T(\mathbf{u}) + c_2T(\mathbf{v}) = \mathbf{0}$ for some weights c_1 and c_2, not both zero. Use this equation.]

43. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation that reflects each vector $\mathbf{x} = (x_1, x_2, x_3)$ through the plane $x_3 = 0$ onto $T(\mathbf{x}) = (x_1, x_2, -x_3)$. Show that T is a linear transformation. [See Example 4 for ideas.]

44. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation that projects each vector $\mathbf{x} = (x_1, x_2, x_3)$ onto the plane $x_2 = 0$, so $T(\mathbf{x}) = (x_1, 0, x_3)$. Show that T is a linear transformation.

T In Exercises 45 and 46, the given matrix determines a linear transformation T. Find all $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{0}$.

45. $\begin{bmatrix} 4 & -2 & 5 & -5 \\ -9 & 7 & -8 & 0 \\ -6 & 4 & 5 & 3 \\ 5 & -3 & 8 & -4 \end{bmatrix}$ **46.** $\begin{bmatrix} -9 & -4 & -9 & 4 \\ 5 & -8 & -7 & 6 \\ 7 & 11 & 16 & -9 \\ 9 & -7 & -4 & 5 \end{bmatrix}$

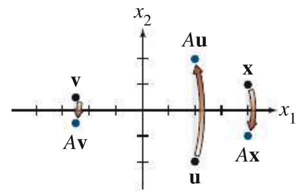
T 47. Let $\mathbf{b} = \begin{bmatrix} 7 \\ 5 \\ 9 \\ 7 \end{bmatrix}$ and let A be the matrix in Exercise 45. Is $\mathbf{b}$ in the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$? If so, find an $\mathbf{x}$ whose image under the transformation is $\mathbf{b}$.

T 48. Let $\mathbf{b} = \begin{bmatrix} -7 \\ -7 \\ 13 \\ -5 \end{bmatrix}$ and let A be the matrix in Exercise 46. Is $\mathbf{b}$ in the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$? If so, find an $\mathbf{x}$ whose image under the transformation is $\mathbf{b}$.

STUDY GUIDE offers additional resources for mastering linear transformations.

Solutions to Practice Problems

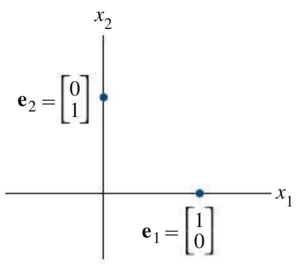
1. A must have five columns for $A\mathbf{x}$ to be defined. A must have two rows for the codomain of T to be $\mathbb{R}^2$.
2. Plot some random points (vectors) on graph paper to see what happens. A point such as $(4, 1)$ maps into $(4, -1)$. The transformation $\mathbf{x} \mapsto A\mathbf{x}$ reflects points through the x-axis (or x_1-axis).
3. Let $\mathbf{x} = t\mathbf{u}$ for some t such that $0 \le t \le 1$. Since T is linear, $T(t\mathbf{u}) = t\,T(\mathbf{u})$, which is a point on the line segment between $\mathbf{0}$ and $T(\mathbf{u})$.



The transformation $\mathbf{x} \mapsto A\mathbf{x}$.

1.9 The Matrix of a Linear Transformation

Whenever a linear transformation T arises geometrically or is described in words, we usually want a "formula" for $T(\mathbf{x})$. The discussion that follows shows that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is actually a matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$ and that important properties of T are intimately related to familiar properties of A. The key to finding A is to observe that T is completely determined by what it does to the columns of the $n \times n$ identity matrix I_n.



EXAMPLE 1 The columns of $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Suppose T is a linear transformation from $\mathbb{R}^2$ into $\mathbb{R}^3$ such that

$$T(\mathbf{e}_1) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}$$

With no additional information, find a formula for the image of an arbitrary $\mathbf{x}$ in $\mathbb{R}^2$.

SOLUTION Write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 \tag{1}$$

Since T is a *linear* transformation,

$$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) \tag{2}$$

$$= x_1\begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} + x_2\begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 5x_1 - 3x_2 \\ -7x_1 + 8x_2 \\ 2x_1 + 0 \end{bmatrix}$$ ■

The step from equation (1) to equation (2) explains why knowledge of $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ is sufficient to determine $T(\mathbf{x})$ for any $\mathbf{x}$. Moreover, since (2) expresses $T(\mathbf{x})$ as a linear combination of vectors, we can put these vectors into the columns of a matrix A and write (2) as

$$T(\mathbf{x}) = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\mathbf{x}$$

THEOREM 10

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that

$$T(\mathbf{x}) = A\mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n$$

In fact, A is the $m \times n$ matrix whose jth column is the vector $T(\mathbf{e}_j)$, where $\mathbf{e}_j$ is the jth column of the identity matrix in $\mathbb{R}^n$:

$$A = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \tag{3}$$

PROOF Write $\mathbf{x} = I_n\mathbf{x} = [\,\mathbf{e}_1 \ \cdots \ \mathbf{e}_n\,]\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, and use the linearity of T to compute

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + \cdots + x_nT(\mathbf{e}_n) \\ &= \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x} \end{aligned}$$

The uniqueness of A is treated in Exercise 41. ■

The matrix A in (3) is called the **standard matrix for the linear transformation** T.

We know now that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be viewed as a matrix transformation, and vice versa. The term *linear transformation* focuses on a property of a mapping, while *matrix transformation* describes how such a mapping is implemented, as Examples 2 and 3 illustrate.

EXAMPLE 2 Find the standard matrix A for the dilation transformation $T(\mathbf{x}) = 3\mathbf{x}$, for $\mathbf{x}$ in $\mathbb{R}^2$.

SOLUTION Write

$$T(\mathbf{e}_1) = 3\mathbf{e}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = 3\mathbf{e}_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$ ■

EXAMPLE 3 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates each point in $\mathbb{R}^2$ about the origin through an angle φ, with counterclockwise rotation for a positive angle. We could show geometrically that such a transformation is linear. (See Figure 6 in Section 1.8.) Find the standard matrix A of this transformation.

SOLUTION $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ rotates into $\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}$, and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ rotates into $\begin{bmatrix} -\sin\varphi \\ \cos\varphi \end{bmatrix}$. See Figure 1. By Theorem 10,

$$A = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

Example 5 in Section 1.8 is a special case of this transformation, with $\varphi = \pi/2$. ■

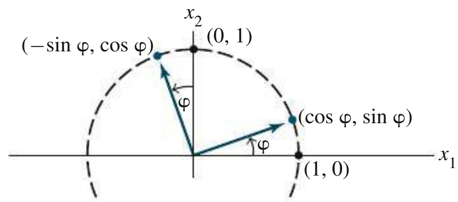


FIGURE 1 A rotation transformation.

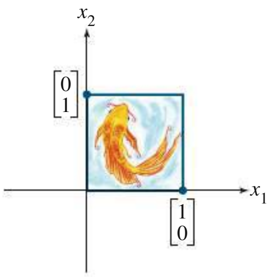


FIGURE 2
The unit square.

Geometric Linear Transformations of $\mathbb{R}^2$

Examples 2 and 3 illustrate linear transformations that are described geometrically. Tables 1–4 illustrate other common geometric linear transformations of the plane. Because the transformations are linear, they are determined completely by what they do to the columns of I_2. Instead of showing only the images of $\mathbf{e}_1$ and $\mathbf{e}_2$, the tables show what a transformation does to the unit square (Figure 2).

Other transformations can be constructed from those listed in Tables 1–4 by applying one transformation after another. For instance, a horizontal shear could be followed by a reflection in the x_2-axis. Section 2.1 will show that such a *composition* of linear transformations is linear. (Also, see Exercise 44.)

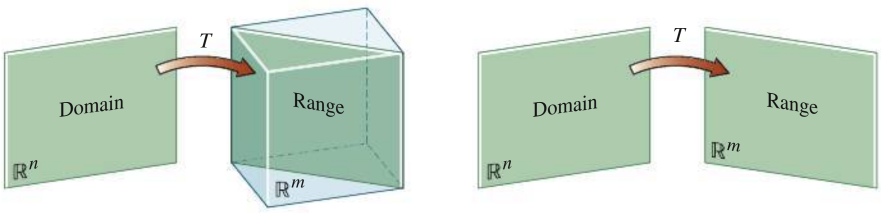
Existence and Uniqueness Questions

The concept of a linear transformation provides a new way to understand the existence and uniqueness questions asked earlier. The next two definitions give the appropriate terminology for transformations.

DEFINITION

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at least one* $\mathbf{x}$ in $\mathbb{R}^n$.

Equivalently, T is onto $\mathbb{R}^m$ when the range of T is all of the codomain $\mathbb{R}^m$. That is, T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if, for each $\mathbf{b}$ in the codomain $\mathbb{R}^m$, there exists at least one solution of $T(\mathbf{x}) = \mathbf{b}$. "Does T map $\mathbb{R}^n$ onto $\mathbb{R}^m$?" is an existence question. The mapping T is *not* onto when there is some $\mathbf{b}$ in $\mathbb{R}^m$ for which the equation $T(\mathbf{x}) = \mathbf{b}$ has no solution. See Figure 3.



T is *not* onto $\mathbb{R}^m$ T is onto $\mathbb{R}^m$

FIGURE 3 Is the range of T all of $\mathbb{R}^m$?

TABLE 1 Reflections

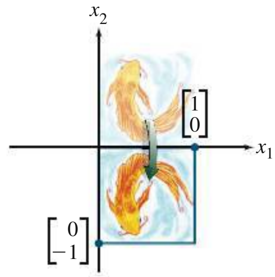
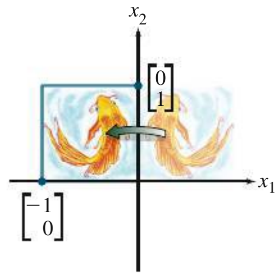
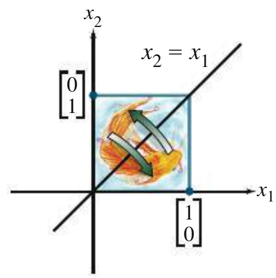
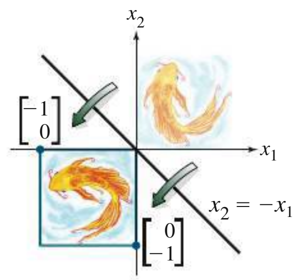
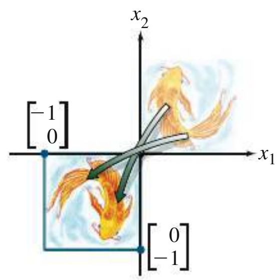
Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$
Reflection through the x_2-axis		$\begin{bmatrix}-1 & 0\\0 & 1\end{bmatrix}$
Reflection through the line $x_2 = x_1$		$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$
Reflection through the line $x_2 = -x_1$		$\begin{bmatrix}0 & -1\\-1 & 0\end{bmatrix}$
Reflection through the origin		$\begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}$

TABLE 2 Contractions and Expansions

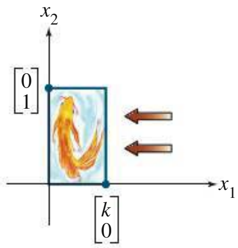
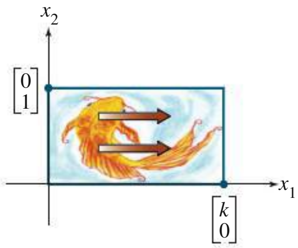
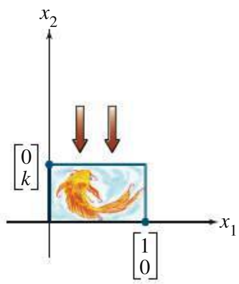
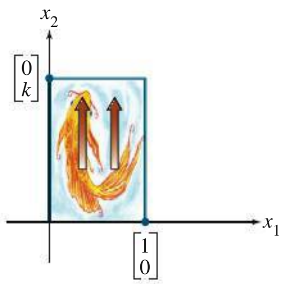
Transformation	Image of the Unit Square		Standard Matrix
Horizontal contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix}k & 0\\0 & 1\end{bmatrix}$
Vertical contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix}1 & 0\\0 & k\end{bmatrix}$

TABLE 3 Shears

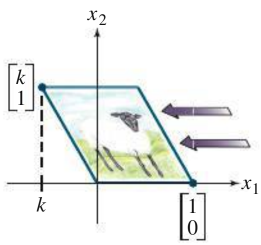
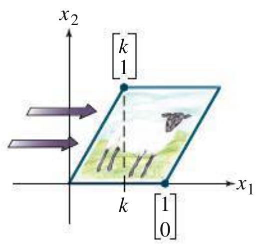
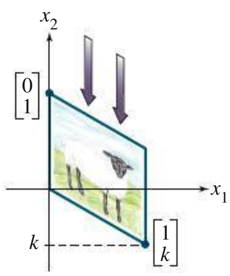
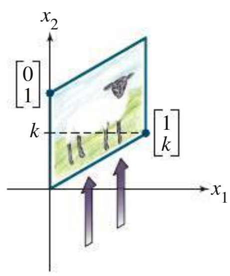
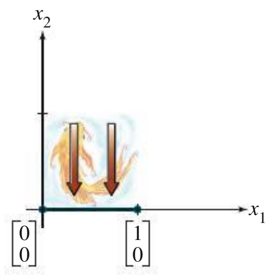
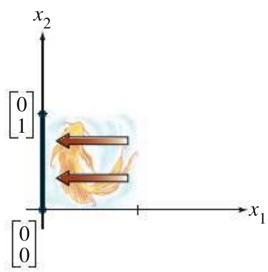
Transformation	Image of the Unit Square		Standard Matrix
Horizontal shear	 $k < 0$	 $k > 0$	$\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$
Vertical shear	 $k < 0$	 $k > 0$	$\begin{bmatrix}1 & 0\\k & 1\end{bmatrix}$

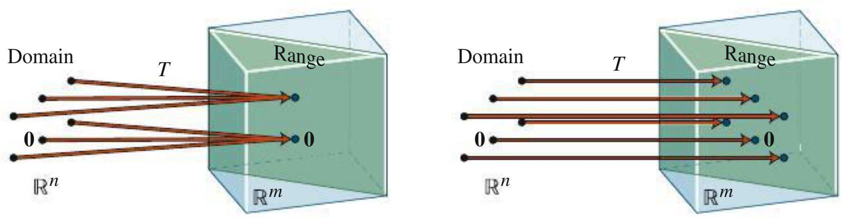
TABLE 4 Projections

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

DEFINITION

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **one-to-one** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at most one* $\mathbf{x}$ in $\mathbb{R}^n$.

Equivalently, T is one-to-one if, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has either a unique solution or none at all. "Is T one-to-one?" is a uniqueness question. The mapping T is *not* one-to-one when some $\mathbf{b}$ in $\mathbb{R}^m$ is the image of more than one vector in $\mathbb{R}^n$. If there is no such $\mathbf{b}$, then T is one-to-one. See Figure 4.



T is *not* one-to-one $\qquad$ T is one-to-one

FIGURE 4 Is every $\mathbf{b}$ the image of at most one vector?

The projection transformations shown in Table 4 are *not* one-to-one and do *not* map $\mathbb{R}^2$ onto $\mathbb{R}^2$. The transformations in Tables 1, 2, and 3 are one-to-one *and* do map $\mathbb{R}^2$ onto $\mathbb{R}^2$. Other possibilities are shown in the two examples below.

Example 4 and the theorems that follow show how the function properties of being one-to-one and mapping onto are related to important concepts studied earlier in this chapter.

EXAMPLE 4 Let T be the linear transformation whose standard matrix is

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

Does T map $\mathbb{R}^4$ onto $\mathbb{R}^3$? Is T a one-to-one mapping?

SOLUTION Since A happens to be in echelon form, we can see at once that A has a pivot position in each row. By Theorem 4 in Section 1.4, for each $\mathbf{b}$ in $\mathbb{R}^3$, the equation $A\mathbf{x} = \mathbf{b}$ is consistent. In other words, the linear transformation T maps $\mathbb{R}^4$ (its domain) onto $\mathbb{R}^3$. However, since the equation $A\mathbf{x} = \mathbf{b}$ has a free variable (because there are four variables and only three basic variables), each $\mathbf{b}$ is the image of more than one $\mathbf{x}$. That is, T is *not* one-to-one. ■

THEOREM 11

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if the equation $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Remark: To prove a theorem that says "statement P is true if and only if statement Q is true," one must establish two things: (1) If P is true, then Q is true and (2) If Q is true, then P is true. The second requirement can also be established by showing (2a): If P is false, then Q is false. (This is called contrapositive reasoning.) This proof uses (1) and (2a) to show that P and Q are either both true or both false.

PROOF Since T is linear, $T(\mathbf{0}) = \mathbf{0}$. If T is one-to-one, then the equation $T(\mathbf{x}) = \mathbf{0}$ has at most one solution and hence only the trivial solution. If T is not one-to-one, then there is a $\mathbf{b}$ that is the image of at least two different vectors in $\mathbb{R}^n$—say, $\mathbf{u}$ and $\mathbf{v}$. That is, $T(\mathbf{u}) = \mathbf{b}$ and $T(\mathbf{v}) = \mathbf{b}$. But then, since T is linear,

$$T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

The vector $\mathbf{u} - \mathbf{v}$ is not zero, since $\mathbf{u} \neq \mathbf{v}$. Hence the equation $T(\mathbf{x}) = \mathbf{0}$ has more than one solution. So, either the two conditions in the theorem are both true or they are both false. ■

THEOREM 12

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let A be the standard matrix for T. Then:

a. T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$;

b. T is one-to-one if and only if the columns of A are linearly independent.

Remark: "If and only if" statements can be linked together. For example if "P if and only if Q" is known and "Q if and only if R" is known, then one can conclude "P if and only if R." This strategy is used repeatedly in this proof.

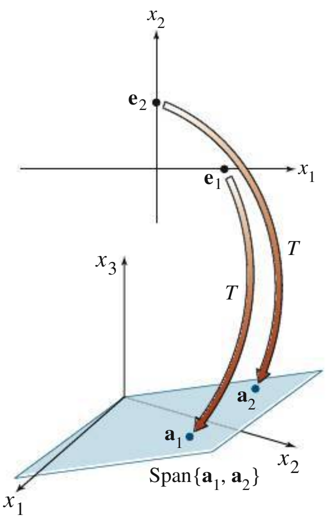
PROOF

a. By Theorem 4 in Section 1.4, the columns of A span $\mathbb{R}^m$ if and only if for each $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ is consistent—in other words, if and only if for every $\mathbf{b}$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at least one solution. This is true if and only if T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.

b. The equations $T(\mathbf{x}) = \mathbf{0}$ and $A\mathbf{x} = \mathbf{0}$ are the same except for notation. So, by Theorem 11, T is one-to-one if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This happens if and only if the columns of A are linearly independent, as was already noted in the boxed statement (3) in Section 1.7. ■

Statement (a) in Theorem 12 is equivalent to the statement "T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if every vector in $\mathbb{R}^m$ is a linear combination of the columns of A." See Theorem 4 in Section 1.4.

In the next example and in some exercises that follow, column vectors are written in rows, such as $\mathbf{x} = (x_1, x_2)$, and $T(\mathbf{x})$ is written as $T(x_1, x_2)$ instead of the more formal $T((x_1, x_2))$.



The transformation T is not onto $\mathbb{R}^3$.

EXAMPLE 5 Let $T(x_1, x_2) = (3x_1 + x_2, 5x_1 + 7x_2, x_1 + 3x_2)$. Show that T is a one-to-one linear transformation. Does T map $\mathbb{R}^2$ onto $\mathbb{R}^3$?

SOLUTION When $\mathbf{x}$ and $T(\mathbf{x})$ are written as column vectors, you can determine the standard matrix of T by inspection, visualizing the row–vector computation of each entry in $A\mathbf{x}$.

$$T(\mathbf{x}) = \begin{bmatrix} 3x_1 + x_2 \\ 5x_1 + 7x_2 \\ x_1 + 3x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} ? & ? \\ ? & ? \\ ? & ? \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4}$$

So T is indeed a linear transformation, with its standard matrix A shown in (4). The columns of A are linearly independent because they are not multiples. By Theorem 12(b), T is one-to-one. To decide if T is onto $\mathbb{R}^3$, examine the span of the columns of A. Since A is 3×2, the columns of A span $\mathbb{R}^3$ if and only if A has 3 pivot positions, by Theorem 4. This is impossible, since A has only 2 columns. So the columns of A do not span $\mathbb{R}^3$, and the associated linear transformation is not onto $\mathbb{R}^3$. ■

Practice Problems

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that first performs a horizontal shear that maps $\mathbf{e}_2$ into $\mathbf{e}_2 - .5\mathbf{e}_1$ (but leaves $\mathbf{e}_1$ unchanged) and then reflects the result through the x_2-axis. Assuming that T is linear, find its standard matrix. [*Hint:* Determine the final location of the images of $\mathbf{e}_1$ and $\mathbf{e}_2$.]
2. Suppose A is a 7×5 matrix with 5 pivots. Let $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation from $\mathbb{R}^5$ into $\mathbb{R}^7$. Is T a one-to-one linear transformation? Is T onto $\mathbb{R}^7$?

1.9 Exercises

In Exercises 1–10, assume that T is a linear transformation. Find the standard matrix of T.

1. $T : \mathbb{R}^2 \to \mathbb{R}^4$, $T(\mathbf{e}_1) = (2, 1, 2, 1)$ and $T(\mathbf{e}_2) = (-5, 2, 0, 0)$, where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$.
2. $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T(\mathbf{e}_1) = (1, 3)$, $T(\mathbf{e}_2) = (4, 2)$, and $T(\mathbf{e}_3) = (-5, 4)$, where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are the columns of the 3×3 identity matrix.
3. $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates points (about the origin) through $3\pi/2$ radians (in the counterclockwise direction).

4. $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates points (about the origin) through $-\pi/4$ radians (since the number is negative, the actual rotation is clockwise). [*Hint:* $T(\mathbf{e}_1) = (1/\sqrt{2}, -1/\sqrt{2})$.]

5. $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a vertical shear transformation that maps $\mathbf{e}_1$ into $\mathbf{e}_1 - 2\mathbf{e}_2$ but leaves the vector $\mathbf{e}_2$ unchanged.

6. $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a horizontal shear transformation that leaves $\mathbf{e}_1$ unchanged and maps $\mathbf{e}_2$ into $\mathbf{e}_2 + 3\mathbf{e}_1$.

7. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first rotates points through $-3\pi/4$ radians (since the number is negative, the actual rotation is clockwise) and then reflects points through the horizontal x_1-axis. [*Hint:* $T(\mathbf{e}_1) = (-1/\sqrt{2}, 1/\sqrt{2})$.]

8. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the horizontal x_1-axis and then reflects points through the line $x_2 = x_1$.

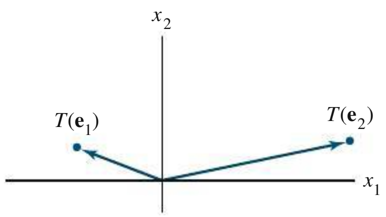
9. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first performs a horizontal shear that transforms $\mathbf{e}_2$ into $\mathbf{e}_2 - 3\mathbf{e}_1$ (leaving $\mathbf{e}_1$ unchanged) and then reflects points through the line $x_2 = -x_1$.

10. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the vertical x_2-axis and then rotates points $3\pi/2$ radians.

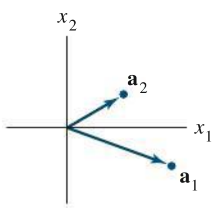
11. A linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the x_1-axis and then reflects points through the x_2-axis. Show that T can also be described as a linear transformation that rotates points about the origin. What is the angle of that rotation?

12. Show that the transformation in Exercise 8 is merely a rotation about the origin. What is the angle of the rotation?

13. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation such that $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ are the vectors shown in the figure. Using the figure, sketch the vector $T(2, 1)$.



14. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with standard matrix $A = [\mathbf{a}_1 \ \ \mathbf{a}_2]$, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are shown in the figure. Using the figure, draw the image of $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ under the transformation T.



In Exercises 15 and 16, fill in the missing entries of the matrix, assuming that the equation holds for all values of the variables.

15. $$\begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 3x_3 \\ 4x_1 \\ x_1 - x_2 + x_3 \end{bmatrix}$$

16. $$\begin{bmatrix} ? & ? \\ ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ -2x_1 + x_2 \\ x_1 \end{bmatrix}$$

In Exercises 17–20, show that T is a linear transformation by finding a matrix that implements the mapping. Note that $x_1, x_2, \ldots$ are not vectors but are entries in vectors.

17. $T(x_1, x_2, x_3, x_4) = (0, x_1 + x_2, x_2 + x_3, x_3 + x_4)$

18. $T(x_1, x_2) = (2x_2 - 3x_1, x_1 - 4x_2, 0, x_2)$

19. $T(x_1, x_2, x_3) = (x_1 - 5x_2 + 4x_3, x_2 - 6x_3)$

20. $T(x_1, x_2, x_3, x_4) = 2x_1 + 3x_3 - 4x_4 \qquad (T : \mathbb{R}^4 \to \mathbb{R})$

21. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T(x_1, x_2) = (x_1 + x_2, 4x_1 + 5x_2)$. Find $\mathbf{x}$ such that $T(\mathbf{x}) = (3, 8)$.

22. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation such that $T(x_1, x_2) = (x_1 - 2x_2, -x_1 + 3x_2, 3x_1 - 2x_2)$. Find $\mathbf{x}$ such that $T(\mathbf{x}) = (-1, 4, 9)$.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is completely determined by its effect on the columns of the $n \times n$ identity matrix.

24. **(T/F)** A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one if each vector in $\mathbb{R}^n$ maps onto a unique vector in $\mathbb{R}^m$.

25. **(T/F)** If $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates vectors about the origin through an angle ϕ, then T is a linear transformation.

26. **(T/F)** The columns of the standard matrix for a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ are the images of the columns of the $n \times n$ identity matrix.

27. **(T/F)** When two linear transformations are performed one after another, the combined effect may not always be a linear transformation.

28. **(T/F)** Not every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a matrix transformation.

29. **(T/F)** A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is onto $\mathbb{R}^m$ if every vector $\mathbf{x}$ in $\mathbb{R}^n$ maps onto some vector in $\mathbb{R}^m$.

30. **(T/F)** The standard matrix of a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ that reflects points through the horizontal axis, the vertical axis, or the origin has the form $\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$, where a and d are ± 1.

31. **(T/F)** A is a 3×2 matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot be one-to-one.

32. **(T/F)** A is a 3×2 matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot map $\mathbb{R}^2$ onto $\mathbb{R}^3$.

In Exercises 33–36, determine if the specified linear transformation is (a) one-to-one and (b) onto. Justify each answer.

33. The transformation in Exercise 17

34. The transformation in Exercise 2

35. The transformation in Exercise 19

36. The transformation in Exercise 14

In Exercises 37 and 38, describe the possible echelon forms of the standard matrix for a linear transformation T. Use the notation of Example 1 in Section 1.2.

37. $T : \mathbb{R}^3 \to \mathbb{R}^4$ is one-to-one.

38. $T : \mathbb{R}^4 \to \mathbb{R}^3$ is onto.

39. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with A its standard matrix. Complete the following statement to make it true: "T is one-to-one if and only if A has ____ pivot columns." Explain why the statement is true. [*Hint:* Look in the exercises for Section 1.7.]

40. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with A its standard matrix. Complete the following statement to make it true: "T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if A has ____ pivot columns." Find some theorems that explain why the statement is true.

41. Verify the uniqueness of A in Theorem 10. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation such that $T(\mathbf{x}) = B\mathbf{x}$ for some $m \times n$ matrix B. Show that if A is the standard matrix for T, then $A = B$. [*Hint:* Show that A and B have the same columns.]

42. Why is the question "Is the linear transformation T onto?" an existence question?

43. If a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$, can you give a relation between m and n? If T is one-to-one, what can you say about m and n?

44. Let $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Show that the mapping $\mathbf{x} \mapsto T(S(\mathbf{x}))$ is a linear transformation (from $\mathbb{R}^p$ to $\mathbb{R}^m$). [*Hint:* Compute $T(S(c\mathbf{u} + d\mathbf{v}))$ for $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^p$ and scalars c and d. Justify each step of the computation, and explain why this computation gives the desired conclusion.]

T In Exercises 45–48, let T be the linear transformation whose standard matrix is given. In Exercises 45 and 46, decide if T is a one-to-one mapping. In Exercises 47 and 48, decide if T maps $\mathbb{R}^5$ onto $\mathbb{R}^5$. Justify your answers.

45. $\begin{bmatrix} -5 & 10 & -5 & 4 \\ 8 & 3 & -4 & 7 \\ 4 & -9 & 5 & -3 \\ -3 & -2 & 5 & 4 \end{bmatrix}$

46. $\begin{bmatrix} 7 & 5 & 4 & -9 \\ 10 & 6 & 16 & -4 \\ 12 & 8 & 12 & 7 \\ -8 & -6 & -2 & 5 \end{bmatrix}$

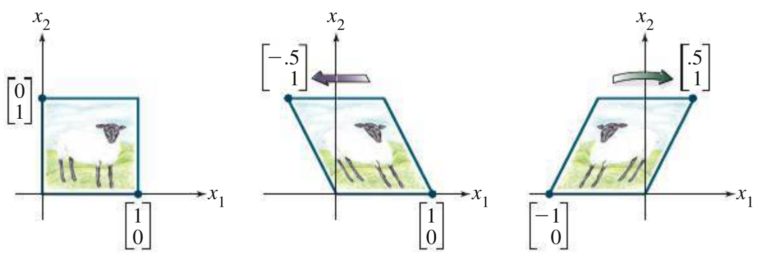
47. $\begin{bmatrix} 4 & -7 & 3 & 7 & 5 \\ 6 & -8 & 5 & 12 & -8 \\ -7 & 10 & -8 & -9 & 14 \\ 3 & -5 & 4 & 2 & -6 \\ -5 & 6 & -6 & -7 & 3 \end{bmatrix}$

48. $\begin{bmatrix} 9 & 13 & 5 & 6 & -1 \\ 14 & 15 & -7 & -6 & 4 \\ -8 & -9 & 12 & -5 & -9 \\ -5 & -6 & -8 & 9 & 8 \\ 13 & 14 & 15 & 2 & 11 \end{bmatrix}$

STUDY GUIDE offers additional resources for mastering existence and uniqueness.

Solution to Practice Problems

1. Follow what happens to $\mathbf{e}_1$ and $\mathbf{e}_2$. See Figure 5. First, $\mathbf{e}_1$ is unaffected by the shear and then is reflected into $-\mathbf{e}_1$. So $T(\mathbf{e}_1) = -\mathbf{e}_1$. Second, $\mathbf{e}_2$ goes to $\mathbf{e}_2 - .5\mathbf{e}_1$ by the shear transformation. Since reflection through the x_2-axis changes $\mathbf{e}_1$ into $-\mathbf{e}_1$ and leaves $\mathbf{e}_2$ unchanged, the vector $\mathbf{e}_2 - .5\mathbf{e}_1$ goes to $\mathbf{e}_2 + .5\mathbf{e}_1$. So $T(\mathbf{e}_2) = \mathbf{e}_2 + .5\mathbf{e}_1$.



Shear transformation   Reflection through the x_2-axis

FIGURE 5 The composition of two transformations.

Thus the standard matrix of T is

$$\begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -\mathbf{e}_1 & \mathbf{e}_2 + .5\mathbf{e}_1 \end{bmatrix} = \begin{bmatrix} -1 & .5 \\ 0 & 1 \end{bmatrix}$$

2. The standard matrix representation of T is the matrix A. Since A has 5 columns and 5 pivots, there is a pivot in every column so the columns are linearly independent. By Theorem 12, T is one-to-one. Since A has 7 rows and only 5 pivots, there is not a pivot in every row hence the columns of A do not span $\mathbb{R}^7$. By Theorem 12, and T is not onto.

1.10 Linear Models in Business, Science, and Engineering

The mathematical models in this section are all *linear*; that is, each describes a problem by means of a linear equation, usually in vector or matrix form. The first model concerns nutrition but actually is representative of a general technique in linear programming problems. The second model comes from electrical engineering. The third model introduces the concept of a *linear difference equation*, a powerful mathematical tool for studying dynamic processes in a wide variety of fields such as engineering, ecology, economics, telecommunications, and the management sciences. Linear models are important because natural phenomena are often linear or nearly linear when the variables involved are held within reasonable bounds. Also, linear models are more easily adapted for computer calculation than are complex nonlinear models.

As you read about each model, pay attention to how its linearity reflects some property of the system being modeled.

Constructing a Nutritious Weight-Loss Diet

The formula for the Cambridge Diet, a popular diet in the 1980s, was based on years of research. A team of scientists headed by Dr. Alan H. Howard developed this diet at Cambridge University after more than eight years of clinical work with obese patients.[1] The very low-calorie powdered formula diet combines a precise balance of carbohydrate, high-quality protein, and fat, together with vitamins, minerals, trace elements, and electrolytes. Millions of persons have used the diet to achieve rapid and substantial weight loss.

To achieve the desired amounts and proportions of nutrients, Dr. Howard had to incorporate a large variety of foodstuffs in the diet. Each foodstuff supplied several of the required ingredients, but not in the correct proportions. For instance, nonfat milk was a major source of protein but contained too much calcium. So soy flour was used for part of the protein because soy flour contains little calcium. However, soy flour contains proportionally too much fat, so whey was added since it supplies less fat in relation to calcium. Unfortunately, whey contains too much carbohydrate. . . .

The following example illustrates the problem on a small scale. Listed in Table 1 are three of the ingredients in the diet, together with the amounts of certain nutrients supplied by 100 grams (g) of each ingredient.[2]

[1] The first announcement of this rapid weight-loss regimen was given in the *International Journal of Obesity* (1978) **2**, 321–332.

[2] Ingredients in the diet as of 1984; nutrient data for ingredients adapted from USDA Agricultural Handbooks No. 8-1 and 8-6, 1976.

TABLE 1 The Cambridge Diet

Amounts (g) Supplied per 100 g of Ingredient				Amounts (g) Supplied by Cambridge Diet in One Day
Nutrient	**Nonfat milk**	**Soy flour**	**Whey**	
Protein	36	51	13	33
Carbohydrate	52	34	74	45
Fat	0	7	1.1	3

EXAMPLE 1 If possible, find some combination of nonfat milk, soy flour, and whey to provide the exact amounts of protein, carbohydrate, and fat supplied by the diet in one day (Table 1).

SOLUTION Let x_1, x_2, and x_3, respectively, denote the number of units (100 g) of these foodstuffs. One approach to the problem is to derive equations for each nutrient separately. For instance, the product

$$\left\{\begin{matrix} x_1 \text{ units of} \\ \text{nonfat milk} \end{matrix}\right\} \cdot \left\{\begin{matrix} \text{protein per unit} \\ \text{of nonfat milk} \end{matrix}\right\}$$

gives the amount of protein supplied by x_1 units of nonfat milk. To this amount, we would then add similar products for soy flour and whey and set the resulting sum equal to the amount of protein we need. Analogous calculations would have to be made for each nutrient.

A more efficient method, and one that is conceptually simpler, is to consider a "nutrient vector" for each foodstuff and build just one vector equation. The amount of nutrients supplied by x_1 units of nonfat milk is the scalar multiple

$$\overset{\text{Scalar}}{\left\{\begin{matrix} x_1 \text{ units of} \\ \text{nonfat milk} \end{matrix}\right\}} \cdot \overset{\text{Vector}}{\left\{\begin{matrix} \text{nutrients per unit} \\ \text{of nonfat milk} \end{matrix}\right\}} = x_1\mathbf{a}_1 \tag{1}$$

where $\mathbf{a}_1$ is the first column in Table 1. Let $\mathbf{a}_2$ and $\mathbf{a}_3$ be the corresponding vectors for soy flour and whey, respectively, and let $\mathbf{b}$ be the vector that lists the total nutrients required (the last column of the table). Then $x_2\mathbf{a}_2$ and $x_3\mathbf{a}_3$ give the nutrients supplied by x_2 units of soy flour and x_3 units of whey, respectively. So the relevant equation is

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b} \tag{2}$$

Row reduction of the augmented matrix for the corresponding system of equations shows that

$$\begin{bmatrix} 36 & 51 & 13 & 33 \\ 52 & 34 & 74 & 45 \\ 0 & 7 & 1.1 & 3 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 0 & .277 \\ 0 & 1 & 0 & .392 \\ 0 & 0 & 1 & .233 \end{bmatrix}$$

To three significant digits, the diet requires .277 units of nonfat milk, .392 units of soy flour, and .233 units of whey in order to provide the desired amounts of protein, carbohydrate, and fat. ■

It is important that the values of x_1, x_2, and x_3 found above are nonnegative. This is necessary for the solution to be physically feasible. (How could you use $-.233$ units of whey, for instance?) With a large number of nutrient requirements, it may be necessary to use a larger number of foodstuffs in order to produce a system of equations with a "nonnegative" solution. Thus many, many different combinations of foodstuffs may need to be examined in order to find a system of equations with such a solution. In fact, the manufacturer of the Cambridge Diet was able to supply 31 nutrients in precise amounts using only 33 ingredients.

The diet construction problem leads to the *linear* equation (2) because the amount of nutrients supplied by each foodstuff can be written as a scalar multiple of a vector, as in (1). That is, the nutrients supplied by a foodstuff are *proportional* to the amount of the foodstuff added to the diet mixture. Also, each nutrient in the mixture is the *sum* of the amounts from the various foodstuffs.

Problems of formulating specialized diets for humans and livestock occur frequently. Usually they are treated by linear programming techniques. Our method of constructing vector equations often simplifies the task of formulating such problems.

Linear Equations and Electrical Networks

Current flow in a simple electrical network can be described by a system of linear equations. A voltage source such as a battery forces a current of electrons to flow through the network. When the current passes through a resistor (such as a lightbulb or motor), some of the voltage is "used up"; by Ohm's law, this "voltage drop" across a resistor is given by

$$V = RI$$

where the voltage V is measured in *volts*, the resistance R in *ohms* (denoted by Ω), and the current flow I in *amperes* (*amps*, for short).

The network in Figure 1 contains three closed loops. The currents flowing in loops 1, 2, and 3 are denoted by I_1, I_2, and I_3, respectively. The designated directions of such *loop currents* are arbitrary. If a current turns out to be negative, then the actual direction of current flow is opposite to that chosen in the figure. If the current direction shown is away from the positive (longer) side of a battery (⊣⊢) around to the negative (shorter) side, the voltage is positive; otherwise, the voltage is negative.

Current flow in a loop is governed by the following rule.

KIRCHHOFF'S VOLTAGE LAW

The algebraic sum of the RI voltage drops in one direction around a loop equals the algebraic sum of the voltage sources in the same direction around the loop.

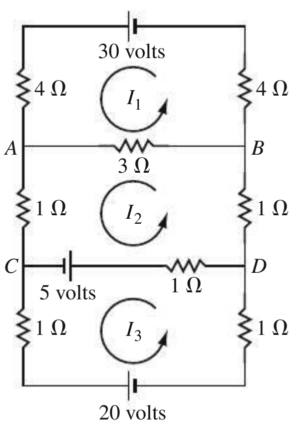


FIGURE 1

EXAMPLE 2 Determine the loop currents in the network in Figure 1.

SOLUTION For loop 1, the current I_1 flows through three resistors, and the sum of the RI voltage drops is

$$4I_1 + 4I_1 + 3I_1 = (4 + 4 + 3)I_1 = 11I_1$$

Current from loop 2 also flows in part of loop 1, through the short *branch* between A and B. The associated RI drop there is $3I_2$ volts. However, the current direction for the branch AB in loop 1 is opposite to that chosen for the flow in loop 2, so the algebraic sum of all RI drops for loop 1 is $11I_1 - 3I_2$. Since the voltage in loop 1 is $+30$ volts, Kirchhoff's voltage law implies that

$$11I_1 - 3I_2 = 30$$

The equation for loop 2 is

$$-3I_1 + 6I_2 - I_3 = 5$$

The term $-3I_1$ comes from the flow of the loop 1 current through the branch AB (with a negative voltage drop because the current flow there is opposite to the flow in loop 2). The term $6I_2$ is the sum of all resistances in loop 2, multiplied by the loop current. The term $-I_3 = -1 \cdot I_3$ comes from the loop 3 current flowing through the 1-ohm resistor in branch CD, in the direction opposite to the flow in loop 2. The loop 3 equation is

$$-I_2 + 3I_3 = -25$$

Note that the 5-volt battery in branch CD is counted as part of both loop 2 and loop 3, but it is -5 volts for loop 3 because of the direction chosen for the current in loop 3. The 20-volt battery is negative for the same reason.

The loop currents are found by solving the system

$$\begin{aligned} 11I_1 - 3I_2 \qquad &= 30 \\ -3I_1 + 6I_2 - I_3 &= 5 \\ - I_2 + 3I_3 &= -25 \end{aligned} \tag{3}$$

Row operations on the augmented matrix lead to the solution: $I_1 = 3$ amps, $I_2 = 1$ amp, and $I_3 = -8$ amps. The negative value of I_3 indicates that the actual current in loop 3 flows in the direction opposite to that shown in Figure 1. ■

It is instructive to look at system (3) as a vector equation:

$$I_1 \underbrace{\begin{bmatrix} 11 \\ -3 \\ 0 \end{bmatrix}}_{\mathbf{r}_1} + I_2 \underbrace{\begin{bmatrix} -3 \\ 6 \\ -1 \end{bmatrix}}_{\mathbf{r}_2} + I_3 \underbrace{\begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix}}_{\mathbf{r}_3} = \underbrace{\begin{bmatrix} 30 \\ 5 \\ -25 \end{bmatrix}}_{\mathbf{v}} \tag{4}$$

The first entry of each vector concerns the first loop, and similarly for the second and third entries. The first resistor vector $\mathbf{r}_1$ lists the resistance in the various loops through which current I_1 flows. A resistance is written negatively when I_1 flows against the flow direction in another loop. Examine Figure 1 and see how to compute the entries in $\mathbf{r}_1$; then do the same for $\mathbf{r}_2$ and $\mathbf{r}_3$. The matrix form of equation (4),

$$R\mathbf{i} = \mathbf{v}, \quad \text{where} \quad R = [\,\mathbf{r}_1 \;\; \mathbf{r}_2 \;\; \mathbf{r}_3\,] \quad \text{and} \quad \mathbf{i} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}$$

provides a matrix version of Ohm's law. If all loop currents are chosen in the same direction (say, counterclockwise), then all entries off the main diagonal of R will be negative.

The matrix equation $R\mathbf{i} = \mathbf{v}$ makes the linearity of this model easy to see at a glance. For instance, if the voltage vector is doubled, then the current vector must double. Also, a *superposition principle* holds. That is, the solution of equation (4) is the sum of the solutions of the equations

$$R\mathbf{i} = \begin{bmatrix} 30 \\ 0 \\ 0 \end{bmatrix}, \qquad R\mathbf{i} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}, \quad \text{and} \quad R\mathbf{i} = \begin{bmatrix} 0 \\ 0 \\ -25 \end{bmatrix}$$

Each equation here corresponds to the circuit with only one voltage source (the other sources being replaced by wires that close each loop). The model for current flow is *linear* precisely because Ohm's law and Kirchhoff's law are linear: The voltage drop across a resistor is *proportional* to the current flowing through it (Ohm), and the *sum* of the voltage drops in a loop equals the sum of the voltage sources in the loop (Kirchhoff).

Loop currents in a network can be used to determine the current in any branch of the network. If only one loop current passes through a branch, such as from B to D in Figure 1, the branch current equals the loop current. If more than one loop current passes through a branch, such as from A to B, the branch current is the algebraic sum of the loop currents in the branch (*Kirchhoff's current law*). For instance, the current in branch AB is $I_1 - I_2 = 3 - 1 = 2$ amps, in the direction of I_1. The current in branch CD is $I_2 - I_3 = 9$ amps.

Difference Equations

In many fields, such as ecology, economics, and engineering, a need arises to model mathematically a dynamic system that changes over time. Several features of the system are each measured at discrete time intervals, producing a sequence of vectors $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2, \ldots$. The entries in $\mathbf{x}_k$ provide information about the *state* of the system at the time of the kth measurement.

If there is a matrix A such that $\mathbf{x}_1 = A\mathbf{x}_0$, $\mathbf{x}_2 = A\mathbf{x}_1$, and, in general,

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \ldots \tag{5}$$

then (5) is called a **linear difference equation** (or **recurrence relation**). Given such an equation, one can compute $\mathbf{x}_1$, $\mathbf{x}_2$, and so on, provided $\mathbf{x}_0$ is known. Sections 4.8 and several sections in Chapter 5 will develop formulas for $\mathbf{x}_k$ and describe what can happen to $\mathbf{x}_k$ as k increases indefinitely. The discussion below illustrates how a difference equation might arise.

A subject of interest to demographers is the movement of populations or groups of people from one region to another. The simple model here considers the changes in the population of a certain city and its surrounding suburbs over a period of years.

Fix an initial year—say, 2020—and denote the populations of the city and suburbs that year by r_0 and s_0, respectively. Let $\mathbf{x}_0$ be the population vector

$$\mathbf{x}_0 = \begin{bmatrix} r_0 \\ s_0 \end{bmatrix} \begin{array}{l} \text{City population, 2020} \\ \text{Suburban population, 2020} \end{array}$$

For 2021 and subsequent years, denote the populations of the city and suburbs by the vectors

$$\mathbf{x}_1 = \begin{bmatrix} r_1 \\ s_1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} r_2 \\ s_2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} r_3 \\ s_3 \end{bmatrix}, \ldots$$

Our goal is to describe mathematically how these vectors might be related.

Suppose demographic studies show that each year about 5% of the city's population moves to the suburbs (and 95% remains in the city), while 3% of the suburban population moves to the city (and 97% remains in the suburbs). See Figure 2.

After 1 year, the original r_0 persons in the city are now distributed between city and suburbs as

$$\begin{bmatrix} .95r_0 \\ .05r_0 \end{bmatrix} = r_0 \begin{bmatrix} .95 \\ .05 \end{bmatrix} \begin{array}{l} \text{Remain in city} \\ \text{Move to suburbs} \end{array} \tag{6}$$

The s_0 persons in the suburbs in 2020 are distributed 1 year later as

$$s_0 \begin{bmatrix} .03 \\ .97 \end{bmatrix} \begin{array}{l} \text{Move to city} \\ \text{Remain in suburbs} \end{array} \tag{7}$$

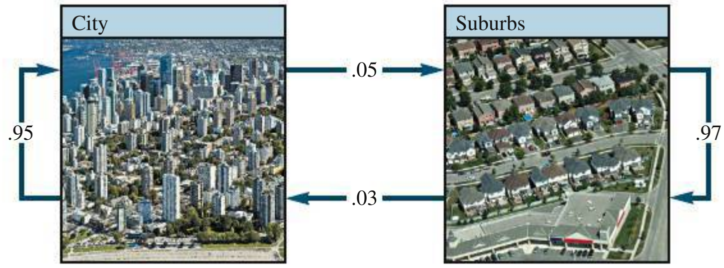


FIGURE 2 Annual percentage migration between city and suburbs.

The vectors in (6) and (7) account for all of the population in 2021.[3] Thus

$$\begin{bmatrix} r_1 \\ s_1 \end{bmatrix} = r_0 \begin{bmatrix} .95 \\ .05 \end{bmatrix} + s_0 \begin{bmatrix} .03 \\ .97 \end{bmatrix} = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} r_0 \\ s_0 \end{bmatrix}$$

That is,

$$\mathbf{x}_1 = M\mathbf{x}_0 \tag{8}$$

where M is the **migration matrix** determined by the following table:

$$\begin{array}{cc} \text{From:} & \\ \begin{array}{cc} \text{City} & \text{Suburbs} \end{array} & \text{To:} \\ \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} & \begin{array}{l} \text{City} \\ \text{Suburbs} \end{array} \end{array}$$

Equation (8) describes how the population changes from 2020 to 2021. If the migration percentages remain constant, then the change from 2021 to 2022 is given by

$$\mathbf{x}_2 = M\mathbf{x}_1$$

and similarly for 2022 to 2023 and subsequent years. In general,

$$\mathbf{x}_{k+1} = M\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \ldots \tag{9}$$

The sequence of vectors $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots\}$ describes the population of the city/suburban region over a period of years.

EXAMPLE 3 Compute the population of the region just described for the years 2021 and 2022, given that the population in 2020 was 600,000 in the city and 400,000 in the suburbs.

SOLUTION The initial population in 2020 is $\mathbf{x}_0 = \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$. For 2021,

$$\mathbf{x}_1 = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix} = \begin{bmatrix} 582{,}000 \\ 418{,}000 \end{bmatrix}$$

For 2022,

$$\mathbf{x}_2 = M\mathbf{x}_1 = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} 582{,}000 \\ 418{,}000 \end{bmatrix} = \begin{bmatrix} 565{,}440 \\ 434{,}560 \end{bmatrix}$$ ■

[3] For simplicity, we ignore other influences on the population such as births, deaths, and migration into and out of the city/suburban region.

The model for population movement in (9) is *linear* because the correspondence $\mathbf{x}_k \mapsto \mathbf{x}_{k+1}$ is a linear transformation. The linearity depends on two facts: the number of people who chose to move from one area to another is *proportional* to the number of people in that area, as shown in (6) and (7), and the cumulative effect of these choices is found by *adding* the movement of people from the different areas.

Practice Problem

Find a matrix A and vectors $\mathbf{x}$ and $\mathbf{b}$ such that the problem in Example 1 amounts to solving the equation $A\mathbf{x} = \mathbf{b}$.

1.10 Exercises

1. The container of a breakfast cereal usually lists the number of calories and the amounts of protein, carbohydrate, and fat contained in one serving of the cereal. The amounts for two common cereals are given below. Suppose a mixture of these two cereals is to be prepared that contains exactly 295 calories, 9 g of protein, 48 g of carbohydrate, and 8 g of fat.
   a. Set up a vector equation for this problem. Include a statement of what the variables in your equation represent.
   b. Write an equivalent matrix equation, and then determine if the desired mixture of the two cereals can be prepared.

Nutrition Information per Serving

Nutrient	General Mills Cheerios®	Quaker® 100% Natural Cereal
Calories	110	130
Protein (g)	4	3
Carbohydrate (g)	20	18
Fat (g)	2	5

2. One serving of Post Shredded Wheat® supplies 160 calories, 5 g of protein, 6 g of fiber, and 1 g of fat. One serving of Crispix® supplies 110 calories, 2 g of protein, .1 g of fiber, and .4 g of fat.
   a. Set up a matrix B and a vector $\mathbf{u}$ such that $B\mathbf{u}$ gives the amounts of calories, protein, fiber, and fat contained in a mixture of three servings of Shredded Wheat and two servings of Crispix.
   T b. Suppose that you want a cereal with more fiber than Crispix but fewer calories than Shredded Wheat. Is it possible for a mixture of the two cereals to supply 130 calories, 3.20 g of protein, 2.46 g of fiber, and .64 g of fat? If so, what is the mixture?

3. After taking a nutrition class, a big Annie's® Mac and Cheese fan decides to improve the levels of protein and fiber in her favorite lunch by adding broccoli and canned chicken. The nutritional information for the foods referred to in this are given in the table.

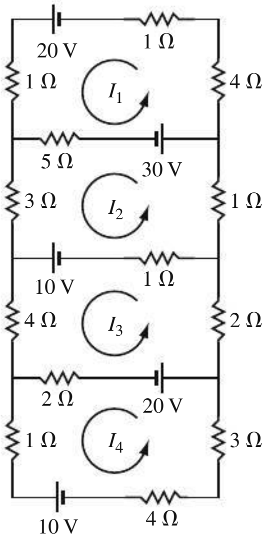
Nutrition Information per Serving

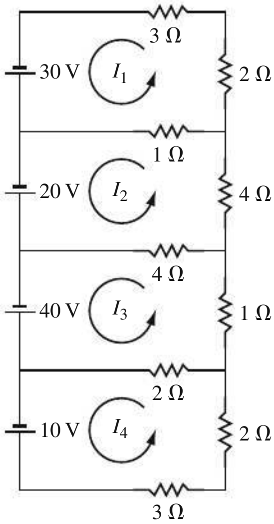
Nutrient	Mac and Cheese	Broccoli	Chicken	Shells
Calories	270	51	70	260
Protein (g)	10	5.4	15	9
Fiber (g)	2	5.2	0	5

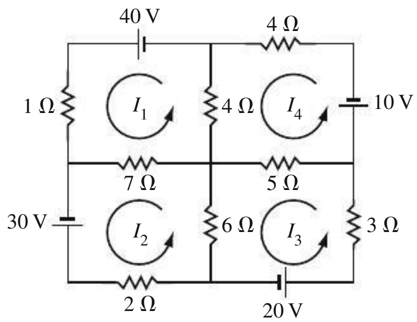
   T a. If she wants to limit her lunch to 400 calories but get 30 g of protein and 10 g of fiber, what proportions of servings of Mac and Cheese, broccoli, and chicken should she use?
   T b. She found that there was too much broccoli in the proportions from part (a), so she decided to switch from classical Mac and Cheese to Annie's® Whole Wheat Shells and White Cheddar. What proportions of servings of each food should she use to meet the same goals as in part (a)?

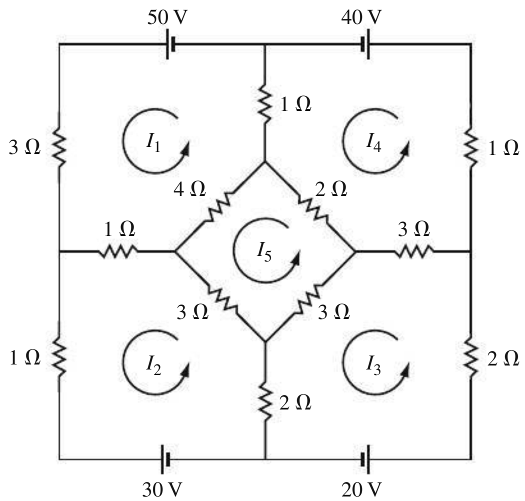
4. The Cambridge Diet supplies .8 g of calcium per day, in addition to the nutrients listed in Table 1 for Example 1. The amounts of calcium per unit (100 g) supplied by the three ingredients in the Cambridge Diet are as follows: 1.26 g from nonfat milk, .19 g from soy flour, and .8 g from whey. Another ingredient in the diet mixture is isolated soy protein, which provides the following nutrients in each unit: 80 g of protein, 0 g of carbohydrate, 3.4 g of fat, and .18 g of calcium.
   a. Set up a matrix equation whose solution determines the amounts of nonfat milk, soy flour, whey, and isolated soy protein necessary to supply the precise amounts of protein, carbohydrate, fat, and calcium in the Cambridge Diet. State what the variables in the equation represent.
   T b. Solve the equation in (a) and discuss your answer.

T In Exercises 5–8, write a matrix equation that determines the loop currents. If MATLAB or another matrix program is available, solve the system for the loop currents.

5. 

6. 

7. 

8. 

9. In a certain region, about 7% of a city's population moves to the surrounding suburbs each year, and about 5% of the suburban population moves into the city. In 2020, there were 800,000 residents in the city and 500,000 in the suburbs. Set up a difference equation that describes this situation, where $\mathbf{x}_0$ is the initial population in 2020. Then estimate the populations in the city and in the suburbs two years later, in 2022. (Ignore other factors that might influence the population sizes.)

10. In a certain region, about 6% of a city's population moves to the surrounding suburbs each year, and about 4% of the suburban population moves into the city. In 2020, there were 10,000,000 residents in the city and 800,000 in the suburbs. Set up a difference equation that describes this situation, where $\mathbf{x}_0$ is the initial population in 2020. Then estimate the populations in the city and in the suburbs two years later, in 2022.

T 11. College Moving Truck Rental has a fleet of 20, 100, and 200 trucks in Pullman, Spokane, and Seattle, respectively. A truck rented at one location may be returned to any of the three locations. The various fractions of trucks returned to the three locations each month are shown in the matrix below. What will be the approximate distribution of the trucks after three months?

Trucks Rented From:

$$\begin{array}{c l} \begin{array}{ccc} \text{Pullman} & \text{Spokane} & \text{Seattle} \end{array} & \text{Returned To:} \\ \begin{bmatrix} .30 & .15 & .05 \\ .30 & .70 & .05 \\ .40 & .15 & .90 \end{bmatrix} & \begin{array}{l} \text{Airport} \\ \text{East} \\ \text{West} \end{array} \end{array}$$

T 12. Budget® Rent a Car in Wichita, Kansas, has a fleet of about 500 cars, at three locations. A car rented at one location may be returned to any of the three locations. The various fractions of cars returned to the three locations are shown in the matrix below. Suppose that on Monday there are 295 cars at the airport (or rented from there), 55 cars at the east side office, and 150 cars at the west side office. What will be the approximate distribution of cars on Wednesday?

Cars Rented From:

$$\begin{array}{c l} \begin{array}{ccc} \text{Airport} & \text{East} & \text{West} \end{array} & \text{Returned To:} \\ \begin{bmatrix} .97 & .05 & .10 \\ .00 & .90 & .05 \\ .03 & .05 & .85 \end{bmatrix} & \begin{array}{l} \text{Airport} \\ \text{East} \\ \text{West} \end{array} \end{array}$$

T 13. Let M and $\mathbf{x}_0$ be as in Example 3.

a. Compute the population vectors $\mathbf{x}_k$ for $k = 1, \dots, 20$. Discuss what you find.

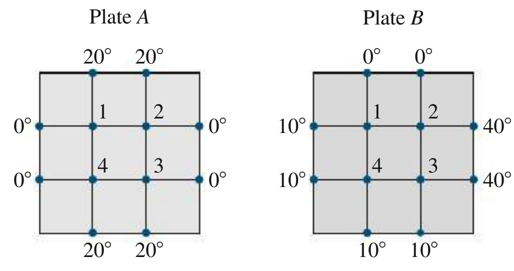
b. Repeat part (a) with an initial population of 350,000 in the city and 650,000 in the suburbs. What do you find?

T 14. Study how changes in boundary temperatures on a steel plate affect the temperatures at interior points on the plate.

a. Begin by estimating the temperatures T_1, T_2, T_3, T_4 at each of the sets of four points on the steel plate shown in the figure. In each case, the value of T_k is approximated by the average of the temperatures at the four closest points. See Exercises 43 and 44 in Section 1.1, where the values (in degrees) turn out to be (20, 27.5, 30, 22.5). How is this list of values related to your results for the points in set (a) and set (b)?

b. Without making any computations, guess the interior temperatures in (a) when the boundary temperatures are all multiplied by 3. Check your guess.

c. Finally, make a general conjecture about the correspondence from the list of eight boundary temperatures to the list of four interior temperatures.



(a) (b)

Solution to Practice Problem

$$A = \begin{bmatrix} 36 & 51 & 13 \\ 52 & 34 & 74 \\ 0 & 7 & 1.1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 33 \\ 45 \\ 3 \end{bmatrix}$$

CHAPTER 1 PROJECTS

Chapter 1 projects are available online at `bit.ly/30IM8gT`.

A. *Interpolating Polynomials*: This project shows how to use a system of linear equations to fit a polynomial through a set of points.

B. *Splines*: This project also shows how to use a system of linear equations to fit a piecewise polynomial curve through a set of points.

C. *Network Flows*: The purpose of this project is to show how systems of linear equations may be used to model flow through a network.

D. *The Art of Linear Transformations*: In this project, it is illustrated how to graph a polygon and then use linear transformations to change its shape and create a design.

E. *Loop Currents*: The purpose of this project is to provide more and larger examples of loop currents.

F. *Diet*: The purpose of this project is to provide examples of vector equations that result from balancing nutrients in a diet.

CHAPTER 1 SUPPLEMENTARY EXERCISES

Mark each statement True or False (**T/F**). Justify each answer. (If true, cite appropriate facts or theorems. If false, explain why or give a counterexample that shows why the statement is not true in every case.

1. **(T/F)** Every matrix is row equivalent to a unique matrix in echelon form.

2. **(T/F)** Any system of n linear equations in n variables has at most n solutions.

3. **(T/F)** If a system of linear equations has two different solutions, it must have infinitely many solutions.

4. **(T/F)** If a system of linear equations has no free variables, then it has a unique solution.

5. **(T/F)** If an augmented matrix $[\, A \quad \mathbf{b}\,]$ is transformed into $[\, C \quad \mathbf{d}\,]$ by elementary row operations, then the equations $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ have exactly the same solution sets.

6. **(T/F)** If a system $A\mathbf{x} = \mathbf{b}$ has more than one solution, then so does the system $A\mathbf{x} = \mathbf{0}$.

7. **(T/F)** If A is an $m \times n$ matrix and the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some $\mathbf{b}$, then the columns of A span $\mathbb{R}^m$.

8. **(T/F)** If an augmented matrix $[\, A \quad \mathbf{b}\,]$ can be transformed by elementary row operations into reduced echelon form, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

9. **(T/F)** If matrices A and B are row equivalent, they have the same reduced echelon form.

10. **(T/F)** The equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution if and only if there are no free variables.

11. **(T/F)** If A is an $m \times n$ matrix and the equation $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$ in $\mathbb{R}^m$, then A has m pivot columns.

12. **(T/F)** If an $m \times n$ matrix A has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

13. **(T/F)** If an $n \times n$ matrix A has n pivot positions, then the reduced echelon form of A is the $n \times n$ identity matrix.

14. **(T/F)** If 3×3 matrices A and B each have three pivot positions, then A can be transformed into B by elementary row operations.

15. **(T/F)** If A is an $m \times n$ matrix, if the equation $A\mathbf{x} = \mathbf{b}$ has at least two different solutions, and if the equation $A\mathbf{x} = \mathbf{c}$ is consistent, then the equation $A\mathbf{x} = \mathbf{c}$ has many solutions.

16. **(T/F)** If A and B are row equivalent $m \times n$ matrices and if the columns of A span $\mathbb{R}^m$, then so do the columns of B.

17. **(T/F)** If none of the vectors in the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in $\mathbb{R}^3$ is a multiple of one of the other vectors, then S is linearly independent.

18. **(T/F)** If $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent, then $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are not in $\mathbb{R}^2$.

19. **(T/F)** In some cases, it is possible for four vectors to span $\mathbb{R}^5$.

20. **(T/F)** If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^m$, then $-\mathbf{u}$ is in Span$\{\mathbf{u}, \mathbf{v}\}$.

21. **(T/F)** If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are nonzero vectors in $\mathbb{R}^2$, then $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

22. **(T/F)** If $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, then $\mathbf{u}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$.

23. **(T/F)** Suppose that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are in $\mathbb{R}^5$, $\mathbf{v}_2$ is not a multiple of $\mathbf{v}_1$, and $\mathbf{v}_3$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. Then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

24. **(T/F)** A linear transformation is a function.

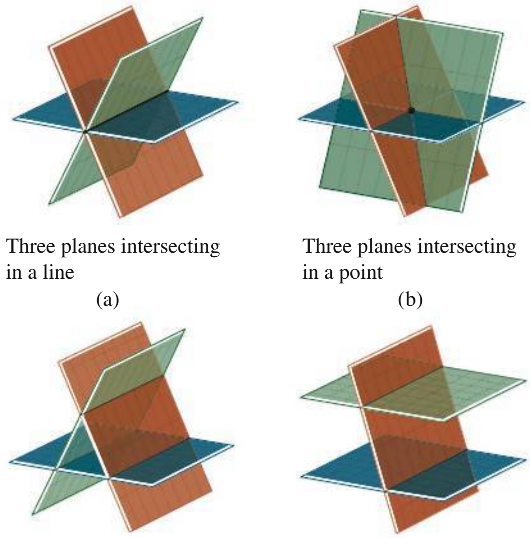
25. **(T/F)** If A is a 6×5 matrix, the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot map $\mathbb{R}^5$ onto $\mathbb{R}^6$.

26. Let a and b represent real numbers. Describe the possible solution sets of the (linear) equation $ax = b$. [*Hint:* The number of solutions depends upon a and b.]

27. The solutions (x, y, z) of a single linear equation

$$ax + by + cz = d$$

form a plane in $\mathbb{R}^3$ when a, b, and c are not all zero. Construct sets of three linear equations whose graphs (a) intersect in a single line, (b) intersect in a single point, and (c) have no points in common. Typical graphs are illustrated in the figure.



Three planes intersecting in a line
(a)

Three planes intersecting in a point
(b)

Three planes with no intersection
(c)

Three planes with no intersection
(c')

28. Suppose the coefficient matrix of a linear system of three equations in three variables has a pivot position in each column. Explain why the system has a unique solution.

29. Determine h and k such that the solution set of the system (i) is empty, (ii) contains a unique solution, and (iii) contains infinitely many solutions.

a. $\begin{aligned} x_1 + 3x_2 &= k \\ 4x_1 + hx_2 &= 8 \end{aligned}$ b. $\begin{aligned} -2x_1 + hx_2 &= 1 \\ 6x_1 + kx_2 &= -2 \end{aligned}$

30. Consider the problem of determining whether the following system of equations is consistent:

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

a. Define appropriate vectors, and restate the problem in terms of linear combinations. Then solve that problem.

b. Define an appropriate matrix, and restate the problem using the phrase "columns of A."

c. Define an appropriate linear transformation T using the matrix in (b), and restate the problem in terms of T.

31. Consider the problem of determining whether the following system of equations is consistent for all b_1, b_2, b_3:

$$\begin{aligned} 2x_1 - 4x_2 - 2x_3 &= b_1 \\ -5x_1 + x_2 + x_3 &= b_2 \\ 7x_1 - 5x_2 - 3x_3 &= b_3 \end{aligned}$$

a. Define appropriate vectors, and restate the problem in terms of Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Then solve that problem.

b. Define an appropriate matrix, and restate the problem using the phrase "columns of A."

c. Define an appropriate linear transformation T using the matrix in (b), and restate the problem in terms of T.

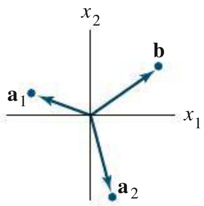
32. Describe the possible echelon forms of the matrix A. Use the notation of Example 1 in Section 1.2.

a. A is a 2×3 matrix whose columns span $\mathbb{R}^2$.

b. A is a 3×3 matrix whose columns span $\mathbb{R}^3$.

33. Write the vector $\begin{bmatrix} 5 \\ 6 \end{bmatrix}$ as the sum of two vectors, one on the line $\{(x, y) : y = 2x\}$ and one on the line $\{(x, y) : y = x/2\}$.

34. Let $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{b}$ be the vectors in $\mathbb{R}^2$ shown in the figure, and let $A = [\mathbf{a}_1 \;\; \mathbf{a}_2]$. Does the equation $A\mathbf{x} = \mathbf{b}$ have a solution? If so, is the solution unique? Explain.



35. Construct a 2×3 matrix A, not in echelon form, such that the solution of $A\mathbf{x} = \mathbf{0}$ is a line in $\mathbb{R}^3$.

36. Construct a 2×3 matrix A, not in echelon form, such that the solution of $A\mathbf{x} = \mathbf{0}$ is a plane in $\mathbb{R}^3$.

37. Write the *reduced* echelon form of a 3×3 matrix A such that the first two columns of A are pivot columns and $A\begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

38. Determine the value(s) of a such that $\left\{ \begin{bmatrix} 1 \\ a \end{bmatrix}, \begin{bmatrix} a \\ a+2 \end{bmatrix} \right\}$ is linearly independent.

39. In (a) and (b), suppose the vectors are linearly independent. What can you say about the numbers $a, \ldots, f$? Justify your answers. [*Hint:* Use a theorem for (b).]

a. $\begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} b \\ c \\ 0 \end{bmatrix}, \begin{bmatrix} d \\ e \\ f \end{bmatrix}$ b. $\begin{bmatrix} a \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} b \\ c \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} d \\ e \\ f \\ 1 \end{bmatrix}$

40. Use Theorem 7 in Section 1.7 to explain why the columns of the matrix A are linearly independent.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 \\ 3 & 6 & 8 & 0 \\ 4 & 7 & 9 & 10 \end{bmatrix}$$

41. Explain why a set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ in $\mathbb{R}^5$ must be linearly independent when $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent and $\mathbf{v}_4$ is *not* in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

42. Suppose $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set in $\mathbb{R}^n$. Show that $\{\mathbf{v}_1, \mathbf{v}_1 + \mathbf{v}_2\}$ is also linearly independent.

43. Suppose $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are distinct points on one line in $\mathbb{R}^3$. The line need not pass through the origin. Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

44. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and suppose $T(\mathbf{u}) = \mathbf{v}$. Show that $T(-\mathbf{u}) = -\mathbf{v}$.

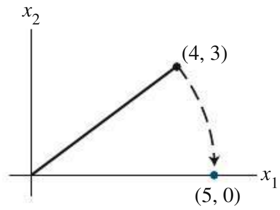
45. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation that reflects each vector through the plane $x_2 = 0$. That is, $T(x_1, x_2, x_3) = (x_1, -x_2, x_3)$. Find the standard matrix of T.

46. Let A be a 3×3 matrix with the property that the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^3$ onto $\mathbb{R}^3$. Explain why the transformation must be one-to-one.

47. A *Givens rotation* is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ used in computer programs to create a zero entry in a vector (usually a column of a matrix). The standard matrix of a Givens rotation in $\mathbb{R}^2$ has the form

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \qquad a^2 + b^2 = 1$$

Find a and b such that $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$ is rotated into $\begin{bmatrix} 5 \\ 0 \end{bmatrix}$.



A Givens rotation in $\mathbb{R}^2$.

48. The following equation describes a Givens rotation in $\mathbb{R}^3$. Find a and b.

$$\begin{bmatrix} a & 0 & -b \\ 0 & 1 & 0 \\ b & 0 & a \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2\sqrt{5} \\ 3 \\ 0 \end{bmatrix}, \qquad a^2 + b^2 = 1$$

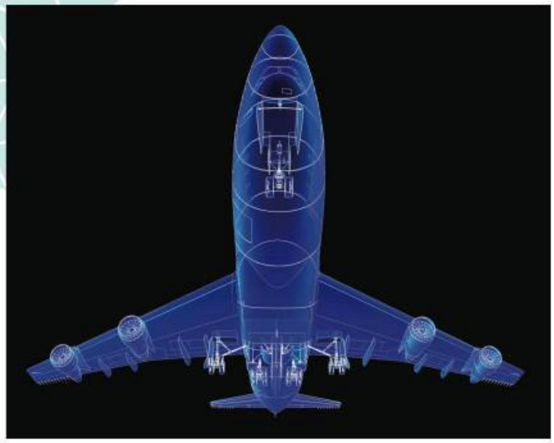
49. A large apartment building is to be built using modular construction techniques. The arrangement of apartments on any particular floor is to be chosen from one of three basic floor plans. Plan A has 18 apartments on one floor, including 3 three-bedroom units, 7 two-bedroom units, and 8 one-bedroom units. Each floor of plan B includes 4 three-bedroom units, 4 two-bedroom units, and 8 one-bedroom units. Each floor of plan C includes 5 three-bedroom units, 3 two-bedroom units, and 9 one-bedroom units. Suppose the building contains a total of x_1 floors of plan A, x_2 floors of plan B, and x_3 floors of plan C.

a. What interpretation can be given to the vector $x_1 \begin{bmatrix} 3 \\ 7 \\ 8 \end{bmatrix}$?

b. Write a formal linear combination of vectors that expresses the total numbers of three-, two-, and one-bedroom apartments contained in the building.

T c. Is it possible to design the building with exactly 66 three-bedroom units, 74 two-bedroom units, and 136 one-bedroom units? If so, is there more than one way to do it? Explain your answer.

2 Matrix Algebra



Introductory Example

COMPUTER MODELS IN AIRCRAFT DESIGN

To design the next generation of commercial and military aircraft, engineers at Boeing's Phantom Works use 3D modeling and computational fluid dynamics (CFD). They study the airflow around a virtual airplane to answer important design questions before physical models are created. This has drastically reduced design cycle times and cost—and linear algebra plays a crucial role in the process.

The virtual airplane begins as a mathematical "wire-frame" model that exists only in computer memory and on graphics display terminals. (Model of a Boeing 747 is shown.) This mathematical model organizes and influences each step of the design and manufacture of the airplane—both the exterior and interior. The CFD analysis concerns the exterior surface.

Although the finished skin of a plane may seem smooth, the geometry of the surface is complicated. In addition to wings and a fuselage, an aircraft has nacelles, stabilizers, slats, flaps, and ailerons. The way air flows around these structures determines how the plane moves through the sky. Equations that describe the airflow are complicated, and they must account for engine intake, engine exhaust, and the wakes left by the wings of the plane. To study the airflow, engineers need a highly refined description of the plane's surface.

A computer creates a model of the surface by first superimposing a three-dimensional grid of "boxes" on the original wire-frame model. Boxes in this grid lie either completely inside or completely outside the plane, or they intersect the surface of the plane. The computer selects the boxes that intersect the surface and subdivides them, retaining only the smaller boxes that still intersect the surface. The subdividing process is repeated until the grid is extremely fine. A typical grid can include more than 400,000 boxes.

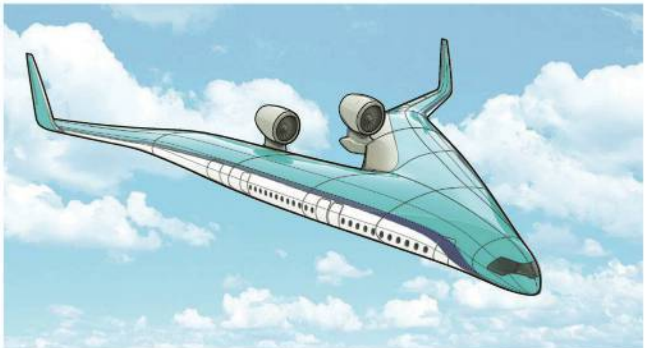
The process for finding the airflow around the plane involves repeatedly solving a system of linear equations $A\mathbf{x} = \mathbf{b}$ that may involve up to 2 million equations and variables. The vector $\mathbf{b}$ changes each time, based on data from the grid and solutions of previous equations. Using the fastest computers available commercially, a Phantom Works team can spend from a few hours to several days setting up and solving a single airflow problem. After the team analyzes the solution, they may make small changes to the airplane surface and begin the whole process again. Thousands of CFD runs may be required.

This chapter presents two important concepts that assist in the solution of such massive systems of equations:

- *Partitioned matrices:* A typical CFD system of equations has a "sparse" coefficient matrix with mostly zero entries. Grouping the variables correctly leads to a partitioned matrix with many zero blocks. Section 2.4 introduces such matrices and describes some of their applications.

- *Matrix factorizations:* Even when written with partitioned matrices, the system of equations is complicated. To further simplify the computations, the CFD software at Boeing uses what is called an LU factorization of the coefficient matrix. Section 2.5 discusses LU and other useful matrix factorizations. Further details about factorizations appear at several points later in the text.

To analyze a solution of an airflow system, engineers want to visualize the airflow over the surface of the plane. They use computer graphics, and linear algebra provides the engine for the graphics. The wire-frame model of the plane's surface is stored as data in many matrices. Once the image has been rendered on a computer screen, engineers can change its scale, zoom in or out of small regions, and rotate the image to see parts that may be hidden from view. Each of these operations is accomplished by appropriate matrix multiplications. Section 2.7 explains the basic ideas.



TU-Delft and Air France-KLM are investigating a flying V aircraft design because of its potential for significantly better fuel economy.

Our ability to analyze and solve equations will be greatly enhanced when we can perform algebraic operations with matrices. Furthermore, the definitions and theorems in this chapter provide some basic tools for handling the many applications of linear algebra that involve two or more matrices. For $n \times n$ matrices, the Invertible Matrix Theorem in Section 2.3 ties together most of the concepts treated earlier in the text. Sections 2.4 and 2.5 examine partitioned matrices and matrix factorizations, which appear in most modern uses of linear algebra. Sections 2.6 and 2.7 describe two interesting applications of matrix algebra: to economics and to computer graphics. Sections 2.8 and 2.9 provide readers enough information about subspaces to move directly into Chapters 5, 6, and 7, without covering Chapter 4. You may want to omit these two sections if you plan to cover Chapter 4 before moving to Chapter 5.

2.1 Matrix Operations

If A is an $m \times n$ matrix—that is, a matrix with m rows and n columns—then the scalar entry in the ith row and jth column of A is denoted by a_{ij} and is called the (i, j)-entry of A. See Figure 1. For instance, the $(3, 2)$-entry is the number a_{32} in the third row, second column. Each column of A is a list of m real numbers, which identifies a vector in $\mathbb{R}^m$. Often, these columns are denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and the matrix A is written as

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

Observe that the number a_{ij} is the ith entry (from the top) of the jth column vector $\mathbf{a}_j$.

The **diagonal entries** in an $m \times n$ matrix $A = [\,a_{ij}\,]$ are $a_{11}, a_{22}, a_{33}, \ldots,$ and they form the **main diagonal** of A. A **diagonal matrix** is a square $n \times n$ matrix whose nondiagonal entries are zero. An example is the $n \times n$ identity matrix, I_n. An $m \times n$ matrix whose entries are all zero is a **zero matrix** and is written as 0. The size of a zero matrix is usually clear from the context.

$$
\begin{array}{ccccccc}
 & & & \text{Column} & & & \\
 & & & j & & & \\
 & \left[\begin{array}{c} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{m1} \end{array}\right. & \begin{array}{c} \cdots \\ \\ \cdots \\ \\ \cdots \end{array} & \begin{array}{c} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{mj} \end{array} & \begin{array}{c} \cdots \\ \\ \cdots \\ \\ \cdots \end{array} & \left.\begin{array}{c} a_{1n} \\ \vdots \\ a_{in} \\ \vdots \\ a_{mn} \end{array}\right] & = A \\
\text{Row } i & & & & & & \\
 & \uparrow & & \uparrow & & \uparrow & \\
 & \mathbf{a}_1 & & \mathbf{a}_j & & \mathbf{a}_n &
\end{array}
$$

FIGURE 1 Matrix notation.

Sums and Scalar Multiples

The arithmetic for vectors described earlier has a natural extension to matrices. We say that two matrices are **equal** if they have the same size (i.e., the same number of rows and the same number of columns) and if their corresponding columns are equal, which amounts to saying that their corresponding entries are equal. If A and B are $m \times n$ matrices, then the **sum** $A + B$ is the $m \times n$ matrix whose columns are the sums of the corresponding columns in A and B. Since vector addition of the columns is done entrywise, each entry in $A + B$ is the sum of the corresponding entries in A and B. The sum $A + B$ is defined only when A and B are the same size.

EXAMPLE 1 Let

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

Then

$$A + B = \begin{bmatrix} 5 & 1 & 6 \\ 2 & 8 & 9 \end{bmatrix}$$

but $A + C$ is not defined because A and C have different sizes. ∎

If r is a scalar and A is a matrix, then the **scalar multiple** rA is the matrix whose columns are r times the corresponding columns in A. As with vectors, $-A$ stands for $(-1)A$, and $A - B$ is the same as $A + (-1)B$.

EXAMPLE 2 If A and B are the matrices in Example 1, then

$$2B = 2\begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 6 & 10 & 14 \end{bmatrix}$$

$$A - 2B = \begin{bmatrix} 4 & 0 & 5 \\ -1 & 3 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 2 \\ 6 & 10 & 14 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 3 \\ -7 & -7 & -12 \end{bmatrix}$$ ∎

It was unnecessary in Example 2 to compute $A - 2B$ as $A + (-1)2B$ because the usual rules of algebra apply to sums and scalar multiples of matrices, as the following theorem shows.

THEOREM 1

Let A, B, and C be matrices of the same size, and let r and s be scalars.

a. $A + B = B + A$
b. $(A + B) + C = A + (B + C)$
c. $A + 0 = A$
d. $r(A + B) = rA + rB$
e. $(r + s)A = rA + sA$
f. $r(sA) = (rs)A$

Each equality in Theorem 1 is verified by showing that the matrix on the left side has the same size as the matrix on the right and that corresponding columns are equal. Size is no problem because A, B, and C are equal in size. The equality of columns follows immediately from analogous properties of vectors. For instance, if the jth columns of A, B, and C are $\mathbf{a}_j$, $\mathbf{b}_j$, and $\mathbf{c}_j$, respectively, then the jth columns of $(A + B) + C$ and $A + (B + C)$ are

$$(\mathbf{a}_j + \mathbf{b}_j) + \mathbf{c}_j \quad \text{and} \quad \mathbf{a}_j + (\mathbf{b}_j + \mathbf{c}_j)$$

respectively. Since these two vector sums are equal for each j, property (b) is verified.

Because of the associative property of addition, we can simply write $A + B + C$ for the sum, which can be computed either as $(A + B) + C$ or as $A + (B + C)$. The same applies to sums of four or more matrices.

Matrix Multiplication

When a matrix B multiplies a vector $\mathbf{x}$, it transforms $\mathbf{x}$ into the vector $B\mathbf{x}$. If this vector is then multiplied in turn by a matrix A, the resulting vector is $A(B\mathbf{x})$. See Figure 2.

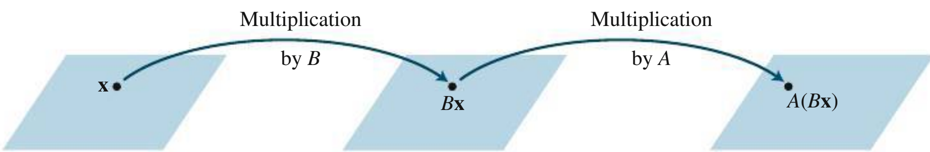


FIGURE 2 Multiplication by B and then A.

Thus $A(B\mathbf{x})$ is produced from $\mathbf{x}$ by a *composition* of mappings—the linear transformations studied in Section 1.8. Our goal is to represent this composite mapping as multiplication by a single matrix, denoted by AB, so that

$$A(B\mathbf{x}) = (AB)\mathbf{x} \tag{1}$$

See Figure 3.

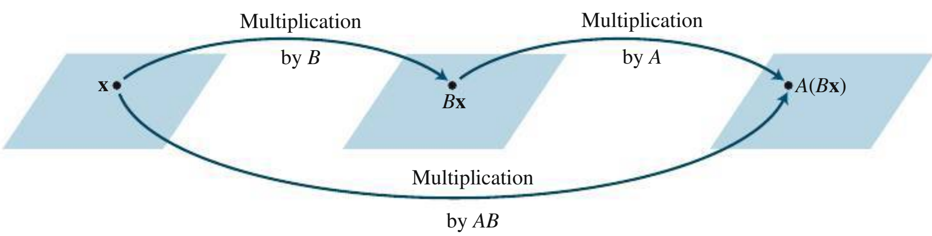


FIGURE 3 Multiplication by AB.

If A is $m \times n$, B is $n \times p$, and $\mathbf{x}$ is in $\mathbb{R}^p$, denote the columns of B by $\mathbf{b}_1, \dots, \mathbf{b}_p$ and the entries in $\mathbf{x}$ by $x_1, \dots, x_p$. Then

$$B\mathbf{x} = x_1\mathbf{b}_1 + \cdots + x_p\mathbf{b}_p$$

By the linearity of multiplication by A,

$$\begin{aligned} A(B\mathbf{x}) &= A(x_1\mathbf{b}_1) + \cdots + A(x_p\mathbf{b}_p) \\ &= x_1A\mathbf{b}_1 + \cdots + x_pA\mathbf{b}_p \end{aligned}$$

The vector $A(B\mathbf{x})$ is a linear combination of the vectors $A\mathbf{b}_1, \dots, A\mathbf{b}_p$, using the entries in $\mathbf{x}$ as weights. In matrix notation, this linear combination is written as

$$A(B\mathbf{x}) = [\, A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_p \,]\mathbf{x}$$

Thus multiplication by $[\, A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_p \,]$ transforms $\mathbf{x}$ into $A(B\mathbf{x})$. We have found the matrix we sought!

DEFINITION

If A is an $m \times n$ matrix, and if B is an $n \times p$ matrix with columns $\mathbf{b}_1, \dots, \mathbf{b}_p$, then the product AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, \dots, A\mathbf{b}_p$. That is,

$$AB = A\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix}$$

This definition makes equation (1) true for all $\mathbf{x}$ in $\mathbb{R}^p$. Equation (1) proves that the composite mapping in Figure 3 is a linear transformation and that its standard matrix is AB. *Multiplication of matrices corresponds to composition of linear transformations.*

EXAMPLE 3 Compute AB, where $A = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix}$.

SOLUTION Write $B = [\, \mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \,]$, and compute:

$$A\mathbf{b}_1 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad A\mathbf{b}_2 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix}, \quad A\mathbf{b}_3 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 6 \\ 3 \end{bmatrix}$$
$$= \begin{bmatrix} 11 \\ -1 \end{bmatrix} \qquad\qquad = \begin{bmatrix} 0 \\ 13 \end{bmatrix} \qquad\qquad = \begin{bmatrix} 21 \\ -9 \end{bmatrix}$$

Then

$$AB = A[\, \mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \,] = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}$$
$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{array}$$

■

Notice that since the first column of AB is $A\mathbf{b}_1$, this column is a linear combination of the columns of A using the entries in $\mathbf{b}_1$ as weights. A similar statement is true for each column of AB.

Each column of AB is a linear combination of the columns of A using weights from the corresponding column of B.

Obviously, the number of columns of A must match the number of rows in B in order for a linear combination such as $A\mathbf{b}_1$ to be defined. Also, the definition of AB shows that AB *has the same number of rows as A and the same number of columns as B.*

EXAMPLE 4 If A is a 3×5 matrix and B is a 5×2 matrix, what are the sizes of AB and BA, if they are defined?

SOLUTION Since A has 5 columns and B has 5 rows, the product AB is defined and is a 3×2 matrix:

$$\underset{3\times 5}{\overset{A}{\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}}}\underset{5\times 2}{\overset{B}{\begin{bmatrix} * & * \\ * & * \\ * & * \\ * & * \\ * & * \end{bmatrix}}} = \underset{3\times 2}{\overset{AB}{\begin{bmatrix} * & * \\ * & * \\ * & * \end{bmatrix}}}$$

Match

Size of AB

The product BA is *not* defined because the 2 columns of B do not match the 3 rows of A. ■

The definition of AB is important for theoretical work and applications, but the following rule provides a more efficient method for calculating the individual entries in AB when working small problems by hand.

ROW–COLUMN RULE FOR COMPUTING AB

If the product AB is defined, then the entry in row i and column j of AB is the sum of the products of corresponding entries from row i of A and column j of B. If $(AB)_{ij}$ denotes the (i, j)-entry in AB, and if A is an $m \times n$ matrix, then

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

To verify this rule, let $B = [\,\mathbf{b}_1 \ \cdots \ \mathbf{b}_p\,]$. Column j of AB is $A\mathbf{b}_j$, and we can compute $A\mathbf{b}_j$ by the row–vector rule for computing $A\mathbf{x}$ from Section 1.4. The ith entry in $A\mathbf{b}_j$ is the sum of the products of corresponding entries from row i of A and the vector $\mathbf{b}_j$, which is precisely the computation described in the rule for computing the (i, j)-entry of AB.

EXAMPLE 5 Use the row–column rule to compute two of the entries in AB for the matrices in Example 3. An inspection of the numbers involved will make it clear how the two methods for calculating AB produce the same matrix.

SOLUTION To find the entry in row 1 and column 3 of AB, consider row 1 of A and column 3 of B. Multiply corresponding entries and add the results, as shown below:

$$AB = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \square & \square & 2(6)+3(3) \\ \square & \square & \square \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & \square & \square \end{bmatrix}$$

For the entry in row 2 and column 2 of AB, use row 2 of A and column 2 of B:

$$\begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & 1(3)+-5(-2) & \square \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & 13 & \square \end{bmatrix}$$

■

EXAMPLE 6 Find the entries in the second row of AB, where

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

SOLUTION By the row–column rule, the entries of the second row of AB come from row 2 of A (and the columns of B):

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

$$= \begin{bmatrix} \square & \square \\ -4 + 21 - 12 & 6 + 3 - 8 \\ \square & \square \\ \square & \square \end{bmatrix} = \begin{bmatrix} \square & \square \\ 5 & 1 \\ \square & \square \\ \square & \square \end{bmatrix}$$

Notice that since Example 6 requested only the second row of AB, we could have written just the second row of A to the left of B and computed

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

This observation about rows of AB is true in general and follows from the row–column rule. Let $\text{row}_i(A)$ denote the ith row of a matrix A. Then

$$\text{row}_i(AB) = \text{row}_i(A) \cdot B \tag{2}$$

Properties of Matrix Multiplication

The following theorem lists the standard properties of matrix multiplication. Recall that I_m represents the $m \times m$ identity matrix and $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^m$.

THEOREM 2

Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B + C) = AB + AC$ (left distributive law)

c. $(B + C)A = BA + CA$ (right distributive law)

d. $r(AB) = (rA)B = A(rB)$ for any scalar r

e. $I_mA = A = AI_n$ (identity for matrix multiplication)

PROOF Properties (b)–(e) are considered in the exercises. Property (a) follows from the fact that matrix multiplication corresponds to composition of linear transformations (which are functions), and it is known (or easy to check) that the composition of functions is associative. Here is another proof of (a) that rests on the "column definition" of the product of two matrices. Let

$$C = [\,\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_p\,]$$

By the definition of matrix multiplication,

$$\begin{aligned} BC &= [\,B\mathbf{c}_1 \quad \cdots \quad B\mathbf{c}_p\,] \\ A(BC) &= [\,A(B\mathbf{c}_1) \quad \cdots \quad A(B\mathbf{c}_p)\,] \end{aligned}$$

Recall from equation (1) that the definition of AB makes $A(B\mathbf{x}) = (AB)\mathbf{x}$ for all $\mathbf{x}$, so

$$A(BC) = [\,(AB)\mathbf{c}_1 \quad \cdots \quad (AB)\mathbf{c}_p\,] = (AB)C$$ ■

The associative and distributive laws in Theorems 1 and 2 say essentially that pairs of parentheses in matrix expressions can be inserted and deleted in the same way as in the algebra of real numbers. In particular, we can write ABC for the product, which can be computed either as $A(BC)$ or as $(AB)C$.[1] Similarly, a product $ABCD$ of four matrices can be computed as $A(BCD)$ or $(ABC)D$ or $A(BC)D$, and so on. It does not matter how we group the matrices when computing the product, so long as the left-to-right order of the matrices is preserved.

The left-to-right order in products is critical because AB and BA are usually not the same. This is not surprising, because the columns of AB are linear combinations of the columns of A, whereas the columns of BA are constructed from the columns of B. The position of the factors in the product AB is emphasized by saying that A is *right-multiplied* by B or that B is *left-multiplied* by A. If $AB = BA$, we say that A and B **commute** with one another.

EXAMPLE 7 Let $A = \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}$. Show that these matrices do not commute. That is, verify that $AB \neq BA$.

SOLUTION

$$\begin{aligned} AB &= \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 3 \\ -2 & -6 \end{bmatrix} \\ BA &= \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 29 & -2 \end{bmatrix} \end{aligned}$$ ■

Example 7 illustrates the first of the following list of important differences between matrix algebra and the ordinary algebra of real numbers. See Exercises 9–12 for examples of these situations.

***Warnings*:**

1. In general, $AB \neq BA$.
2. The cancellation laws do *not* hold for matrix multiplication. That is, if $AB = AC$, then it is *not* true in general that $B = C$. (See Exercise 10.)
3. If a product AB is the zero matrix, you *cannot* conclude in general that either $A = 0$ or $B = 0$. (See Exercise 12.)

[1] When B is square and C has fewer columns than A has rows, it is more efficient to compute $A(BC)$ than $(AB)C$.

Powers of a Matrix

If A is an $n \times n$ matrix and if k is a positive integer, then A^k denotes the product of k copies of A:

$$A^k = \underbrace{A \cdots A}_{k}$$

If A is nonzero and if $\mathbf{x}$ is in $\mathbb{R}^n$, then $A^k\mathbf{x}$ is the result of left-multiplying $\mathbf{x}$ by A repeatedly k times. If $k = 0$, then $A^0\mathbf{x}$ should be $\mathbf{x}$ itself. Thus A^0 is interpreted as the identity matrix. Matrix powers are useful in both theory and applications (Sections 2.6, 5.9, and later in the text).

The Transpose of a Matrix

Given an $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A.

EXAMPLE 8 Let

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

Then

$$A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}, \quad B^T = \begin{bmatrix} -5 & 1 & 0 \\ 2 & -3 & 4 \end{bmatrix}, \quad C^T = \begin{bmatrix} 1 & -3 \\ 1 & 5 \\ 1 & -2 \\ 1 & 7 \end{bmatrix}$$

■

THEOREM 3

Let A and B denote matrices whose sizes are appropriate for the following sums and products.

a. $(A^T)^T = A$
b. $(A + B)^T = A^T + B^T$
c. For any scalar r, $(rA)^T = rA^T$
d. $(AB)^T = B^T A^T$

Proofs of (a)–(c) are straightforward and are omitted. For (d), see Exercise 41. Usually, $(AB)^T$ is not equal to $A^T B^T$, even when A and B have sizes such that the product $A^T B^T$ is defined.

The generalization of Theorem 3(d) to products of more than two factors can be stated in words as follows:

> The transpose of a product of matrices equals the product of their transposes in the *reverse* order.

The exercises contain numerical examples that illustrate properties of transposes.

Artificial intelligence (AI) involves having a computer learn to recognize important information about anything that can be presented in a digitized format. One important area of AI is identifying whether the object in a picture matches a chosen object such as a number, fingerprint, or face.

In the next example, matrix transposition and matrix multiplication are used to tell whether or not a 2×2 block of colored squares matches the chosen checkerboard pattern in Figure 4.

EXAMPLE 9 In order to feed a 2×2 colored block into the computer, it first gets converted into a 4×1 vector by assigning a 1 to each block that is blue and a 0 to each block that is white. Then, the computer converts the block of numbers into a vector by placing the numbers in each column below the numbers in the column to its left.

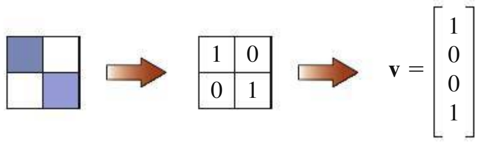
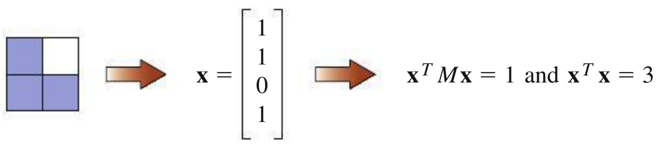


FIGURE 4

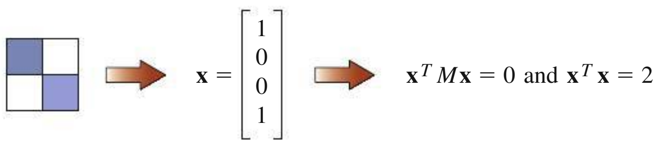
Let $M = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$.

Notice that $\mathbf{v}^T M \mathbf{v} = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 0,$

and $\mathbf{w}^T M \mathbf{w} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0$, where $\mathbf{w}$ is the vector generated by a 2×2 block of all white squares. It can be verified that for any other vector $\mathbf{x}$ generated from a 2×2 block of white and blue squares, if $\mathbf{x}$ is not $\mathbf{v}$ or $\mathbf{w}$, then the product $\mathbf{x}^T M \mathbf{x}$ is nonzero. Thus, if a computer checks the value of $\mathbf{x}^T M \mathbf{x}$ and finds it is nonzero, the computer knows that the pattern corresponding to $\mathbf{x}$ is not the checkerboard with a blue square in the top left corner.



This pattern is not the checkerboard pattern since $\mathbf{x}^T M \mathbf{x} \neq 0$.



This pattern is the checkerboard pattern since $\mathbf{x}^T M \mathbf{x} = 0$, but $\mathbf{x}^T \mathbf{x} \neq 0$.

FIGURE 5

However, if the computer finds that $\mathbf{x}^T M\mathbf{x} = 0$, then $\mathbf{x}$ could be either $\mathbf{v}$ or $\mathbf{w}$. To distinguish between the two, the computer can calculate the product $\mathbf{x}^T\mathbf{x}$, for $\mathbf{x}^T\mathbf{x}$ is zero if and only if $\mathbf{x}$ is $\mathbf{w}$.[2] Thus, to conclude that $\mathbf{x}$ is equal to $\mathbf{v}$, the computer must have $\mathbf{x}^T M\mathbf{x} = 0$ and $\mathbf{x}^T\mathbf{x} \neq 0$. ■

Example 5 of Section 6.3 illustrates one way to choose a matrix M so that matrix multiplication and transposition can be used to identify a particular pattern of colored squares.

Another important aspect of AI starts even before the data is fed to the machine. In Section 1.9, it is illustrated how matrix multiplication can be used to move vectors around in space. In the next example, matrix multiplication is used to *scrub* data and prepare it for processing.

EXAMPLE 10 The dates of ground crew accidents for January and February of 2020 are listed in the columns of matrix T for Toronto Pearson Airport and matrix C for Chicago O'Hare Airport:

$$\text{Toronto: } T = \begin{bmatrix} 1 & 12 & 14 & 15 & 21 & 22 & 23 & 1 & 2 & 3 & 12 & 15 & 17 & 19 & 26 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

$$\text{Chicago: } C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 1 & 11 & 22 & 23 & 24 & 1 & 2 & 5 & 20 & 21 \end{bmatrix}$$

Clearly the data is listed differently in the two matrices. Canada and the United states have different traditions for whether the month or day comes first when writing a date. For matrix T, the day is listed in the first row and the month is listed in the second row. For matrix C, the month is listed in the first row and the day is listed in the second row. In order to use this data, the first and second rows need to be swapped in one of the matrices. Reviewing the effects of matrix multiplication in Table 1 of Section 1.9, notice that the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ switches the x_1 and x_2 coordinates of any vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ it is applied to and indeed

$$AT = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 12 & 14 & 15 & 21 & 22 & 23 & 1 & 2 & 3 & 12 & 15 & 17 & 19 & 26 \end{bmatrix}$$

has the data listed in the same order as it is listed in matrix C. The matrices AT and C can now be fed into the same machine. ■

In Exercises 51 and 52 you will be asked to *scrub* further data for this project.[3]

[2] To see why $\mathbf{x}^T\mathbf{x}$ is zero if and only if $\mathbf{x}$ is $\mathbf{w}$, let $\mathbf{x}^T = [x_1\ x_2\ x_3\ x_4]$. Then $\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + x_3^2 + x_4^2$ and this sum is zero if and only if the coordinates of $\mathbf{x}$ are all zero. That is, if and only if $\mathbf{x} = \mathbf{w}$.

[3] Although the data in this example and the corresponding exercises are fictitious, Data Analytics students at Washington State University identified scrubbing the data they received as an important first step in their actual analysis of ground crew accidents at three major airports in the United States.

Numerical Notes

1. The fastest way to obtain AB on a computer depends on the way in which the computer stores matrices in its memory. The standard high-performance algorithms, such as in LAPACK, calculate AB by columns, as in our definition of the product. (A version of LAPACK written in C++ calculates AB by rows.)
2. The definition of AB lends itself well to parallel processing on a computer. The columns of B are assigned individually or in groups to different processors, which independently and hence simultaneously compute the corresponding columns of AB.

Practice Problems

1. Since vectors in $\mathbb{R}^n$ may be regarded as $n \times 1$ matrices, the properties of transposes in Theorem 3 apply to vectors, too. Let
$$A = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$
Compute $(A\mathbf{x})^T$, $\mathbf{x}^T A^T$, $\mathbf{x}\mathbf{x}^T$, and $\mathbf{x}^T\mathbf{x}$. Is $A^T\mathbf{x}^T$ defined?
2. Let A be a 4×4 matrix and let $\mathbf{x}$ be a vector in $\mathbb{R}^4$. What is the fastest way to compute $A^2\mathbf{x}$? Count the multiplications.
3. Suppose A is an $m \times n$ matrix, all of whose rows are identical. Suppose B is an $n \times p$ matrix, all of whose columns are identical. What can be said about the entries in AB?

2.1 Exercises

In Exercises 1 and 2, compute each matrix sum or product if it is defined. If an expression is undefined, explain why. Let

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

1. $-2A$, $B - 2A$, AC, CD
2. $A + 2B$, $3C - E$, CB, EB

In the rest of this exercise set and in those to follow, you should assume that each matrix expression is defined. That is, the sizes of the matrices (and vectors) involved "match" appropriately.

3. Let $A = \begin{bmatrix} 4 & -1 \\ 5 & -2 \end{bmatrix}$. Compute $3I_2 - A$ and $(3I_2)A$.
4. Compute $A - 5I_3$ and $(5I_3)A$, when
$$A = \begin{bmatrix} 9 & -1 & 3 \\ -8 & 7 & -3 \\ -4 & 1 & 8 \end{bmatrix}.$$

In Exercises 5 and 6, compute the product AB in two ways: (a) by the definition, where $A\mathbf{b}_1$ and $A\mathbf{b}_2$ are computed separately, and (b) by the row–column rule for computing AB.

5. $A = \begin{bmatrix} -1 & 2 \\ 5 & 4 \\ 2 & -3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & -4 \\ -2 & 1 \end{bmatrix}$
6. $A = \begin{bmatrix} 4 & -2 \\ -3 & 0 \\ 3 & 5 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 3 \\ 4 & -1 \end{bmatrix}$
7. If a matrix A is 5×3 and the product AB is 5×7, what is the size of B?
8. How many rows does B have if BC is a 3×4 matrix?
9. Let $A = \begin{bmatrix} 2 & 5 \\ -3 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & -5 \\ 3 & k \end{bmatrix}$. What value(s) of k, if any, will make $AB = BA$?
10. Let $A = \begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \end{bmatrix}$, and $C = \begin{bmatrix} 5 & -2 \\ 3 & 1 \end{bmatrix}$. Verify that $AB = AC$ and yet $B \neq C$.
11. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 5 \end{bmatrix}$ and $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$. Compute AD and DA. Explain how the columns or rows of A change when A is multiplied by D on the right or on the left. Find a 3×3 matrix B, not the identity matrix or the zero matrix, such that $AB = BA$.

12. Let $A = \begin{bmatrix} 3 & -6 \\ -1 & 2 \end{bmatrix}$. Construct a 2×2 matrix B such that AB is the zero matrix. Use two different nonzero columns for B.

13. Let $\mathbf{r}_1, \ldots, \mathbf{r}_p$ be vectors in $\mathbb{R}^n$, and let Q be an $m \times n$ matrix. Write the matrix $[\, Q\mathbf{r}_1 \cdots Q\mathbf{r}_p \,]$ as a *product* of two matrices (neither of which is an identity matrix).

14. Let U be the 3×2 cost matrix described in Example 6 of Section 1.8. The first column of U lists the costs per dollar of output for manufacturing product B, and the second column lists the costs per dollar of output for product C. (The costs are categorized as materials, labor, and overhead.) Let $\mathbf{q}_1$ be a vector in $\mathbb{R}^2$ that lists the output (measured in dollars) of products B and C manufactured during the first quarter of the year, and let $\mathbf{q}_2, \mathbf{q}_3$, and $\mathbf{q}_4$ be the analogous vectors that list the amounts of products B and C manufactured in the second, third, and fourth quarters, respectively. Give an economic description of the data in the matrix UQ, where $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3 \quad \mathbf{q}_4]$.

Exercises 15–24 concern arbitrary matrices A, B, and C for which the indicated sums and products are defined. Mark each statement True or False **(T/F)**. Justify each answer.

15. **(T/F)** If A and B are 2×2 with columns $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{b}_1, \mathbf{b}_2$, respectively, then $AB = [\mathbf{a}_1\mathbf{b}_1 \quad \mathbf{a}_2\mathbf{b}_2]$.

16. **(T/F)** If A and B are 3×3 and $B = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3]$, then $AB = [A\mathbf{b}_1 + A\mathbf{b}_2 + A\mathbf{b}_3]$.

17. **(T/F)** Each column of AB is a linear combination of the columns of B using weights from the corresponding column of A.

18. **(T/F)** The second row of AB is the second row of A multiplied on the right by B.

19. **(T/F)** $AB + AC = A(B + C)$

20. **(T/F)** $A^T + B^T = (A + B)^T$

21. **(T/F)** $(AB)C = (AC)B$

22. **(T/F)** $(AB)^T = A^T B^T$

23. **(T/F)** The transpose of a product of matrices equals the product of their transposes in the same order.

24. **(T/F)** The transpose of a sum of matrices equals the sum of their transposes.

25. If $A = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$ and $AB = \begin{bmatrix} -1 & 2 & -1 \\ 6 & -9 & 3 \end{bmatrix}$, determine the first and second columns of B.

26. Suppose the first two columns, $\mathbf{b}_1$ and $\mathbf{b}_2$, of B are equal. What can you say about the columns of AB (if AB is defined)? Why?

27. Suppose the third column of B is the sum of the first two columns. What can you say about the third column of AB? Why?

28. Suppose the second column of B is all zeros. What can you say about the second column of AB?

29. Suppose the last column of AB is all zeros, but B itself has no column of zeros. What can you say about the columns of A?

30. Show that if the columns of B are linearly dependent, then so are the columns of AB.

31. Suppose $CA = I_n$ (the $n \times n$ identity matrix). Show that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Explain why A cannot have more columns than rows.

32. Suppose $AD = I_m$ (the $m \times m$ identity matrix). Show that for any $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution. [*Hint:* Think about the equation $AD\mathbf{b} = \mathbf{b}$.] Explain why A cannot have more rows than columns.

33. Suppose A is an $m \times n$ matrix and there exist $n \times m$ matrices C and D such that $CA = I_n$ and $AD = I_m$. Prove that $m = n$ and $C = D$. [*Hint:* Think about the product CAD.]

34. Suppose A is a $3 \times n$ matrix whose columns span $\mathbb{R}^3$. Explain how to construct an $n \times 3$ matrix D such that $AD = I_3$.

In Exercises 35 and 36, view vectors in $\mathbb{R}^n$ as $n \times 1$ matrices. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, called the **scalar product**, or **inner product**, of $\mathbf{u}$ and $\mathbf{v}$. It is usually written as a single real number without brackets. The matrix product $\mathbf{u}\mathbf{v}^T$ is an $n \times n$ matrix, called the **outer product** of $\mathbf{u}$ and $\mathbf{v}$. The products $\mathbf{u}^T\mathbf{v}$ and $\mathbf{u}\mathbf{v}^T$ will appear later in the text.

35. Let $\mathbf{u} = \begin{bmatrix} -2 \\ 3 \\ -4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Compute $\mathbf{u}^T\mathbf{v}$, $\mathbf{v}^T\mathbf{u}$, $\mathbf{u}\mathbf{v}^T$, and $\mathbf{v}\mathbf{u}^T$.

36. If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^n$, how are $\mathbf{u}^T\mathbf{v}$ and $\mathbf{v}^T\mathbf{u}$ related? How are $\mathbf{u}\mathbf{v}^T$ and $\mathbf{v}\mathbf{u}^T$ related?

37. Prove Theorem 2(b) and 2(c). Use the row–column rule. The (i, j)-entry in $A(B + C)$ can be written as

$$a_{i1}(b_{1j} + c_{1j}) + \cdots + a_{in}(b_{nj} + c_{nj}) \text{ or } \sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})$$

38. Prove Theorem 2(d). [*Hint:* The (i, j)-entry in $(rA)B$ is $(ra_{i1})b_{1j} + \cdots + (ra_{in})b_{nj}$.]

39. Show that $I_m A = A$ when A is an $m \times n$ matrix. You can assume $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^m$.

40. Show that $AI_n = A$ when A is an $m \times n$ matrix. [*Hint:* Use the (column) definition of AI_n.]

41. Prove Theorem 3(d). [*Hint:* Consider the jth row of $(AB)^T$.]

42. Give a formula for $(AB\mathbf{x})^T$, where $\mathbf{x}$ is a vector and A and B are matrices of appropriate sizes.

T 43. Use a web search engine such as Google to find documentation for your matrix program, and write the commands that will produce the following matrices (without keying in each entry of the matrix).

a. A 5×6 matrix of zeros

b. A 3×5 matrix of ones

c. The 6×6 identity matrix

d. A 5×5 diagonal matrix, with diagonal entries 3, 5, 7, 2, 4

A useful way to test new ideas in matrix algebra, or to make conjectures, is to make calculations with matrices selected at random. Checking a property for a few matrices does not prove that the property holds in general, but it makes the property more believable. Also, if the property is actually false, you may discover this when you make a few calculations.

T **44.** Write the command(s) that will create a 6×4 matrix with random entries. In what range of numbers do the entries lie? Tell how to create a 3×3 matrix with random integer entries between -9 and 9. [*Hint:* If x is a random number such that $0 < x < 1$, then $-9.5 < 19(x - .5) < 9.5$.]

T **45.** Construct a random 4×4 matrix A and test whether $(A + I)(A - I) = A^2 - I$. The best way to do this is to compute $(A + I)(A - I) - (A^2 - I)$ and verify that this difference is the zero matrix. Do this for three random matrices. Then test $(A + B)(A - B) = A^2 - B^2$ the same way for three pairs of random 4×4 matrices. Report your conclusions.

T **46.** Use at least three pairs of random 4×4 matrices A and B to test the equalities $(A + B)^T = A^T + B^T$ and $(AB)^T = A^T B^T$. (See Exercise 45.) Report your conclusions. [*Note:* Most matrix programs use A' for A^T.]

T **47.** Let

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

Compute S^k for $k = 2, \dots, 6$.

T **48.** Describe in words what happens when you compute A^5, A^{10}, A^{20}, and A^{30} for

$$A = \begin{bmatrix} 1/6 & 1/2 & 1/3 \\ 1/2 & 1/4 & 1/4 \\ 1/3 & 1/4 & 5/12 \end{bmatrix}$$

T **49.** The matrix M can detect a particular 2×2 colored pattern like in Example 9. Create a nonzero 4×1 vector $\mathbf{x}$ by choosing each entry to be a zero or one. Test to see if $\mathbf{x}$ corresponds to the right pattern by calculating $\mathbf{x}^T M \mathbf{x}$. If $\mathbf{x}^T M \mathbf{x} = 0$, then $\mathbf{x}$ is the pattern identified by M. If $\mathbf{x}^T M \mathbf{x} \neq 0$, try a different nonzero vector of zeros and ones. You may want to be systematic in the way that you choose each $\mathbf{x}$ in order to avoid testing the same vector twice. You are using "guess and check" to determine which pattern of 2×2 colored squares the matrix M detects.

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

T **50.** Repeat Exercise 49 with the matrix

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

T **51.** Use the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ to switch the first and second rows of the matrix M containing dates of accidents at the Montreal Trudeau Airport.

Montreal:

$$M = \begin{bmatrix} 2 & 3 & 16 & 24 & 25 & 26 & 6 & 7 & 19 & 26 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \end{bmatrix}$$

This data in matrix M has been scrubbed in matrix AM and can be fed into the same machine as the other data from Example 10.

T **52.** Use the matrix $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ to remove the last row from the matrix N containing dates of accidents at the New York JFK Airport.

New York:

$$N = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 12 & 21 & 22 & 3 & 20 & 21 \\ 2020 & 2020 & 2020 & 2020 & 2020 & 2020 & 2020 \end{bmatrix}$$

The data in matrix N has been scrubbed in matrix BN and can be fed into the same machine as the other data from Example 10.

Solutions to Practice Problems

1. $A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$. So $(A\mathbf{x})^T = \begin{bmatrix} -4 & 2 \end{bmatrix}$. Also,

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

The quantities $(A\mathbf{x})^T$ and $\mathbf{x}^T A^T$ are equal, by Theorem 3(d). Next,

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 5 \\ 3 \end{bmatrix} \begin{bmatrix} 5 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 15 \\ 15 & 9 \end{bmatrix}$$

$$\mathbf{x}^T\mathbf{x} = \begin{bmatrix} 5 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} = [\,25 + 9\,] = 34$$

A 1×1 matrix such as $\mathbf{x}^T\mathbf{x}$ is usually written without the brackets. Finally, $A^T\mathbf{x}^T$ is not defined, because $\mathbf{x}^T$ does not have two rows to match the two columns of A^T.

2. The fastest way to compute $A^2\mathbf{x}$ is to compute $A(A\mathbf{x})$. The product $A\mathbf{x}$ requires 16 multiplications, 4 for each entry, and $A(A\mathbf{x})$ requires 16 more. In contrast, the product A^2 requires 64 multiplications, 4 for each of the 16 entries in A^2. After that, $A^2\mathbf{x}$ takes 16 more multiplications, for a total of 80.
3. First observe that by the definition of matrix multiplication,

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

so the columns of AB are identical. Next, recall that $\text{row}_i(AB) = \text{row}_i(A) \cdot B$. Since all the rows of A are identical, all the rows of AB are identical. Putting this information about the rows and columns together, it follows that all the entries in AB are the same.

2.2 The Inverse of a Matrix

Matrix algebra provides tools for manipulating matrix equations and creating various useful formulas in ways similar to doing ordinary algebra with real numbers. This section investigates the matrix analogue of the reciprocal, or multiplicative inverse, of a nonzero number.

Recall that the multiplicative inverse of a number such as 5 is 1/5 or 5^{-1}. This inverse satisfies the equations

$$5^{-1}(5) = 1 \quad \text{and} \quad 5(5^{-1}) = 1$$

The matrix generalization requires *both* equations and avoids the slanted-line notation (for division) because matrix multiplication is not commutative. Furthermore, a full generalization is possible only if the matrices involved are square.[1]

An $n \times n$ matrix A is said to be **invertible** if there is an $n \times n$ matrix C such that

$$CA = I \quad \text{and} \quad AC = I$$

where $I = I_n$, the $n \times n$ identity matrix. In this case, C is an **inverse** of A. In fact, C is uniquely determined by A, because if B were another inverse of A, then $B = BI = B(AC) = (BA)C = IC = C$. This unique inverse is denoted by A^{-1}, so that

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I$$

A matrix that is *not* invertible is sometimes called a **singular matrix**, and an invertible matrix is called a **nonsingular matrix**.

[1] One could say that an $m \times n$ matrix A is invertible if there exist $n \times m$ matrices C and D such that $CA = I_n$ and $AD = I_m$. However, these equations imply that A is square and $C = D$. Thus, A is invertible as defined above. See Exercises 31–33 in Section 2.1.

EXAMPLE 1 If $A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$ and $C = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}$, then

$$AC = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}\begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and}$$
$$CA = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus $C = A^{-1}$. ■

Here is a simple formula for the inverse of a 2×2 matrix, along with a test to tell if the inverse exists.

THEOREM 4

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If $ad - bc = 0$, then A is not invertible.

The simple proof of Theorem 4 is outlined in Exercises 35 and 36. The quantity $ad - bc$ is called the **determinant** of A, and we write

$$\det A = ad - bc$$

Theorem 4 says that a 2×2 matrix A is invertible if and only if $\det A \neq 0$.

EXAMPLE 2 Find the inverse of $A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$.

SOLUTION Since $\det A = 3(6) - 4(5) = -2 \neq 0$, A is invertible, and

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} = \begin{bmatrix} 6/(-2) & -4/(-2) \\ -5/(-2) & 3/(-2) \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix}$$ ■

Invertible matrices are indispensable in linear algebra—mainly for algebraic calculations and formula derivations, as in the next theorem. There are also occasions when an inverse matrix provides insight into a mathematical model of a real-life situation, as in Example 3.

THEOREM 5

If A is an invertible $n \times n$ matrix, then for each $\mathbf{b}$ in $\mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

PROOF Take any $\mathbf{b}$ in $\mathbb{R}^n$. A solution exists because if $A^{-1}\mathbf{b}$ is substituted for $\mathbf{x}$, then $A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I\mathbf{b} = \mathbf{b}$. So $A^{-1}\mathbf{b}$ is a solution. To prove that the solution is unique, show that if $\mathbf{u}$ is any solution, then $\mathbf{u}$, in fact, must be $A^{-1}\mathbf{b}$. Indeed, if $A\mathbf{u} = \mathbf{b}$, we can multiply both sides by A^{-1} and obtain

$$A^{-1}A\mathbf{u} = A^{-1}\mathbf{b}, \quad I\mathbf{u} = A^{-1}\mathbf{b}, \quad \text{and} \quad \mathbf{u} = A^{-1}\mathbf{b}$$ ■

EXAMPLE 3 A horizontal elastic beam is supported at each end and is subjected to forces at points 1, 2, and 3, as shown in Figure 1. Let $\mathbf{f}$ in $\mathbb{R}^3$ list the forces at these points, and let $\mathbf{y}$ in $\mathbb{R}^3$ list the amounts of deflection (that is, movement) of the beam at the three points. Using Hooke's law from physics, it can be shown that

$$\mathbf{y} = D\mathbf{f}$$

where D is a *flexibility matrix*. Its inverse is called the *stiffness matrix*. Describe the physical significance of the columns of D and D^{-1}.

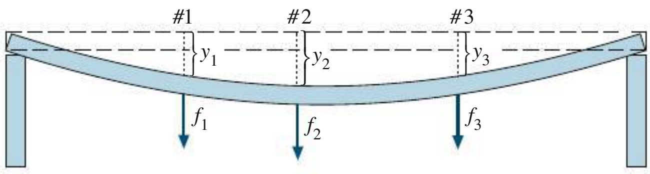


FIGURE 1 Deflection of an elastic beam.

SOLUTION Write $I_3 = [\, \mathbf{e}_1 \quad \mathbf{e}_2 \quad \mathbf{e}_3 \,]$ and observe that

$$D = DI_3 = [\, D\mathbf{e}_1 \quad D\mathbf{e}_2 \quad D\mathbf{e}_3 \,]$$

Interpret the vector $\mathbf{e}_1 = (1, 0, 0)$ as a unit force applied downward at point 1 on the beam (with zero force at the other two points). Then $D\mathbf{e}_1$, the first column of D, lists the beam deflections due to a unit force at point 1. Similar descriptions apply to the second and third columns of D.

To study the stiffness matrix D^{-1}, observe that the equation $\mathbf{f} = D^{-1}\mathbf{y}$ computes a force vector $\mathbf{f}$ when a deflection vector $\mathbf{y}$ is given. Write

$$D^{-1} = D^{-1}I_3 = [\, D^{-1}\mathbf{e}_1 \quad D^{-1}\mathbf{e}_2 \quad D^{-1}\mathbf{e}_3 \,]$$

Now interpret $\mathbf{e}_1$ as a deflection vector. Then $D^{-1}\mathbf{e}_1$ lists the forces that create the deflection. That is, the first column of D^{-1} lists the forces that must be applied at the three points to produce a unit deflection at point 1 and zero deflections at the other points. Similarly, columns 2 and 3 of D^{-1} list the forces required to produce unit deflections at points 2 and 3, respectively. In each column, one or two of the forces must be negative (point upward) to produce a unit deflection at the desired point and zero deflections at the other two points. If the flexibility is measured, for example, in inches of deflection per pound of load, then the stiffness matrix entries are given in pounds of load per inch of deflection. ■

The formula in Theorem 5 is seldom used to solve an equation $A\mathbf{x} = \mathbf{b}$ numerically because row reduction of $[\, A \quad \mathbf{b} \,]$ is nearly always faster. (Row reduction is usually more accurate, too, when computations involve rounding off numbers.) One possible exception is the 2×2 case. In this case, mental computations to solve $A\mathbf{x} = \mathbf{b}$ are sometimes easier using the formula for A^{-1}, as in the next example.

EXAMPLE 4 Use the inverse of the matrix A in Example 2 to solve the system

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7 \end{aligned}$$

SOLUTION This system is equivalent to $A\mathbf{x} = \mathbf{b}$, so

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$ ■

The next theorem provides three useful facts about invertible matrices.

THEOREM 6

a. If A is an invertible matrix, then A^{-1} is invertible and

$$(A^{-1})^{-1} = A$$

b. If A and B are $n \times n$ invertible matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is,

$$(AB)^{-1} = B^{-1}A^{-1}$$

c. If A is an invertible matrix, then so is A^T, and the inverse of A^T is the transpose of A^{-1}. That is,

$$(A^T)^{-1} = (A^{-1})^T$$

PROOF To verify statement (a), find a matrix C such that

$$A^{-1}C = I \quad \text{and} \quad CA^{-1} = I$$

In fact, these equations are satisfied with A in place of C. Hence A^{-1} is invertible, and A is its inverse. Next, to prove statement (b), compute:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

A similar calculation shows that $(B^{-1}A^{-1})(AB) = I$. For statement (c), use Theorem 3(d), read from right to left, $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$. Similarly, $A^T(A^{-1})^T = I^T = I$. Hence A^T is invertible, and its inverse is $(A^{-1})^T$. ■

Remark: Part (b) illustrates the important role that definitions play in proofs. The theorem claims that $B^{-1}A^{-1}$ is the inverse of AB. The proof establishes this by showing that $B^{-1}A^{-1}$ satisfies the definition of what it means to be the inverse of AB. Now, the inverse of AB is a matrix that when multiplied on the left (or right) by AB, the product is the identity matrix I. So the proof consists of showing that $B^{-1}A^{-1}$ has this property.

The following generalization of Theorem 6(b) is needed later.

The product of $n \times n$ invertible matrices is invertible, and the inverse is the product of their inverses in the reverse order.

There is an important connection between invertible matrices and row operations that leads to a method for computing inverses. As we shall see, an invertible matrix A is row equivalent to an identity matrix, and we can find A^{-1} by *watching the row reduction of A to I*.

Elementary Matrices

An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix. The next example illustrates the three kinds of elementary matrices.

EXAMPLE 5 Let

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix},$$

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Compute E_1A, E_2A, and E_3A, and describe how these products can be obtained by elementary row operations on A.

SOLUTION Verify that

$$E_1A = \begin{bmatrix} a & b & c \\ d & e & f \\ g-4a & h-4b & i-4c \end{bmatrix}, \quad E_2A = \begin{bmatrix} d & e & f \\ a & b & c \\ g & h & i \end{bmatrix},$$

$$E_3A = \begin{bmatrix} a & b & c \\ d & e & f \\ 5g & 5h & 5i \end{bmatrix}.$$

Addition of -4 times row 1 of A to row 3 produces E_1A. (This is a row replacement operation.) An interchange of rows 1 and 2 of A produces E_2A, and multiplication of row 3 of A by 5 produces E_3A. ■

Left-multiplication (that is, multiplication on the left) by E_1 in Example 5 has the same effect on any $3 \times n$ matrix. It adds -4 times row 1 to row 3. In particular, since $E_1 \cdot I = E_1$, we see that E_1 *itself* is produced by this same row operation on the identity. Thus Example 5 illustrates the following general fact about elementary matrices. See Exercises 37 and 38.

If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Since row operations are reversible, as shown in Section 1.1, elementary matrices are invertible, for if E is produced by a row operation on I, then there is another row operation of the same type that changes E back into I. Hence there is an elementary matrix F such that $FE = I$. Since E and F correspond to reverse operations, $EF = I$, too.

Each elementary matrix E is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

EXAMPLE 6 Find the inverse of $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$.

SOLUTION To transform E_1 into I, add $+4$ times row 1 to row 3. The elementary matrix that does this is

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ +4 & 0 & 1 \end{bmatrix}$$ ■

The following theorem provides the best way to "visualize" an invertible matrix, and the theorem leads immediately to a method for finding the inverse of a matrix.

THEOREM 7

An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

Remark: The comment on the proof of Theorem 11 in Chapter 1 noted that "P if and only if Q" is equivalent to two statements: (1) "If P then Q" and (2) "If Q then P." The second statement is called the *converse* of the first and explains the use of the word *conversely* in the second paragraph of this proof.

PROOF Suppose that A is invertible. Then, since the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ (Theorem 5), A has a pivot position in every row (Theorem 4 in Section 1.4). Because A is square, the n pivot positions must be on the diagonal, which implies that the reduced echelon form of A is I_n. That is, $A \sim I_n$.

Now suppose, conversely, that $A \sim I_n$. Then, since each step of the row reduction of A corresponds to left-multiplication by an elementary matrix, there exist elementary matrices $E_1, \dots, E_p$ such that

$$A \sim E_1A \sim E_2(E_1A) \sim \cdots \sim E_p(E_{p-1}\cdots E_1A) = I_n$$

That is,

$$E_p \cdots E_1 A = I_n \tag{1}$$

Since the product $E_p \cdots E_1$ of invertible matrices is invertible, (1) leads to

$$\begin{aligned}(E_p \cdots E_1)^{-1}(E_p \cdots E_1)A &= (E_p \cdots E_1)^{-1} I_n \\ A &= (E_p \cdots E_1)^{-1}\end{aligned}$$

Thus A is invertible, as it is the inverse of an invertible matrix (Theorem 6). Also,

$$A^{-1} = \left[\, (E_p \cdots E_1)^{-1} \,\right]^{-1} = E_p \cdots E_1$$

Then $A^{-1} = E_p \cdots E_1 I_n$, which says that A^{-1} results from applying $E_1, \dots, E_p$ successively to I_n. This is the same sequence in (1) that reduced A to I_n. ■

An Algorithm for Finding A^{-1}

If we place A and I side by side to form an augmented matrix $[\,A \quad I\,]$, then row operations on this matrix produce identical operations on A and on I. By Theorem 7, either there are row operations that transform A to I_n and I_n to A^{-1} or else A is not invertible.

ALGORITHM FOR FINDING A^{-1}

Row reduce the augmented matrix $[\,A \quad I\,]$. If A is row equivalent to I, then $[\,A \quad I\,]$ is row equivalent to $[\,I \quad A^{-1}\,]$. Otherwise, A does not have an inverse.

EXAMPLE 7 Find the inverse of the matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$, if it exists.

SOLUTION

$$[\,A \quad I\,] = \begin{bmatrix} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix}$$

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

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

$$\sim \begin{bmatrix} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{bmatrix}$$

Theorem 7 shows, since $A \sim I$, that A is invertible, and

$$A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}$$

■

Reasonable Answers

Once you have found a candidate for the inverse of a matrix, you can check that your answer is correct by finding the product of A with A^{-1}. For the inverse found for matrix A in Example 7, notice

$$AA^{-1} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

confirming that answer is correct. It is not necessary to check that $A^{-1}A = I$ since A is invertible.

Another View of Matrix Inversion

Denote the columns of I_n by $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Then row reduction of $[\,A \quad I\,]$ to $[\,I \quad A^{-1}\,]$ can be viewed as the simultaneous solution of the n systems

$$A\mathbf{x} = \mathbf{e}_1, \quad A\mathbf{x} = \mathbf{e}_2, \quad \ldots, \quad A\mathbf{x} = \mathbf{e}_n \tag{2}$$

where the "augmented columns" of these systems have all been placed next to A to form $[\,A \quad \mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_n\,] = [\,A \quad I\,]$. The equation $AA^{-1} = I$ and the definition of matrix multiplication show that the columns of A^{-1} are precisely the solutions of the systems in (2). This observation is useful because some applied problems may require finding only one or two columns of A^{-1}. In this case, only the corresponding systems in (2) need to be solved.

Numerical Note

In practical work, A^{-1} is seldom computed, unless the entries of A^{-1} are needed. Computing both A^{-1} and $A^{-1}\mathbf{b}$ takes about three times as many arithmetic operations as solving $A\mathbf{x} = \mathbf{b}$ by row reduction, and row reduction may be more accurate.

Practice Problems

1. Use determinants to determine which of the following matrices are invertible.

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

2. Find the inverse of the matrix $A = \begin{bmatrix} 1 & -2 & -1 \\ -1 & 5 & 6 \\ 5 & -4 & 5 \end{bmatrix}$, if it exists.

3. If A is an invertible matrix, prove that $5A$ is an invertible matrix.

2.2 Exercises

Find the inverses of the matrices in Exercises 1–4.

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

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

3. $\begin{bmatrix} 8 & 3 \\ -7 & -3 \end{bmatrix}$

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

5. Verify that the inverse you found in Exercise 1 is correct.

6. Verify that the inverse you found in Exercise 2 is correct.

7. Use the inverse found in Exercise 1 to solve the system

$$\begin{aligned} 8x_1 + 3x_2 &= 2 \\ 5x_1 + 2x_2 &= -1 \end{aligned}$$

8. Use the inverse found in Exercise 2 to solve the system

$$\begin{aligned} 3x_1 + x_2 &= -2 \\ 7x_1 + 2x_2 &= 3 \end{aligned}$$

9. Let $A = \begin{bmatrix} 1 & 2 \\ 5 & 12 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$, and $\mathbf{b}_4 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

   a. Find A^{-1}, and use it to solve the four equations $A\mathbf{x} = \mathbf{b}_1$, $A\mathbf{x} = \mathbf{b}_2$, $A\mathbf{x} = \mathbf{b}_3$, $A\mathbf{x} = \mathbf{b}_4$

   b. The four equations in part (a) can be solved by the same set of row operations, since the coefficient matrix is the same in each case. Solve the four equations in part (a) by row reducing the augmented matrix $\begin{bmatrix} A & \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix}$

10. Use matrix algebra to show that if A is invertible and D satisfies $AD = I$, then $D = A^{-1}$.

In Exercises 11–20, mark each statement True or False **(T/F)**. Justify each answer.

11. **(T/F)** In order for a matrix B to be the inverse of A, both equations $AB = I$ and $BA = I$ must be true.

12. **(T/F)** A product of invertible $n \times n$ matrices is invertible, and the inverse of the product is the product of their inverses in the same order.

13. **(T/F)** If A and B are $n \times n$ and invertible, then $A^{-1}B^{-1}$ is the inverse of AB.

14. **(T/F)** If A is invertible, then the inverse of A^{-1} is A itself.

15. **(T/F)** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ab - cd \neq 0$, then A is invertible.

16. **(T/F)** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad = bc$, then A is not invertible.

17. **(T/F)** If A is an invertible $n \times n$ matrix, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent for *each* $\mathbf{b}$ in $\mathbb{R}^n$.

18. **(T/F)** If A can be row reduced to the identity matrix, then A must be invertible.

19. **(T/F)** Each elementary matrix is invertible.

20. **(T/F)** If A is invertible, then the elementary row operations that reduce A to the identity I_n also reduce A^{-1} to I_n.

21. Let A be an invertible $n \times n$ matrix, and let B be an $n \times p$ matrix. Show that the equation $AX = B$ has a unique solution $A^{-1}B$.

22. Let A be an invertible $n \times n$ matrix, and let B be an $n \times p$ matrix. Explain why $A^{-1}B$ can be computed by row reduction:

$$\text{If } [A \ \ B] \sim \cdots \sim [I \ \ X], \text{ then } X = A^{-1}B.$$

If A is larger than 2×2, then row reduction of $[A \ \ B]$ is much faster than computing both A^{-1} and $A^{-1}B$.

23. Suppose $AB = AC$, where B and C are $n \times p$ matrices and A is invertible. Show that $B = C$. Is this true, in general, when A is not invertible?

24. Suppose $(B - C)D = 0$, where B and C are $m \times n$ matrices and D is invertible. Show that $B = C$.

25. Suppose A, B, and C are invertible $n \times n$ matrices. Show that ABC is also invertible by producing a matrix D such that $(ABC)\,D = I$ and $D\,(ABC) = I$.

26. Suppose A and B are $n \times n$, B is invertible, and AB is invertible. Show that A is invertible. [*Hint:* Let $C = AB$, and solve this equation for A.]

27. Solve the equation $AB = BC$ for A, assuming that A, B, and C are square and B is invertible.

28. Suppose P is invertible and $A = PBP^{-1}$. Solve for B in terms of A.

29. If A, B, and C are $n \times n$ invertible matrices, does the equation $C^{-1}(A + X)B^{-1} = I_n$ have a solution, X? If so, find it.

30. Suppose A, B, and X are $n \times n$ matrices with A, X, and $A - AX$ invertible, and suppose

$$(A - AX)^{-1} = X^{-1}B \tag{3}$$

a. Explain why B is invertible.

b. Solve (3) for X. If you need to invert a matrix, explain why that matrix is invertible.

31. Explain why the columns of an $n \times n$ matrix A are linearly independent when A is invertible.

32. Explain why the columns of an $n \times n$ matrix A span $\mathbb{R}^n$ when A is invertible. [*Hint:* Review Theorem 4 in Section 1.4.]

33. Suppose A is $n \times n$ and the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Explain why A has n pivot columns and A is row equivalent to I_n. By Theorem 7, this shows that A must be invertible. (This exercise and Exercise 34 will be cited in Section 2.3.)

34. Suppose A is $n \times n$ and the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^n$. Explain why A must be invertible. [*Hint:* Is A row equivalent to I_n?]

Exercises 35 and 36 prove Theorem 4 for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

35. Show that if $ad - bc = 0$, then the equation $A\mathbf{x} = \mathbf{0}$ has more than one solution. Why does this imply that A is not invertible? [*Hint:* First, consider $a = b = 0$. Then, if a and b are not both zero, consider the vector $\mathbf{x} = \begin{bmatrix} -b \\ a \end{bmatrix}$.]

36. Show that if $ad - bc \neq 0$, the formula for A^{-1} works.

Exercises 37 and 38 prove special cases of the facts about elementary matrices stated in the box following Example 5. Here A is a 3×3 matrix and $I = I_3$. (A general proof would require slightly more notation.)

37. a. Use equation (1) from Section 2.1 to show that $\text{row}_i(A) = \text{row}_i(I) \cdot A$, for $i = 1, 2, 3$.

b. Show that if rows 1 and 2 of A are interchanged, then the result may be written as EA, where E is an elementary matrix formed by interchanging rows 1 and 2 of I.

c. Show that if row 3 of A is multiplied by 5, then the result may be written as EA, where E is formed by multiplying row 3 of I by 5.

38. Show that if row 3 of A is replaced by $\text{row}_3(A) - 4\text{row}_1(A)$, the result is EA, where E is formed from I by replacing $\text{row}_3(I)$ by $\text{row}_3(I) - 4\text{row}_1(I)$.

Find the inverses of the matrices in Exercises 39–42, if they exist. Use the algorithm introduced in this section.

39. $\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix}$

40. $\begin{bmatrix} 5 & 10 \\ 4 & 7 \end{bmatrix}$

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

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

43. Use the algorithm from this section to find the inverses of

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Let A be the corresponding $n \times n$ matrix, and let B be its inverse. Guess the form of B, and then prove that $AB = I$ and $BA = I$.

44. Repeat the strategy of Exercise 43 to guess the inverse of $A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 0 & & 0 \\ 1 & 2 & 3 & & 0 \\ \vdots & & & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}$. Prove that your guess is correct.

45. Let $A = \begin{bmatrix} -2 & -7 & -9 \\ 2 & 5 & 6 \\ 1 & 3 & 4 \end{bmatrix}$. Find the third column of A^{-1} without computing the other columns.

T 46. Let $A = \begin{bmatrix} -25 & -9 & -27 \\ 546 & 180 & 537 \\ 154 & 50 & 149 \end{bmatrix}$. Find the second and third columns of A^{-1} without computing the first column.

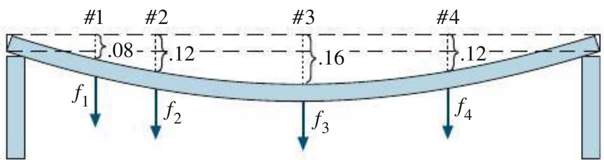
47. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}$. Construct a 2×3 matrix C (by trial and error) using only 1, -1, and 0 as entries, such that $CA = I_2$. Compute AC and note that $AC \neq I_3$.

48. Let $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$. Construct a 4×2 matrix D using only 1 and 0 as entries, such that $AD = I_2$. Is it possible that $CA = I_4$ for some 4×2 matrix C? Why or why not?

49. Let $D = \begin{bmatrix} .005 & .002 & .001 \\ .002 & .004 & .002 \\ .001 & .002 & .005 \end{bmatrix}$ be a flexibility matrix, with flexibility measured in inches per pound. Suppose that forces of 30, 50, and 20 lb are applied at points 1, 2, and 3, respectively, in Figure 1 of Example 3. Find the corresponding deflections.

T **50.** Compute the stiffness matrix D^{-1} for D in Exercise 49. List the forces needed to produce a deflection of .04 in. at point 3, with zero deflections at the other points.

T **51.** Let $D = \begin{bmatrix} .0040 & .0030 & .0010 & .0005 \\ .0030 & .0050 & .0030 & .0010 \\ .0010 & .0030 & .0050 & .0030 \\ .0005 & .0010 & .0030 & .0040 \end{bmatrix}$ be a flexibility matrix for an elastic beam with four points at which force is applied. Units are centimeters per newton of force. Measurements at the four points show deflections of .08, .12, .16, and .12 cm. Determine the forces at the four points.



Deflection of elastic beam in Exercises 51 and 52.

T **52.** With D as in Exercise 51, determine the forces that produce a deflection of .24 cm at the second point on the beam, with zero deflections at the other three points. How is the answer related to the entries in D^{-1}? [*Hint:* First answer the question when the deflection is 1 cm at the second point.]

Solutions to Practice Problems

1. a. $\det \begin{bmatrix} 3 & -9 \\ 2 & 6 \end{bmatrix} = 3 \cdot 6 - (-9) \cdot 2 = 18 + 18 = 36$. The determinant is nonzero, so the matrix is invertible.

b. $\det \begin{bmatrix} 4 & -9 \\ 0 & 5 \end{bmatrix} = 4 \cdot 5 - (-9) \cdot 0 = 20 \neq 0$. The matrix is invertible.

c. $\det \begin{bmatrix} 6 & -9 \\ -4 & 6 \end{bmatrix} = 6 \cdot 6 - (-9)(-4) = 36 - 36 = 0$. The matrix is not invertible.

2.
$$[\,A \quad I\,] \sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ -1 & 5 & 6 & 0 & 1 & 0 \\ 5 & -4 & 5 & 0 & 0 & 1 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 6 & 10 & -5 & 0 & 1 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 0 & 0 & -7 & -2 & 1 \end{bmatrix}$$

So $[\,A \quad I\,]$ is row equivalent to a matrix of the form $[\,B \quad D\,]$, where B is square and has a row of zeros. Further row operations will not transform B into I, so we stop. A does not have an inverse.

3. Since A is an invertible matrix, there exists a matrix C such that $AC = I = CA$. The goal is to find a matrix D so that $(5A)D = I = D(5A)$. Set $D = 1/5\,C$. Applying Theorem 2 from Section 2.1 establishes that $(5A)(1/5\,C) = (5)(1/5)(AC) = 1\,I = I$, and $(1/5\,C)(5A) = (1/5)(5)(CA) = 1\,I = I$. Thus $1/5\,C$ is indeed the inverse of A, proving that A is invertible.

2.3 Characterizations of Invertible Matrices

This section provides a review of most of the concepts introduced in Chapter 1, in relation to systems of n linear equations in n unknowns and to *square* matrices. The main result is Theorem 8.

THEOREM 8

The Invertible Matrix Theorem

Let A be a square $n \times n$ matrix. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

a. A is an invertible matrix.
b. A is row equivalent to the $n \times n$ identity matrix.
c. A has n pivot positions.
d. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
e. The columns of A form a linearly independent set.
f. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
g. The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.
h. The columns of A span $\mathbb{R}^n$.
i. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
j. There is an $n \times n$ matrix C such that $CA = I$.
k. There is an $n \times n$ matrix D such that $AD = I$.
l. A^T is an invertible matrix.

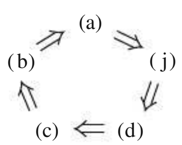
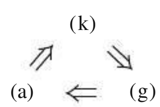


FIGURE 1

First, we need some notation. If the truth of statement (a) always implies that statement (j) is true, we say that (a) *implies* (j) and write (a) $\Rightarrow$ (j). The proof will establish the "circle" of implications shown in Figure 1. If any one of these five statements is true, then so are the others. Finally, the proof will link the remaining statements of the theorem to the statements in this circle.

PROOF If statement (a) is true, then A^{-1} works for C in (j), so (a) $\Rightarrow$ (j). Next, (j) $\Rightarrow$ (d) by Exercise 31 in Section 2.1. (Turn back and read the exercise.) Also, (d) $\Rightarrow$ (c) by Exercise 33 in Section 2.2. If A is square and has n pivot positions, then the pivots must lie on the main diagonal, in which case the reduced echelon form of A is I_n. Thus (c) $\Rightarrow$ (b). Also, (b) $\Rightarrow$ (a) by Theorem 7 in Section 2.2. This completes the circle in Figure 1.



Next, (a) $\Rightarrow$ (k) because A^{-1} works for D. Also, (k) $\Rightarrow$ (g) by Exercise 32 in Section 2.1, and (g) $\Rightarrow$ (a) by Exercise 34 in Section 2.2. So (k) and (g) are linked to the circle. Further, (g), (h), and (i) are equivalent for any matrix, by Theorem 4 in Section 1.4 and Theorem 12(a) in Section 1.9. Thus, (h) and (i) are linked through (g) to the circle.

$$(g) \Leftrightarrow (h) \Leftrightarrow (i)$$

Since (d) is linked to the circle, so are (e) and (f), because (d), (e), and (f) are all equivalent for *any* matrix A. (See Section 1.7 and Theorem 12(b) in Section 1.9.) Finally, (a) $\Rightarrow$ (l) by Theorem 6(c) in Section 2.2, and (l) $\Rightarrow$ (a) by the same theorem with A and A^T interchanged. This completes the proof. ■

$$(d) \Leftrightarrow (e) \Leftrightarrow (f)$$

$$(a) \Leftrightarrow (l)$$

Because of Theorem 5 in Section 2.2, statement (g) in Theorem 8 could also be written as "The equation $A\mathbf{x} = \mathbf{b}$ has a *unique* solution for each $\mathbf{b}$ in $\mathbb{R}^n$." This statement certainly implies (b) and hence implies that A is invertible.

The next fact follows from Theorem 8 and Exercise 10 in Section 2.2.

Let A and B be square matrices. If $AB = I$, then A and B are both invertible, with $B = A^{-1}$ and $A = B^{-1}$.

The Invertible Matrix Theorem divides the set of all $n \times n$ matrices into two disjoint classes: the invertible (nonsingular) matrices, and the noninvertible (singular) matrices. Each statement in the theorem describes a property of every $n \times n$ invertible matrix. The *negation* of a statement in the theorem describes a property of every $n \times n$ singular matrix. For instance, an $n \times n$ singular matrix is *not* row equivalent to I_n, does *not* have n pivot positions, and has linearly *dependent* columns. Negations of other statements are considered in the exercises.

EXAMPLE 1 Use the Invertible Matrix Theorem to decide if A is invertible:

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

SOLUTION

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

So A has three pivot positions and hence is invertible, by the Invertible Matrix Theorem, statement (c). ■

STUDY GUIDE offers an expanded table for the Invertible Matrix Theorem.

The power of the Invertible Matrix Theorem lies in the connections it provides among so many important concepts, such as linear independence of columns of a matrix A and the existence of solutions to equations of the form $A\mathbf{x} = \mathbf{b}$. It should be emphasized, however, that the Invertible Matrix Theorem *applies only to square matrices*. For example, if the columns of a 4×3 matrix are linearly independent, we cannot use the Invertible Matrix Theorem to conclude anything about the existence or nonexistence of solutions to equations of the form $A\mathbf{x} = \mathbf{b}$.

Invertible Linear Transformations

Recall from Section 2.1 that matrix multiplication corresponds to composition of linear transformations. When a matrix A is invertible, the equation $A^{-1}A\mathbf{x} = \mathbf{x}$ can be viewed as a statement about linear transformations. See Figure 2.

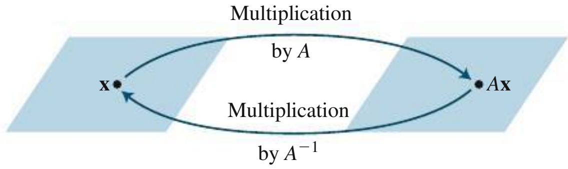


FIGURE 2 A^{-1} transforms $A\mathbf{x}$ back to $\mathbf{x}$.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be **invertible** if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$S(T(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{1}$$

$$T(S(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{2}$$

The next theorem shows that if such an S exists, it is unique and must be a linear transformation. We call S the **inverse** of T and write it as T^{-1}.

THEOREM 9

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let A be the standard matrix for T. Then T is invertible if and only if A is an invertible matrix. In that case, the linear transformation S given by $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique function satisfying equations (1) and (2).

Remark: See the comment on the proof of Theorem 7.

PROOF Suppose that T is invertible. Then (2) shows that T is onto $\mathbb{R}^n$, for if $\mathbf{b}$ is in $\mathbb{R}^n$ and $\mathbf{x} = S(\mathbf{b})$, then $T(\mathbf{x}) = T(S(\mathbf{b})) = \mathbf{b}$, so each $\mathbf{b}$ is in the range of T. Thus A is invertible, by the Invertible Matrix Theorem, statement (i).

Conversely, suppose that A is invertible, and let $S(\mathbf{x}) = A^{-1}\mathbf{x}$. Then, S is a linear transformation, and S obviously satisfies (1) and (2). For instance,

$$S(T(\mathbf{x})) = S(A\mathbf{x}) = A^{-1}(A\mathbf{x}) = \mathbf{x}$$

Thus T is invertible. The proof that S is unique is outlined in Exercise 47. ■

EXAMPLE 2 What can you say about a one-to-one linear transformation T from $\mathbb{R}^n$ into $\mathbb{R}^n$?

SOLUTION The columns of the standard matrix A of T are linearly independent (by Theorem 12 in Section 1.9). So A is invertible, by the Invertible Matrix Theorem, and T maps $\mathbb{R}^n$ *onto* $\mathbb{R}^n$. Also, T is invertible, by Theorem 9. ■

Numerical Notes

In practical work, you might occasionally encounter a "nearly singular" or **ill-conditioned** matrix—an invertible matrix that can become singular if some of its entries are changed ever so slightly. In this case, row reduction may produce fewer than n pivot positions, as a result of roundoff error. Also, roundoff error can sometimes make a singular matrix appear to be invertible.

Some matrix programs will compute a **condition number** for a square matrix. The larger the condition number, the closer the matrix is to being singular. The condition number of the identity matrix is 1. A singular matrix has an infinite condition number. In extreme cases, a matrix program may not be able to distinguish between a singular matrix and an ill-conditioned matrix.

Exercises 49–53 show that matrix computations can produce substantial error when a condition number is large.

Practice Problems

1. Determine if $A = \begin{bmatrix} 2 & 3 & 4 \\ 2 & 3 & 4 \\ 2 & 3 & 4 \end{bmatrix}$ is invertible.
2. Suppose that for a certain $n \times n$ matrix A, statement (g) of the Invertible Matrix Theorem is *not* true. What can you say about equations of the form $A\mathbf{x} = \mathbf{b}$?
3. Suppose that A and B are $n \times n$ matrices and the equation $AB\mathbf{x} = \mathbf{0}$ has a nontrivial solution. What can you say about the matrix AB?

2.3 Exercises

Unless otherwise specified, assume that all matrices in these exercises are $n \times n$. Determine which of the matrices in Exercises 1–10 are invertible. Use as few calculations as possible. Justify your answers.

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

2. $\begin{bmatrix} -4 & 6 \\ 6 & -9 \end{bmatrix}$

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

4. $\begin{bmatrix} -7 & 0 & 4 \\ 3 & 0 & -1 \\ 2 & 0 & 9 \end{bmatrix}$

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

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

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

8. $\begin{bmatrix} 1 & 3 & 7 & 4 \\ 0 & 5 & 9 & 6 \\ 0 & 0 & 2 & 8 \\ 0 & 0 & 0 & 10 \end{bmatrix}$

T 9. $\begin{bmatrix} 4 & 0 & -7 & -7 \\ -6 & 1 & 11 & 9 \\ 7 & -5 & 10 & 19 \\ -1 & 2 & 3 & -1 \end{bmatrix}$

T 10. $\begin{bmatrix} 5 & 3 & 1 & 7 & 9 \\ 6 & 4 & 2 & 8 & -8 \\ 7 & 5 & 3 & 10 & 9 \\ 9 & 6 & 4 & -9 & -5 \\ 8 & 5 & 2 & 11 & 4 \end{bmatrix}$

In Exercises 11–20, the matrices are all $n \times n$. Each part of the exercises is an *implication* of the form "If 'statement 1', then 'statement 2'." Mark an implication as True if the truth of "statement 2" *always* follows whenever "statement 1" happens to be true. An implication is False if there is an instance in which "statement 2" is false but "statement 1" is true. Justify each answer.

11. **(T/F)** If the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then A is row equivalent to the $n \times n$ identity matrix.

12. **(T/F)** If there is an $n \times n$ matrix D such that $AD = I$, then there is also an $n \times n$ matrix C such that $CA = I$.

13. **(T/F)** If the columns of A span $\mathbb{R}^n$, then the columns are linearly independent.

14. **(T/F)** If the columns of A are linearly independent, then the columns of A span $\mathbb{R}^n$.

15. **(T/F)** If A is an $n \times n$ matrix, then the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.

16. **(T/F)** If the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$, then the solution is unique for each $\mathbf{b}$.

17. **(T/F)** If the equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, then A has fewer than n pivot positions.

18. **(T/F)** If the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ into $\mathbb{R}^n$, then A has n pivot positions.

19. **(T/F)** If A^T is not invertible, then A is not invertible.

20. **(T/F)** If there is a $\mathbf{b}$ in $\mathbb{R}^n$ such that the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is not one-to-one.

21. An $m \times n$ **upper triangular matrix** is one whose entries *below* the main diagonal are 0's (as in Exercise 8). When is a square upper triangular matrix invertible? Justify your answer.

22. An $m \times n$ **lower triangular matrix** is one whose entries *above* the main diagonal are 0's (as in Exercise 3). When is a square lower triangular matrix invertible? Justify your answer.

23. Can a square matrix with two identical columns be invertible? Why or why not?

24. Is it possible for a 5×5 matrix to be invertible when its columns do not span $\mathbb{R}^5$? Why or why not?

25. If A is invertible, then the columns of A^{-1} are linearly independent. Explain why.

26. If C is 6×6 and the equation $C\mathbf{x} = \mathbf{v}$ is consistent for every $\mathbf{v}$ in $\mathbb{R}^6$, is it possible that for some $\mathbf{v}$, the equation $C\mathbf{x} = \mathbf{v}$ has more than one solution? Why or why not?

27. If the columns of a 7×7 matrix D are linearly independent, what can you say about solutions of $D\mathbf{x} = \mathbf{b}$? Why?

28. If $n \times n$ matrices E and F have the property that $EF = I$, then E and F commute. Explain why.

29. If the equation $G\mathbf{x} = \mathbf{y}$ has more than one solution for some $\mathbf{y}$ in $\mathbb{R}^n$, can the columns of G span $\mathbb{R}^n$? Why or why not?

30. If the equation $H\mathbf{x} = \mathbf{c}$ is inconsistent for some $\mathbf{c}$ in $\mathbb{R}^n$, what can you say about the equation $H\mathbf{x} = \mathbf{0}$? Why?

31. If an $n \times n$ matrix K cannot be row reduced to I_n, what can you say about the columns of K? Why?

32. If L is $n \times n$ and the equation $L\mathbf{x} = \mathbf{0}$ has the trivial solution, do the columns of L span $\mathbb{R}^n$? Why?

33. Verify the boxed statement preceding Example 1.

34. Explain why the columns of A^2 span $\mathbb{R}^n$ whenever the columns of A are linearly independent.

35. Show that if AB is invertible, so is A. You cannot use Theorem 6(b), because you cannot *assume* that A and B are invertible. [*Hint:* There is a matrix W such that $ABW = I$. Why?]

36. Show that if AB is invertible, so is B.

37. If A is an $n \times n$ matrix and the equation $A\mathbf{x} = \mathbf{b}$ has more than one solution for some $\mathbf{b}$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is

not one-to-one. What else can you say about this transformation? Justify your answer.

38. If A is an $n \times n$ matrix and the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one, what else can you say about this transformation? Justify your answer.

39. Suppose A is an $n \times n$ matrix with the property that the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$. Without using Theorems 5 or 8, explain why each equation $A\mathbf{x} = \mathbf{b}$ has in fact exactly one solution.

40. Suppose A is an $n \times n$ matrix with the property that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Without using the Invertible Matrix Theorem, explain directly why the equation $A\mathbf{x} = \mathbf{b}$ must have a solution for each $\mathbf{b}$ in $\mathbb{R}^n$.

In Exercises 41 and 42, T is a linear transformation from $\mathbb{R}^2$ into $\mathbb{R}^2$. Show that T is invertible and find a formula for T^{-1}.

41. $T(x_1, x_2) = (-5x_1 + 9x_2, 4x_1 - 7x_2)$

42. $T(x_1, x_2) = (6x_1 - 8x_2, -5x_1 + 7x_2)$

43. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Explain why T is both one-to-one and onto $\mathbb{R}^n$. Use equations (1) and (2). Then give a second explanation using one or more theorems.

44. Let T be a linear transformation that maps $\mathbb{R}^n$ onto $\mathbb{R}^n$. Show that T^{-1} exists and maps $\mathbb{R}^n$ onto $\mathbb{R}^n$. Is T^{-1} also one-to-one?

45. Suppose T and U are linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that $T(U\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. Is it true that $U(T\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$? Why or why not?

46. Suppose a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ has the property that $T(\mathbf{u}) = T(\mathbf{v})$ for some pair of distinct vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$. Can T map $\mathbb{R}^n$ onto $\mathbb{R}^n$? Why or why not?

47. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation, and let S and U be functions from $\mathbb{R}^n$ into $\mathbb{R}^n$ such that $S(T(\mathbf{x})) = \mathbf{x}$ and $U(T(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. Show that $U(\mathbf{v}) = S(\mathbf{v})$ for all $\mathbf{v}$ in $\mathbb{R}^n$. This will show that T has a unique inverse, as asserted in Theorem 9. [*Hint:* Given any $\mathbf{v}$ in $\mathbb{R}^n$, we can write $\mathbf{v} = T(\mathbf{x})$ for some $\mathbf{x}$. Why? Compute $S(\mathbf{v})$ and $U(\mathbf{v})$.]

48. Suppose T and S satisfy the invertibility equations (1) and (2), where T is a linear transformation. Show directly that S is a linear transformation. [*Hint:* Given $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$, let $\mathbf{x} = S(\mathbf{u})$, $\mathbf{y} = S(\mathbf{v})$. Then $T(\mathbf{x}) = \mathbf{u}$, $T(\mathbf{y}) = \mathbf{v}$. Why? Apply S to both sides of the equation $T(\mathbf{x}) + T(\mathbf{y}) = T(\mathbf{x} + \mathbf{y})$. Also, consider $T(c\mathbf{x}) = cT(\mathbf{x})$.]

T 49. Suppose an experiment leads to the following system of equations:

$$\begin{aligned} 4.5x_1 + 3.1x_2 &= 19.249 \\ 1.6x_1 + 1.1x_2 &= 6.843 \end{aligned} \tag{3}$$

a. Solve system (3), and then solve system (4), below, in which the data on the right have been rounded to two decimal places. In each case, find the *exact* solution.

$$\begin{aligned} 4.5x_1 + 3.1x_2 &= 19.25 \\ 1.6x_1 + 1.1x_2 &= 6.84 \end{aligned} \tag{4}$$

b. The entries in (4) differ from those in (3) by less than .05%. Find the percentage error when using the solution of (4) as an approximation for the solution of (3).

c. Use your matrix program to produce the condition number of the coefficient matrix in (3).

Exercises 50–52 show how to use the condition number of a matrix A to estimate the accuracy of a computed solution of $A\mathbf{x} = \mathbf{b}$. If the entries of A and $\mathbf{b}$ are accurate to about r significant digits and if the condition number of A is approximately 10^k (with k a positive integer), then the computed solution of $A\mathbf{x} = \mathbf{b}$ should usually be accurate to at least $r - k$ significant digits.

T 50. Find the condition number of the matrix A in Exercise 9. Construct a random vector $\mathbf{x}$ in $\mathbb{R}^4$ and compute $\mathbf{b} = A\mathbf{x}$. Then use your matrix program to compute the solution $\mathbf{x}_1$ of $A\mathbf{x} = \mathbf{b}$. To how many digits do $\mathbf{x}$ and $\mathbf{x}_1$ agree? Find out the number of digits your matrix program stores accurately, and report how many digits of accuracy are lost when $\mathbf{x}_1$ is used in place of the exact solution $\mathbf{x}$.

T 51. Repeat Exercise 50 for the matrix in Exercise 10.

T 52. Solve an equation $A\mathbf{x} = \mathbf{b}$ for a suitable $\mathbf{b}$ to find the last column of the inverse of the *fifth-order Hilbert matrix*

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

How many digits in each entry of $\mathbf{x}$ do you expect to be correct? Explain. [*Note:* The exact solution is (630, −12600, 56700, −88200, 44100).]

T 53. Some matrix programs, such as MATLAB, have a command to create Hilbert matrices of various sizes. If possible, use an inverse command to compute the inverse of a twelfth-order or larger Hilbert matrix, A. Compute AA^{-1}. Report what you find.

STUDY GUIDE offers additional resources for reviewing and reflecting on what you have learned.

Solutions to Practice Problems

1. The columns of A are obviously linearly dependent because columns 2 and 3 are multiples of column 1. Hence, A cannot be invertible (by the Invertible Matrix Theorem).

Solutions to Practice Problems (Continued)

2. If statement (g) is *not* true, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for at least one $\mathbf{b}$ in $\mathbb{R}^n$.

3. Apply the Invertible Matrix Theorem to the matrix AB in place of A. Then statement (d) becomes: $AB\mathbf{x} = \mathbf{0}$ has only the trivial solution. This is not true. So AB is not invertible.

2.4 Partitioned Matrices

A key feature of our work with matrices has been the ability to regard a matrix A as a list of column vectors rather than just a rectangular array of numbers. This point of view has been so useful that we wish to consider other **partitions** of A, indicated by horizontal and vertical dividing rules, as in Example 1 below. Partitioned matrices appear in most modern applications of linear algebra because the notation highlights essential structures in matrix analysis, as in the chapter introductory example on aircraft design. This section provides an opportunity to review matrix algebra and use the Invertible Matrix Theorem.

EXAMPLE 1 The matrix

$$A = \left[\begin{array}{rrr|rr|r} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \\ \hline -8 & -6 & 3 & 1 & 7 & -4 \end{array}\right]$$

can also be written as the 2×3 **partitioned** (or **block**) **matrix**

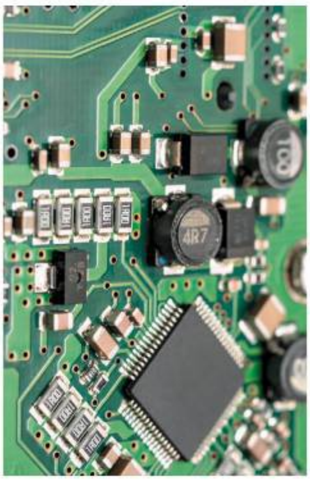
$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

whose entries are the *blocks* (or *submatrices*)

$$A_{11} = \begin{bmatrix} 3 & 0 & -1 \\ -5 & 2 & 4 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 5 & 9 \\ 0 & -3 \end{bmatrix}, \quad A_{13} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$A_{21} = \begin{bmatrix} -8 & -6 & 3 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 1 & 7 \end{bmatrix}, \quad A_{23} = \begin{bmatrix} -4 \end{bmatrix}$$

■



EXAMPLE 2 When a matrix A appears in a mathematical model of a physical system such as an electrical network, a transportation system, or a large corporation, it may be natural to regard A as a partitioned matrix. For instance, if a microcomputer circuit board consists mainly of three VLSI (very large-scale integrated) microchips, then the matrix for the circuit board might have the general form

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

The submatrices on the "diagonal" of A—namely A_{11}, A_{22}, and A_{33}—concern the three VLSI chips, while the other submatrices depend on the interconnections among those microchips. ■

Addition and Scalar Multiplication

If matrices A and B are the same size and are partitioned in exactly the same way, then it is natural to make the same partition of the ordinary matrix sum $A + B$. In this case, each block of $A + B$ is the (matrix) sum of the corresponding blocks of A and B. Multiplication of a partitioned matrix by a scalar is also computed block by block.

Multiplication of Partitioned Matrices

Partitioned matrices can be multiplied by the usual row–column rule as if the block entries were scalars, provided that for a product AB, the column partition of A matches the row partition of B.

EXAMPLE 3 Let

$$A = \left[\begin{array}{rrr|rr} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ \hline 0 & -4 & -2 & 7 & -1 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \left[\begin{array}{rr} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ \hline -1 & 3 \\ 5 & 2 \end{array}\right] = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

The 5 columns of A are partitioned into a set of 3 columns and then a set of 2 columns. The 5 rows of B are partitioned in the same way—into a set of 3 rows and then a set of 2 rows. We say that the partitions of A and B are **conformable** for **block multiplication**. It can be shown that the ordinary product AB can be written as

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix} = \left[\begin{array}{rr} -5 & 4 \\ -6 & 2 \\ \hline 2 & 1 \end{array}\right]$$

It is important for each smaller product in the expression for AB to be written with the submatrix from A on the left, since matrix multiplication is not commutative. For instance,

$$A_{11}B_1 = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 5 & -2 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \end{bmatrix} = \begin{bmatrix} 15 & 12 \\ 2 & -5 \end{bmatrix}$$

$$A_{12}B_2 = \begin{bmatrix} 0 & -4 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} -20 & -8 \\ -8 & 7 \end{bmatrix}$$

Hence the top block in AB is

$$A_{11}B_1 + A_{12}B_2 = \begin{bmatrix} 15 & 12 \\ 2 & -5 \end{bmatrix} + \begin{bmatrix} -20 & -8 \\ -8 & 7 \end{bmatrix} = \begin{bmatrix} -5 & 4 \\ -6 & 2 \end{bmatrix} \quad \blacksquare$$

The row–column rule for multiplication of block matrices provides the most general way to regard the product of two matrices. Each of the following views of a product has already been described using simple partitions of matrices: (1) the definition of $A\mathbf{x}$ using the columns of A, (2) the column definition of AB, (3) the row–column rule for computing AB, and (4) the rows of AB as products of the rows of A and the matrix B. A fifth view of AB, again using partitions, follows in Theorem 10.

The calculations in the next example prepare the way for Theorem 10. Here $\text{col}_k(A)$ is the kth column of A, and $\text{row}_k(B)$ is the kth row of B.

EXAMPLE 4 Let $A = \begin{bmatrix} -3 & 1 & 2 \\ 1 & -4 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$. Verify that

$$AB = \text{col}_1(A)\,\text{row}_1(B) + \text{col}_2(A)\,\text{row}_2(B) + \text{col}_3(A)\,\text{row}_3(B)$$

SOLUTION Each term in the preceding equation is an *outer product*. (See Exercises 35 and 36 in Section 2.1.) By the row–column rule for computing a matrix product,

$$\mathrm{col}_1(A)\,\mathrm{row}_1(B) = \begin{bmatrix} -3 \\ 1 \end{bmatrix}\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} -3a & -3b \\ a & b \end{bmatrix}$$
$$\mathrm{col}_2(A)\,\mathrm{row}_2(B) = \begin{bmatrix} 1 \\ -4 \end{bmatrix}\begin{bmatrix} c & d \end{bmatrix} = \begin{bmatrix} c & d \\ -4c & -4d \end{bmatrix}$$
$$\mathrm{col}_3(A)\,\mathrm{row}_3(B) = \begin{bmatrix} 2 \\ 5 \end{bmatrix}\begin{bmatrix} e & f \end{bmatrix} = \begin{bmatrix} 2e & 2f \\ 5e & 5f \end{bmatrix}$$

Thus

$$\sum_{k=1}^{3} \mathrm{col}_k(A)\,\mathrm{row}_k(B) = \begin{bmatrix} -3a + c + 2e & -3b + d + 2f \\ a - 4c + 5e & b - 4d + 5f \end{bmatrix}$$

This matrix is obviously AB. Notice that the $(1, 1)$-entry in AB is the sum of the $(1, 1)$-entries in the three outer products, the $(1, 2)$-entry in AB is the sum of the $(1, 2)$-entries in the three outer products, and so on. ■

THEOREM 10

Column–Row Expansion of AB

If A is $m \times n$ and B is $n \times p$, then

$$\begin{aligned} AB &= \begin{bmatrix} \mathrm{col}_1(A) & \mathrm{col}_2(A) & \cdots & \mathrm{col}_n(A) \end{bmatrix} \begin{bmatrix} \mathrm{row}_1(B) \\ \mathrm{row}_2(B) \\ \vdots \\ \mathrm{row}_n(B) \end{bmatrix} \qquad (1) \\ &= \mathrm{col}_1(A)\,\mathrm{row}_1(B) + \cdots + \mathrm{col}_n(A)\,\mathrm{row}_n(B) \end{aligned}$$

PROOF For each row index i and column index j, the (i, j)-entry in $\mathrm{col}_k(A)\,\mathrm{row}_k(B)$ is the product of a_{ik} from $\mathrm{col}_k(A)$ and b_{kj} from $\mathrm{row}_k(B)$. Hence the (i, j)-entry in the sum shown in equation (1) is

$$\underset{(k=1)}{a_{i1}b_{1j}} + \underset{(k=2)}{a_{i2}b_{2j}} + \cdots + \underset{(k=n)}{a_{in}b_{nj}}$$

This sum is also the (i, j)-entry in AB, by the row–column rule. ■

Inverses of Partitioned Matrices

The next example illustrates calculations involving inverses and partitioned matrices.

EXAMPLE 5 A matrix of the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

is said to be *block upper triangular*. Assume that A_{11} is $p \times p$, A_{22} is $q \times q$, and A is invertible. Find a formula for A^{-1}.

SOLUTION Denote A^{-1} by B and partition B so that

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix} \tag{2}$$

This matrix equation provides four equations that will lead to the unknown blocks $B_{11}, \dots, B_{22}$. Compute the product on the left side of equation (2), and equate each entry with the corresponding block in the identity matrix on the right. That is, set

$$A_{11}B_{11} + A_{12}B_{21} = I_p \tag{3}$$
$$A_{11}B_{12} + A_{12}B_{22} = 0 \tag{4}$$
$$A_{22}B_{21} = 0 \tag{5}$$
$$A_{22}B_{22} = I_q \tag{6}$$

By itself, equation (6) does not show that A_{22} is invertible. However, since A_{22} is square, the Invertible Matrix Theorem and (6) together show that A_{22} is invertible and $B_{22} = A_{22}^{-1}$. Next, left-multiply both sides of (5) by A_{22}^{-1} and obtain

$$B_{21} = A_{22}^{-1}0 = 0$$

so that (3) simplifies to

$$A_{11}B_{11} + 0 = I_p$$

Since A_{11} is square, this shows that A_{11} is invertible and $B_{11} = A_{11}^{-1}$. Finally, use these results with (4) to find that

$$A_{11}B_{12} = -A_{12}B_{22} = -A_{12}A_{22}^{-1} \quad \text{and} \quad B_{12} = -A_{11}^{-1}A_{12}A_{22}^{-1}$$

Thus

$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}$$ ■

A **block diagonal matrix** is a partitioned matrix with zero blocks off the main diagonal (of blocks). Such a matrix is invertible if and only if each block on the diagonal is invertible. See Exercises 15 and 16.

Numerical Notes

1. When matrices are too large to fit in a computer's high-speed memory, partitioning permits the computer to work with only two or three submatrices at a time. For instance, one linear programming research team simplified a problem by partitioning the matrix into 837 rows and 51 columns. The problem's solution took about 4 minutes on a Cray supercomputer.[1]
2. Some high-speed computers, particularly those with vector pipeline architecture, perform matrix calculations more efficiently when the algorithms use partitioned matrices.[2]
3. Professional software for high-performance numerical linear algebra, such as LAPACK, makes intensive use of partitioned matrix calculations.

[1] The solution time doesn't sound too impressive until you learn that each of the 51 block columns contained about 250,000 individual columns. The original problem had 837 equations and more than 12,750,000 variables! Nearly 100 million of the more than 10 billion entries in the matrix were nonzero. See Robert E. Bixby et al., "Very Large-Scale Linear Programming: A Case Study in Combining Interior Point and Simplex Methods," *Operations Research*, 40, no. 5 (1992): 885–897.

[2] The importance of block matrix algorithms for computer calculations is described in *Matrix Computations*, 3rd ed., by Gene H. Golub and Charles F. van Loan (Baltimore: Johns Hopkins University Press, 1996).

The exercises that follow give practice with matrix algebra and illustrate typical calculations found in applications.

Practice Problems

1. Show that $\begin{bmatrix} I & 0 \\ A & I \end{bmatrix}$ is invertible and find its inverse.
2. Compute $X^T X$, where X is partitioned as $\begin{bmatrix} X_1 & X_2 \end{bmatrix}$.

2.4 Exercises

In Exercises 1–9, assume that the matrices are partitioned conformably for block multiplication. Compute the products shown in Exercises 1–4.

1. $\begin{bmatrix} I & 0 \\ E & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}$

2. $\begin{bmatrix} E & 0 \\ 0 & F \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}$

3. $\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}\begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$

4. $\begin{bmatrix} I & 0 \\ -X & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}$

In Exercises 5–8, find formulas for X, Y, and Z in terms of A, B, and C, and justify your calculations. In some cases, you may need to make assumptions about the size of a matrix in order to produce a formula. [*Hint:* Compute the product on the left, and set it equal to the right side.]

5. $\begin{bmatrix} A & B \\ C & 0 \end{bmatrix}\begin{bmatrix} I & 0 \\ X & Y \end{bmatrix} = \begin{bmatrix} 0 & I \\ Z & 0 \end{bmatrix}$

6. $\begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}\begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$

7. $\begin{bmatrix} X & 0 & 0 \\ Y & 0 & I \end{bmatrix}\begin{bmatrix} A & Z \\ 0 & 0 \\ B & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$

8. $\begin{bmatrix} A & B \\ 0 & I \end{bmatrix}\begin{bmatrix} X & Y & Z \\ 0 & 0 & I \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & I \end{bmatrix}$

9. Suppose A_{11} is an invertible matrix. Find matrices X and Y such that the product below has the form indicated. Also, compute B_{22}. [*Hint:* Compute the product on the left, and set it equal to the right side.]

$$\begin{bmatrix} I & 0 & 0 \\ X & I & 0 \\ Y & 0 & I \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \\ 0 & B_{32} \end{bmatrix}$$

10. The inverse of $\begin{bmatrix} I & 0 & 0 \\ C & I & 0 \\ A & B & I \end{bmatrix}$ is $\begin{bmatrix} I & 0 & 0 \\ Z & I & 0 \\ X & Y & I \end{bmatrix}$.

Find X, Y, and Z.

In Exercises 11–14, mark each statement True or False **(T/F)**. Justify each answer.

11. (T/F) If $A = [A_1 \quad A_2]$ and $B = [B_1 \quad B_2]$, with A_1 and A_2 the same sizes as B_1 and B_2, respectively, then $A + B = [A_1 + B_1 \quad A_2 + B_2]$.

12. (T/F) The definition of the matrix–vector product $A\mathbf{x}$ is a special case of block multiplication.

13. (T/F) If $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ and $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$, then the partitions of A and B are conformable for block multiplication.

14. (T/F) If A_1, A_2, B_1, and B_2 are $n \times n$ matrices, $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$, and $B = \begin{bmatrix} B_1 & B_2 \end{bmatrix}$, then the product BA is defined, but AB is not.

15. Let $A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$, where B and C are square. Show that A is invertible if and only if both B and C are invertible.

16. Show that the block upper triangular matrix A in Example 5 is invertible if and only if both A_{11} and A_{22} are invertible. [*Hint:* If A_{11} and A_{22} are invertible, the formula for A^{-1} given in Example 5 actually works as the inverse of A.] This fact about A is an important part of several computer algorithms that estimate eigenvalues of matrices. Eigenvalues are discussed in Chapter 5.

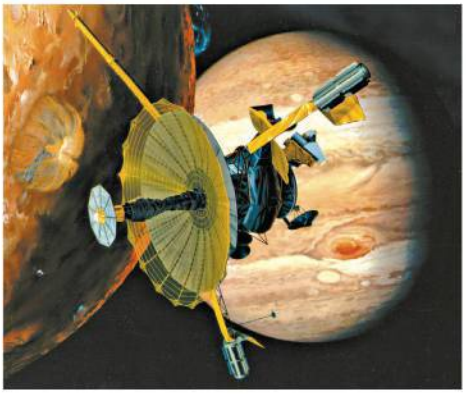
17. Suppose A_{11} is invertible. Find X and Y such that

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix}\begin{bmatrix} A_{11} & 0 \\ 0 & S \end{bmatrix}\begin{bmatrix} I & Y \\ 0 & I \end{bmatrix} \tag{7}$$

where $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$. The matrix S is called the **Schur complement** of A_{11}. Likewise, if A_{22} is invertible, the matrix $A_{11} - A_{12}A_{22}^{-1}A_{21}$ is called the Schur complement of A_{22}. Such expressions occur frequently in the theory of systems engineering, and elsewhere.

18. Suppose the block matrix A on the left side of (7) is invertible and A_{11} is invertible. Show that the Schur complement S of A_{11} is invertible. [*Hint:* The outside factors on the right side of (7) are always invertible. Verify this.] When A and A_{11} are both invertible, (7) leads to a formula for A^{-1}, using S^{-1}, A_{11}^{-1}, and the other entries in A.

19. When a deep space probe is launched, corrections may be necessary to place the probe on a precisely calculated trajectory. Radio telemetry provides a stream of vectors, $\mathbf{x}_1, \dots, \mathbf{x}_k$, giving information at different times about how the probe's position compares with its planned trajectory. Let X_k be the matrix $[\mathbf{x}_1 \cdots \mathbf{x}_k]$. The matrix $G_k = X_k X_k^T$ is computed as the radar data are analyzed. When $\mathbf{x}_{k+1}$ arrives, a new G_{k+1} must be computed. Since the data vectors arrive at high speed, the computational burden could be severe. But partitioned matrix multiplication helps tremendously. Compute the column–row expansions of G_k and G_{k+1}, and describe what must be computed in order to *update* G_k to form G_{k+1}.



The probe Galileo was launched October 18, 1989, and arrived near Jupiter in early December 1995.

20. Let X be an $m \times n$ data matrix such that $X^T X$ is invertible, and let $M = I_m - X(X^T X)^{-1} X^T$. Add a column $\mathbf{x}_0$ to the data and form

$$W = [X \quad \mathbf{x}_0]$$

Compute $W^T W$. The (1, 1)-entry is $X^T X$. Show that the Schur complement (Exercise 17) of $X^T X$ can be written in the form $\mathbf{x}_0^T M \mathbf{x}_0$. It can be shown that the quantity $(\mathbf{x}_0^T M \mathbf{x}_0)^{-1}$ is the (2, 2)-entry in $(W^T W)^{-1}$. This entry has a useful statistical interpretation, under appropriate hypotheses.

In the study of engineering control of physical systems, a standard set of differential equations is transformed by Laplace transforms into the following system of linear equations:

$$\begin{bmatrix} A - sI_n & B \\ C & I_m \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix} \tag{8}$$

where A is $n \times n$, B is $n \times m$, C is $m \times n$, and s is a variable. The vector $\mathbf{u}$ in $\mathbb{R}^m$ is the "input" to the system, $\mathbf{y}$ in $\mathbb{R}^m$ is the "output," and $\mathbf{x}$ in $\mathbb{R}^n$ is the "state" vector. (Actually, the vectors $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{y}$ are functions of s, but we suppress this fact because it does not affect the algebraic calculations in Exercises 21 and 22.)

21. Assume $A - sI_n$ is invertible and view (8) as a system of two matrix equations. Solve the top equation for $\mathbf{x}$ and substitute into the bottom equation. The result is an equation of the form $W(s)\mathbf{u} = \mathbf{y}$, where $W(s)$ is a matrix that depends on s. $W(s)$ is called the *transfer function* of the system because it transforms the input $\mathbf{u}$ into the output $\mathbf{y}$. Find $W(s)$ and describe how it is related to the partitioned *system matrix* on the left side of (8). See Exercise 17.

22. Suppose the transfer function $W(s)$ in Exercise 21 is invertible for some s. It can be shown that the inverse transfer function $W(s)^{-1}$, which transforms outputs into inputs, is the Schur complement of $A - BC - sI_n$ for the matrix below. Find this Schur complement. See Exercise 17.

$$\begin{bmatrix} A - BC - sI_n & B \\ -C & I_m \end{bmatrix}$$

23. a. Verify that $A^2 = I$ when $A = \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}$.

b. Use partitioned matrices to show that $M^2 = I$ when

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

24. Generalize the idea of Exercise 23(a) [not 23(b)] by constructing a 5×5 matrix $M = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix}$ such that $M^2 = I$. Make C a nonzero 2×3 matrix. Show that your construction works.

25. Use partitioned matrices to prove by induction that the product of two lower triangular matrices is also lower triangular. [*Hint:* A $(k+1) \times (k+1)$ matrix A_1 can be written in the form below, where a is a scalar, $\mathbf{v}$ is in $\mathbb{R}^k$, and A is a $k \times k$ lower triangular matrix. See the *Study Guide* for help with induction.]

$$A_1 = \begin{bmatrix} a & \mathbf{0}^T \\ \mathbf{v} & A \end{bmatrix}$$

26. Use partitioned matrices to prove by induction that for $n = 2, 3, \dots$, the $n \times n$ matrix A shown below is invertible and B is its inverse.

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & & 0 \\ 1 & 1 & 1 & & 0 \\ \vdots & & & \ddots & \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & & \dots & -1 & 1 \end{bmatrix}$$

For the induction step, assume A and B are $(k+1) \times (k+1)$ matrices, and partition A and B in a form similar to that displayed in Exercise 25.

27. Without using row reduction, find the inverse of

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

T 28. For block operations, it may be necessary to access or enter submatrices of a large matrix. Describe the functions or commands of your matrix program that accomplish the following tasks. Suppose A is a 20×30 matrix.

a. Display the submatrix of A from rows 15 to 20 and columns 5 to 10.

b. Insert a 5×10 matrix B into A, beginning at row 10 and column 20.

c. Create a 50×50 matrix of the form $B = \begin{bmatrix} A & 0 \\ 0 & A^T \end{bmatrix}$.

[*Note:* It may not be necessary to specify the zero blocks in B.]

T 29. Suppose memory or size restrictions prevent your matrix program from working with matrices having more than 32 rows and 32 columns, and suppose some project involves 50×50 matrices A and B. Describe the commands or operations of your matrix program that accomplish the following tasks.

a. Compute $A + B$.

b. Compute AB.

c. Solve $A\mathbf{x} = \mathbf{b}$ for some vector $\mathbf{b}$ in $\mathbb{R}^{50}$, assuming that A can be partitioned into a 2×2 block matrix $[A_{ij}]$, with A_{11} an invertible 20×20 matrix, A_{22} an invertible 30×30 matrix, and A_{12} a zero matrix. [*Hint:* Describe appropriate smaller systems to solve, without using any matrix inverses.]

Solutions to Practice Problems

1. If $\begin{bmatrix} I & 0 \\ A & I \end{bmatrix}$ is invertible, its inverse has the form $\begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$. Verify that

$$\begin{bmatrix} I & 0 \\ A & I \end{bmatrix}\begin{bmatrix} W & X \\ Y & Z \end{bmatrix} = \begin{bmatrix} W & X \\ AW + Y & AX + Z \end{bmatrix}$$

So W, X, Y, and Z must satisfy $W = I$, $X = 0$, $AW + Y = 0$, and $AX + Z = I$. It follows that $Y = -A$ and $Z = I$. Hence

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

The product in the reverse order is also the identity, so the block matrix is invertible, and its inverse is $\begin{bmatrix} I & 0 \\ -A & I \end{bmatrix}$. (You could also appeal to the Invertible Matrix Theorem.)

2. $X^T X = \begin{bmatrix} X_1^T \\ X_2^T \end{bmatrix}\begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{bmatrix}$. The partitions of X^T and X are automatically conformable for block multiplication because the columns of X^T are the rows of X. This partition of $X^T X$ is used in several computer algorithms for matrix computations.

2.5 Matrix Factorizations

A *factorization* of a matrix A is an equation that expresses A as a product of two or more matrices. Whereas matrix multiplication involves a *synthesis* of data (combining the effects of two or more linear transformations into a single matrix), matrix factorization is an *analysis* of data. In the language of computer science, the expression of A as a product amounts to a *preprocessing* of the data in A, organizing that data into two or more parts whose structures are more useful in some way, perhaps more accessible for computation.

Matrix factorizations and, later, factorizations of linear transformations will appear at a number of key points throughout the text. This section focuses on a factorization that lies at the heart of several important computer programs widely used in applications, such as the airflow problem described in the chapter introduction. Several other factorizations, to be studied later, are introduced in the exercises.

The LU Factorization

The LU factorization, described below, is motivated by the fairly common industrial and business problem of solving a sequence of equations, all with the same coefficient matrix:

$$A\mathbf{x} = \mathbf{b}_1, \quad A\mathbf{x} = \mathbf{b}_2, \quad \ldots, \quad A\mathbf{x} = \mathbf{b}_p \tag{1}$$

See Exercise 32, for example. Also see Section 5.8, where the inverse power method is used to estimate eigenvalues of a matrix by solving equations like those in sequence (1), one at a time.

When A is invertible, one could compute A^{-1} and then compute $A^{-1}\mathbf{b}_1$, $A^{-1}\mathbf{b}_2$, and so on. However, it is more efficient to solve the first equation in sequence (1) by row reduction and obtain an LU factorization of A at the same time. Thereafter, the remaining equations in sequence (1) are solved with the LU factorization.

At first, assume that A is an $m \times n$ matrix that can be row reduced to echelon form, *without row interchanges*. (Later, we will treat the general case.) Then A can be written in the form $A = LU$, where L is an $m \times m$ lower triangular matrix with 1's on the diagonal and U is an $m \times n$ echelon form of A. For instance, see Figure 1. Such a factorization is called an **LU factorization** of A. The matrix L is invertible and is called a *unit* lower triangular matrix.

$$A = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{U}$$

FIGURE 1 An LU factorization.

Before studying how to construct L and U, we should look at why they are so useful. When $A = LU$, the equation $A\mathbf{x} = \mathbf{b}$ can be written as $L(U\mathbf{x}) = \mathbf{b}$. Writing $\mathbf{y}$ for $U\mathbf{x}$, we can find $\mathbf{x}$ by solving the *pair* of equations

$$\boxed{\begin{aligned} L\mathbf{y} &= \mathbf{b} \\ U\mathbf{x} &= \mathbf{y} \end{aligned}} \tag{2}$$

First solve $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$, and then solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$. See Figure 2. Each equation is easy to solve because L and U are triangular.

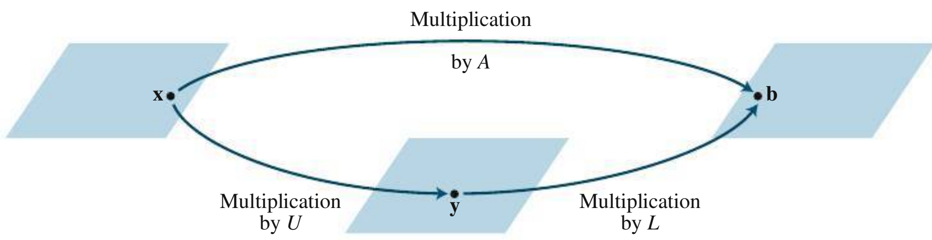


FIGURE 2 Factorization of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

EXAMPLE 1 It can be verified that

$$A = \begin{bmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 0 \\ -3 & 8 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & -1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} = LU$$

Use this LU factorization of A to solve $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = \begin{bmatrix} -9 \\ 5 \\ 7 \\ 11 \end{bmatrix}$.

SOLUTION The solution of $L\mathbf{y} = \mathbf{b}$ needs only 6 multiplications and 6 additions, because the arithmetic takes place only in column 5. (The zeros below each pivot in L are created automatically by the choice of row operations.)

$$\begin{bmatrix} L & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ -1 & 1 & 0 & 0 & 5 \\ 2 & -5 & 1 & 0 & 7 \\ -3 & 8 & 3 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} I & \mathbf{y} \end{bmatrix}$$

Then, for $U\mathbf{x} = \mathbf{y}$, the "backward" phase of row reduction requires 4 divisions, 6 multiplications, and 6 additions. (For instance, creating the zeros in column 4 of $\begin{bmatrix} U & \mathbf{y} \end{bmatrix}$ requires 1 division in row 4 and 3 multiplication–addition pairs to add multiples of row 4 to the rows above.)

$$\begin{bmatrix} U & \mathbf{y} \end{bmatrix} = \begin{bmatrix} 3 & -7 & -2 & 2 & -9 \\ 0 & -2 & -1 & 2 & -4 \\ 0 & 0 & -1 & 1 & 5 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ -6 \\ -1 \end{bmatrix}$$

To find $\mathbf{x}$ requires 28 arithmetic operations, or "flops" (floating point operations), excluding the cost of finding L and U. In contrast, row reduction of $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to $\begin{bmatrix} I & \mathbf{x} \end{bmatrix}$ takes 62 operations. ■

The computational efficiency of the LU factorization depends on knowing L and U. The next algorithm shows that the row reduction of A to an echelon form U amounts to an LU factorization because it produces L with essentially no extra work. After the first row reduction, L and U are available for solving additional equations whose coefficient matrix is A.

An LU Factorization Algorithm

Suppose A can be reduced to an echelon form U using only row replacements that add a multiple of one row to another row *below it*. In this case, there exist unit lower triangular elementary matrices $E_1, \ldots, E_p$ such that

$$E_p \cdots E_1 A = U \tag{3}$$

Then

$$A = (E_p \cdots E_1)^{-1} U = LU$$

where

$$L = (E_p \cdots E_1)^{-1} \tag{4}$$

It can be shown that products and inverses of unit lower triangular matrices are also unit lower triangular. (For instance, see Exercise 19.) Thus L is unit lower triangular.

Note that the row operations in equation (3), which reduce A to U, also reduce the L in equation (4) to I, because $E_p \cdots E_1 L = (E_p \cdots E_1)(E_p \cdots E_1)^{-1} = I$. This observation is the key to *constructing* L.

ALGORITHM FOR AN LU FACTORIZATION

1. Reduce A to an echelon form U by a sequence of row replacement operations, if possible.
2. Place entries in L such that the *same sequence of row operations* reduces L to I.

Step 1 is not always possible, but when it is, the argument above shows that an LU factorization exists. Example 2 will show how to implement step 2. By construction, L will satisfy

$$(E_p \cdots E_1)L = I$$

using the same $E_1, \dots, E_p$ as in equation (3). Thus L will be invertible, by the Invertible Matrix Theorem, with $(E_p \cdots E_1) = L^{-1}$. From (3), $L^{-1}A = U$, and $A = LU$. So step 2 will produce an acceptable L.

EXAMPLE 2 Find an LU factorization of

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

SOLUTION Since A has four rows, L should be 4×4. The first column of L is the first column of A divided by the top pivot entry:

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

Compare the first columns of A and L. *The row operations that create zeros in the first column of A will also create zeros in the first column of L.* To make this same correspondence of row operations on A hold for the rest of L, watch a row reduction of A to an echelon form U. That is, *highlight the entries* in each matrix that are used to determine the sequence of row operations that transform A into U. [See the highlighted entries in equation (5).]

$$A = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & -9 & -3 & -4 & 10 \\ 0 & 12 & 4 & 12 & -5 \end{bmatrix} = A_1 \tag{5}$$

$$\sim A_2 = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 4 & 7 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix} = U$$

These highlighted entries determine the row reduction of A to U. At each pivot column, divide the highlighted entries by the pivot and place the result into L:

$$\begin{bmatrix} 2 \\ -4 \\ 2 \\ -6 \end{bmatrix} \begin{bmatrix} 3 \\ -9 \\ 12 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 5 \end{bmatrix}$$

$$\div 2 \qquad \div 3 \qquad \div 2 \qquad \div 5$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$

$$\begin{bmatrix} 1 & & & \\ -2 & 1 & & \\ 1 & -3 & 1 & \\ -3 & 4 & 2 & 1 \end{bmatrix}, \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ -3 & 4 & 2 & 1 \end{bmatrix}$$

An easy calculation verifies that this L and U satisfy $LU = A$. ■

In practical work, row interchanges are nearly always needed, because partial pivoting is used for high accuracy. (Recall that this procedure selects, among the possible choices for a pivot, an entry in the column having the largest absolute value.) To handle row interchanges, the LU factorization above can be modified easily to produce an L that is *permuted lower triangular*, in the sense that a rearrangement (called a permutation) of the rows of L can make L (unit) lower triangular. The resulting *permuted LU factorization* solves $A\mathbf{x} = \mathbf{b}$ in the same way as before, except that the reduction of $[\,L \quad \mathbf{b}\,]$ to $[\,I \quad \mathbf{y}\,]$ follows the order of the pivots in L from left to right, starting with the pivot in the first column. A reference to an "LU factorization" usually includes the possibility that L might be permuted lower triangular. For details, see the *Study Guide*.

STUDY GUIDE offers information about permuted LU factorizations.

Numerical Notes

The following operation counts apply to an $n \times n$ dense matrix A (with most entries nonzero) for n moderately large, say, $n \geq 30$.[1]

1. Computing an LU factorization of A takes about $2n^3/3$ flops (about the same as row reducing $[\,A \quad \mathbf{b}\,]$), whereas finding A^{-1} requires about $2n^3$ flops.
2. Solving $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$ requires about $2n^2$ flops, because any $n \times n$ triangular system can be solved in about n^2 flops.
3. Multiplication of $\mathbf{b}$ by A^{-1} also requires about $2n^2$ flops, but the result may not be as accurate as that obtained from L and U (because of roundoff error when computing both A^{-1} and $A^{-1}\mathbf{b}$).
4. If A is sparse (with mostly zero entries), then L and U may be sparse, too, whereas A^{-1} is likely to be dense. In this case, a solution of $A\mathbf{x} = \mathbf{b}$ with an LU factorization is *much* faster than using A^{-1}. See Exercise 31.

A Matrix Factorization in Electrical Engineering

Matrix factorization is intimately related to the problem of constructing an electrical network with specified properties. The following discussion gives just a glimpse of the connection between factorization and circuit design.

[1] See Section 3.8 in *Applied Linear Algebra*, 3rd ed., by Ben Noble and James W. Daniel (Englewood Cliffs, NJ: Prentice-Hall, 1988). Recall that for our purposes, a *flop* is $+$, $-$, $\times$, or $\div$.

Suppose the box in Figure 3 represents some sort of electric circuit, with an input and output. Record the input voltage and current by $\begin{bmatrix} v_1 \\ i_1 \end{bmatrix}$ (with voltage v in volts and current i in amps), and record the output voltage and current by $\begin{bmatrix} v_2 \\ i_2 \end{bmatrix}$. Frequently, the transformation $\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} \mapsto \begin{bmatrix} v_2 \\ i_2 \end{bmatrix}$ is linear. That is, there is a matrix A, called the *transfer matrix*, such that

$$\begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = A\begin{bmatrix} v_1 \\ i_1 \end{bmatrix}$$

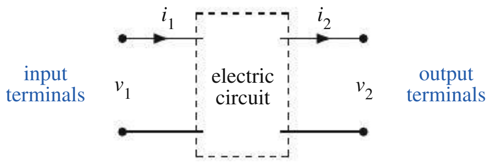


FIGURE 3 A circuit with input and output terminals.

Figure 4 shows a *ladder network*, where two circuits (there could be more) are connected in series, so that the output of one circuit becomes the input of the next circuit. The left circuit in Figure 4 is called a *series circuit*, with resistance R_1 (in ohms).

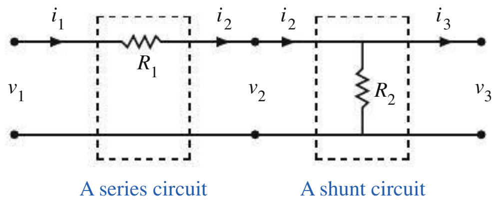


FIGURE 4 A ladder network.

The right circuit in Figure 4 is a *shunt circuit*, with resistance R_2. Using Ohm's law and Kirchhoff's laws, one can show that the transfer matrices of the series and shunt circuits, respectively, are

$$\underset{\text{Transfer matrix of series circuit}}{\begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix}} \quad \text{and} \quad \underset{\text{Transfer matrix of shunt circuit}}{\begin{bmatrix} 1 & 0 \\ -1/R_2 & 1 \end{bmatrix}}$$

EXAMPLE 3

a. Compute the transfer matrix of the ladder network in Figure 4.

b. Design a ladder network whose transfer matrix is $\begin{bmatrix} 1 & -8 \\ -.5 & 5 \end{bmatrix}$.

SOLUTION

a. Let A_1 and A_2 be the transfer matrices of the series and shunt circuits, respectively. Then an input vector $\mathbf{x}$ is transformed first into $A_1\mathbf{x}$ and then into $A_2(A_1\mathbf{x})$. The series connection of the circuits corresponds to composition of linear transformations, and the transfer matrix of the ladder network is (note the order)

$$A_2A_1 = \begin{bmatrix} 1 & 0 \\ -1/R_2 & 1 \end{bmatrix}\begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -R_1 \\ -1/R_2 & 1 + R_1/R_2 \end{bmatrix} \tag{6}$$

b. To factor the matrix $\begin{bmatrix} 1 & -8 \\ -.5 & 5 \end{bmatrix}$ into the product of transfer matrices, as in equation (6), look for R_1 and R_2 in Figure 4 to satisfy

$$\begin{bmatrix} 1 & -R_1 \\ -1/R_2 & 1 + R_1/R_2 \end{bmatrix} = \begin{bmatrix} 1 & -8 \\ -.5 & 5 \end{bmatrix}$$

From the $(1, 2)$-entries, $R_1 = 8$ ohms, and from the $(2, 1)$-entries, $1/R_2 = .5$ ohm and $R_2 = 1/.5 = 2$ ohms. With these values, the network in Figure 4 has the desired transfer matrix. ■

A network transfer matrix summarizes the input–output behavior (the design specifications) of the network without reference to the interior circuits. To physically build a network with specified properties, an engineer first determines if such a network can be constructed (or *realized*). Then the engineer tries to factor the transfer matrix into matrices corresponding to smaller circuits that perhaps are already manufactured and ready for assembly. In the common case of alternating current, the entries in the transfer matrix are usually rational complex-valued functions. (See Exercises 21 and 22 in Section 2.4.) A standard problem is to find a *minimal realization* that uses the smallest number of electrical components.

Practice Problem

Find an LU factorization of $A = \begin{bmatrix} 2 & -4 & -2 & 3 \\ 6 & -9 & -5 & 8 \\ 2 & -7 & -3 & 9 \\ 4 & -2 & -2 & -1 \\ -6 & 3 & 3 & 4 \end{bmatrix}$. [*Note:* It will turn out that A has only three pivot columns, so the method of Example 2 will produce only the first three columns of L. The remaining two columns of L come from I_5.]

2.5 Exercises

In Exercises 1–6, solve the equation $A\mathbf{x} = \mathbf{b}$ by using the LU factorization given for A. In Exercises 1 and 2, also solve $A\mathbf{x} = \mathbf{b}$ by ordinary row reduction.

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

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

2. $A = \begin{bmatrix} 4 & 3 & -5 \\ -4 & -5 & 7 \\ 8 & 6 & -8 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix}$

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

3. $A = \begin{bmatrix} 2 & -1 & 2 \\ -6 & 0 & -2 \\ 8 & -1 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$

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

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

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

5. $A = \begin{bmatrix} 1 & -2 & -4 & -3 \\ 2 & -7 & -7 & -6 \\ -1 & 2 & 6 & 4 \\ -4 & -1 & 9 & 8 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 7 \\ 0 \\ 3 \end{bmatrix}$

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

6. $A = \begin{bmatrix} 1 & 3 & 4 & 0 \\ -3 & -6 & -7 & 2 \\ 3 & 3 & 0 & -4 \\ -5 & -3 & 2 & 9 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}$

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

Find an LU factorization of the matrices in Exercises 7–16 (with L unit lower triangular). Note that MATLAB will usually produce a permuted LU factorization because it uses partial pivoting for numerical accuracy.

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

8. $\begin{bmatrix} 6 & 9 \\ 4 & 5 \end{bmatrix}$

9. $\begin{bmatrix} 3 & -1 & 2 \\ -3 & -2 & 10 \\ 9 & -5 & 6 \end{bmatrix}$

10. $\begin{bmatrix} -5 & 3 & 4 \\ 10 & -8 & -9 \\ 15 & 1 & 2 \end{bmatrix}$

11. $\begin{bmatrix} 3 & -6 & 3 \\ 6 & -7 & 2 \\ -1 & 7 & 0 \end{bmatrix}$

12. $\begin{bmatrix} 2 & -4 & 2 \\ 1 & 5 & -4 \\ -6 & -2 & 4 \end{bmatrix}$

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

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

15. $\begin{bmatrix} 2 & -4 & 4 & -2 \\ 6 & -9 & 7 & -3 \\ -1 & -4 & 8 & 0 \end{bmatrix}$

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

17. When A is invertible, MATLAB finds A^{-1} by factoring $A = LU$ (where L may be permuted lower triangular), inverting L and U, and then computing $U^{-1}L^{-1}$. Use this method to compute the inverse of A in Exercise 2. (Apply the algorithm of Section 2.2 to L and to U.)

18. Find A^{-1} as in Exercise 17, using A from Exercise 3.

19. Let A be a lower triangular $n \times n$ matrix with nonzero entries on the diagonal. Show that A is invertible and A^{-1} is lower triangular. [*Hint:* Explain why A can be changed into I using only row replacements and scaling. (Where are the pivots?) Also, explain why the row operations that reduce A to I change I into a lower triangular matrix.]

20. Let $A = LU$ be an LU factorization. Explain why A can be row reduced to U using only replacement operations. (This fact is the converse of what was proved in the text.)

21. Suppose $A = BC$, where B is invertible. Show that any sequence of row operations that reduces B to I also reduces A to C. The converse is not true, since the zero matrix may be factored as $0 = B(0)$.

Exercises 22–26 provide a glimpse of some widely used matrix factorizations, some of which are discussed later in the text.

22. (*Reduced LU Factorization*) With A as in the Practice Problem, find a 5×3 matrix B and a 3×4 matrix C such that $A = BC$. Generalize this idea to the case where A is $m \times n$, $A = LU$, and U has only three nonzero rows.

23. (*Rank Factorization*) Suppose an $m \times n$ matrix A admits a factorization $A = CD$ where C is $m \times 4$ and D is $4 \times n$.

a. Show that A is the sum of four outer products. (See Section 2.4.)

b. Let $m = 400$ and $n = 100$. Explain why a computer programmer might prefer to store the data from A in the form of two matrices C and D.

24. (*QR Factorization*) Suppose $A = QR$, where Q and R are $n \times n$, R is invertible and upper triangular, and Q has the property that $Q^TQ = I$. Show that for each $\mathbf{b}$ in $\mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution. What computations with Q and R will produce the solution?

25. (*Singular Value Decomposition*) Suppose $A = UDV^T$, where U and V are $n \times n$ matrices with the property that $U^TU = I$ and $V^TV = I$, and where D is a diagonal matrix with positive numbers $\sigma_1, \ldots, \sigma_n$ on the diagonal. Show that A is invertible, and find a formula for A^{-1}.

26. (*Spectral Factorization*) Suppose a 3×3 matrix A admits a factorization as $A = PDP^{-1}$, where P is some invertible 3×3 matrix and D is the diagonal matrix

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

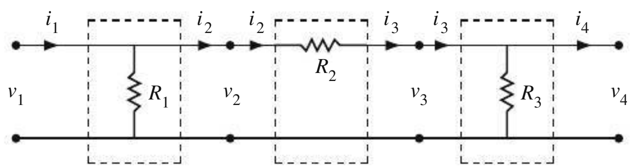
Show that this factorization is useful when computing high powers of A. Find fairly simple formulas for A^2, A^3, and A^k (k a positive integer), using P and the entries in D.

27. Design two different ladder networks that each output 9 volts and 4 amps when the input is 12 volts and 6 amps.

28. Show that if three shunt circuits (with resistances R_1, R_2, R_3) are connected in series, the resulting network has the same transfer matrix as a single shunt circuit. Find a formula for the resistance in that circuit.

29. a. Compute the transfer matrix of the network in the figure.

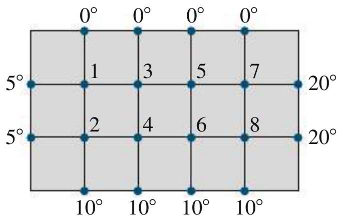
b. Let $A = \begin{bmatrix} 4/3 & -12 \\ -1/4 & 3 \end{bmatrix}$. Design a ladder network whose transfer matrix is A by finding a suitable matrix factorization of A.



30. Find a different factorization of the A in Exercise 29, and thereby design a different ladder network whose transfer matrix is A.

T 31. The solution to the steady-state heat flow problem for the plate in the figure is approximated by the solution to the equation $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (5, 15, 0, 10, 0, 10, 20, 30)$ and

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



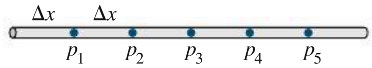
(Refer to Exercise 43 of Section 1.1.) The missing entries in A are zeros. The nonzero entries of A lie within a band along the main diagonal. Such *band matrices* occur in a variety of applications and often are extremely large (with thousands of rows and columns but relatively narrow bands).

a. Use the method of Example 2 to construct an LU factorization of A, and note that both factors are band matrices (with two nonzero diagonals below or above the main diagonal). Compute $LU - A$ to check your work.

b. Use the LU factorization to solve $A\mathbf{x} = \mathbf{b}$.

c. Obtain A^{-1} and note that A^{-1} is a dense matrix with no band structure. When A is large, L and U can be stored in much less space than A^{-1}. This fact is another reason for preferring the LU factorization of A to A^{-1} itself.

T 32. The band matrix A shown below can be used to estimate the unsteady conduction of heat in a rod when the temperatures at points $p_1, \dots, p_5$ on the rod change with time.[2]



The constant C in the matrix depends on the physical nature of the rod, the distance Δx between the points on the rod, and the length of time Δt between successive temperature measurements. Suppose that for $k = 0, 1, 2, \dots$, a vector $\mathbf{t}_k$ in $\mathbb{R}^5$ lists the temperatures at time $k\Delta t$. If the two ends of the rod are maintained at 0°, then the temperature vectors satisfy the equation $A\mathbf{t}_{k+1} = \mathbf{t}_k$ $(k = 0, 1, \dots)$, where

$$A = \begin{bmatrix} (1+2C) & -C & & & \\ -C & (1+2C) & -C & & \\ & -C & (1+2C) & -C & \\ & & -C & (1+2C) & -C \\ & & & -C & (1+2C) \end{bmatrix}$$

a. Find the LU factorization of A when $C = 1$. A matrix such as A with three nonzero diagonals is called a *tridiagonal matrix*. The L and U factors are *bidiagonal matrices*.

b. Suppose $C = 1$ and $\mathbf{t}_0 = (10, 12, 12, 12, 10)$. Use the LU factorization of A to find the temperature distributions $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_3$, and $\mathbf{t}_4$.

[2] See Biswa N. Datta, *Numerical Linear Algebra and Applications* (Pacific Grove, CA: Brooks/Cole, 1994), pp. 200–201.

Solution to Practice Problem

$$A = \begin{bmatrix} 2 & -4 & -2 & 3 \\ 6 & -9 & -5 & 8 \\ 2 & -7 & -3 & 9 \\ 4 & -2 & -2 & -1 \\ -6 & 3 & 3 & 4 \end{bmatrix} \sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & -3 & -1 & 6 \\ 0 & 6 & 2 & -7 \\ 0 & -9 & -3 & 13 \end{bmatrix}$$

$$\sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 10 \end{bmatrix} \sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

Divide the entries in each highlighted column by the pivot at the top. The resulting columns form the first three columns in the lower half of L. This suffices to make row reduction of L to I correspond to reduction of A to U. Use the last two columns of I_5

to make L unit lower triangular.

$$\begin{array}{ccc} \begin{bmatrix} 2\\ 6\\ 2\\ 4\\ -6 \end{bmatrix} & \begin{bmatrix} 3\\ -3\\ 6\\ -9 \end{bmatrix} & \begin{bmatrix} 5\\ -5\\ 10 \end{bmatrix} \\ \div 2 & \div 3 & \div 5 \\ \downarrow & \downarrow & \downarrow \end{array}$$

$$\begin{bmatrix} 1 & & & \\ 3 & 1 & & \\ 1 & -1 & 1 & \cdots \\ 2 & 2 & -1 & \\ -3 & -3 & 2 & \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 2 & 2 & -1 & 1 & 0 \\ -3 & -3 & 2 & 0 & 1 \end{bmatrix}$$

2.6 The Leontief Input–Output Model

Linear algebra played an essential role in the Nobel prize–winning work of Wassily Leontief, as mentioned at the beginning of Chapter 1. The economic model described in this section is the basis for more elaborate models used in many parts of the world.

Suppose a nation's economy is divided into n sectors that produce goods or services, and let $\mathbf{x}$ be a **production vector** in $\mathbb{R}^n$ that lists the output of each sector for one year. Also, suppose another part of the economy (called the *open sector*) does not produce goods or services but only consumes them, and let $\mathbf{d}$ be a **final demand vector** (or **bill of final demands**) that lists the values of the goods and services demanded from the various sectors by the nonproductive part of the economy. The vector $\mathbf{d}$ can represent consumer demand, government consumption, surplus production, exports, or other external demands.

As the various sectors produce goods to meet consumer demand, the producers themselves create additional **intermediate demand** for goods they need as inputs for their own production. The interrelations between the sectors are very complex, and the connection between the final demand and the production is unclear. Leontief asked if there is a production level $\mathbf{x}$ such that the amounts produced (or "supplied") will exactly balance the total demand for that production, so that

$$\left\{\begin{matrix}\text{amount}\\ \text{produced}\\ \mathbf{x}\end{matrix}\right\} = \left\{\begin{matrix}\text{intermediate}\\ \text{demand}\end{matrix}\right\} + \left\{\begin{matrix}\text{final}\\ \text{demand}\\ \mathbf{d}\end{matrix}\right\} \tag{1}$$

The basic assumption of Leontief's input–output model is that for each sector, there is a **unit consumption vector** in $\mathbb{R}^n$ that lists the inputs needed *per unit of output* of the sector. All input and output units are measured in millions of dollars, rather than in quantities such as tons or bushels. (Prices of goods and services are held constant.)

As a simple example, suppose the economy consists of three sectors—manufacturing, agriculture, and services—with unit consumption vectors $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$, as shown in the table that follows.

	Inputs Consumed per Unit of Output		
Purchased from	**Manufacturing**	**Agriculture**	**Services**
Manufacturing	.50	.40	.20
Agriculture	.20	.30	.10
Services	.10	.10	.30
	↑ $\mathbf{c}_1$	↑ $\mathbf{c}_2$	↑ $\mathbf{c}_3$

EXAMPLE 1 What amounts will be consumed by the manufacturing sector if it decides to produce 100 units?

SOLUTION Compute

$$100\mathbf{c}_1 = 100\begin{bmatrix} .50 \\ .20 \\ .10 \end{bmatrix} = \begin{bmatrix} 50 \\ 20 \\ 10 \end{bmatrix}$$

To produce 100 units, manufacturing will order (i.e., "demand") and consume 50 units from other parts of the manufacturing sector, 20 units from agriculture, and 10 units from services. ■

If manufacturing decides to produce x_1 units of output, then $x_1\mathbf{c}_1$ represents the *intermediate demands* of manufacturing, because the amounts in $x_1\mathbf{c}_1$ will be consumed in the process of creating the x_1 units of output. Likewise, if x_2 and x_3 denote the planned outputs of the agriculture and services sectors, $x_2\mathbf{c}_2$ and $x_3\mathbf{c}_3$ list their corresponding intermediate demands. The total intermediate demand from all three sectors is given by

$$\begin{aligned} \{\text{intermediate demand}\} &= x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + x_3\mathbf{c}_3 \\ &= C\mathbf{x} \end{aligned} \tag{2}$$

where C is the **consumption matrix** $[\,\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3\,]$, namely

$$C = \begin{bmatrix} .50 & .40 & .20 \\ .20 & .30 & .10 \\ .10 & .10 & .30 \end{bmatrix} \tag{3}$$

Equations (1) and (2) yield Leontief's model.

THE LEONTIEF INPUT–OUTPUT MODEL, OR PRODUCTION EQUATION

$$\underset{\text{Amount produced}}{\mathbf{x}} = \underset{\text{Intermediate demand}}{C\mathbf{x}} + \underset{\text{Final demand}}{\mathbf{d}} \tag{4}$$

Equation (4) may also be written as $I\mathbf{x} - C\mathbf{x} = \mathbf{d}$, or

$$(I - C)\mathbf{x} = \mathbf{d} \tag{5}$$

EXAMPLE 2 Consider the economy whose consumption matrix is given by (3). Suppose the final demand is 50 units for manufacturing, 30 units for agriculture, and 20 units for services. Find the production level $\mathbf{x}$ that will satisfy this demand.

SOLUTION The coefficient matrix in (5) is

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

To solve (5), row reduce the augmented matrix

$$\begin{bmatrix} .5 & -.4 & -.2 & 50 \\ -.2 & .7 & -.1 & 30 \\ -.1 & -.1 & .7 & 20 \end{bmatrix} \sim \begin{bmatrix} 5 & -4 & -2 & 500 \\ -2 & 7 & -1 & 300 \\ -1 & -1 & 7 & 200 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 0 & 226 \\ 0 & 1 & 0 & 119 \\ 0 & 0 & 1 & 78 \end{bmatrix}$$

The last column is rounded to the nearest whole unit. Manufacturing must produce approximately 226 units, agriculture 119 units, and services only 78 units. ■

If the matrix $I - C$ is invertible, then we can apply Theorem 5 in Section 2.2, with A replaced by $(I - C)$, and from the equation $(I - C)\mathbf{x} = \mathbf{d}$ obtain $\mathbf{x} = (I - C)^{-1}\mathbf{d}$. The theorem below shows that in most practical cases, $I - C$ *is* invertible and the production vector $\mathbf{x}$ is economically feasible, in the sense that the entries in $\mathbf{x}$ are nonnegative.

In the theorem, the term **column sum** denotes the sum of the entries in a column of a matrix. Under ordinary circumstances, the column sums of a consumption matrix are less than 1 because a sector should require less than one unit's worth of inputs to produce one unit of output.

THEOREM 11

Let C be the consumption matrix for an economy, and let $\mathbf{d}$ be the final demand. If C and $\mathbf{d}$ have nonnegative entries and if each column sum of C is less than 1, then $(I - C)^{-1}$ exists and the production vector

$$\mathbf{x} = (I - C)^{-1}\mathbf{d}$$

has nonnegative entries and is the unique solution of

$$\mathbf{x} = C\mathbf{x} + \mathbf{d}$$

The following discussion will suggest why the theorem is true and will lead to a new way to compute $(I - C)^{-1}$.

A Formula for $(I - C)^{-1}$

Imagine that the demand represented by $\mathbf{d}$ is presented to the various industries at the beginning of the year, and the industries respond by setting their production levels at $\mathbf{x} = \mathbf{d}$, which will exactly meet the final demand. As the industries prepare to produce $\mathbf{d}$, they send out orders for their raw materials and other inputs. This creates an intermediate demand of $C\mathbf{d}$ for inputs.

To meet the additional demand of $C\mathbf{d}$, the industries will need as additional inputs the amounts in $C(C\mathbf{d}) = C^2\mathbf{d}$. Of course, this creates a second round of intermediate demand, and when the industries decide to produce even more to meet this new demand, they create a third round of demand, namely $C(C^2\mathbf{d}) = C^3\mathbf{d}$. And so it goes.

Theoretically, this process could continue indefinitely, although in real life it would not take place in such a rigid sequence of events. We can diagram this hypothetical situation as follows:

	Demand That Must Be Met	Inputs Needed to Meet This Demand
Final demand	$\mathbf{d}$	$C\mathbf{d}$
Intermediate demand		
1st round	$C\mathbf{d}$	$C(C\mathbf{d}) = C^2\mathbf{d}$
2nd round	$C^2\mathbf{d}$	$C(C^2\mathbf{d}) = C^3\mathbf{d}$
3rd round	$C^3\mathbf{d}$	$C(C^3\mathbf{d}) = C^4\mathbf{d}$
	$\vdots$	$\vdots$

The production level $\mathbf{x}$ that will meet all of this demand is

$$\begin{aligned} \mathbf{x} &= \mathbf{d} + C\mathbf{d} + C^2\mathbf{d} + C^3\mathbf{d} + \cdots \\ &= (I + C + C^2 + C^3 + \cdots)\mathbf{d} \end{aligned} \tag{6}$$

To make sense of equation (6), consider the following algebraic identity:

$$(I - C)(I + C + C^2 + \cdots + C^m) = I - C^{m+1} \tag{7}$$

It can be shown that if the column sums in C are all strictly less than 1, then $I - C$ is invertible, C^m approaches the zero matrix as m gets arbitrarily large, and $I - C^{m+1} \to I$. (This fact is analogous to the fact that if a positive number t is less than 1, then $t^m \to 0$ as m increases.) Using equation (7), write

$$\begin{gathered} (I - C)^{-1} \approx I + C + C^2 + C^3 + \cdots + C^m \\ \text{when the column sums of } C \text{ are less than 1.} \end{gathered} \tag{8}$$

The approximation in (8) means that the right side can be made as close to $(I - C)^{-1}$ as desired by taking m sufficiently large.

In actual input–output models, powers of the consumption matrix approach the zero matrix rather quickly. So (8) really provides a practical way to compute $(I - C)^{-1}$. Likewise, for any $\mathbf{d}$, the vectors $C^m\mathbf{d}$ approach the zero vector quickly, and (6) is a practical way to solve $(I - C)\mathbf{x} = \mathbf{d}$. If the entries in C and $\mathbf{d}$ are nonnegative, then (6) shows that the entries in $\mathbf{x}$ are nonnegative, too.

The Economic Importance of Entries in $(I - C)^{-1}$

The entries in $(I - C)^{-1}$ are significant because they can be used to predict how the production $\mathbf{x}$ will have to change when the final demand $\mathbf{d}$ changes. In fact, the entries in column j of $(I - C)^{-1}$ are the *increased* amounts the various sectors will have to produce in order to satisfy *an increase of 1 unit* in the final demand for output from sector j. See Exercise 8.

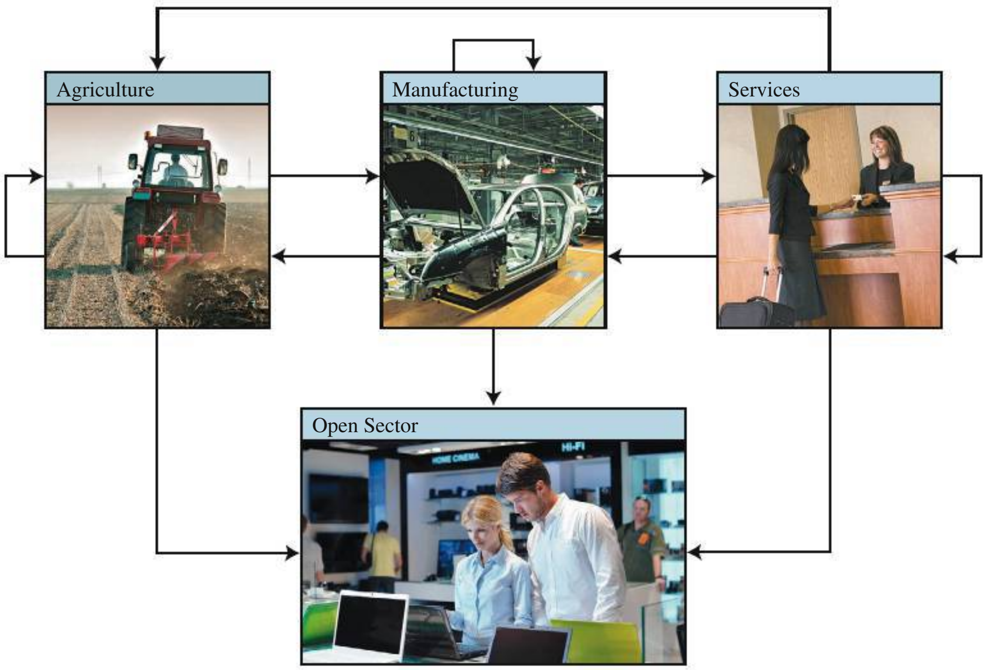
Numerical Note

In any applied problem (not just in economics), an equation $A\mathbf{x} = \mathbf{b}$ can always be written as $(I - C)\mathbf{x} = \mathbf{b}$, with $C = I - A$. If the system is large and *sparse* (with mostly zero entries), it can happen that the column sums of the absolute values in C are less than 1. In this case, $C^m \to 0$. If C^m approaches zero quickly enough, (6) and (8) will provide practical formulas for solving $A\mathbf{x} = \mathbf{b}$ and finding A^{-1}.

Practice Problem

Suppose an economy has two sectors: goods and services. One unit of output from goods requires inputs of .2 unit from goods and .5 unit from services. One unit of output from services requires inputs of .4 unit from goods and .3 unit from services. There is a final demand of 20 units of goods and 30 units of services. Set up the Leontief input–output model for this situation.

2.6 Exercises



Exercises 1–4 refer to an economy that is divided into three sectors—manufacturing, agriculture, and services. For each unit of output, manufacturing requires .10 unit from other companies in that sector, .30 unit from agriculture, and .30 unit from services. For each unit of output, agriculture uses .20 unit of its own output, .60 unit from manufacturing, and .10 unit from services. For each unit of output, the services sector consumes .10 unit from services, .60 unit from manufacturing, but no agricultural products.

1. Construct the consumption matrix for this economy, and determine what intermediate demands are created if agriculture plans to produce 100 units.

2. Determine the production levels needed to satisfy a final demand of 18 units for agriculture, with no final demand for the other sectors. (Do not compute an inverse matrix.)

3. Determine the production levels needed to satisfy a final demand of 18 units for manufacturing, with no final demand for the other sectors. (Do not compute an inverse matrix.)

4. Determine the production levels needed to satisfy a final demand of 18 units for manufacturing, 18 units for agriculture, and 0 units for services.

5. Consider the production model $\mathbf{x} = C\mathbf{x} + \mathbf{d}$ for an economy with two sectors, where

$$C = \begin{bmatrix} .0 & .5 \\ .6 & .2 \end{bmatrix}, \qquad \mathbf{d} = \begin{bmatrix} 50 \\ 30 \end{bmatrix}$$

Use an inverse matrix to determine the production level necessary to satisfy the final demand.

6. Repeat Exercise 5 with $C = \begin{bmatrix} .1 & .6 \\ .5 & .2 \end{bmatrix}$, and $\mathbf{d} = \begin{bmatrix} 18 \\ 11 \end{bmatrix}$.

7. Let C and $\mathbf{d}$ be as in Exercise 5.

   a. Determine the production level necessary to satisfy a final demand for 1 unit of output from sector 1.

b. Use an inverse matrix to determine the production level necessary to satisfy a final demand of $\begin{bmatrix} 51 \\ 30 \end{bmatrix}$.

c. Use the fact that $\begin{bmatrix} 51 \\ 30 \end{bmatrix} = \begin{bmatrix} 50 \\ 30 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to explain how and why the answers to parts (a) and (b) and to Exercise 5 are related.

8. Let C be an $n \times n$ consumption matrix whose column sums are less than 1. Let $\mathbf{x}$ be the production vector that satisfies a final demand $\mathbf{d}$, and let $\Delta\mathbf{x}$ be a production vector that satisfies a different final demand $\Delta\mathbf{d}$.

a. Show that if the final demand changes from $\mathbf{d}$ to $\mathbf{d} + \Delta\mathbf{d}$, then the new production level must be $\mathbf{x} + \Delta\mathbf{x}$. Thus $\Delta\mathbf{x}$ gives the amounts by which production must *change* in order to accommodate the *change* $\Delta\mathbf{d}$ in demand.

b. Let $\Delta\mathbf{d}$ be the vector in $\mathbb{R}^n$ with 1 as the first entry and 0's elsewhere. Explain why the corresponding production $\Delta\mathbf{x}$ is the first column of $(I - C)^{-1}$. This shows that the first column of $(I - C)^{-1}$ gives the amounts the various sectors must produce to satisfy an increase of 1 unit in the final demand for output from sector 1.

9. Solve the Leontief production equation for an economy with three sectors, given that

$$C = \begin{bmatrix} .2 & .2 & .0 \\ .3 & .1 & .3 \\ .1 & .0 & .2 \end{bmatrix} \quad \text{and} \quad \mathbf{d} = \begin{bmatrix} 40 \\ 60 \\ 80 \end{bmatrix}$$

10. The consumption matrix C for the U.S. economy in 1972 has the property that *every entry* in the matrix $(I - C)^{-1}$ is nonzero (and positive).[1] What does that say about the effect of raising the demand for the output of just one sector of the economy?

11. The Leontief production equation, $\mathbf{x} = C\mathbf{x} + \mathbf{d}$, is usually accompanied by a dual **price equation**,

$$\mathbf{p} = C^T\mathbf{p} + \mathbf{v}$$

where $\mathbf{p}$ is a **price vector** whose entries list the price per unit for each sector's output, and $\mathbf{v}$ is a **value added vector** whose entries list the value added per unit of output. (Value added includes wages, profit, depreciation, etc.) An important fact in economics is that the gross domestic product (GDP) can be expressed in two ways:

$$\{\text{gross domestic product}\} = \mathbf{p}^T\mathbf{d} = \mathbf{v}^T\mathbf{x}$$

Verify the second equality. [*Hint:* Compute $\mathbf{p}^T\mathbf{x}$ in two ways.]

[1] Wassily W. Leontief, "The World Economy of the Year 2000," *Scientific American*, September 1980, pp. 206–231.

12. Let C be a consumption matrix such that $C^m \to 0$ as $m \to \infty$, and for $m = 1, 2, \dots$, let $D_m = I + C + \cdots + C^m$. Find a difference equation that relates D_m and D_{m+1} and thereby obtain an iterative procedure for computing formula (8) for $(I - C)^{-1}$.

T 13. The consumption matrix C below is based on input–output data for the U.S. economy in 1958, with data for 81 sectors grouped into 7 larger sectors: (1) nonmetal household and personal products, (2) final metal products (such as motor vehicles), (3) basic metal products and mining, (4) basic nonmetal products and agriculture, (5) energy, (6) services, and (7) entertainment and miscellaneous products.[2] Find the production levels needed to satisfy the final demand $\mathbf{d}$. (Units are in millions of dollars.)

$$\begin{bmatrix} .1588 & .0064 & .0025 & .0304 & .0014 & .0083 & .1594 \\ .0057 & .2645 & .0436 & .0099 & .0083 & .0201 & .3413 \\ .0264 & .1506 & .3557 & .0139 & .0142 & .0070 & .0236 \\ .3299 & .0565 & .0495 & .3636 & .0204 & .0483 & .0649 \\ .0089 & .0081 & .0333 & .0295 & .3412 & .0237 & .0020 \\ .1190 & .0901 & .0996 & .1260 & .1722 & .2368 & .3369 \\ .0063 & .0126 & .0196 & .0098 & .0064 & .0132 & .0012 \end{bmatrix},$$

$$\mathbf{d} = \begin{bmatrix} 74{,}000 \\ 56{,}000 \\ 10{,}500 \\ 25{,}000 \\ 17{,}500 \\ 196{,}000 \\ 5{,}000 \end{bmatrix}$$

T 14. The demand vector in Exercise 13 is reasonable for 1958 data, but Leontief's discussion of the economy in the reference cited there used a demand vector closer to 1964 data:

$$\mathbf{d} = (99640,\ 75548,\ 14444,\ 33501,\ 23527,\ 263985,\ 6526)$$

Find the production levels needed to satisfy this demand.

T 15. Use equation (6) to solve the problem in Exercise 13. Set $\mathbf{x}^{(0)} = \mathbf{d}$, and for $k = 1, 2, \dots,$ compute $\mathbf{x}^{(k)} = \mathbf{d} + C\mathbf{x}^{(k-1)}$. How many steps are needed to obtain the answer in Exercise 13 to four significant figures?

[2] Wassily W. Leontief, "The Structure of the U.S. Economy," *Scientific American*, April 1965, pp. 30–32.

Solution to Practice Problem

The following data are given:

	Inputs Needed per Unit of Output		
Purchased from	**Goods**	**Services**	**External Demand**
Goods	.2	.4	20
Services	.5	.3	30

The Leontief input–output model is $\mathbf{x} = C\mathbf{x} + \mathbf{d}$, where

$$C = \begin{bmatrix} .2 & .4 \\ .5 & .3 \end{bmatrix}, \qquad \mathbf{d} = \begin{bmatrix} 20 \\ 30 \end{bmatrix}$$

2.7 Applications to Computer Graphics

Computer graphics are images displayed or animated on a computer screen. Applications of computer graphics are widespread and growing rapidly. For instance, computer-aided design (CAD) is an integral part of many engineering processes such as the aircraft design process described in the chapter introduction. The entertainment industry has made the most spectacular use of computer graphics—from the special effects in *Amazing Spider-Man 2* to PlayStation 4 and Xbox One.

Most interactive computer software for business and industry makes use of computer graphics in the screen displays and for other functions, such as graphical display of data, desktop publishing, and slide production for commercial and educational presentations. Consequently, anyone studying a computer language invariably spends time learning how to use at least two-dimensional (2D) graphics.

This section examines some of the basic mathematics used to manipulate and display graphical images such as a wire-frame model of an airplane. Such an image (or picture) consists of a number of points, connecting lines or curves, and information about how to fill in closed regions bounded by the lines and curves. Often, curved lines are approximated by short straight-line segments, and a figure is defined mathematically by a list of points.

Among the simplest 2D graphics symbols are letters used for labels on the screen. Some letters are stored as wire-frame objects; others that have curved portions are stored with additional mathematical formulas for the curves.

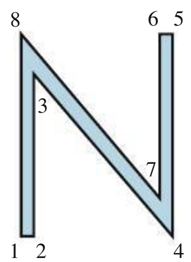


FIGURE 1
Regular *N*.

EXAMPLE 1 The capital letter N in Figure 1 is determined by eight points, or *vertices*. The coordinates of the points can be stored in a data matrix, D.

$$\begin{array}{rc} & \text{Vertex:} \\ & \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{array}{r} x\text{-coordinate} \\ y\text{-coordinate} \end{array} & \begin{bmatrix} 0 & .5 & .5 & 6 & 6 & 5.5 & 5.5 & 0 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{bmatrix} \end{array} = D$$

In addition to D, it is necessary to specify which vertices are connected by lines, but we omit this detail. ■

The main reason graphical objects are described by collections of straight-line segments is that the standard transformations in computer graphics map line segments onto other line segments. (For instance, see Exercise 35 in Section 1.8.) Once the vertices that describe an object have been transformed, their images can be connected with the appropriate straight lines to produce the complete image of the original object.

EXAMPLE 2 Given $A = \begin{bmatrix} 1 & .25 \\ 0 & 1 \end{bmatrix}$, describe the effect of the shear transformation $\mathbf{x} \mapsto A\mathbf{x}$ on the letter N in Example 1.

SOLUTION By definition of matrix multiplication, the columns of the product AD contain the images of the vertices of the letter N.

$$AD = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 0 & .5 & 2.105 & 6 & 8 & 7.5 & 5.895 & 2 \\ 0 & 0 & 6.420 & 0 & 8 & 8 & 1.580 & 8 \end{bmatrix} \end{array}$$

The transformed vertices are plotted in Figure 2, along with connecting line segments that correspond to those in the original figure. ■

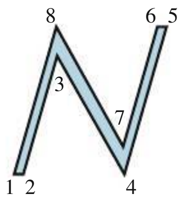


FIGURE 2
Slanted *N*.

The italic N in Figure 2 looks a bit too wide. To compensate, shrink the width by a scale transformation that affects the x-coordinates of the points.

EXAMPLE 3 Compute the matrix of the transformation that performs a shear transformation, as in Example 2, and then scales all x-coordinates by a factor of .75.

SOLUTION The matrix that multiplies the x-coordinate of a point by .75 is

$$S = \begin{bmatrix} .75 & 0 \\ 0 & 1 \end{bmatrix}$$

So the matrix of the composite transformation is

$$\begin{aligned} SA &= \begin{bmatrix} .75 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & .25 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} .75 & .1875 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

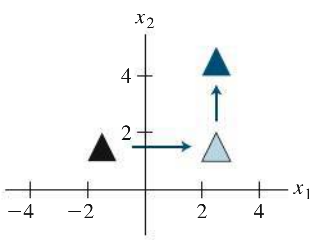
FIGURE 3
Composite transformation of *N*.

The result of this composite transformation is shown in Figure 3. ■

The mathematics of computer graphics is intimately connected with matrix multiplication. Unfortunately, translating an object on a screen does not correspond directly to matrix multiplication because translation is not a linear transformation. The standard way to avoid this difficulty is to introduce what are called *homogeneous coordinates*.

Homogeneous Coordinates

Each point (x, y) in $\mathbb{R}^2$ can be identified with the point $(x, y, 1)$ on the plane in $\mathbb{R}^3$ that lies one unit above the xy-plane. We say that (x, y) has *homogeneous coordinates* $(x, y, 1)$. For instance, the point $(0, 0)$ has homogeneous coordinates $(0, 0, 1)$. Homogeneous coordinates for points are not added or multiplied by scalars, but they can be transformed via multiplication by 3×3 matrices.



Translation by $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$.

EXAMPLE 4 A translation of the form $(x, y) \mapsto (x + h, y + k)$ is written in homogeneous coordinates as $(x, y, 1) \mapsto (x + h, y + k, 1)$. This transformation can be computed via matrix multiplication:

$$\begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x + h \\ y + k \\ 1 \end{bmatrix}$$ ■

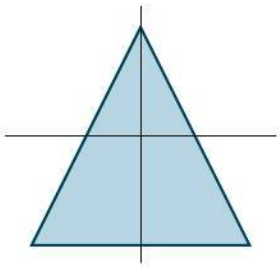
EXAMPLE 5 Any linear transformation on $\mathbb{R}^2$ is represented with respect to homogeneous coordinates by a partitioned matrix of the form $\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}$, where A is a 2×2 matrix. Typical examples are

$$\underset{\substack{\text{Counterclockwise}\\ \text{rotation about the}\\ \text{origin, angle } \varphi}}{\begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}}, \quad \underset{\substack{\text{Reflection}\\ \text{through } y = x}}{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}, \quad \underset{\substack{\text{Scale } x \text{ by } s\\ \text{and } y \text{ by } t}}{\begin{bmatrix} s & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & 1 \end{bmatrix}}$$

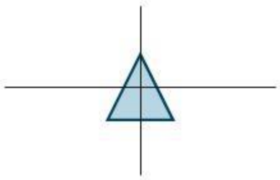
■

Composite Transformations

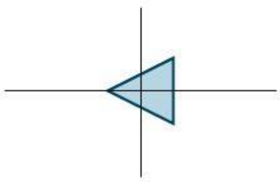
The movement of a figure on a computer screen often requires two or more basic transformations. The composition of such transformations corresponds to matrix multiplication when homogeneous coordinates are used.



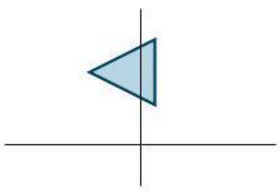
Original Figure



After Scaling



After Rotating



After Translating

EXAMPLE 6 Find the 3×3 matrix that corresponds to the composite transformation of a scaling by .3, a rotation of 90° about the origin, and finally a translation that adds $(-.5, 2)$ to each point of a figure.

SOLUTION If $\varphi = \pi/2$, then $\sin\varphi = 1$ and $\cos\varphi = 0$. From Examples 4 and 5, we have

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \xrightarrow{\text{Scale}} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\xrightarrow{\text{Rotate}} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\xrightarrow{\text{Translate}} \begin{bmatrix} 1 & 0 & -.5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

The matrix for the composite transformation is

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

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

■

3D Computer Graphics

Some of the newest and most exciting work in computer graphics is connected with molecular modeling. With 3D (three-dimensional) graphics, a biologist can examine a simulated protein molecule and search for active sites that might accept a drug molecule. The biologist can rotate and translate an experimental drug and attempt to attach it to the protein. This ability to *visualize* potential chemical reactions is vital to modern drug and cancer research. In fact, advances in drug design depend to some extent upon progress in the ability of computer graphics to construct realistic simulations of molecules and their interactions.[1]

Current research in molecular modeling is focused on *virtual reality*, an environment in which a researcher can see and *feel* the drug molecule slide into the protein. In Figure 4, such tactile feedback is provided by a force-displaying remote manipulator.

FIGURE 4 Molecular modeling in virtual reality.

Another design for virtual reality involves a helmet and glove that detect head, hand, and finger movements. The helmet contains two tiny computer screens, one for each eye. Making this virtual environment more realistic is a challenge to engineers, scientists, and mathematicians. The mathematics we examine here barely opens the door to this interesting field of research.

Homogeneous 3D Coordinates

By analogy with the 2D case, we say that $(x, y, z, 1)$ are homogeneous coordinates for the point (x, y, z) in $\mathbb{R}^3$. In general, (X, Y, Z, H) are **homogeneous coordinates** for (x, y, z) if $H \neq 0$ and

$$x = \frac{X}{H}, \qquad y = \frac{Y}{H}, \qquad \text{and} \qquad z = \frac{Z}{H} \tag{1}$$

Each nonzero scalar multiple of $(x, y, z, 1)$ gives a set of homogeneous coordinates for (x, y, z). For instance, both $(10, -6, 14, 2)$ and $(-15, 9, -21, -3)$ are homogeneous coordinates for $(5, -3, 7)$.

The next example illustrates the transformations used in molecular modeling to move a drug into a protein molecule.

EXAMPLE 7 Give 4×4 matrices for the following transformations:

a. Rotation about the y-axis through an angle of 30°. (By convention, a positive angle is the counterclockwise direction when looking toward the origin from the positive half of the axis of rotation—in this case, the y-axis.)

b. Translation by the vector $\mathbf{p} = (-6, 4, 5)$.

SOLUTION

a. First, construct the 3×3 matrix for the rotation. The vector $\mathbf{e}_1$ rotates down toward the negative z-axis, stopping at $(\cos 30°, 0, -\sin 30°) = (\sqrt{3}/2, 0, -.5)$. The vector $\mathbf{e}_2$ on the y-axis does not move, but $\mathbf{e}_3$ on the z-axis rotates down toward the positive

[1] Robert Pool, "Computing in Science," *Science* **256**, 3 April 1992, p. 45.

x-axis, stopping at $(\sin 30°, 0, \cos 30°) = (.5, 0, \sqrt{3}/2)$. See Figure 5. From Section 1.9, the standard matrix for this rotation is

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

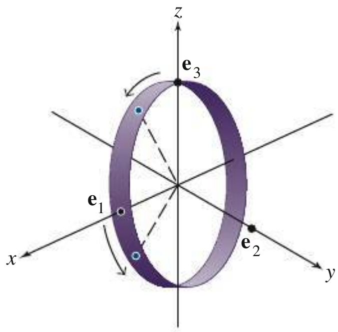


FIGURE 5

So the rotation matrix for homogeneous coordinates is

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

b. We want $(x, y, z, 1)$ to map to $(x - 6, y + 4, z + 5, 1)$. The matrix that does this is

$$\begin{bmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

■

Perspective Projections

A three-dimensional object is represented on the two-dimensional computer screen by projecting the object onto a *viewing plane*. (We ignore other important steps, such as selecting the portion of the viewing plane to display on the screen.) For simplicity, let the xy-plane represent the computer screen, and imagine that the eye of a viewer is along the positive z-axis, at a point $(0, 0, d)$. A *perspective projection* maps each point (x, y, z) onto an image point $(x^*, y^*, 0)$ so that the two points and the eye position, called the *center of projection*, are on a line. See Figure 6(a).

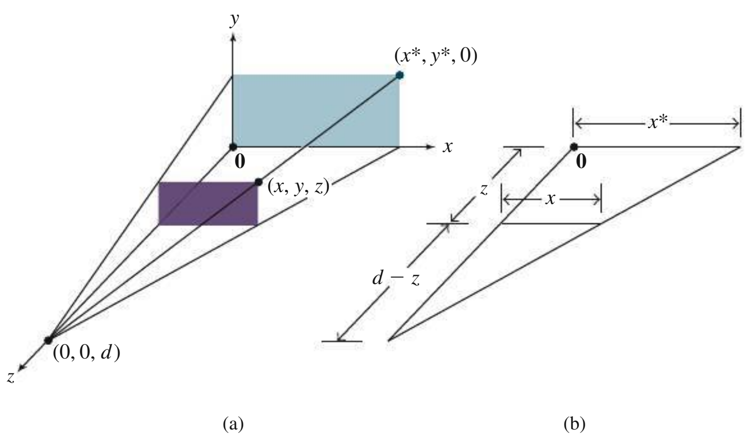


FIGURE 6 Perspective projection of (x, y, z) onto $(x^*, y^*, 0)$.

The triangle in the xz-plane in Figure 6(a) is redrawn in part (b) showing the lengths of line segments. Similar triangles show that

$$\frac{x^*}{d} = \frac{x}{d - z} \quad \text{and} \quad x^* = \frac{dx}{d - z} = \frac{x}{1 - z/d}$$

Similarly,

$$y^* = \frac{y}{1 - z/d}$$

Using homogeneous coordinates, we can represent the perspective projection by a matrix, say, P. We want $(x, y, z, 1)$ to map into $\left(\frac{x}{1-z/d}, \frac{y}{1-z/d}, 0, 1\right)$. Scaling these coordinates by $1 - z/d$, we can also use $(x, y, 0, 1 - z/d)$ as homogeneous coordinates for the image. Now it is easy to display P. In fact,

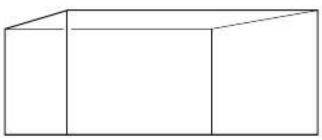
$$P\begin{bmatrix}x\\y\\z\\1\end{bmatrix} = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&-1/d&1\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix} = \begin{bmatrix}x\\y\\0\\1-z/d\end{bmatrix}$$

EXAMPLE 8 Let S be the box with vertices $(3, 1, 5)$, $(5, 1, 5)$, $(5, 0, 5)$, $(3, 0, 5)$, $(3, 1, 4)$, $(5, 1, 4)$, $(5, 0, 4)$, and $(3, 0, 4)$. Find the image of S under the perspective projection with center of projection at $(0, 0, 10)$.

SOLUTION Let P be the projection matrix, and let D be the data matrix for S using homogeneous coordinates. The data matrix for the image of S is

$$PD = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&-1/10&1\end{bmatrix}\begin{array}{c}\text{Vertex:}\\ \begin{array}{cccccccc}1&2&3&4&5&6&7&8\end{array}\\ \begin{bmatrix}3&5&5&3&3&5&5&3\\1&1&0&0&1&1&0&0\\5&5&5&5&4&4&4&4\\1&1&1&1&1&1&1&1\end{bmatrix}\end{array}$$

$$= \begin{bmatrix}3&5&5&3&3&5&5&3\\1&1&0&0&1&1&0&0\\0&0&0&0&0&0&0&0\\.5&.5&.5&.5&.6&.6&.6&.6\end{bmatrix}$$



S under the perspective transformation.

To obtain $\mathbb{R}^3$ coordinates, use equation (1) before Example 7, and divide the top three entries in each column by the corresponding entry in the fourth row:

$$\begin{array}{c}\text{Vertex:}\\ \begin{array}{cccccccc}1&2&3&4&5&6&7&8\end{array}\\ \begin{bmatrix}6&10&10&6&5&8.3&8.3&5\\2&2&0&0&1.7&1.7&0&0\\0&0&0&0&0&0&0&0\end{bmatrix}\end{array}$$ ■

This text's web site has some interesting applications of computer graphics, including a further discussion of perspective projections. One of the chapter projects involves simple animation.

Numerical Note

Continuous movement of graphical 3D objects requires intensive computation with 4×4 matrices, particularly when the surfaces are *rendered* to appear realistic, with texture and appropriate lighting. *High-end computer graphics boards* have 4×4 matrix operations and graphics algorithms embedded in their microchips and circuitry. Such boards can perform the billions of matrix multiplications per second needed for realistic color animation in 3D gaming programs.[2]

Further Reading

James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes, *Computer Graphics: Principles and Practice*, 3rd ed. (Boston, MA: Addison-Wesley, 2002), Chapters 5 and 6.

Practice Problem

Rotation of a figure about a point $\mathbf{p}$ in $\mathbb{R}^2$ is accomplished by first translating the figure by $-\mathbf{p}$, rotating about the origin, and then translating back by $\mathbf{p}$. See Figure 7. Construct the 3×3 matrix that rotates points $-30°$ about the point $(-2, 6)$, using homogeneous coordinates.

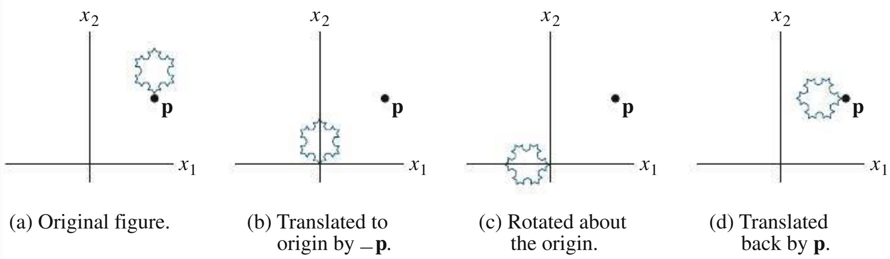


FIGURE 7 Rotation of figure about point $\mathbf{p}$.

2.7 Exercises

1. What 3×3 matrix will have the same effect on homogeneous coordinates for $\mathbb{R}^2$ that the shear matrix A has in Example 2?

2. Use matrix multiplication to find the image of the triangle with data matrix $D = \begin{bmatrix} 5 & 2 & 4 \\ 0 & 2 & 3 \end{bmatrix}$ under the transformation that reflects points through the y-axis. Sketch both the original triangle and its image.

In Exercises 3–8, find the 3×3 matrices that produce the described composite 2D transformations, using homogeneous coordinates.

3. Translate by (3, 1), and then rotate 45° about the origin.

4. Translate by $(-2, 3)$, and then scale the x-coordinate by .8 and the y-coordinated by 1.2.

5. Reflect points through the x-axis, and then rotate 30° about the origin.

6. Rotate points 30°, and then reflect through the x-axis.

7. Rotate points through 60° about the point (6, 8).

8. Rotate points through 45° about the point (3, 7).

9. A 2×200 data matrix D contains the coordinates of 200 points. Compute the number of multiplications required to transform these points using two arbitrary 2×2 matrices A and B. Consider the two possibilities $A(BD)$ and $(AB)D$. Discuss the implications of your results for computer graphics calculations.

10. Consider the following geometric 2D transformations: D, a dilation (in which x-coordinates and y-coordinates are scaled by the same factor); R, a rotation; and T, a translation. Does D *commute* with R? That is, is $D(R(\mathbf{x})) = R(D(\mathbf{x}))$ for all $\mathbf{x}$ in $\mathbb{R}^2$? Does D commute with T? Does R commute with T?

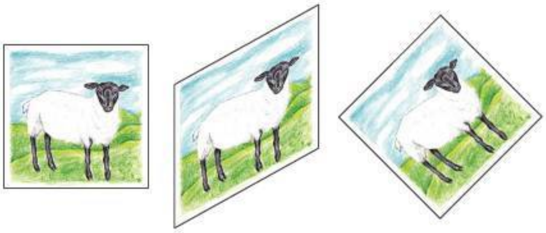
[2] See Jan Ozer, "High-Performance Graphics Boards," *PC Magazine* **19**, September 1, 2000, pp. 187–200. Also, "The Ultimate Upgrade Guide: Moving On Up," *PC Magazine* **21**, January 29, 2002, pp. 82–91.

11. A rotation on a computer screen is sometimes implemented as the product of two shear-and-scale transformations, which can speed up calculations that determine how a graphic image actually appears in terms of screen pixels. (The screen consists of rows and columns of small dots, called *pixels*.) The first transformation A_1 shears vertically and then compresses each column of pixels; the second transformation A_2 shears horizontally and then stretches each row of pixels. Let

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} \sec\varphi & -\tan\varphi & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Show that the composition of the two transformations is a rotation in $\mathbb{R}^2$.



12. A rotation in $\mathbb{R}^2$ usually requires four multiplications. Compute the product below, and show that the matrix for a rotation can be factored into three shear transformations (each of which requires only one multiplication).

$$\begin{bmatrix} 1 & -\tan\varphi/2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ \sin\varphi & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} 1 & -\tan\varphi/2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

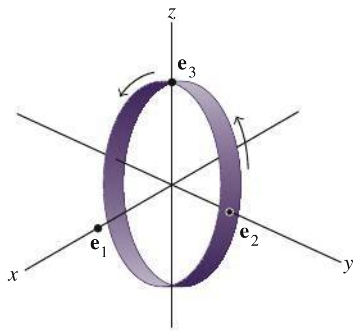
13. The usual transformations on homogeneous coordinates for 2D computer graphics involve 3×3 matrices of the form $\begin{bmatrix} A & \mathbf{p} \\ \mathbf{0}^T & 1 \end{bmatrix}$ where A is a 2×2 matrix and $\mathbf{p}$ is in $\mathbb{R}^2$. Show that such a transformation amounts to a linear transformation on $\mathbb{R}^2$ followed by a translation. [*Hint:* Find an appropriate matrix factorization involving partitioned matrices.]

14. Show that the transformation in Exercise 7 is equivalent to a rotation about the origin followed by a translation by $\mathbf{p}$. Find $\mathbf{p}$.

15. What vector in $\mathbb{R}^3$ has homogeneous coordinates $\left(\frac{1}{2}, -\frac{1}{4}, \frac{1}{8}, \frac{1}{24}\right)$?

16. Are $(1, -2, 3, 4)$ and $(10, -20, 30, 40)$ homogeneous coordinates for the same point in $\mathbb{R}^3$? Why or why not?

17. Give the 4×4 matrix that rotates points in $\mathbb{R}^3$ about the x-axis through an angle of $60°$. (See the figure.)



18. Give the 4×4 matrix that rotates points in $\mathbb{R}^3$ about the z-axis through an angle of $-30°$, and then translates by $\mathbf{p} = (5, -2, 1)$.

19. Let S be the triangle with vertices $(4.2, 1.2, 4), (6, 4, 2), (2, 2, 6)$. Find the image of S under the perspective projection with center of projection at $(0, 0, 10)$.

20. Let S be the triangle with vertices $(9, 3, -5), (12, 8, 2), (1.8, 2.7, 1)$. Find the image of S under the perspective projection with center of projection at $(0, 0, 10)$.

Exercises 21 and 22 concern the way in which color is specified for display in computer graphics. A color on a computer screen is encoded by three numbers (R, G, B) that list the amount of energy an electron gun must transmit to red, green, and blue phosphor dots on the computer screen. (A fourth number specifies the luminance or intensity of the color.)

T 21. The actual color a viewer sees on a screen is influenced by the specific type and amount of phosphors on the screen. So each computer screen manufacturer must convert between the (R, G, B) data and an international CIE standard for color, which uses three primary colors, called X, Y, and Z. A typical conversion for short-persistence phosphors is

$$\begin{bmatrix} .61 & .29 & .150 \\ .35 & .59 & .063 \\ .04 & .12 & .787 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

A computer program will send a stream of color information to the screen, using standard CIE data (X, Y, Z). Find the equation that converts these data to the (R, G, B) data needed for the screen's electron gun.

T 22. The signal broadcast by commercial television describes each color by a vector (Y, I, Q). If the screen is black and white, only the Y-coordinate is used. (This gives a better monochrome picture than using CIE data for colors.) The correspondence between *YIQ* and a "standard" *RGB* color is given by

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} .299 & .587 & .114 \\ .596 & -.275 & -.321 \\ .212 & -.528 & .311 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

(A screen manufacturer would change the matrix entries to work for its *RGB* screens.) Find the equation that converts the *YIQ* data transmitted by the television station to the *RGB* data needed for the television screen.

Solution to Practice Problem

Assemble the matrices right-to-left for the three operations. Using $\mathbf{p} = (-2, 6)$, $\cos(-30°) = \sqrt{3}/2$, and $\sin(-30°) = -.5$, we have

$$\underbrace{\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Translate back by } p} \underbrace{\begin{bmatrix} \sqrt{3}/2 & 1/2 & 0 \\ -1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Rotate around the origin}} \underbrace{\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Translate by } -p}$$

$$= \begin{bmatrix} \sqrt{3}/2 & 1/2 & \sqrt{3}-5 \\ -1/2 & \sqrt{3}/2 & -3\sqrt{3}+5 \\ 0 & 0 & 1 \end{bmatrix}$$

2.8 Subspaces of $\mathbb{R}^n$

This section focuses on important sets of vectors in $\mathbb{R}^n$ called *subspaces*. Often subspaces arise in connection with some matrix A, and they provide useful information about the equation $A\mathbf{x} = \mathbf{b}$. The concepts and terminology in this section will be used repeatedly throughout the rest of the book.[1]

DEFINITION

A **subspace** of $\mathbb{R}^n$ is any set H in $\mathbb{R}^n$ that has three properties:

a. The zero vector is in H.

b. For each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H.

c. For each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is in H.

In words, a subspace is *closed* under addition and scalar multiplication. As you will see in the next few examples, most sets of vectors discussed in Chapter 1 are subspaces. For instance, a plane through the origin is the standard way to visualize the subspace in Example 1. See Figure 1.

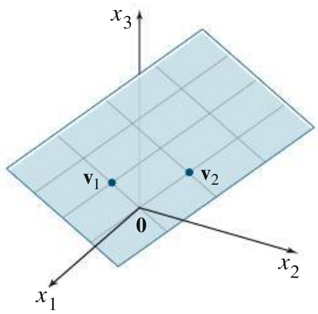


FIGURE 1 Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ as a plane through the origin.

EXAMPLE 1 If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in $\mathbb{R}^n$ and $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, then H is a subspace of $\mathbb{R}^n$. To verify this statement, note that the zero vector is in H (because $0\mathbf{v}_1 + 0\mathbf{v}_2$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$). Now take two arbitrary vectors in H, say,

$$\mathbf{u} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 \quad \text{and} \quad \mathbf{v} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

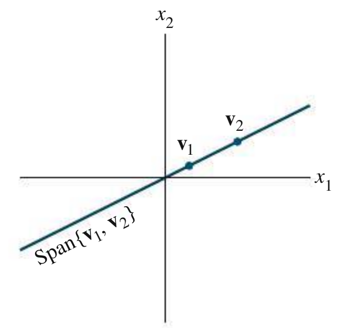
Then

$$\mathbf{u} + \mathbf{v} = (s_1 + t_1)\mathbf{v}_1 + (s_2 + t_2)\mathbf{v}_2$$

which shows that $\mathbf{u} + \mathbf{v}$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ and hence is in H. Also, for any scalar c, the vector $c\mathbf{u}$ is in H, because $c\mathbf{u} = c(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) = (cs_1)\mathbf{v}_1 + (cs_2)\mathbf{v}_2$. ■

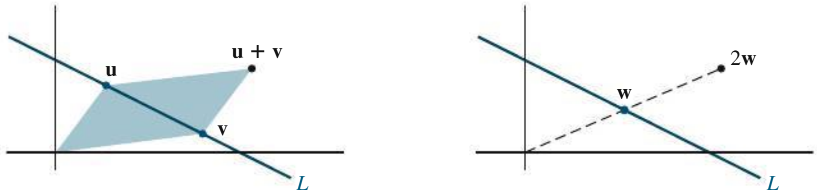
If $\mathbf{v}_1$ is not zero and if $\mathbf{v}_2$ is a multiple of $\mathbf{v}_1$, then $\mathbf{v}_1$ and $\mathbf{v}_2$ simply span a *line* through the origin. So a line through the origin is another example of a subspace.

[1] Sections 2.8 and 2.9 are included here to permit readers to postpone the study of most or all of the next two chapters and to skip directly to Chapter 5, if so desired. *Omit* these two sections if you plan to work through Chapter 4 before beginning Chapter 5.



$\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2 = k\mathbf{v}_1$.

EXAMPLE 2 A line L *not* through the origin is *not* a subspace, because it does not contain the origin, as required. Also, Figure 2 shows that L is not closed under addition or scalar multiplication. ■



$\mathbf{u} + \mathbf{v}$ is not on L $\qquad$ $2\mathbf{w}$ is not on L

FIGURE 2

EXAMPLE 3 For $\mathbf{v}_1, \dots, \mathbf{v}_p$ in $\mathbb{R}^n$, the set of all linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_p$ is a subspace of $\mathbb{R}^n$. The verification of this statement is similar to the argument given in Example 1. We shall now refer to Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ as **the subspace spanned** (or **generated**) by $\mathbf{v}_1, \dots, \mathbf{v}_p$. ■

Note that $\mathbb{R}^n$ is a subspace of itself because it has the three properties required for a subspace. Another special subspace is the set consisting of only the zero vector in $\mathbb{R}^n$. This set, called the **zero subspace**, also satisfies the conditions for a subspace.

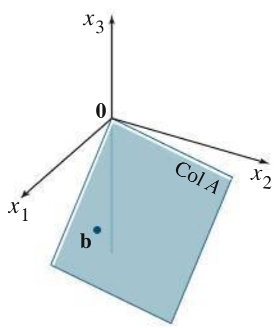
Column Space and Null Space of a Matrix

Subspaces of $\mathbb{R}^n$ usually occur in applications and theory in one of two ways. In both cases, the subspace can be related to a matrix.

DEFINITION

The **column space** of a matrix A is the set Col A of all linear combinations of the columns of A.

If $A = [\,\mathbf{a}_1 \ \cdots \ \mathbf{a}_n\,]$, with the columns in $\mathbb{R}^m$, then Col A is the same as Span $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$. Example 4 shows that the **column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$**. Note that Col A equals $\mathbb{R}^m$ only when the columns of A span $\mathbb{R}^m$. Otherwise, Col A is only part of $\mathbb{R}^m$.



EXAMPLE 4 Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$. Determine whether $\mathbf{b}$ is in the column space of A.

SOLUTION The vector $\mathbf{b}$ is a linear combination of the columns of A if and only if $\mathbf{b}$ can be written as $A\mathbf{x}$ for some $\mathbf{x}$, that is, if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution. Row reducing the augmented matrix $[\,A \ \ \mathbf{b}\,]$,

$$\begin{bmatrix} 1 & -3 & -4 & 3 \\ -4 & 6 & -2 & 3 \\ -3 & 7 & 6 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & -4 & 3 \\ 0 & -6 & -18 & 15 \\ 0 & -2 & -6 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & -4 & 3 \\ 0 & -6 & -18 & 15 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we conclude that $A\mathbf{x} = \mathbf{b}$ is consistent and $\mathbf{b}$ is in Col A. ■

The solution of Example 4 shows that when a system of linear equations is written in the form $A\mathbf{x} = \mathbf{b}$, the column space of A is the set of all $\mathbf{b}$ for which the system has a solution.

DEFINITION

The **null space** of a matrix A is the set Nul A of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

When A has n columns, the solutions of $A\mathbf{x} = \mathbf{0}$ belong to $\mathbb{R}^n$, and the null space of A is a subset of $\mathbb{R}^n$. In fact, Nul A has the properties of a *subspace* of $\mathbb{R}^n$.

THEOREM 12

The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions of a system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

PROOF The zero vector is in Nul A (because $A\mathbf{0} = \mathbf{0}$). To show that Nul A satisfies the other two properties required for a subspace, take any $\mathbf{u}$ and $\mathbf{v}$ in Nul A. That is, suppose $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. Then, by a property of matrix multiplication,

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Thus $\mathbf{u} + \mathbf{v}$ satisfies $A\mathbf{x} = \mathbf{0}$, and so $\mathbf{u} + \mathbf{v}$ is in Nul A. Also, for any scalar c, $A(c\mathbf{u}) = c(A\mathbf{u}) = c(\mathbf{0}) = \mathbf{0}$, which shows that $c\mathbf{u}$ is in Nul A. ■

To test whether a given vector $\mathbf{v}$ is in Nul A, just compute $A\mathbf{v}$ to see whether $A\mathbf{v}$ is the zero vector. Because Nul A is described by a condition that must be checked for each vector, we say that the null space is defined *implicitly*. In contrast, the column space is defined *explicitly*, because vectors in Col A can be constructed (by linear combinations) from the columns of A. To create an explicit description of Nul A, solve the equation $A\mathbf{x} = \mathbf{0}$ and write the solution in parametric vector form. (See Example 6.)[2]

Basis for a Subspace

Because a subspace typically contains an infinite number of vectors, some problems involving a subspace are handled best by working with a small finite set of vectors that span the subspace. The smaller the set, the better. It can be shown that the smallest possible spanning set must be linearly independent.

DEFINITION

A **basis** for a subspace H of $\mathbb{R}^n$ is a linearly independent set in H that spans H.

EXAMPLE 5 The columns of an invertible $n \times n$ matrix form a basis for all of $\mathbb{R}^n$ because they are linearly independent and span $\mathbb{R}^n$, by the Invertible Matrix Theorem. One such matrix is the $n \times n$ identity matrix. Its columns are denoted by $\mathbf{e}_1, \dots, \mathbf{e}_n$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$. See Figure 3. ■

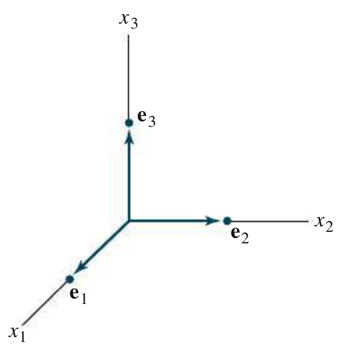


FIGURE 3
The standard basis for $\mathbb{R}^3$.

[2] The contrast between Nul A and Col A is discussed further in Section 4.2.

The next example shows that the standard procedure for writing the solution set of $A\mathbf{x} = \mathbf{0}$ in parametric vector form actually identifies a basis for Nul A. This fact will be used throughout Chapter 5.

EXAMPLE 6 Find a basis for the null space of the matrix

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

SOLUTION First, write the solution of $A\mathbf{x} = \mathbf{0}$ in parametric vector form:

$$\begin{bmatrix} A & \mathbf{0} \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 - 2x_2 \qquad - x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The general solution is $x_1 = 2x_2 + x_4 - 3x_5$, $x_3 = -2x_4 + 2x_5$, with x_2, x_4, and x_5 free.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \underset{\substack{\uparrow \\ \mathbf{u}}}{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}} + x_4 \underset{\substack{\uparrow \\ \mathbf{v}}}{\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}} + x_5 \underset{\substack{\uparrow \\ \mathbf{w}}}{\begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}}$$

$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w} \tag{1}$$

Equation (1) shows that Nul A coincides with the set of all linear combinations of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. That is, $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ generates Nul A. In fact, this construction of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ automatically makes them linearly independent, because equation (1) shows that $\mathbf{0} = x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$ only if the weights x_2, x_4, and x_5 are all zero. (Examine entries 2, 4, and 5 in the vector $x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$.) So $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a *basis* for Nul A. ■

Finding a basis for the column space of a matrix is actually less work than finding a basis for the null space. However, the method requires some explanation. Let's begin with a simple case.

EXAMPLE 7 Find a basis for the column space of the matrix

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

SOLUTION Denote the columns of B by $\mathbf{b}_1, \ldots, \mathbf{b}_5$ and note that $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ and $\mathbf{b}_4 = 5\mathbf{b}_1 - \mathbf{b}_2$. The fact that $\mathbf{b}_3$ and $\mathbf{b}_4$ are combinations of the pivot columns means that any combination of $\mathbf{b}_1, \ldots, \mathbf{b}_5$ is actually just a combination of $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_5$. Indeed, if $\mathbf{v}$ is any vector in Col B, say,

$$\mathbf{v} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + c_3\mathbf{b}_3 + c_4\mathbf{b}_4 + c_5\mathbf{b}_5$$

then, substituting for $\mathbf{b}_3$ and $\mathbf{b}_4$, we can write $\mathbf{v}$ in the form

$$\mathbf{v} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + c_3(-3\mathbf{b}_1 + 2\mathbf{b}_2) + c_4(5\mathbf{b}_1 - \mathbf{b}_2) + c_5\mathbf{b}_5$$

which is a linear combination of $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_5$. So $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_5\}$ spans Col B. Also, $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_5$ are linearly independent, because they are columns from an identity matrix. So the pivot columns of B form a basis for Col B. ■

The matrix B in Example 7 is in reduced echelon form. To handle a general matrix A, recall that linear dependence relations among the columns of A can be expressed in the form $A\mathbf{x} = \mathbf{0}$ for some $\mathbf{x}$. (If some columns are not involved in a particular dependence relation, then the corresponding entries in $\mathbf{x}$ are zero.) When A is row reduced to echelon form B, the columns are drastically changed, but the equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same set of solutions. That is, the columns of A have *exactly the same linear dependence relationships* as the columns of B.

EXAMPLE 8 It can be verified that the matrix

$$A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_5\,] = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}$$

is row equivalent to the matrix B in Example 7. Find a basis for Col A.

SOLUTION From Example 7, the pivot columns of A are columns 1, 2, and 5. Also, $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ and $\mathbf{b}_4 = 5\mathbf{b}_1 - \mathbf{b}_2$. Since row operations do not affect linear dependence relations among the columns of the matrix, we should have

$$\mathbf{a}_3 = -3\mathbf{a}_1 + 2\mathbf{a}_2 \quad \text{and} \quad \mathbf{a}_4 = 5\mathbf{a}_1 - \mathbf{a}_2$$

Check that this is true! By the argument in Example 7, $\mathbf{a}_3$ and $\mathbf{a}_4$ are not needed to generate the column space of A. Also, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5\}$ must be linearly independent, because any dependence relation among $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_5$ would imply the same dependence relation among $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_5$. Since $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_5\}$ is linearly independent, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5\}$ is also linearly independent and hence is a basis for Col A. ■

The argument in Example 8 can be adapted to prove the following theorem.

THEOREM 13

The pivot columns of a matrix A form a basis for the column space of A.

Warning: Be careful to use *pivot columns of A itself* for the basis of Col A. The columns of an echelon form B are often not in the column space of A. (For instance, in Examples 7 and 8, the columns of B all have zeros in their last entries and cannot generate the columns of A.)

Practice Problems

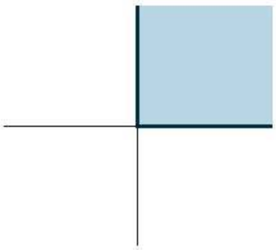
1. Let $A = \begin{bmatrix} 1 & -1 & 5 \\ 2 & 0 & 7 \\ -3 & -5 & -3 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} -7 \\ 3 \\ 2 \end{bmatrix}$. Is $\mathbf{u}$ in Nul A? Is $\mathbf{u}$ in Col A? Justify each answer.
2. Given $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, find a vector in Nul A and a vector in Col A.
3. Suppose an $n \times n$ matrix A is invertible. What can you say about Col A? About Nul A?

STUDY GUIDE offers additional resources for mastering the concepts of subspace, column space, and null space.

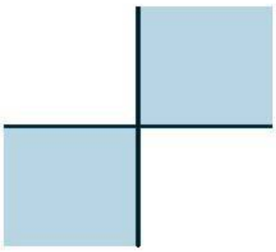
2.8 Exercises

Exercises 1–4 display sets in $\mathbb{R}^2$. Assume the sets include the bounding lines. In each case, give a specific reason why the set H is *not* a subspace of $\mathbb{R}^2$. (For instance, find two vectors in H whose sum is *not* in H, or find a vector in H with a scalar multiple that is not in H. Draw a picture.)

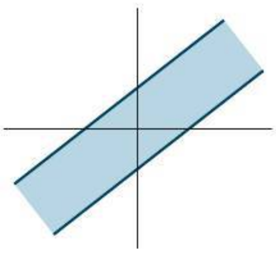
1.



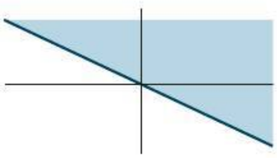
2.



3.



4.



5. Let $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ -5 \\ 8 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 8 \\ 2 \\ -9 \end{bmatrix}$. Determine if $\mathbf{w}$ is in the subspace of $\mathbb{R}^3$ generated by $\mathbf{v}_1$ and $\mathbf{v}_2$.

6. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ -7 \\ 9 \\ 7 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 5 \\ -8 \\ 6 \\ 5 \end{bmatrix}$, and $\mathbf{u} = \begin{bmatrix} -4 \\ 10 \\ -7 \\ -5 \end{bmatrix}$. Determine if $\mathbf{u}$ is in the subspace of $\mathbb{R}^4$ generated by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

7. Let $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -3 \\ 8 \\ -7 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -4 \\ 6 \\ -7 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} 6 \\ -10 \\ 11 \end{bmatrix}$, and $A = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]$.

a. How many vectors are in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

b. How many vectors are in Col A?

c. Is $\mathbf{p}$ in Col A? Why or why not?

8. Let $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 0 \\ 6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0 \\ -6 \\ 3 \end{bmatrix}$, and $\mathbf{p} = \begin{bmatrix} 1 \\ 14 \\ -9 \end{bmatrix}$. Determine if $\mathbf{p}$ is in Col A, where $A = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]$.

9. With A and $\mathbf{p}$ as in Exercise 7, determine if $\mathbf{p}$ is in Nul A.

10. With $\mathbf{u} = (-2, 3, 1)$ and A as in Exercise 8, determine if $\mathbf{u}$ is in Nul A.

In Exercises 11 and 12, give integers p and q such that Nul A is a subspace of $\mathbb{R}^p$ and Col A is a subspace of $\mathbb{R}^q$.

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

12. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ -5 & -1 & 0 \\ 2 & 7 & 11 \end{bmatrix}$

13. For A as in Exercise 11, find a nonzero vector in Nul A and a nonzero vector in Col A.

14. For A as in Exercise 12, find a nonzero vector in Nul A and a nonzero vector in Col A.

Determine which sets in Exercises 15–20 are bases for $\mathbb{R}^2$ or $\mathbb{R}^3$. Justify each answer.

15. $\begin{bmatrix} 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 10 \\ -3 \end{bmatrix}$

16. $\begin{bmatrix} -4 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \end{bmatrix}$

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

18. $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -5 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \\ -5 \end{bmatrix}$

19. $\begin{bmatrix} 3 \\ -8 \\ 1 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \\ -5 \end{bmatrix}$

20. $\begin{bmatrix} 1 \\ -6 \\ -7 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \\ 7 \end{bmatrix}, \begin{bmatrix} -2 \\ 7 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 8 \\ 9 \end{bmatrix}$

In Exercises 21–30, mark each statement Ture or False **(T/F)**. Justify each answer.

21. **(T/F)** A subspace of $\mathbb{R}^n$ is any set H such that (i) the zero vector is in H, (ii) $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u}+\mathbf{v}$ are in H, and (iii) c is a scalar and $c\mathbf{u}$ is in H.

22. **(T/F)** A subset H of $\mathbb{R}^n$ is a subspace if the zero vector is in H.

23. **(T/F)** If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in $\mathbb{R}^n$, then Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is the same as the column space of the matrix $[\mathbf{v}_1 \ \dots \ \mathbf{v}_p]$.

24. **(T/F)** Given vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ in $\mathbb{R}^n$, the set of all linear combinations of these vectors is a subspace of $\mathbb{R}^n$.

25. **(T/F)** The set of all solutions of a system of m homogeneous equations in n unknowns is a subspace of $\mathbb{R}^m$.

26. **(T/F)** The null space of an $m \times n$ matrix is a subspace of $\mathbb{R}^n$.

27. **(T/F)** The columns of an invertible $n \times n$ matrix form a basis for $\mathbb{R}^n$.

28. **(T/F)** The column space of a matrix A is the set of solutions of $A\mathbf{x} = \mathbf{b}$.

29. **(T/F)** Row operations do not affect linear dependence relations among the columns of a matrix.

30. **(T/F)** If B is an echelon form of a matrix A, then the pivot columns of B form a basis for Col A.

Exercises 31–34 display a matrix A and an echelon form of A. Find a basis for Col A and a basis for Nul A.

31. $A = \begin{bmatrix} 4 & 5 & 9 & -2 \\ 6 & 5 & 1 & 12 \\ 3 & 4 & 8 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 6 & -5 \\ 0 & 1 & 5 & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

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

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

$\sim \begin{bmatrix} 1 & 4 & 8 & 0 & 5 \\ 0 & 2 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

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

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

35. Construct a nonzero 3×3 matrix A and a nonzero vector $\mathbf{b}$ such that $\mathbf{b}$ is in Col A, but $\mathbf{b}$ is not the same as any one of the columns of A.

36. Construct a nonzero 3×3 matrix A and a vector $\mathbf{b}$ such that $\mathbf{b}$ is *not* in Col A.

37. Construct a nonzero 3×3 matrix A and a nonzero vector $\mathbf{b}$ such that $\mathbf{b}$ is in Nul A.

38. Suppose the columns of a matrix $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_p]$ are linearly independent. Explain why $\{\mathbf{a}_1, \dots, \mathbf{a}_p\}$ is a basis for Col A.

In Exercises 39–44, respond as comprehensively as possible, and justify your answer.

39. Suppose F is a 5×5 matrix whose column space is not equal to $\mathbb{R}^5$. What can you say about Nul F?

40. If R is a 6×6 matrix and Nul R is *not* the zero subspace, what can you say about Col R?

41. If Q is a 4×4 matrix and Col $Q = \mathbb{R}^4$, what can you say about solutions of equations of the form $Q\mathbf{x} = \mathbf{b}$ for $\mathbf{b}$ in $\mathbb{R}^4$?

42. If P is a 5×5 matrix and Nul P is the zero subspace, what can you say about solutions of equations of the form $P\mathbf{x} = \mathbf{b}$ for $\mathbf{b}$ in $\mathbb{R}^5$?

43. What can you say about Nul B when B is a 5×4 matrix with linearly independent columns?

44. What can you say about the shape of an $m \times n$ matrix A when the columns of A form a basis for $\mathbb{R}^m$?

In Exercises 45 and 46, construct bases for the column space and the null space of the given matrix A. Justify your work.

T 45. $A = \begin{bmatrix} 3 & -5 & 0 & -1 & 3 \\ -7 & 9 & -4 & 9 & -11 \\ -5 & 7 & -2 & 5 & -7 \\ 3 & -7 & -3 & 4 & 0 \end{bmatrix}$

T 46. $A = \begin{bmatrix} 5 & 2 & 0 & -8 & -8 \\ 4 & 1 & 2 & -8 & -9 \\ 5 & 1 & 3 & 5 & 19 \\ -8 & -5 & 6 & 8 & 5 \end{bmatrix}$

Solutions to Practice Problems

1. To determine whether $\mathbf{u}$ is in Nul A, simply compute

$$A\mathbf{u} = \begin{bmatrix} 1 & -1 & 5 \\ 2 & 0 & 7 \\ -3 & -5 & -3 \end{bmatrix}\begin{bmatrix} -7 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The result shows that $\mathbf{u}$ is in Nul A. Deciding whether $\mathbf{u}$ is in Col A requires more work. Reduce the augmented matrix $[A \quad \mathbf{u}]$ to echelon form to determine whether the equation $A\mathbf{x} = \mathbf{u}$ is consistent:

$$\begin{bmatrix} 1 & -1 & 5 & -7 \\ 2 & 0 & 7 & 3 \\ -3 & -5 & -3 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 5 & -7 \\ 0 & 2 & -3 & 17 \\ 0 & -8 & 12 & -19 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 5 & -7 \\ 0 & 2 & -3 & 17 \\ 0 & 0 & 0 & 49 \end{bmatrix}$$

The equation $A\mathbf{x} = \mathbf{u}$ has no solution, so $\mathbf{u}$ is not in Col A.

2. In contrast to Practice Problem 1, finding a vector in Nul A requires more work than testing whether a specified vector is in Nul A. However, since A is already in reduced echelon form, the equation $A\mathbf{x} = \mathbf{0}$ shows that if $\mathbf{x} = (x_1, x_2, x_3)$, then $x_2 = 0$, $x_3 = 0$, and x_1 is a free variable. Thus, a basis for Nul A is $\mathbf{v} = (1, 0, 0)$. Finding just one vector in Col A is trivial, since each column of A is in Col A. In this particular case, the same vector $\mathbf{v}$ is in both Nul A *and* Col A. For most $n \times n$ matrices, the zero vector of $\mathbb{R}^n$ is the only vector in both Nul A and Col A.

3. If A is invertible, then the columns of A span $\mathbb{R}^n$, by the Invertible Matrix Theorem. By definition, the columns of any matrix always span the column space, so in this case Col A is all of $\mathbb{R}^n$. In symbols, Col $A = \mathbb{R}^n$. Also, since A is invertible, the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This means that Nul A is the zero subspace. In symbols, Nul $A = \{\mathbf{0}\}$.

2.9 Dimension and Rank

This section continues the discussion of subspaces and bases for subspaces, beginning with the concept of a coordinate system. The definition and example below should make a useful new term, *dimension*, seem quite natural, at least for subspaces of $\mathbb{R}^3$.

Coordinate Systems

The main reason for selecting a basis for a subspace H, instead of merely a spanning set, is that each vector in H can be written in only one way as a linear combination of the basis vectors. To see why, suppose $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for H, and suppose a vector $\mathbf{x}$ in H can be generated in two ways, say,

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p \quad \text{and} \quad \mathbf{x} = d_1\mathbf{b}_1 + \cdots + d_p\mathbf{b}_p \tag{1}$$

Then, subtracting gives

$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (c_1 - d_1)\mathbf{b}_1 + \cdots + (c_p - d_p)\mathbf{b}_p \tag{2}$$

Since $\mathcal{B}$ is linearly independent, the weights in (2) must all be zero. That is, $c_j = d_j$ for $1 \le j \le p$, which shows that the two representations in (1) are actually the same.

DEFINITION

Suppose the set $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for a subspace H. For each $\mathbf{x}$ in H, the **coordinates of x relative to the basis** $\mathcal{B}$ are the weights $c_1, \dots, c_p$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p$, and the vector in $\mathbb{R}^p$

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

is called the **coordinate vector of x (relative to** $\mathcal{B}$**)** or the $\mathcal{B}$**-coordinate vector of x**.[1]

EXAMPLE 1 Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Then $\mathcal{B}$ is a basis for $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ because $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. Determine if $\mathbf{x}$ is in H, and if it is, find the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$.

SOLUTION If $\mathbf{x}$ is in H, then the following vector equation is consistent:

$$c_1 \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$$

The scalars c_1 and c_2, if they exist, are the $\mathcal{B}$-coordinates of $\mathbf{x}$. Row operations show that

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

Thus $c_1 = 2$, $c_2 = 3$, and $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. The basis $\mathcal{B}$ determines a "coordinate system" on H, which can be visualized by the grid shown in Figure 1. ■

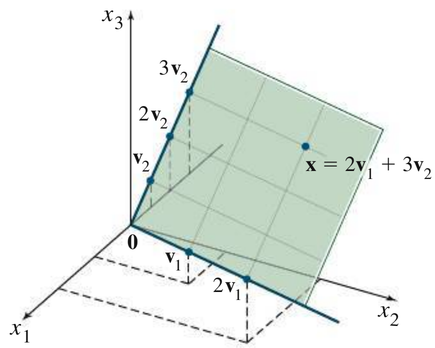


FIGURE 1 A coordinate system on a plane H in $\mathbb{R}^3$.

Notice that although points in H are also in $\mathbb{R}^3$, they are completely determined by their coordinate vectors, which belong to $\mathbb{R}^2$. The grid on the plane in Figure 1

[1] It is important that the elements of $\mathcal{B}$ are numbered because the entries in $[\mathbf{x}]_{\mathcal{B}}$ depend on the order of the vectors in $\mathcal{B}$.

makes H "look" like $\mathbb{R}^2$. The correspondence $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$ is a one-to-one correspondence between H and $\mathbb{R}^2$ that preserves linear combinations. We call such a correspondence an *isomorphism*, and we say that H is *isomorphic* to $\mathbb{R}^2$.

In general, if $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for H, then the mapping $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one correspondence that makes H look and act the same as $\mathbb{R}^p$ (even though the vectors in H themselves may have more than p entries). (Section 4.4 has more details.)

The Dimension of a Subspace

It can be shown that if a subspace H has a basis of p vectors, then every basis of H must consist of exactly p vectors. (See Exercises 35 and 36.) Thus the following definition makes sense.

DEFINITION The **dimension** of a nonzero subspace H, denoted by $\dim H$, is the number of vectors in any basis for H. The dimension of the zero subspace $\{\mathbf{0}\}$ is defined to be zero.[2]

The space $\mathbb{R}^n$ has dimension n. Every basis for $\mathbb{R}^n$ consists of n vectors. A plane through $\mathbf{0}$ in $\mathbb{R}^3$ is two-dimensional, and a line through $\mathbf{0}$ is one-dimensional.

EXAMPLE 2 Recall that the null space of the matrix A in Example 6 in Section 2.8 had a basis of 3 vectors. So the dimension of Nul A in this case is 3. Observe how each basis vector corresponds to a free variable in the equation $A\mathbf{x} = \mathbf{0}$. Our construction always produces a basis in this way. So, to find the dimension of Nul A, simply identify and count the number of free variables in $A\mathbf{x} = \mathbf{0}$. ■

DEFINITION The **rank** of a matrix A, denoted by rank A, is the dimension of the column space of A.

Since the pivot columns of A form a basis for Col A, the rank of A is just the number of pivot columns in A.

EXAMPLE 3 Determine the rank of the matrix

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

SOLUTION Reduce A to echelon form:

$$A \sim \begin{bmatrix} 2 & 5 & -3 & -4 & 8 \\ 0 & -3 & 2 & 5 & -7 \\ 0 & -6 & 4 & 14 & -20 \\ 0 & -9 & 6 & 5 & -6 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 2 & 5 & -3 & -4 & 8 \\ 0 & -3 & 2 & 5 & -7 \\ 0 & 0 & 0 & 4 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Pivot columns

The matrix A has 3 pivot columns, so rank $A = 3$. ■

[2] The zero subspace has *no* basis (because the zero vector by itself forms a linearly dependent set).

The row reduction in Example 3 reveals that there are two free variables in $A\mathbf{x} = \mathbf{0}$, because two of the five columns of A are *not* pivot columns. (The nonpivot columns correspond to the free variables in $A\mathbf{x} = \mathbf{0}$.) Since the number of pivot columns plus the number of nonpivot columns is exactly the number of columns, the dimensions of Col A and Nul A have the following useful connection. (See the Rank Theorem in Section 4.6 for additional details.)

THEOREM 14

The Rank Theorem

If a matrix A has n columns, then rank A + dim Nul $A = n$.

The following theorem is important for applications and will be needed in Chapters 5 and 6. The theorem (proved in Section 4.5) is certainly plausible, if you think of a p-dimensional subspace as isomorphic to $\mathbb{R}^p$. The Invertible Matrix Theorem shows that p vectors in $\mathbb{R}^p$ are linearly independent if and only if they also span $\mathbb{R}^p$.

THEOREM 15

The Basis Theorem

Let H be a p-dimensional subspace of $\mathbb{R}^n$. Any linearly independent set of exactly p elements in H is automatically a basis for H. Also, any set of p elements of H that spans H is automatically a basis for H.

Rank and the Invertible Matrix Theorem

The various vector space concepts associated with a matrix provide several more statements for the Invertible Matrix Theorem. They are presented below to follow the statements in the original theorem in Section 2.3.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix.

m. The columns of A form a basis of $\mathbb{R}^n$.
n. Col $A = \mathbb{R}^n$
o. rank $A = n$
p. dim Nul $A = 0$
q. Nul $A = \{\mathbf{0}\}$

PROOF Statement (m) is logically equivalent to statements (e) and (h) regarding linear independence and spanning. The other four statements are linked to the earlier ones of the theorem by the following chain of almost trivial implications:

$$(g) \Rightarrow (n) \Rightarrow (o) \Rightarrow (p) \Rightarrow (q) \Rightarrow (d)$$

Statement (g), which says that the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$, implies statement (n), because Col A is precisely the set of all $\mathbf{b}$ such that the equation $A\mathbf{x} = \mathbf{b}$ is consistent. The implications (n) $\Rightarrow$ (o) $\Rightarrow$ (p) follow from the definitions of *dimension* and *rank*. If the rank of A is n, the number of columns of A,

STUDY GUIDE offers an expanded Invertible Matrix Theorem Table.

then dim Nul $A = 0$, by the Rank Theorem, and so Nul $A = \{\mathbf{0}\}$. Thus (p) $\Rightarrow$ (q). Also, statement (q) implies that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, which is statement (d). Since statements (d) and (g) are already known to be equivalent to the statement that A is invertible, the proof is complete. ■

Numerical Notes

Many algorithms discussed in this text are useful for understanding concepts and making simple computations by hand. However, the algorithms are often unsuitable for large-scale problems in real life.

Rank determination is a good example. It would seem easy to reduce a matrix to echelon form and count the pivots. But unless exact arithmetic is performed on a matrix whose entries are specified exactly, row operations can change the apparent rank of a matrix. For instance, if the value of x in the matrix $\begin{bmatrix} 5 & 7 \\ 5 & x \end{bmatrix}$ is not stored exactly as 7 in a computer, then the rank may be 1 or 2, depending on whether the computer treats $x - 7$ as zero.

In practical applications, the effective rank of a matrix A is often determined from the singular value decomposition of A, to be discussed in Section 7.4.

Practice Problems

1. Determine the dimension of the subspace H of $\mathbb{R}^3$ spanned by the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. (First, find a basis for H.)

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ -7 \\ -1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 6 \\ -7 \end{bmatrix}$$

2. Consider the basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ .2 \end{bmatrix}, \begin{bmatrix} .2 \\ 1 \end{bmatrix} \right\}$$

for $\mathbb{R}^2$. If $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, what is $\mathbf{x}$?

3. Could $\mathbb{R}^3$ possibly contain a four-dimensional subspace? Explain.

2.9 Exercises

In Exercises 1 and 2, find the vector $\mathbf{x}$ determined by the given coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ and the given basis $\mathcal{B}$. Illustrate your answer with a figure, as in the solution of Practice Problem 2.

1. $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}, [\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

2. $\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}, [\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

In Exercises 3–6, the vector $\mathbf{x}$ is in a subspace H with a basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Find the $\mathcal{B}$-coordinate vector of $\mathbf{x}$.

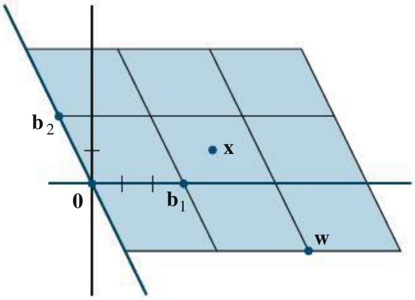
3. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -2 \\ 7 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} -3 \\ 7 \end{bmatrix}$

4. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} -7 \\ 5 \end{bmatrix}$

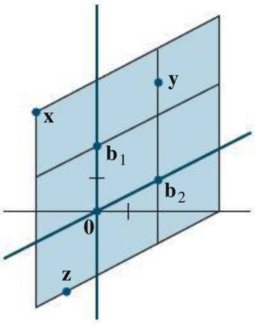
5. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -3 \\ -7 \\ 5 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 4 \\ 10 \\ -7 \end{bmatrix}$

6. $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 1 \\ -4 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 7 \\ 5 \\ -6 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 11 \\ 0 \\ 7 \end{bmatrix}$

7. Let $\mathbf{b}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 7 \\ -2 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Use the figure to estimate $[\mathbf{w}]_{\mathcal{B}}$ and $[\mathbf{x}]_{\mathcal{B}}$. Confirm your estimate of $[\mathbf{x}]_{\mathcal{B}}$ by using it and $\{\mathbf{b}_1, \mathbf{b}_2\}$ to compute $\mathbf{x}$.



8. Let $\mathbf{b}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$, $\mathbf{z} = \begin{bmatrix} -1 \\ -2.5 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Use the figure to estimate $[\mathbf{x}]_{\mathcal{B}}$, $[\mathbf{y}]_{\mathcal{B}}$, and $[\mathbf{z}]_{\mathcal{B}}$. Confirm your estimates of $[\mathbf{y}]_{\mathcal{B}}$ and $[\mathbf{z}]_{\mathcal{B}}$ by using them and $\{\mathbf{b}_1, \mathbf{b}_2\}$ to compute $\mathbf{y}$ and $\mathbf{z}$.



Exercises 9–12 display a matrix A and an echelon form of A. Find bases for Col A and Nul A, and then state the dimensions of these subspaces.

9. $A = \begin{bmatrix} 1 & -3 & 2 & -4 \\ -3 & 9 & -1 & 5 \\ 2 & -6 & 4 & -3 \\ -4 & 12 & 2 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 2 & -4 \\ 0 & 0 & 5 & -7 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

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

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

11. $A = \begin{bmatrix} 1 & 2 & -5 & 0 & -1 \\ 2 & 5 & -8 & 4 & 3 \\ -3 & -9 & 9 & -7 & -2 \\ 3 & 10 & -7 & 11 & 7 \end{bmatrix}$

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

12. $A = \begin{bmatrix} 1 & 2 & -4 & 3 & 3 \\ 5 & 10 & -9 & -7 & 8 \\ 4 & 8 & -9 & -2 & 7 \\ -2 & -4 & 5 & 0 & -6 \end{bmatrix}$

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

In Exercises 13 and 14, find a basis for the subspace spanned by the given vectors. What is the dimension of the subspace?

13. $\begin{bmatrix} 1 \\ -3 \\ 2 \\ -4 \end{bmatrix}, \begin{bmatrix} -3 \\ 9 \\ -6 \\ 12 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 5 \\ -3 \\ 7 \end{bmatrix}$

14. $\begin{bmatrix} 1 \\ -1 \\ -2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ -1 \\ 6 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -6 \\ 8 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ -7 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ -8 \\ 9 \\ -5 \end{bmatrix}$

15. Suppose a 3×5 matrix A has three pivot columns. Is Col $A = \mathbb{R}^3$? Is Nul $A = \mathbb{R}^2$? Explain your answers.

16. Suppose a 4×7 matrix A has three pivot columns. Is Col $A = \mathbb{R}^3$? What is the dimension of Nul A? Explain your answers.

In Exercises 17–26, mark each statement True or False **(T/F)**. Justify each answer. Here A is an $m \times n$ matrix.

17. **(T/F)** If $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis for a subspace H and if $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$, then $c_1, \dots, c_p$ are the coordinates of $\mathbf{x}$ relative to the basis $\mathcal{B}$.

18. **(T/F)** If $\mathcal{B}$ is a basis for a subspace H, then each vector in H can be written in only one way as a linear combination of the vectors in $\mathcal{B}$.

19. **(T/F)** Each line in $\mathbb{R}^n$ is a one-dimensional subspace of $\mathbb{R}^n$.

20. **(T/F)** If $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis for a subspace H of $\mathbb{R}^n$, then the correspondence $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ makes H look and act the same as $\mathbb{R}^p$.

21. **(T/F)** The dimension of Col A is the number of pivot columns of A.

22. **(T/F)** The dimension of Nul A is the number of variables in the equation $A\mathbf{x} = \mathbf{0}$.

23. **(T/F)** The dimensions of Col A and Nul A add up to the number of columns of A.

24. **(T/F)** The dimension of the column space of A is rank A.

25. **(T/F)** If a set of p vectors spans a p-dimensional subspace H of $\mathbb{R}^n$, then these vectors form a basis for H.

26. **(T/F)** If H is a p-dimensional subspace of $\mathbb{R}^n$, then a linearly independent set of p vectors in H is a basis for H.

In Exercises 27–32, justify each answer or construction.

27. If the subspace of all solutions of $A\mathbf{x} = \mathbf{0}$ has a basis consisting of three vectors and if A is a 5×7 matrix, what is the rank of A?

28. What is the rank of a 4×5 matrix whose null space is three-dimensional?

29. If the rank of a 7×6 matrix A is 4, what is the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$?

30. Show that a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_5\}$ in $\mathbb{R}^n$ is linearly dependent when dim Span $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_5\} = 4$.

31. If possible, construct a 3×4 matrix A such that dim Nul $A = 2$ and dim Col $A = 2$.

32. Construct a 4×3 matrix with rank 1.

33. Let A be an $n \times p$ matrix whose column space is p-dimensional. Explain why the columns of A must be linearly independent.

34. Suppose columns 1, 3, 5, and 6 of a matrix A are linearly independent (but are not necessarily pivot columns) and the rank of A is 4. Explain why the four columns mentioned must be a basis for the column space of A.

35. Suppose vectors $\mathbf{b}_1, \dots, \mathbf{b}_p$ span a subspace W, and let $\{\mathbf{a}_1, \dots, \mathbf{a}_q\}$ be any set in W containing more than p vectors. Fill in the details of the following argument to show that $\{\mathbf{a}_1, \dots, \mathbf{a}_q\}$ must be linearly dependent. First, let $B = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_p]$ and $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_q]$.
    a. Explain why for each vector $\mathbf{a}_j$, there exists a vector $\mathbf{c}_j$ in $\mathbb{R}^p$ such that $\mathbf{a}_j = B\mathbf{c}_j$.
    b. Let $C = [\mathbf{c}_1 \ \cdots \ \mathbf{c}_q]$. Explain why there is a nonzero vector $\mathbf{u}$ such that $C\mathbf{u} = \mathbf{0}$.
    c. Use B and C to show that $A\mathbf{u} = \mathbf{0}$. This shows that the columns of A are linearly dependent.

36. Use Exercise 35 to show that if $\mathcal{A}$ and $\mathcal{B}$ are bases for a subspace W of $\mathbb{R}^n$, then $\mathcal{A}$ cannot contain more vectors than $\mathcal{B}$, and, conversely, $\mathcal{B}$ cannot contain more vectors than $\mathcal{A}$.

T 37. Let $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Show that $\mathbf{x}$ is in H, and find the $\mathcal{B}$-coordinate vector of $\mathbf{x}$, when

$$\mathbf{v}_1 = \begin{bmatrix} 11 \\ -5 \\ 10 \\ 7 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 14 \\ -8 \\ 13 \\ 10 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 19 \\ -13 \\ 18 \\ 15 \end{bmatrix}$$

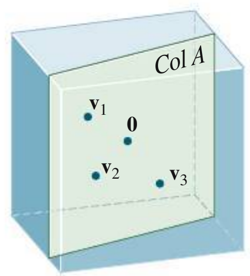
T 38. Let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Show that $\mathcal{B}$ is a basis for H and $\mathbf{x}$ is in H, and find the $\mathcal{B}$-coordinate vector of $\mathbf{x}$, when

$$\mathbf{v}_1 = \begin{bmatrix} -6 \\ 4 \\ -9 \\ 4 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 8 \\ -3 \\ 7 \\ -3 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -9 \\ 5 \\ -8 \\ 3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 4 \\ 7 \\ -8 \\ 3 \end{bmatrix}$$

STUDY GUIDE offers additional resources for mastering the concepts of dimension and rank.

Solutions to Practice Problems

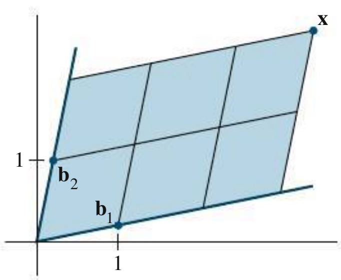
1. Construct $A = [\mathbf{v}_1 \ \ \mathbf{v}_2 \ \ \mathbf{v}_3]$ so that the subspace spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is the column space of A. A basis for this space is provided by the pivot columns of A.



$$A = \begin{bmatrix} 2 & 3 & -1 \\ -8 & -7 & 6 \\ 6 & -1 & -7 \end{bmatrix} \sim \begin{bmatrix} 2 & 3 & -1 \\ 0 & 5 & 2 \\ 0 & -10 & -4 \end{bmatrix} \sim \begin{bmatrix} 2 & 3 & -1 \\ 0 & 5 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

The first two columns of A are pivot columns and form a basis for H. Thus dim $H = 2$.

2. If $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, then $\mathbf{x}$ is formed from a linear combination of the basis vectors using weights 3 and 2:



$$\mathbf{x} = 3\mathbf{b}_1 + 2\mathbf{b}_2 = 3\begin{bmatrix} 1 \\ .2 \end{bmatrix} + 2\begin{bmatrix} .2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3.4 \\ 2.6 \end{bmatrix}$$

The basis $\{\mathbf{b}_1, \mathbf{b}_2\}$ determines a *coordinate system* for $\mathbb{R}^2$, illustrated by the grid in the figure. Note how $\mathbf{x}$ is 3 units in the $\mathbf{b}_1$-direction and 2 units in the $\mathbf{b}_2$-direction.

3. A four-dimensional subspace would contain a basis of four linearly independent vectors. This is impossible inside $\mathbb{R}^3$. Since any linearly independent set in $\mathbb{R}^3$ has no more than three vectors, any subspace of $\mathbb{R}^3$ has dimension no more than 3. The space $\mathbb{R}^3$ itself is the only three-dimensional subspace of $\mathbb{R}^3$. Other subspaces of $\mathbb{R}^3$ have dimension 2, 1, or 0.

CHAPTER 2 PROJECTS

Chapter 2 projects are available online at `bit.ly/30IM8gT`.

A. *Other Matrix Products*: This project introduces two new operations on square matrices called the Jordan product and the commutator product and their properties are explored.

B. *Adjacency Matrices*: The purpose of this project is to show how powers of a matrix may be used to investigate graphs.

C. *Dominance Matrices*: The purpose of this project is to apply matrices and their powers to questions concerning various forms of competition between individuals and groups.

D. *Condition Numbers*: The purpose of this project is to show how a condition number of a matrix A may be defined, and how its value affects the accuracy of solutions to systems of equations $A\mathbf{x} = \mathbf{b}$.

E. *Equilibrium Temperature Distributions*: The purpose of this project is to discuss a physical situation in which solving a system of linear equations becomes necessary: that of determining the equilibrium temperature of a thin plate.

F. *The LU and QR Factorizations*: The purpose of this project is to explore a relationship between two matrix factorizations: the LU factorization and the QR factorization.

G. *The Leontief Input–Output Model*: The purpose of this project is to provide three more examples of the Leontief input–output model in action.

H. *The Art of Linear Transformations*: This project illustrates how to graph a polygon and then use linear transformations to move it around in the plane.

CHAPTER 2 SUPPLEMENTARY EXERCISES

Assume that the matrices mentioned in Exercises 1–15 below have appropriate sizes. Mark each statement True or False **(T/F)**. Justify each answer.

1. **(T/F)** If A and B are $m \times n$, then both AB^T and A^TB are defined.

2. **(T/F)** If $AB = C$ and C has 2 columns, then A has 2 columns.

3. **(T/F)** Left-multiplying a matrix B by a diagonal matrix A, with nonzero entries on the diagonal, scales the rows of B.

4. **(T/F)** If $BC = BD$, then $C = D$.

5. **(T/F)** If $AC = 0$, then either $A = 0$ or $C = 0$.

6. **(T/F)** If A and B are $n \times n$, then $(A + B)(A - B) = A^2 - B^2$.

7. **(T/F)** An elementary $n \times n$ matrix has either n or $n + 1$ nonzero entries.

8. **(T/F)** The transpose of an elementary matrix is an elementary matrix.

9. **(T/F)** An elementary matrix must be square.

10. **(T/F)** Every square matrix is a product of elementary matrices.

11. **(T/F)** If A is a 3×3 matrix with three pivot positions, there exist elementary matrices $E_1, \dots, E_p$ such that $E_p \cdots E_1 A = I$.

12. **(T/F)** If $AB = I$, then A is invertible.

13. **(T/F)** If A and B are square and invertible, then AB is invertible, and $(AB)^{-1} = A^{-1}B^{-1}$.

14. **(T/F)** If $AB = BA$ and if A is invertible, then $A^{-1}B = BA^{-1}$.

15. **(T/F)** If A is invertible and if $r \neq 0$, then $(rA)^{-1} = rA^{-1}$.

16. Find the matrix C whose inverse is $C^{-1} = \begin{bmatrix} 4 & 5 \\ 6 & 7 \end{bmatrix}$.

17. Let $A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Show that $A^3 = 0$. Use matrix algebra to compute the product $(I - A)(I + A + A^2)$.

18. Suppose $A^n = 0$ for some $n > 1$. Find an inverse for $I - A$.

19. Suppose an $n \times n$ matrix A satisfies the equation $A^2 - 2A + I = 0$. Show that $A^3 = 3A - 2I$ and $A^4 = 4A - 3I$.

20. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. These are *Pauli spin matrices* used in the study of electron spin in quantum mechanics. Show that $A^2 = I$, $B^2 = I$, and $AB = -BA$. Matrices such that $AB = -BA$ are said to *anticommute*.

21. Let $A = \begin{bmatrix} 1 & 3 & 8 \\ 2 & 4 & 11 \\ 1 & 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} -3 & 5 \\ 1 & 5 \\ 3 & 4 \end{bmatrix}$. Compute $A^{-1}B$ without computing A^{-1}. [*Hint:* $A^{-1}B$ is the solution of the equation $AX = B$.]

22. Find a matrix A such that the transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 7 \end{bmatrix}$ into $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$, respectively. [*Hint:* Write a matrix equation involving A, and solve for A.]

23. Suppose $AB = \begin{bmatrix} 5 & 4 \\ -2 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}$. Find A.

24. Suppose A is invertible. Explain why A^TA is also invertible. Then show that $A^{-1} = (A^TA)^{-1}A^T$.

25. Let $x_1, \dots, x_n$ be fixed numbers. The matrix below, called a *Vandermonde matrix*, occurs in applications such as signal processing, error-correcting codes, and polynomial interpolation.

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

Given $\mathbf{y} = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, suppose $\mathbf{c} = (c_0, \dots, c_{n-1})$ in $\mathbb{R}^n$ satisfies $V\mathbf{c} = \mathbf{y}$, and define the polynomial

$$p(t) = c_0 + c_1t + c_2t^2 + \cdots + c_{n-1}t^{n-1}.$$

a. Show that $p(x_1) = y_1, \dots, p(x_n) = y_n$. We call $p(t)$ an *interpolating polynomial for the points* $(x_1, y_1), \dots, (x_n, y_n)$ because the graph of $p(t)$ passes through the points.

b. Suppose $x_1, \dots, x_n$ are distinct numbers. Show that the columns of V are linearly independent. [*Hint:* How many zeros can a polynomial of degree $n - 1$ have?]

c. Prove: "If $x_1, \dots, x_n$ are distinct numbers, and $y_1, \dots, y_n$ are arbitrary numbers, then there is an interpolating polynomial of degree $\le n - 1$ for $(x_1, y_1), \dots, (x_n, y_n)$."

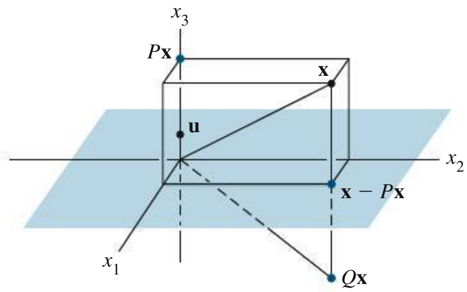
26. Let $A = LU$, where L is an invertible lower triangular matrix and U is upper triangular. Explain why the first column of A is a multiple of the first column of L. How is the second column of A related to the columns of L?

27. Given $\mathbf{u}$ in $\mathbb{R}^n$ with $\mathbf{u}^T\mathbf{u} = 1$, let $P = \mathbf{u}\mathbf{u}^T$ (an outer product) and $Q = I - 2P$. Justify statements (a), (b), and (c).

a. $P^2 = P$ b. $P^T = P$ c. $Q^2 = I$

The transformation $\mathbf{x} \mapsto P\mathbf{x}$ is called a *projection*, and $\mathbf{x} \mapsto Q\mathbf{x}$ is called a *Householder reflection*. Such reflections are used in computer programs to create multiple zeros in a vector (usually a column of a matrix).

28. Let $\mathbf{u} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix}$. Determine P and Q as in Exercise 27, and compute $P\mathbf{x}$ and $Q\mathbf{x}$. The figure shows that $Q\mathbf{x}$ is the reflection of $\mathbf{x}$ through the x_1x_2-plane.



A Householder reflection through the plane $x_3 = 0$.

29. Suppose $C = E_3E_2E_1B$, where E_1, E_2, and E_3 are elementary matrices. Explain why C is row equivalent to B.

30. Let A be an $n \times n$ singular matrix. Describe how to construct an $n \times n$ nonzero matrix B such that $AB = 0$.

31. Let A be a 6×4 matrix and B a 4×6 matrix. Show that the 6×6 matrix AB cannot be invertible.

32. Suppose A is a 5×3 matrix and there exists a 3×5 matrix C such that $CA = I_3$. Suppose further that for some given $\mathbf{b}$ in $\mathbb{R}^5$, the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution. Show that this solution is unique.

T 33. Certain dynamical systems can be studied by examining powers of a matrix, such as those below. Determine what happens to A^k and B^k as k increases (for example, try $k = 2, \dots, 16$). Try to identify what is special about A and B. Investigate large powers of other matrices of this type, and make a conjecture about such matrices.

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

T 34. Let A_n be the $n \times n$ matrix with 0's on the main diagonal and 1's elsewhere. Compute A_n^{-1} for $n = 4, 5$, and 6, and make a conjecture about the general form of A_n^{-1} for larger values of n.

3 Determinants

Introductory Example

WEIGHING DIAMONDS

How is the value of a diamond determined? Jewelers use the four cs: cut, clarity, color, and carats; a carat is a unit of mass equal to 0.2 grams. When a jeweler receives a supply of diamonds, it is vital that they be weighed accurately as part of determining their value. The difference of half a carat can have a large impact on a diamond's value.

When weighing small objects, such as diamonds or other gemstones, one strategy is to weigh the objects individually, but there are more accurate strategies that involve weighing the objects in groups and then deducing the individual weights from the results.

Suppose there are n small objects to be weighed, labeled $s_1, s_2, \cdots, s_n$. One method of determining the weight of each small object uses a two-pan balance. A *weighing* consists of placing some of the small objects in the left pan and the rest in the right pan. The balance records the difference between the weights in the pans.

The jeweler (or other individual weighing small light objects) plans her strategy in advance by creating a design matrix D with entries determined by the following scheme: If gemstone s_j is placed in the left pan during the ith weighing, then $d_{ij} = -1$ and if gemstone s_j is placed in the right pan during the ith weighing the $d_{ij} = 1$. Each row of the matrix D corresponds to a particular weighing. The jth column of D tells you where to put s_j at each weighing. Thus D is an $m \times n$ matrix, where m corresponds to the number of weighings and n corresponds to the number of objects. It has been shown that the accuracy of a weighing design is highest when a design matrix that maximizes the value of the *determinant* of $D^T D$ is chosen.

For example, consider the design matrix $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$ for weighing the gems s_1, s_2, s_3, s_4. For this design, the first weighing has all four gems in the right tray (the first row of D consists of all ones). For the second weighing, gem s_2 is in the left tray and the rest of the gems are in the right tray (the second row of D has a -1 in the second column). For the third weighing, gem s_3 is in the left tray and the rest of the gems are in the right tray (the third row of D has a -1 in the third column). In the last weighing, gem s_4 is in the left tray and the remaining gems are in the right tray (the fourth row of D has a -1 in the fourth column). The determinant of $D^T D$ is 64.

However, this is not the best design for using four weighings to determine the weight of four objects.

If $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix}$, then the determinant of $D^T D = 256$, and hence this is a better design. Notice that the first weighing of this design is the same as the previous one, but then the remaining weighings each have two objects in each pan.

Calculating determinants of matrices and understanding their properties is the theme of this chapter. As you learn more about determinants, you may also come up with strategies for good and bad choices for a weighing design.

Another important use of the determinant is to calculate the area of a parallelogram or the volume of a parallelepiped. In Section 1.9, we saw that matrix multiplication can be used to change the shape of a box or other object. The determinant of the matrix used determines how much the area changes when it is multiplied by a matrix, just as a fish story can transform the size of the fish caught.

Indeed, the determinant has so many uses that a summary of the applications known in the early 1900s filled a four-volume treatise by Thomas Muir. With changes in emphasis and the greatly increased sizes of the matrices used in modem applications, many uses that were important then are no longer critical today. Nevertheless, the determinant still plays many important theoretical and practical roles.

Beyond introducing the determinant in Section 3.1, this chapter presents two important ideas. Section 3.2 derives an invertibility criterion for a square matrix that plays a pivotal role in Chapter 5. Section 3.3 shows how the determinant measures the amount by which a linear transformation changes the area of a figure. When applied locally, this technique answers the question of a map's expansion rate near the poles. This idea plays a critical role in multivariable calculus in the form of the Jacobian.

3.1 Introduction to Determinants

Recall from Section 2.2 that a 2×2 matrix is invertible if and only if its determinant is nonzero. To extend this useful fact to larger matrices, we need a definition for the determinant of an $n \times n$ matrix. We can discover the definition for the 3×3 case by watching what happens when an invertible 3×3 matrix A is row reduced.

Consider $A = [a_{ij}]$ with $a_{11} \neq 0$. If we multiply the second and third rows of A by a_{11} and then subtract appropriate multiples of the first row from the other two rows, we find that A is row equivalent to the following two matrices:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{11}a_{21} & a_{11}a_{22} & a_{11}a_{23} \\ a_{11}a_{31} & a_{11}a_{32} & a_{11}a_{33} \end{bmatrix} \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & a_{11}a_{32} - a_{12}a_{31} & a_{11}a_{33} - a_{13}a_{31} \end{bmatrix} \tag{1}$$

Since A is invertible, either the $(2, 2)$-entry or the $(3, 2)$-entry on the right in (1) is nonzero. Let us suppose that the $(2, 2)$-entry is nonzero. (Otherwise, we can make a row interchange before proceeding.) Multiply row 3 by $a_{11}a_{22} - a_{12}a_{21}$, and then to the new row 3 add $-(a_{11}a_{32} - a_{12}a_{31})$ times row 2. This will show that

$$A \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{bmatrix}$$

where

$$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \qquad (2)$$

Since A is invertible, Δ must be nonzero. The converse is true, too, as we will see in Section 3.2. We call Δ in (2) the **determinant** of the 3×3 matrix A.

Recall that the determinant of a 2×2 matrix, $A = [a_{ij}]$, is the number

$$\det A = a_{11}a_{22} - a_{12}a_{21}$$

For a 1×1 matrix—say, $A = [a_{11}]$—we define $\det A = a_{11}$. To generalize the definition of the determinant to larger matrices, we'll use 2×2 determinants to rewrite the 3×3 determinant Δ described above. Since the terms in Δ can be grouped as $(a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32}) - (a_{12}a_{21}a_{33} - a_{12}a_{23}a_{31}) + (a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31})$,

$$\Delta = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

For brevity, write

$$\Delta = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} \qquad (3)$$

where A_{11}, A_{12}, and A_{13} are obtained from A by deleting the first row and one of the three columns. For any square matrix A, let A_{ij} denote the submatrix formed by deleting the ith row and jth column of A. For instance, if

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

then A_{32} is obtained by crossing out row 3 and column 2,

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

so that

$$A_{32} = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

We can now give a *recursive* definition of a determinant. When $n = 3$, $\det A$ is defined using determinants of the 2×2 submatrices A_{1j}, as in (3) above. When $n = 4$, $\det A$ uses determinants of the 3×3 submatrices A_{1j}. In general, an $n \times n$ determinant is defined by determinants of $(n-1) \times (n-1)$ submatrices.

DEFINITION

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is the sum of n terms of the form $\pm a_{1j} \det A_{1j}$, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, \dots, a_{1n}$ are from the first row of A. In symbols,

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \det A_{1n} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \end{aligned}$$

EXAMPLE 1 Compute the determinant of

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

SOLUTION Compute $\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}$:

$$\begin{aligned} \det A &= 1 \det \begin{bmatrix} 4 & -1 \\ -2 & 0 \end{bmatrix} - 5 \det \begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} + 0 \det \begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix} \\ &= 1(0-2) - 5(0-0) + 0(-4-0) = -2 \end{aligned}$$

■

Another common notation for the determinant of a matrix uses a pair of vertical lines in place of brackets. Thus the calculation in Example 1 can be written as

$$\det A = 1 \begin{vmatrix} 4 & -1 \\ -2 & 0 \end{vmatrix} - 5 \begin{vmatrix} 2 & -1 \\ 0 & 0 \end{vmatrix} + 0 \begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} = \cdots = -2$$

To state the next theorem, it is convenient to write the definition of $\det A$ in a slightly different form. Given $A = [a_{ij}]$, the **(i, j)-cofactor** of A is the number C_{ij} given by

$$C_{ij} = (-1)^{i+j} \det A_{ij} \tag{4}$$

Then

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

This formula is called a **cofactor expansion across the first row** of A. We omit the proof of the following fundamental theorem to avoid a lengthy digression.

THEOREM 1

The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion across any row or down any column. The expansion across the ith row using the cofactors in (4) is

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

The cofactor expansion down the jth column is

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

The plus or minus sign in the (i, j)-cofactor depends on the position of a_{ij} in the matrix, regardless of the sign of a_{ij} itself. The factor $(-1)^{i+j}$ determines the following checkerboard pattern of signs:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

EXAMPLE 2 Use a cofactor expansion across the third row to compute $\det A$, where

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

SOLUTION Compute

$$\begin{aligned}\det A &= a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} \\ &= (-1)^{3+1}a_{31}\det A_{31} + (-1)^{3+2}a_{32}\det A_{32} + (-1)^{3+3}a_{33}\det A_{33} \\ &= 0\begin{vmatrix} 5 & 0 \\ 4 & -1 \end{vmatrix} - (-2)\begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix} + 0\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix} \\ &= 0 + 2(-1) + 0 = -2\end{aligned}$$

Theorem 1 is helpful for computing the determinant of a matrix that contains many zeros. For example, if a row is mostly zeros, then the cofactor expansion across that row has many terms that are zero, and the cofactors in those terms need not be calculated. The same approach works with a column that contains many zeros.

EXAMPLE 3 Compute $\det A$, where

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

SOLUTION The cofactor expansion down the first column of A has all terms equal to zero except the first. Thus

$$\det A = 3\begin{vmatrix} 2 & -5 & 7 & 3 \\ 0 & 1 & 5 & 0 \\ 0 & 2 & 4 & -1 \\ 0 & 0 & -2 & 0 \end{vmatrix} + 0\,C_{21} + 0\,C_{31} + 0\,C_{41} + 0\,C_{51}$$

Henceforth we will omit the zero terms in the cofactor expansion. Next, expand this 4×4 determinant down the first column to take advantage of the zeros there. We have

$$\det A = 3(2)\begin{vmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{vmatrix}$$

This 3×3 determinant was computed in Example 1 and found to equal -2. Hence $\det A = 3(2)(-2) = -12$.

The matrix in Example 3 was nearly triangular. The method in that example is easily adapted to prove the following theorem.

THEOREM 2

If A is a triangular matrix, then $\det A$ is the product of the entries on the main diagonal of A.

The strategy in Example 3 of looking for zeros works extremely well when an entire row or column consists of zeros. In such a case, the cofactor expansion along such a row or column is a sum of zeros! So the determinant is zero. Unfortunately, most cofactor expansions are not so quickly evaluated.

Reasonable Answers

How big can a determinant be? Let A be an $n \times n$ matrix. Notice that taking the determinant of A consists of adding and subtracting terms with n products each. If p is the product of the n largest elements in absolute value (the same number may be repeated if it occurs more than once as a matrix entry), then the determinant must be between $-np$ and np. For example, consider $A = \begin{bmatrix} 6 & 5 \\ -7 & 9 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 6 \\ 7 & -9 \end{bmatrix}$. The largest number in absolute value of each matrix is 9, and the second largest number is 7. In these two cases, $p = 7(9) = 63$ and $np = 126$. The determinant of each of these matrices should be a number between -126 and 126. Notice that $\det A = 6(9) - 5(-7) = 54 + 35 = 89$, $\det B = 7(-9) - 6(7) = -63 - 42 = -105$, illustrating that because the products are added and subtracted, any number in the range between -126 and 126 could turn out to be the determinant.

Next, consider $C = \begin{bmatrix} 7 & 9 \\ 7 & 9 \end{bmatrix}$ and $D = \begin{bmatrix} -9 & 9 \\ 9 & 9 \end{bmatrix}$. In matrices C and D, the number 9 appears twice and so should be selected twice. In this case, $p = 9(9) = 81$ and $np = 162$, so the determinants of C and D should be numbers between -162 and 162. Indeed, $\det C = (7)(9) - (7)(9) = 0$ and $\det D = (-9)(9) - (9)(9) = -162$. Notice that it is important to choose 9 twice as the two largest numbers in matrix D in order to get the correct bounds for the determinant of D.

Numerical Note

By today's standards, a 25×25 matrix is small. Yet it would be impossible to calculate a 25×25 determinant by cofactor expansion. In general, a cofactor expansion requires more than $n!$ multiplications, and 25! is approximately 1.55×10^{25}.

If a computer performs one trillion multiplications per second, it would have to run for almost 500,000 years to compute a 25×25 determinant by this method. Fortunately, there are faster methods, as we'll soon discover.

Exercises 19–38 explore important properties of determinants, mostly for the 2×2 case. The results from Exercises 33–36 will be used in the next section to derive the analogous properties for $n \times n$ matrices.

Practice Problem

Compute $\begin{vmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{vmatrix}$.

3.1 Exercises

Compute the determinants in Exercises 1–8 using a cofactor expansion across the first row. In Exercises 1–4, also compute the determinant by a cofactor expansion down the second column.

1. $\begin{vmatrix} 3 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -1 \end{vmatrix}$ **2.** $\begin{vmatrix} 0 & 4 & 1 \\ 5 & -3 & 0 \\ 2 & 4 & 1 \end{vmatrix}$

3. $\begin{vmatrix} 2 & -2 & 3 \\ 3 & 1 & 2 \\ 1 & 3 & -1 \end{vmatrix}$ **4.** $\begin{vmatrix} 1 & 2 & 4 \\ 3 & 1 & 1 \\ 2 & 4 & 2 \end{vmatrix}$

5. $\begin{vmatrix} 2 & 3 & -3 \\ 4 & 0 & 3 \\ 6 & 1 & 5 \end{vmatrix}$ **6.** $\begin{vmatrix} 5 & -2 & 3 \\ 0 & 3 & -3 \\ 2 & -4 & 7 \end{vmatrix}$

7. $\begin{vmatrix} 4 & 3 & 0 \\ 6 & 5 & 2 \\ 9 & 7 & 3 \end{vmatrix}$ **8.** $\begin{vmatrix} 4 & 1 & 2 \\ 4 & 0 & 3 \\ 3 & -2 & 5 \end{vmatrix}$

Compute the determinants in Exercises 9–14 by cofactor expansions. At each step, choose a row or column that involves the least amount of computation.

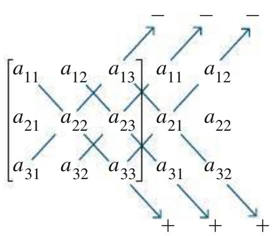
9. $\begin{vmatrix} 4 & 0 & 0 & 5 \\ 1 & 7 & 2 & -5 \\ 3 & 0 & 0 & 0 \\ 8 & 3 & 1 & 7 \end{vmatrix}$ **10.** $\begin{vmatrix} 1 & -2 & 4 & 2 \\ 0 & 0 & 3 & 0 \\ 2 & -4 & -3 & 5 \\ 2 & 0 & 3 & 5 \end{vmatrix}$

11. $\begin{vmatrix} 3 & 5 & -6 & 4 \\ 0 & -2 & 3 & -3 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 3 \end{vmatrix}$ **12.** $\begin{vmatrix} 3 & 0 & 0 & 0 \\ 7 & -2 & 0 & 0 \\ 2 & 6 & 3 & 0 \\ 3 & -8 & 4 & -3 \end{vmatrix}$

13. $\begin{vmatrix} 4 & 0 & -7 & 3 & -5 \\ 0 & 0 & 2 & 0 & 0 \\ 7 & 3 & -6 & 4 & -8 \\ 5 & 0 & 5 & 2 & -3 \\ 0 & 0 & 9 & -1 & 2 \end{vmatrix}$

14. $\begin{vmatrix} 6 & 0 & 2 & 4 & 0 \\ 9 & 0 & -4 & 1 & 0 \\ 8 & -5 & 6 & 7 & 1 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 2 & 3 & 2 & 0 \end{vmatrix}$

The expansion of a 3×3 determinant can be remembered by the following device. Write a second copy of the first two columns to the right of the matrix, and compute the determinant by multiplying entries on six diagonals:



Add the downward diagonal products and subtract the upward products. Use this method to compute the determinants in Exercises 15–18. ***Warning:*** *This trick does not generalize in any reasonable way to* 4×4 *or larger matrices.*

15. $\begin{vmatrix} 1 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -2 \end{vmatrix}$ **16.** $\begin{vmatrix} 0 & 3 & 1 \\ 4 & -5 & 0 \\ 3 & 4 & 1 \end{vmatrix}$

17. $\begin{vmatrix} 2 & -3 & 3 \\ 3 & 2 & 2 \\ 1 & 3 & -1 \end{vmatrix}$ **18.** $\begin{vmatrix} 1 & 3 & 4 \\ 2 & 3 & 2 \\ 3 & 3 & 2 \end{vmatrix}$

In Exercises 19–24, explore the effect of an elementary row operation on the determinant of a matrix. In each case, state the row operation and describe how it affects the determinant.

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

20. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} a & b \\ kc & kd \end{bmatrix}$

21. $\begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ 5+3k & 4+2k \end{bmatrix}$

22. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} a+kc & b+kd \\ c & d \end{bmatrix}$

23. $\begin{bmatrix} a & b & c \\ 3 & 2 & 1 \\ 4 & 5 & 6 \end{bmatrix}, \begin{bmatrix} 3 & 2 & 1 \\ a & b & c \\ 4 & 5 & 6 \end{bmatrix}$

24. $\begin{bmatrix} 1 & 0 & 1 \\ -3 & 4 & -4 \\ 2 & -3 & 1 \end{bmatrix}, \begin{bmatrix} k & 0 & k \\ -3 & 4 & -4 \\ 2 & -3 & 1 \end{bmatrix}$

Compute the determinants of the elementary matrices given in Exercises 25–30. (See Section 2.2, Examples 5 and 6.)

25. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{bmatrix}$ **26.** $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

27. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix}$ **28.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

29. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **30.** $\begin{bmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Use Exercises 25–30 to answer the questions in Exercises 31 and 32. Give reasons for your answers.

31. What is the determinant of an elementary row replacement matrix?

32. What is the determinant of an elementary scaling matrix with k on the diagonal?

In Exercises 33–36, verify that $\det EA = (\det E)(\det A)$, where E is the elementary matrix shown and $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

33. $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ **34.** $\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

35. $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ **36.** $\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

37. Let $A = \begin{bmatrix} 3 & 1 \\ 4 & 2 \end{bmatrix}$. Write $5A$. Is $\det 5A = 5 \det A$?

38. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and let k be a scalar. Find a formula that relates $\det kA$ to k and $\det A$.

In Exercises 39 through 42, A is an $n \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

39. **(T/F)** An $n \times n$ determinant is defined by determinants of $(n-1) \times (n-1)$ submatrices.

40. **(T/F)** The (i, j)-cofactor of a matrix A is the matrix A_{ij} obtained by deleting from A its ith row and jth column.

41. **(T/F)** The cofactor expansion of $\det A$ down a column is equal to the cofactor expansion along a row.

42. **(T/F)** The determinant of a triangular matrix is the sum of the entries on the main diagonal.

43. Let $\mathbf{u} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Compute the area of the parallelogram determined by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, and $\mathbf{0}$, and compute the determinant of $[\,\mathbf{u} \;\; \mathbf{v}\,]$. How do they compare? Replace the first entry of $\mathbf{v}$ by an arbitrary number x, and repeat the problem. Draw a picture and explain what you find.

44. Let $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} c \\ 0 \end{bmatrix}$, where a, b, and c are positive (for simplicity). Compute the area of the parallelogram determined by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, and $\mathbf{0}$, and compute the determinants of the matrices $[\,\mathbf{u} \;\; \mathbf{v}\,]$ and $[\,\mathbf{v} \;\; \mathbf{u}\,]$. Draw a picture and explain what you find.

45. Let A be a 2×2 matrix all of whose entries are numbers that are greater than or equal to -10 and less than or equal to 10. Decide if each of the following is a reasonable answer for $\det A$.

a. 0

b. 202

c. -110

d. 555

46. Let A be a 3×3 matrix all of whose entries are numbers that are greater than or equal to -5 and less than or equal to 5. Decide if each of the following is a reasonable answer for $\det A$.

a. 300

b. -220

c. 1000

d. 10

T 47. Construct a random 4×4 matrix A with integer entries between -9 and 9. How is $\det A^{-1}$ related to $\det A$? Experiment with random $n \times n$ integer matrices for $n = 4$, 5, and 6, and make a conjecture. *Note:* In the unlikely event that you encounter a matrix with a zero determinant, reduce it to echelon form and discuss what you find.

T 48. Is it true that $\det AB = (\det A)(\det B)$? To find out, generate random 5×5 matrices A and B, and compute $\det AB - (\det A \det B)$. Repeat the calculations for three other pairs of $n \times n$ matrices, for various values of n. Report your results.

T 49. Is it true that $\det(A + B) = \det A + \det B$? Experiment with four pairs of random matrices as in Exercise 48, and make a conjecture.

T 50. Construct a random 4×4 matrix A with integer entries between -9 and 9, and compare $\det A$ with $\det A^T$, $\det(-A)$, $\det(2A)$, and $\det(10A)$. Repeat with two other random 4×4 integer matrices, and make conjectures about how these determinants are related. (Refer to Exercise 44 in Section 2.1.) Then check your conjectures with several random 5×5 and 6×6 integer matrices. Modify your conjectures, if necessary, and report your results.

T 51. Recall from the introductory section that the larger the determinant of $D^T D$, where D is the design matrix, the better will be the accuracy of the calculated weights for small light objects. Which of the following matrices corresponds to the best design for four weighings of four objects? Describe which of the objects s_1, s_2, s_3, and s_4 you would put in the left and right pans for each weighing corresponding to the best design matrix.

a. $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$

b. $D = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$

c. $D = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$

T 52. Repeat Exercise 51 for the case of five weighings of four objects and the following design matrices.

a. $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$

b. $D = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$

c. $D = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}$

Solution to Practice Problem

Take advantage of the zeros. Begin with a cofactor expansion down the third column to obtain a 3×3 matrix, which may be evaluated by an expansion down its first column.

$$\begin{vmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{vmatrix} = (-1)^{1+3}(2)\begin{vmatrix} 0 & 3 & -4 \\ -5 & -8 & 3 \\ 0 & 5 & -6 \end{vmatrix}$$

$$= 2\,(-1)^{2+1}(-5)\begin{vmatrix} 3 & -4 \\ 5 & -6 \end{vmatrix} = 20$$

The $(-1)^{2+1}$ in the next-to-last calculation came from the $(2, 1)$-position of the -5 in the 3×3 determinant.

3.2 Properties of Determinants

The secret of determinants lies in how they change when row operations are performed. The following theorem generalizes the results of Exercises 19–24 in Section 3.1. The proof is at the end of this section.

THEOREM 3 **Row Operations**

Let A be a square matrix.

a. If a multiple of one row of A is added to another row to produce a matrix B, then $\det B = \det A$.

b. If two rows of A are interchanged to produce B, then $\det B = -\det A$.

c. If one row of A is multiplied by k to produce B, then $\det B = k \det A$.

The following examples show how to use Theorem 3 to find determinants efficiently.

EXAMPLE 1 Compute $\det A$, where $A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$.

SOLUTION The strategy is to reduce A to echelon form and then to use the fact that the determinant of a triangular matrix is the product of the diagonal entries. The first two row replacements in column 1 do not change the determinant:

$$\det A = \begin{vmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{vmatrix}$$

An interchange of rows 2 and 3 reverses the sign of the determinant, so

$$\det A = -\begin{vmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{vmatrix} = -(1)(3)(-5) = 15$$

A common use of Theorem 3(c) in hand calculations is to *factor out a common multiple of one row* of a matrix. For instance,

$$\begin{vmatrix} * & * & * \\ 5k & -2k & 3k \\ * & * & * \end{vmatrix} = k\begin{vmatrix} * & * & * \\ 5 & -2 & 3 \\ * & * & * \end{vmatrix}$$

where the starred entries are unchanged. We use this step in the next example.

EXAMPLE 2 Compute $\det A$, where $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$.

SOLUTION To simplify the arithmetic, we want a 1 in the upper-left corner. We could interchange rows 1 and 4. Instead, we factor out 2 from the top row, and then proceed with row replacements in the first column:

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{vmatrix} = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & -12 & 10 & 10 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

Next, we could factor out another 2 from row 3 or use the 3 in the second column as a pivot. We choose the latter operation, adding 4 times row 2 to row 3:

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

Finally, adding $-1/2$ times row 3 to row 4, and computing the "triangular" determinant, we find that

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 2\,(1)(3)(-6)(1) = -36$$

$$U = \begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & \blacksquare \end{bmatrix}$$
$\det U \neq 0$

$$U = \begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$\det U = 0$

FIGURE 1
Typical echelon forms of square matrices.

Suppose a square matrix A has been reduced to an echelon form U by row replacements and row interchanges. (This is always possible. See the row reduction algorithm in Section 1.2.) If there are r interchanges, then Theorem 3 shows that

$$\det A = (-1)^r \det U$$

Since U is in echelon form, it is triangular, and so $\det U$ is the product of the diagonal entries $u_{11}, \ldots, u_{nn}$. If A is invertible, the entries u_{ii} are all pivots (because $A \sim I_n$ and the u_{ii} have not been scaled to 1's). Otherwise, at least u_{nn} is zero, and the product $u_{11} \cdots u_{nn}$ is zero. See Figure 1. Thus

$$\det A = \begin{cases} (-1)^r \cdot \begin{pmatrix} \text{product of} \\ \text{pivots in } U \end{pmatrix} & \text{when } A \text{ is invertible} \\ 0 & \text{when } A \text{ is not invertible} \end{cases} \tag{1}$$

It is interesting to note that although the echelon form U described above is not unique (because it is not completely row reduced), and the pivots are not unique, the *product* of the pivots *is* unique, except for a possible minus sign.

Formula (1) not only gives a concrete interpretation of $\det A$ but also proves the main theorem of this section:

THEOREM 4

A square matrix A is invertible if and only if $\det A \neq 0$.

Theorem 4 adds the statement "$\det A \neq 0$" to the Invertible Matrix Theorem. A useful corollary is that $\det A = 0$ when the columns of A are linearly dependent. Also, $\det A = 0$ when the *rows* of A are linearly dependent. (Rows of A are columns of A^T, and linearly dependent columns of A^T make A^T singular. When A^T is singular, so is A, by the Invertible Matrix Theorem.) In practice, linear dependence is obvious when two columns or two rows are the same or a column or a row is zero.

EXAMPLE 3 Compute $\det A$, where $A = \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{bmatrix}$.

SOLUTION Add 2 times row 1 to row 3 to obtain

$$\det A = \det \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 5 & -3 & -6 \\ -5 & -8 & 0 & 9 \end{bmatrix} = 0$$

because the second and third rows of the second matrix are equal. ■

Numerical Notes

1. Most computer programs that compute $\det A$ for a general matrix A use the method of formula (1) above.
2. It can be shown that evaluation of an $n \times n$ determinant using row operations requires about $2n^3/3$ arithmetic operations. Any modern microcomputer can calculate a 25×25 determinant in a fraction of a second, since only about 10,000 operations are required.

Computers can also handle large "sparse" matrices, with special routines that take advantage of the presence of many zeros. Of course, zero entries can speed hand computations, too. The calculations in the next example combine the power of row operations with the strategy from Section 3.1 of using zero entries in cofactor expansions.

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

SOLUTION A good way to begin is to use the 2 in column 1 as a pivot, eliminating the -2 below it. Then use a cofactor expansion to reduce the size of the determinant, followed by another row replacement operation. Thus

$$\det A = \begin{vmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ 0 & 0 & -3 & 1 \end{vmatrix} = -2\begin{vmatrix} 1 & 2 & -1 \\ 3 & 6 & 2 \\ 0 & -3 & 1 \end{vmatrix} = -2\begin{vmatrix} 1 & 2 & -1 \\ 0 & 0 & 5 \\ 0 & -3 & 1 \end{vmatrix}$$

An interchange of rows 2 and 3 would produce a "triangular determinant." Another approach is to make a cofactor expansion down the first column:

$$\det A = (-2)(1)\begin{vmatrix} 0 & 5 \\ -3 & 1 \end{vmatrix} = -2\,(15) = -30$$

Column Operations

We can perform operations on the columns of a matrix in a way that is analogous to the row operations we have considered. The next theorem shows that column operations have the same effects on determinants as row operations.

Remark: The Principle of Mathematical Induction says the following: Let $P(n)$ be a statement that is either true or false for each natural number n. Then $P(n)$ is true for all $n \geq 1$ provided that $P(1)$ is true, and for each natural number k, if $P(k)$ is true, then $P(k+1)$ is true. The Principle of Mathematical Induction is used to prove the next theorem.

THEOREM 5

If A is an $n \times n$ matrix, then $\det A^T = \det A$.

PROOF The theorem is obvious for $n = 1$. Suppose the theorem is true for $k \times k$ determinants and let $n = k + 1$. Then the cofactor of a_{1j} in A equals the cofactor of a_{j1} in A^T, because the cofactors involve $k \times k$ determinants. Hence the cofactor expansion of $\det A$ along the first *row* equals the cofactor expansion of $\det A^T$ down the first *column*. That is, A and A^T have equal determinants. The theorem is true for $n = 1$, and the truth of the theorem for one value of n implies its truth for the next value of n. By the Principle of Mathematical Induction, the theorem is true for all $n \geq 1$.

Because of Theorem 5, each statement in Theorem 3 is true when the word *row* is replaced everywhere by *column*. To verify this property, one merely applies the original Theorem 3 to A^T. A row operation on A^T amounts to a column operation on A.

Column operations are useful for both theoretical purposes and hand computations. However, for simplicity we'll perform only row operations in numerical calculations.

Determinants and Matrix Products

The proof of the following useful theorem is at the end of the section. Applications are in the exercises.

THEOREM 6 **Multiplicative Property**

If A and B are $n \times n$ matrices, then $\det AB = (\det A)(\det B)$.

EXAMPLE 5 Verify Theorem 6 for $A = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$.

SOLUTION

$$AB = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 14 & 13 \end{bmatrix}$$

and

$$\det AB = 25(13) - 20(14) = 325 - 280 = 45$$

Since $\det A = 9$ and $\det B = 5$,

$$(\det A)(\det B) = 9(5) = 45 = \det AB$$

■

Warning: A common misconception is that Theorem 6 has an analogue for *sums* of matrices. However, $\det(A + B)$ is *not* equal to $\det A + \det B$, in general.

A Linearity Property of the Determinant Function

For an $n \times n$ matrix A, we can consider $\det A$ as a function of the n column vectors in A. We will show that if all columns except one are held fixed, then $\det A$ is a *linear function* of that one (vector) variable.

Suppose that the jth column of A is allowed to vary, and write

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

Define a transformation T from $\mathbb{R}^n$ to $\mathbb{R}$ by

$$T(\mathbf{x}) = \det\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

Then,

$$T(c\mathbf{x}) = cT(\mathbf{x}) \quad \text{for all scalars } c \text{ and all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{2}$$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{for all } \mathbf{u}, \mathbf{v} \text{ in } \mathbb{R}^n \tag{3}$$

Property (2) is Theorem 3(c) applied to the columns of A. A proof of property (3) follows from a cofactor expansion of $\det A$ down the jth column. (See Exercise 49.) This (multi-) linearity property of the determinant turns out to have many useful consequences that are studied in more advanced courses.

Proofs of Theorems 3 and 6

It is convenient to prove Theorem 3 when it is stated in terms of the elementary matrices discussed in Section 2.2. We call an elementary matrix E a *row replacement* (*matrix*) if E is obtained from the identity I by adding a multiple of one row to another row; E is an *interchange* if E is obtained by interchanging two rows of I; and E is *a scale by* r if E is obtained by multiplying a row of I by a nonzero scalar r. With this terminology, Theorem 3 can be reformulated as follows:

If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then

$$\det EA = (\det E)(\det A)$$

where

$$\det E = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r \end{cases}$$

PROOF OF THEOREM 3 The proof is by induction on the size of A. The case of a 2×2 matrix was verified in Exercises 33–36 of Section 3.1. Suppose the theorem has been verified for determinants of $k \times k$ matrices with $k \geq 2$, let $n = k + 1$, and let A be $n \times n$. The action of E on A involves either two rows or only one row. So we can expand $\det EA$ across a row that is unchanged by the action of E, say, row i. Let A_{ij} (respectively, B_{ij}) be the matrix obtained by deleting row i and column j from A (respectively, EA). Then the rows of B_{ij} are obtained from the rows of A_{ij} by the same type of elementary row operation that E performs on A. Since these submatrices are only $k \times k$, the induction assumption implies that

$$\det B_{ij} = \alpha \det A_{ij}$$

where $\alpha = 1, -1$, or r, depending on the nature of E. The cofactor expansion across row i is

$$\begin{aligned} \det EA &= a_{i1}(-1)^{i+1} \det B_{i1} + \cdots + a_{in}(-1)^{i+n} \det B_{in} \\ &= \alpha a_{i1}(-1)^{i+1} \det A_{i1} + \cdots + \alpha a_{in}(-1)^{i+n} \det A_{in} \\ &= \alpha \det A \end{aligned}$$

In particular, taking $A = I_n$, we see that $\det E = 1, -1$, or r, depending on the nature of E. Thus the theorem is true for $n = 2$, and the truth of the theorem for one value of n implies its truth for the next value of n. By the principle of induction, the theorem must be true for $n \geq 2$. The theorem is trivially true for $n = 1$. ■

PROOF OF THEOREM 6 If A is not invertible, then neither is AB, by Exercise 35 in Section 2.3. In this case, $\det AB = (\det A)(\det B)$, because both sides are zero, by Theorem 4. If A is invertible, then A and the identity matrix I_n are row equivalent by the Invertible Matrix Theorem. So there exist elementary matrices $E_1, \ldots, E_p$ such that

$$A = E_p E_{p-1} \cdots E_1 I_n = E_p E_{p-1} \cdots E_1$$

For brevity, write $|A|$ for det A. Then repeated application of Theorem 3, as rephrased above, shows that

$$\begin{aligned}|AB| &= |E_p \cdots E_1 B| = |E_p||E_{p-1} \cdots E_1 B| = \cdots \\ &= |E_p| \cdots |E_1||B| = \cdots = |E_p \cdots E_1||B| \\ &= |A||B|\end{aligned}$$ ■

Practice Problems

1. Compute $\begin{vmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 8 \end{vmatrix}$ in as few steps as possible.
2. Use a determinant to decide if $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent, when

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ -7 \\ 9 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 3 \\ -5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ -7 \\ 5 \end{bmatrix}$$

3. Let A be an $n \times n$ matrix such that $A^2 = I$. Show that det $A = \pm 1$.

3.2 Exercises

Each equation in Exercises 1–4 illustrates a property of determinants. State the property.

1. $\begin{vmatrix} 0 & 5 & -2 \\ 1 & -3 & 6 \\ 4 & -1 & 8 \end{vmatrix} = -\begin{vmatrix} 1 & -3 & 6 \\ 0 & 5 & -2 \\ 4 & -1 & 8 \end{vmatrix}$

2. $\begin{vmatrix} 3 & -6 & 9 \\ 3 & 5 & -5 \\ 1 & 3 & 3 \end{vmatrix} = 3\begin{vmatrix} 1 & -2 & 3 \\ 3 & 5 & -5 \\ 1 & 3 & 3 \end{vmatrix}$

3. $\begin{vmatrix} 1 & 2 & 2 \\ 0 & 3 & -4 \\ 2 & 7 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 2 \\ 0 & 3 & -4 \\ 0 & 3 & 0 \end{vmatrix}$

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

Find the determinants in Exercises 5–10 by row reduction to echelon form.

5. $\begin{vmatrix} 1 & 5 & -4 \\ -1 & -4 & 5 \\ -2 & -8 & 7 \end{vmatrix}$

6. $\begin{vmatrix} 2 & 2 & -2 \\ 3 & 4 & -4 \\ 2 & -3 & -5 \end{vmatrix}$

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

8. $\begin{vmatrix} 1 & 3 & 2 & -4 \\ 0 & 1 & 2 & -5 \\ 2 & 7 & 6 & -3 \\ -3 & -10 & -7 & 2 \end{vmatrix}$

9. $\begin{vmatrix} 1 & -1 & -3 & 0 \\ 0 & 1 & 5 & 4 \\ -1 & 0 & 5 & 3 \\ 3 & -3 & -2 & 3 \end{vmatrix}$

10. $\begin{vmatrix} 1 & 3 & -1 & 0 & -2 \\ 0 & 1 & -2 & -1 & -3 \\ -2 & -6 & 2 & 3 & 10 \\ 1 & 5 & -6 & 2 & -3 \\ 0 & 2 & -4 & 5 & 9 \end{vmatrix}$

Combine the methods of row reduction and cofactor expansion to compute the determinants in Exercises 11–14.

11. $\begin{vmatrix} 3 & 4 & -3 & -1 \\ 3 & 0 & 1 & -3 \\ -6 & 0 & -4 & 3 \\ 6 & 8 & -4 & -1 \end{vmatrix}$

12. $\begin{vmatrix} -1 & 2 & 3 & 0 \\ 3 & 4 & 3 & 0 \\ 11 & 4 & 6 & 6 \\ 4 & 2 & 4 & 3 \end{vmatrix}$

13. $\begin{vmatrix} 2 & 5 & 4 & 1 \\ 4 & 7 & 6 & 2 \\ 6 & -2 & -4 & 0 \\ -6 & 7 & 7 & 0 \end{vmatrix}$

14. $\begin{vmatrix} 1 & 5 & 4 & 1 \\ 0 & -3 & -6 & 0 \\ 3 & 5 & 4 & 1 \\ -6 & 5 & 5 & 0 \end{vmatrix}$

Find the determinants in Exercises 15–20, where

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 7.$$

15. $\begin{vmatrix} a & b & c \\ d & e & f \\ 3g & 3h & 3i \end{vmatrix}$ **16.** $\begin{vmatrix} a & b & c \\ d+3g & e+3h & f+3i \\ g & h & i \end{vmatrix}$

17. $\begin{vmatrix} a+d & b+e & c+f \\ d & e & f \\ g & h & i \end{vmatrix}$

18. $\begin{vmatrix} a & b & c \\ 5d & 5e & 5f \\ g & h & i \end{vmatrix}$

19. $\begin{vmatrix} a & b & c \\ 2d+a & 2e+b & 2f+c \\ g & h & i \end{vmatrix}$

20. $\begin{vmatrix} d & e & f \\ a & b & c \\ g & h & i \end{vmatrix}$

In Exercises 21–23, use determinants to find out if the matrix is invertible.

21. $\begin{bmatrix} 2 & 6 & 0 \\ 1 & 3 & 2 \\ 3 & 9 & 2 \end{bmatrix}$ **22.** $\begin{bmatrix} 5 & 1 & -1 \\ 1 & -3 & -2 \\ 0 & 5 & 3 \end{bmatrix}$

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

In Exercises 24–26, use determinants to decide if the set of vectors is linearly independent.

24. $\begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} -6 \\ 0 \\ 6 \end{bmatrix}, \begin{bmatrix} -3 \\ -5 \\ -2 \end{bmatrix}$

25. $\begin{bmatrix} 7 \\ -4 \\ -6 \end{bmatrix}, \begin{bmatrix} -8 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \\ -5 \end{bmatrix}$

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

In Exercises 27–34, A and B are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer.

27. **(T/F)** A row replacement operation does not affect the determinant of a matrix.

28. **(T/F)** If $\det A$ is zero, then two rows or two columns are the same, or a row or a column is zero.

29. **(T/F)** If the columns of A are linearly dependent, then $\det A = 0$.

30. **(T/F)** The determinant of A is the product of the diagonal entries in A.

31. **(T/F)** If three row interchanges are made in succession, then the new determinant equals the old determinant.

32. **(T/F)** The determinant of A is the product of the pivots in any echelon form U of A, multiplied by $(-1)^r$, where r is the number of row interchanges made during row reduction from A to U.

33. **(T/F)** $\det(A+B) = \det A + \det B$.

34. **(T/F)** $\det A^{-1} = (-1)\det A$.

35. Compute $\det B^4$, where $B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.

36. Use Theorem 3 (but not Theorem 4) to show that if two rows of a square matrix A are equal, then $\det A = 0$. The same is true for two columns. Why?

In Exercises 37–42, mention an appropriate theorem in your explanation.

37. Show that if A is invertible, then $\det A^{-1} = \dfrac{1}{\det A}$.

38. Suppose that A is a square matrix such that $\det A^3 = 0$. Explain why A cannot be invertible.

39. Let A and B be square matrices. Show that even though AB and BA may not be equal, it is always true that $\det AB = \det BA$.

40. Let A and P be square matrices, with P invertible. Show that $\det(PAP^{-1}) = \det A$.

41. Let U be a square matrix such that $U^TU = I$. Show that $\det U = \pm 1$.

42. Find a formula for $\det(rA)$ when A is an $n \times n$ matrix.

Verify that $\det AB = (\det A)(\det B)$ for the matrices in Exercises 43 and 44. (Do not use Theorem 6.)

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

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

45. Let A and B be 3×3 matrices, with $\det A = -2$ and $\det B = 3$. Use properties of determinants (in the text and in the preceding exercises) to compute:

a. $\det AB$ b. $\det 5A$ c. $\det B^T$
d. $\det A^{-1}$ e. $\det A^3$

46. Let A and B be 4×4 matrices, with $\det A = 4$ and $\det B = -3$. Compute:

a. $\det AB$ b. $\det B^5$ c. $\det 2A$
d. $\det A^TBA$ e. $\det B^{-1}AB$

47. Verify that $\det A = \det B + \det C$, where

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

48. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Show that $\det(A+B) = \det A + \det B$ if and only if $a + d = 0$.

49. Verify that $\det A = \det B + \det C$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & u_1 + v_1 \\ a_{21} & a_{22} & u_2 + v_2 \\ a_{31} & a_{32} & u_3 + v_3 \end{bmatrix},$$

$$B = \begin{bmatrix} a_{11} & a_{12} & u_1 \\ a_{21} & a_{22} & u_2 \\ a_{31} & a_{32} & u_3 \end{bmatrix},\ C = \begin{bmatrix} a_{11} & a_{12} & v_1 \\ a_{21} & a_{22} & v_2 \\ a_{31} & a_{32} & v_3 \end{bmatrix}$$

Note, however, that A is *not* the same as $B + C$.

50. Right-multiplication by an elementary matrix E affects the *columns* of A in the same way that left-multiplication affects the *rows*. Use Theorems 5 and 3 and the obvious fact that E^T is another elementary matrix to show that

$$\det AE = (\det E)(\det A)$$

Do not use Theorem 6.

51. Suppose A is an $n \times n$ matrix and a computer suggests that $\det A = 5$ and $\det\left(A^{-1}\right) = 1$. Should you trust these answers? Why or why not?

52. Suppose A and B are $n \times n$ matrices and a computer suggests that $\det A = 5$, $\det B = 2$ and $\det AB = 7$. Should you trust these answers? Why or why not?

T 53. Compute $\det A^T A$ and $\det AA^T$ for several random 4×5 matrices and several random 5×6 matrices. What can you say about $A^T A$ and AA^T when A has more columns than rows?

T 54. If $\det A$ is close to zero, is the matrix A nearly singular? Experiment with the nearly singular 4×4 matrix

$$A = \begin{bmatrix} 4 & 0 & -7 & -7 \\ -6 & 1 & 11 & 9 \\ 7 & -5 & 10 & 19 \\ -1 & 2 & 3 & -1 \end{bmatrix}$$

Compute the determinants of A, $10A$, and $0.1A$. In contrast, compute the condition numbers of these matrices. Repeat these calculations when A is the 4×4 identity matrix. Discuss your results.

Solutions to Practice Problems

1. Perform row replacements to create zeros in the first column, and then create a row of zeros.

$$\begin{vmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 8 \end{vmatrix} = \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & -4 & 5 & 1 \\ 0 & 1 & -3 & 2 \end{vmatrix} = \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & -4 & 5 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix} = 0$$

2. $\det[\,\mathbf{v}_1 \ \ \mathbf{v}_2 \ \ \mathbf{v}_3\,] = \begin{vmatrix} 5 & -3 & 2 \\ -7 & 3 & -7 \\ 9 & -5 & 5 \end{vmatrix} = \begin{vmatrix} 5 & -3 & 2 \\ -2 & 0 & -5 \\ 9 & -5 & 5 \end{vmatrix}$ Row 1 added to row 2

$= -(-3)\begin{vmatrix} -2 & -5 \\ 9 & 5 \end{vmatrix} - (-5)\begin{vmatrix} 5 & 2 \\ -2 & -5 \end{vmatrix}$ Cofactors of column 2

$= 3\,(35) + 5\,(-21) = 0$

By Theorem 4, the matrix $[\,\mathbf{v}_1 \ \ \mathbf{v}_2 \ \ \mathbf{v}_3\,]$ is not invertible. The columns are linearly dependent, by the Invertible Matrix Theorem.

3. Recall that $\det I = 1$. By Theorem 6, $\det(AA) = (\det A)(\det A)$. Putting these two observations together results in

$$1 = \det I = \det A^2 = \det(AA) = (\det A)(\det A) = (\det A)^2$$

Taking the square root of both sides establishes that $\det A = \pm 1$.

3.3 Cramer's Rule, Volume, and Linear Transformations

This section applies the theory of the preceding sections to obtain important theoretical formulas and a geometric interpretation of the determinant.

Cramer's Rule

Cramer's rule is needed in a variety of theoretical calculations. For instance, it can be used to study how the solution of $A\mathbf{x} = \mathbf{b}$ is affected by changes in the entries of $\mathbf{b}$. However, the formula is inefficient for hand calculations, except for 2×2 or perhaps 3×3 matrices.

For any $n \times n$ matrix A and any $\mathbf{b}$ in $\mathbb{R}^n$, let $A_i(\mathbf{b})$ be the matrix obtained from A by replacing column i by the vector $\mathbf{b}$.

$$A_i(\mathbf{b}) = [\mathbf{a}_1 \quad \cdots \quad \underset{\substack{\uparrow \\ \text{col } i}}{\mathbf{b}} \quad \cdots \quad \mathbf{a}_n]$$

THEOREM 7

Cramer's Rule

Let A be an invertible $n \times n$ matrix. For any $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \qquad i = 1, 2, \ldots, n \tag{1}$$

PROOF Denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and the columns of the $n \times n$ identity matrix I by $\mathbf{e}_1, \ldots, \mathbf{e}_n$. If $A\mathbf{x} = \mathbf{b}$, the definition of matrix multiplication shows that

$$\begin{aligned} A\,(I_i(\mathbf{x})) = A\begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} &= \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix} = A_i(\mathbf{b}) \end{aligned}$$

By the multiplicative property of determinants,

$$(\det A)(\det I_i(\mathbf{x})) = \det A_i(\mathbf{b})$$

The second determinant on the left is simply x_i. (Make a cofactor expansion along the ith row.) Hence $(\det A)\,x_i = \det A_i(\mathbf{b})$. This proves (1) because A is invertible and $\det A \neq 0$. ■

EXAMPLE 1 Use Cramer's rule to solve the system

$$\begin{aligned} 3x_1 - 2x_2 &= 6 \\ -5x_1 + 4x_2 &= 8 \end{aligned}$$

SOLUTION View the system as $A\mathbf{x} = \mathbf{b}$. Using the notation introduced above,

$$A = \begin{bmatrix} 3 & -2 \\ -5 & 4 \end{bmatrix}, \qquad A_1(\mathbf{b}) = \begin{bmatrix} 6 & -2 \\ 8 & 4 \end{bmatrix}, \qquad A_2(\mathbf{b}) = \begin{bmatrix} 3 & 6 \\ -5 & 8 \end{bmatrix}$$

Since $\det A = 2$, the system has a unique solution. By Cramer's rule,

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A} = \frac{24 + 16}{2} = 20$$
$$x_2 = \frac{\det A_2(\mathbf{b})}{\det A} = \frac{24 + 30}{2} = 27 \quad \blacksquare$$

Application to Engineering

A number of important engineering problems, particularly in electrical engineering and control theory, can be analyzed by *Laplace transforms*. This approach converts an appropriate system of linear differential equations into a system of linear algebraic equations whose coefficients involve a parameter. The next example illustrates the type of algebraic system that may arise.

EXAMPLE 2 Consider the following system in which s is an unspecified parameter. Determine the values of s for which the system has a unique solution, and use Cramer's rule to describe the solution.

$$\begin{aligned} 3sx_1 - 2x_2 &= 4 \\ -6x_1 + sx_2 &= 1 \end{aligned}$$

SOLUTION View the system as $A\mathbf{x} = \mathbf{b}$. Then

$$A = \begin{bmatrix} 3s & -2 \\ -6 & s \end{bmatrix}, \quad A_1(\mathbf{b}) = \begin{bmatrix} 4 & -2 \\ 1 & s \end{bmatrix}, \quad A_2(\mathbf{b}) = \begin{bmatrix} 3s & 4 \\ -6 & 1 \end{bmatrix}$$

Since

$$\det A = 3s^2 - 12 = 3(s + 2)(s - 2)$$

the system has a unique solution precisely when $s \neq \pm 2$. For such an s, the solution is (x_1, x_2), where

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A} = \frac{4s + 2}{3(s + 2)(s - 2)}$$
$$x_2 = \frac{\det A_2(\mathbf{b})}{\det A} = \frac{3s + 24}{3(s + 2)(s - 2)} = \frac{s + 8}{(s + 2)(s - 2)} \quad \blacksquare$$

A Formula for A^{-1}

Cramer's rule leads easily to a general formula for the inverse of an $n \times n$ matrix A. The jth column of A^{-1} is a vector $\mathbf{x}$ that satisfies

$$A\mathbf{x} = \mathbf{e}_j$$

where $\mathbf{e}_j$ is the jth column of the identity matrix, and the ith entry of $\mathbf{x}$ is the (i, j)-entry of A^{-1}. By Cramer's rule,

$$\{(i, j)\text{-entry of } A^{-1}\} = x_i = \frac{\det A_i(\mathbf{e}_j)}{\det A} \tag{2}$$

Recall that A_{ji} denotes the submatrix of A formed by deleting row j and column i. A cofactor expansion down column i of $A_i(\mathbf{e}_j)$ shows that

$$\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji} \tag{3}$$

where C_{ji} is a cofactor of A. By (2), the (i, j)-entry of A^{-1} is the cofactor C_{ji} divided by $\det A$. [Note that the subscripts on C_{ji} are the reverse of (i, j).] Thus

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \tag{4}$$

The matrix of cofactors on the right side of (4) is called the **adjugate** (or **classical adjoint**) of A, denoted by adj A. (The term *adjoint* also has another meaning in advanced texts on linear transformations.) The next theorem simply restates (4).

THEOREM 8

An Inverse Formula

Let A be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A$$

EXAMPLE 3 Find the inverse of the matrix $A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix}$.

SOLUTION The nine cofactors are

$$C_{11} = +\begin{vmatrix} -1 & 1 \\ 4 & -2 \end{vmatrix} = -2, \quad C_{12} = -\begin{vmatrix} 1 & 1 \\ 1 & -2 \end{vmatrix} = 3, \quad C_{13} = +\begin{vmatrix} 1 & -1 \\ 1 & 4 \end{vmatrix} = 5$$

$$C_{21} = -\begin{vmatrix} 1 & 3 \\ 4 & -2 \end{vmatrix} = 14, \quad C_{22} = +\begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} = -7, \quad C_{23} = -\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} = -7$$

$$C_{31} = +\begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} = 4, \quad C_{32} = -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = 1, \quad C_{33} = +\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} = -3$$

The adjugate matrix is the *transpose* of the matrix of cofactors. [For instance, C_{12} goes in the (2, 1) position.] Thus

$$\operatorname{adj} A = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}$$

We could compute $\det A$ directly, but the following computation provides a check on the calculations for adj A *and* produces $\det A$:

$$(\operatorname{adj} A)\,A = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}\begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix} = \begin{bmatrix} 14 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 14 \end{bmatrix} = 14I$$

Since $(\operatorname{adj} A)A = 14I$, Theorem 8 shows that $\det A = 14$ and

$$A^{-1} = \frac{1}{14}\begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} = \begin{bmatrix} -1/7 & 1 & 2/7 \\ 3/14 & -1/2 & 1/14 \\ 5/14 & -1/2 & -3/14 \end{bmatrix}$$ ■

Numerical Notes

Theorem 8 is useful mainly for theoretical calculations. The formula for A^{-1} permits one to deduce properties of the inverse without actually calculating it. Except for special cases, the algorithm in Section 2.2 gives a much better way to compute A^{-1}, if the inverse is really needed.

Cramer's rule is also a theoretical tool. It can be used to study how sensitive the solution of $A\mathbf{x} = \mathbf{b}$ is to changes in an entry in $\mathbf{b}$ or in A (perhaps due to experimental error when acquiring the entries for $\mathbf{b}$ or A). When A is a 3×3 matrix with *complex* entries, Cramer's rule is sometimes selected for hand computation because row reduction of $[\,A \quad \mathbf{b}\,]$ with complex arithmetic can be messy, and the determinants are fairly easy to compute. For a larger $n \times n$ matrix (real or complex), Cramer's rule is hopelessly inefficient. Computing just *one* determinant takes about as much work as solving $A\mathbf{x} = \mathbf{b}$ by row reduction.

Determinants as Area or Volume

In the next application, we verify the geometric interpretation of determinants described in the chapter introduction. Although a general discussion of length and distance in $\mathbb{R}^n$ will not be given until Chapter 6, we assume here that the usual Euclidean concepts of length, area, and volume are already understood for $\mathbb{R}^2$ and $\mathbb{R}^3$.

THEOREM 9

If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det A|$.

STUDY GUIDE provides a geometric proof of the determinant as area.

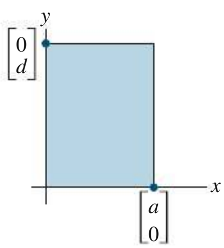


FIGURE 1
Area $= |ad|$.

PROOF The theorem is obviously true for any 2×2 diagonal matrix:

$$\left|\det\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}\right| = |ad| = \left\{\begin{matrix}\text{area of} \\ \text{rectangle}\end{matrix}\right\}$$

See Figure 1. It will suffice to show that any 2×2 matrix $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2\,]$ can be transformed into a diagonal matrix in a way that changes neither the area of the associated parallelogram nor $|\det A|$. From Section 3.2, we know that the absolute value of the determinant is unchanged when two columns are interchanged or a multiple of one column is added to another. And it is easy to see that such operations suffice to transform A into a diagonal matrix. Column interchanges do not change the parallelogram at all. So it suffices to prove the following simple geometric observation that applies to vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$:

Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be nonzero vectors. Then for any scalar c, the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2$ equals the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2 + c\mathbf{a}_1$.

To prove this statement, we may assume that $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_1$, for otherwise the two parallelograms would be degenerate and have zero area. If L is the line through $\mathbf{0}$ and $\mathbf{a}_1$, then $\mathbf{a}_2 + L$ is the line through $\mathbf{a}_2$ parallel to L, and $\mathbf{a}_2 + c\mathbf{a}_1$ is on this line. See Figure 2. The points $\mathbf{a}_2$ and $\mathbf{a}_2 + c\mathbf{a}_1$ have the same perpendicular distance to L. Hence the two parallelograms in Figure 2 have the same area, since they share the base from $\mathbf{0}$ to $\mathbf{a}_1$. This completes the proof for $\mathbb{R}^2$.

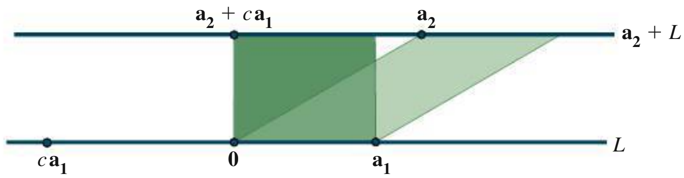


FIGURE 2 Two parallelograms of equal area.

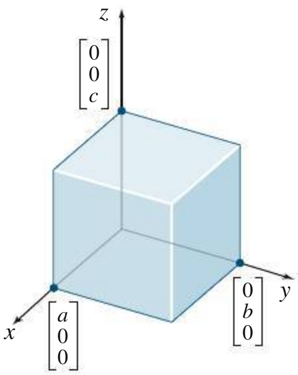


FIGURE 3
Volume $= |abc|$.

The proof for $\mathbb{R}^3$ is similar. The theorem is obviously true for a 3×3 diagonal matrix. See Figure 3. And any 3×3 matrix A can be transformed into a diagonal matrix using column operations that do not change $|\det A|$. (Think about doing row operations on A^T.) So it suffices to show that these operations do not affect the volume of the parallelepiped determined by the columns of A.

A parallelepiped is shown in Figure 4 as a shaded box with two sloping sides. Its volume is the area of the base in the plane Span $\{\mathbf{a}_1, \mathbf{a}_3\}$ times the altitude of $\mathbf{a}_2$ above Span $\{\mathbf{a}_1, \mathbf{a}_3\}$. Any vector $\mathbf{a}_2 + c\mathbf{a}_1$ has the same altitude because $\mathbf{a}_2 + c\mathbf{a}_1$ lies in the plane $\mathbf{a}_2 +$ Span $\{\mathbf{a}_1, \mathbf{a}_3\}$, which is parallel to Span $\{\mathbf{a}_1, \mathbf{a}_3\}$. Hence the volume of the parallelepiped is unchanged when $[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$ is changed to $[\,\mathbf{a}_1 \quad \mathbf{a}_2 + c\mathbf{a}_1 \quad \mathbf{a}_3\,]$. Thus a column replacement operation does not affect the volume of the parallelepiped. Since column interchanges have no effect on the volume, the proof is complete. ■

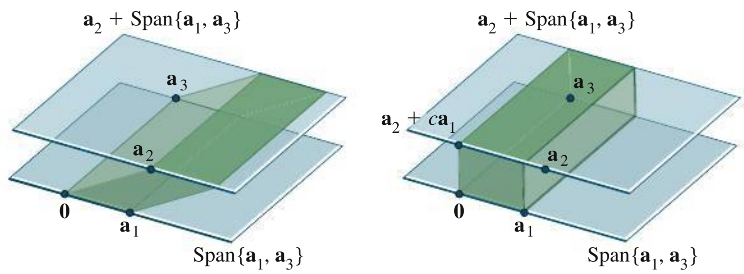


FIGURE 4 Two parallelepipeds of equal volume.

EXAMPLE 4 Calculate the area of the parallelogram determined by the points $(-2, -2)$, $(0, 3)$, $(4, -1)$, and $(6, 4)$. See Figure 5(a).

SOLUTION First translate the parallelogram to one having the origin as a vertex. For example, subtract the vertex $(-2, -2)$ from each of the four vertices. The new parallelogram has the same area, and its vertices are $(0, 0)$, $(2, 5)$, $(6, 1)$, and $(8, 6)$. See Figure 5(b). This parallelogram is determined by the columns of

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

Since $|\det A| = |-28|$, the area of the parallelogram is 28. ■

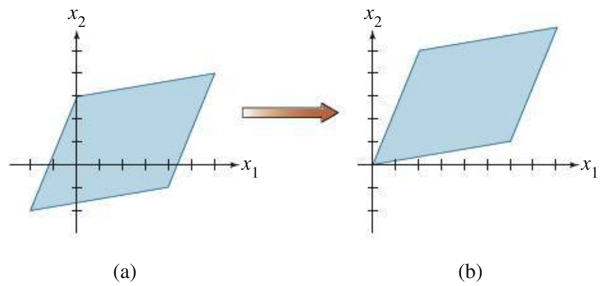


FIGURE 5 Translating a parallelogram does not change its area.

Linear Transformations

Determinants can be used to describe an important geometric property of linear transformations in the plane and in $\mathbb{R}^3$. If T is a linear transformation and S is a set in the domain of T, let $T(S)$ denote the set of images of points in S. We are interested in how the area (or volume) of $T(S)$ compares with the area (or volume) of the original set S. For convenience, when S is a region bounded by a parallelogram, we also refer to S as a parallelogram.

THEOREM 10

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$\{\text{area of } T(S)\} = |\det A| \cdot \{\text{area of } S\} \tag{5}$$

If T is determined by a 3×3 matrix A, and if S is a parallelepiped in $\mathbb{R}^3$, then

$$\{\text{volume of } T(S)\} = |\det A| \cdot \{\text{volume of } S\} \tag{6}$$

PROOF Consider the 2×2 case, with $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2\,]$. A parallelogram at the origin in $\mathbb{R}^2$ determined by vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ has the form

$$S = \{s_1\mathbf{b}_1 + s_2\mathbf{b}_2 : 0 \le s_1 \le 1,\ 0 \le s_2 \le 1\}$$

The image of S under T consists of points of the form

$$\begin{aligned} T(s_1\mathbf{b}_1 + s_2\mathbf{b}_2) &= s_1T(\mathbf{b}_1) + s_2T(\mathbf{b}_2) \\ &= s_1A\mathbf{b}_1 + s_2A\mathbf{b}_2 \end{aligned}$$

where $0 \le s_1 \le 1$, $0 \le s_2 \le 1$. It follows that $T(S)$ is the parallelogram determined by the columns of the matrix $[\,A\mathbf{b}_1 \quad A\mathbf{b}_2\,]$. This matrix can be written as AB, where $B = [\,\mathbf{b}_1 \quad \mathbf{b}_2\,]$. By Theorem 9 and the product theorem for determinants,

$$\begin{aligned} \{\text{area of } T(S)\} &= |\det AB| = |\det A|\,|\det B| \\ &= |\det A| \cdot \{\text{area of } S\} \end{aligned} \tag{7}$$

An arbitrary parallelogram has the form $\mathbf{p} + S$, where $\mathbf{p}$ is a vector and S is a parallelogram at the origin, as seen previously. It is easy to see that T transforms $\mathbf{p} + S$

into $T(\mathbf{p}) + T(S)$. (See Exercise 26.) Since translation does not affect the area of a set,

$$\begin{aligned}
\{\text{area of } T(\mathbf{p}+S)\} &= \{\text{area of } T(\mathbf{p}) + T(S)\} \\
&= \{\text{area of } T(S)\} && \text{Translation} \\
&= |\det A| \cdot \{\text{area of } S\} && \text{By equation (7)} \\
&= |\det A| \cdot \{\text{area of } (\mathbf{p}+S)\} && \text{Translation}
\end{aligned}$$

This shows that (5) holds for all parallelograms in $\mathbb{R}^2$. The proof of (6) for the 3×3 case is analogous. ■

When we attempt to generalize Theorem 10 to a region in $\mathbb{R}^2$ or $\mathbb{R}^3$ that is not bounded by straight lines or planes, we must face the problem of how to define and compute its area or volume. This is a question studied in calculus, and we shall only outline the basic idea for $\mathbb{R}^2$. If R is a planar region that has a finite area, then R can be approximated by a grid of small squares that lie inside R. By making the squares sufficiently small, the area of R may be approximated as closely as desired by the sum of the areas of the small squares. See Figure 6.

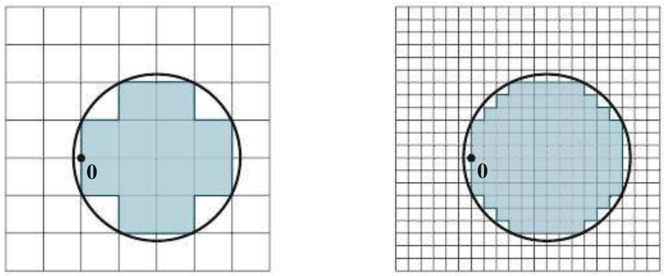


FIGURE 6 Approximating a planar region by a union of squares. The approximation improves as the grid becomes finer.

If T is a linear transformation associated with a 2×2 matrix A, then the image of a planar region R under T is approximated by the images of the small squares inside R. The proof of Theorem 10 shows that each such image is a parallelogram whose area is $|\det A|$ times the area of the square. If R' is the union of the squares inside R, then the area of $T(R')$ is $|\det A|$ times the area of R'. See Figure 7. Also, the area of $T(R')$ is close to the area of $T(R)$. An argument involving a limiting process may be given to justify the following generalization of Theorem 10.

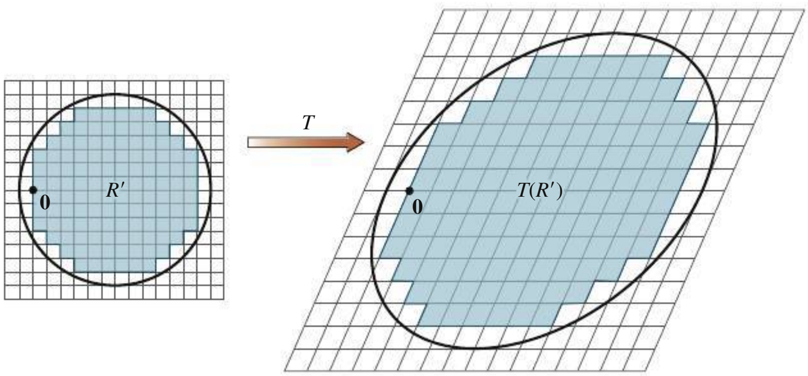
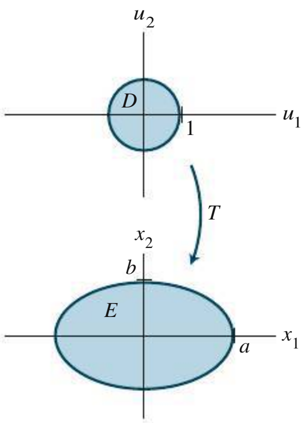


FIGURE 7 Approximating $T(R)$ by a union of parallelograms.

The conclusions of Theorem 10 hold whenever S is a region in $\mathbb{R}^2$ with finite area or a region in $\mathbb{R}^3$ with finite volume.

EXAMPLE 5 Let a and b be positive numbers. Find the area of the region E bounded by the ellipse whose equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$



SOLUTION We claim that E is the image of the unit disk D under the linear transformation T determined by the matrix $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, because if $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and $\mathbf{x} = A\mathbf{u}$, then

$$u_1 = \frac{x_1}{a} \quad \text{and} \quad u_2 = \frac{x_2}{b}$$

It follows that $\mathbf{u}$ is in the unit disk, with $u_1^2 + u_2^2 \leq 1$, if and only if $\mathbf{x}$ is in E, with $(x_1/a)^2 + (x_2/b)^2 \leq 1$. By the generalization of Theorem 10,

$$\begin{aligned} \{\text{area of ellipse}\} &= \{\text{area of } T(D)\} \\ &= |\det A| \cdot \{\text{area of } D\} \\ &= ab\pi(1)^2 = \pi ab \end{aligned}$$

■

Practice Problem

Let S be the parallelogram determined by the vectors $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, and let $A = \begin{bmatrix} 1 & -.1 \\ 0 & 2 \end{bmatrix}$. Compute the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

3.3 Exercises

Use Cramer's rule to compute the solutions of the systems in Exercises 1–6.

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

2. $\begin{aligned} 4x_1 + x_2 &= 6 \\ 3x_1 + 2x_2 &= 5 \end{aligned}$

3. $\begin{aligned} 3x_1 - 2x_2 &= 3 \\ -4x_1 + 6x_2 &= -5 \end{aligned}$

4. $\begin{aligned} -5x_1 + 2x_2 &= 9 \\ 3x_1 - x_2 &= -4 \end{aligned}$

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

6. $\begin{aligned} x_1 + 3x_2 + x_3 &= 8 \\ -x_1 \phantom{{}+3x_2} + 2x_3 &= 4 \\ 3x_1 + x_2 \phantom{{}+x_3} &= 4 \end{aligned}$

In Exercises 7–10, determine the values of the parameter s for which the system has a unique solution, and describe the solution.

7. $\begin{aligned} 6sx_1 + 4x_2 &= 5 \\ 9x_1 + 2sx_2 &= -2 \end{aligned}$

8. $\begin{aligned} 3sx_1 + 5x_2 &= 3 \\ 12x_1 + 5sx_2 &= 2 \end{aligned}$

9. $\begin{aligned} sx_1 + 2sx_2 &= -1 \\ 3x_1 + 6sx_2 &= 4 \end{aligned}$

10. $\begin{aligned} sx_1 - 2x_2 &= 1 \\ 4sx_1 + 4sx_2 &= 2 \end{aligned}$

In Exercises 11–16, compute the adjugate of the given matrix, and then use Theorem 8 to give the inverse of the matrix.

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

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

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

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

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

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

17. Show that if A is 2×2, then Theorem 8 gives the same formula for A^{-1} as that given by Theorem 4 in Section 2.2.

18. Suppose that all the entries in A are integers and $\det A = 1$. Explain why all the entries in A^{-1} are integers.

In Exercises 19–22, find the area of the parallelogram whose vertices are listed.

19. $(0, 0), (5, 2), (6, 4), (11, 6)$

20. $(0, 0), (-2, 4), (6, -5), (4, -1)$

21. $(-2, 0), (0, 3), (1, 3), (-1, 0)$

22. $(0, -2), (5, -2), (-3, 1), (2, 1)$

23. Find the volume of the parallelepiped with one vertex at the origin and adjacent vertices at $(1, 0, -3)$, $(1, 2, 4)$, and $(5, 1, 0)$.

24. Find the volume of the parallelepiped with one vertex at the origin and adjacent vertices at $(1, 3, 0)$, $(-2, 0, 2)$, and $(-1, 3, -1)$.

25. Use the concept of volume to explain why the determinant of a 3×3 matrix A is zero if and only if A is not invertible. Do not appeal to Theorem 4 in Section 3.2. [*Hint:* Think about the columns of A.]

26. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, and let $\mathbf{p}$ be a vector and S a set in $\mathbb{R}^m$. Show that the image of $\mathbf{p} + S$ under T is the translated set $T(\mathbf{p}) + T(S)$ in $\mathbb{R}^n$.

27. Let S be the parallelogram determined by the vectors $\mathbf{b}_1 = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$, and let $A = \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix}$. Compute the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

28. Repeat Exercise 27 with $\mathbf{b}_1 = \begin{bmatrix} 4 \\ -7 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $A = \begin{bmatrix} 5 & 2 \\ 1 & 1 \end{bmatrix}$.

29. Find a formula for the area of the triangle whose vertices are $\mathbf{0}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ in $\mathbb{R}^2$.

30. Let R be the triangle with vertices at (x_1, y_1), (x_2, y_2), and (x_3, y_3). Show that

$$\{\text{area of triangle}\} = \frac{1}{2} \det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix}$$

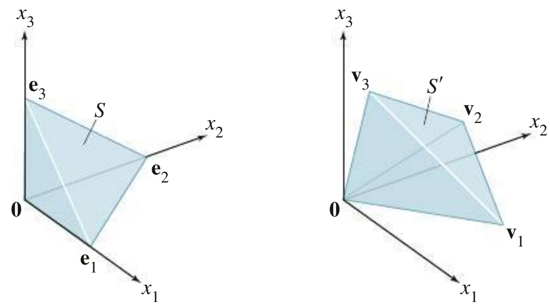
[*Hint:* Translate R to the origin by subtracting one of the vertices, and use Exercise 29.]

31. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation determined by the matrix $A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$, where a, b, and c are positive numbers. Let S be the unit ball, whose bounding surface has the equation $x_1^2 + x_2^2 + x_3^2 = 1$.

a. Show that $T(S)$ is bounded by the ellipsoid with the equation $\dfrac{x_1^2}{a^2} + \dfrac{x_2^2}{b^2} + \dfrac{x_3^2}{c^2} = 1$.

b. Use the fact that the volume of the unit ball is $4\pi/3$ to determine the volume of the region bounded by the ellipsoid in part (a).

32. Let S be the tetrahedron in $\mathbb{R}^3$ with vertices at the vectors $\mathbf{0}$, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, and let S' be the tetrahedron with vertices at vectors $\mathbf{0}$, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. See the figure.



a. Describe a linear transformation that maps S onto S'.

b. Find a formula for the volume of the tetrahedron S' using the fact that

$$\{\text{volume of } S\} = (1/3) \cdot \{\text{area of base}\} \cdot \{\text{height}\}$$

33. Let A be an $n \times n$ matrix. If $A^{-1} = \dfrac{1}{\det A}\, \text{adj}\, A$ is computed, what should AA^{-1} be equal to in order to confirm that A^{-1} has been found correctly?

34. If a parallelogram fits inside a circle radius 1 and $\det A = 2$, where A is the matrix whose columns correspond to the edges of the parallelogram, does it seem like A and its determinant have been calculated correctly to correspond to the area of this parallelogram? Explain why or why not.

In Exercises 35–38, mark each statement as True or False **(T/F)**. Justify each answer.

35. **(T/F)** Two parallelograms with the same base and height have the same area.

36. **(T/F)** Applying a linear transformation to a region does not change its area.

37. **(T/F)** If A is an invertible $n \times n$ matrix, then $A^{-1} = \text{adj}\, A$.

38. **(T/F)** Cramer's rule can only be used for invertible matrices.

T **39.** Test the inverse formula of Theorem 8 for a random 4×4 matrix A. Use your matrix program to compute the cofactors of the 3×3 submatrices, construct the adjugate, and set $B = (\text{adj } A)/(\det A)$. Then compute $B - \text{inv}(A)$, where $\text{inv}(A)$ is the inverse of A as computed by the matrix program. Use floating point arithmetic with the maximum possible number of decimal places. Report your results.

T 40. Test Cramer's rule for a random 4×4 matrix A and a random 4×1 vector $\mathbf{b}$. Compute each entry in the solution of $A\mathbf{x} = \mathbf{b}$, and compare these entries with the entries in $A^{-1}\mathbf{b}$. Write the command (or keystrokes) for your matrix program that uses Cramer's rule to produce the second entry of $\mathbf{x}$.

T 41. If your version of MATLAB has the `flops` command, use it to count the number of floating point operations to compute A^{-1} for a random 30×30 matrix. Compare this number with the number of flops needed to form $(\text{adj } A)/(\det A)$.

Solution to Practice Problem

The area of S is $\left|\det \begin{bmatrix} 1 & 5 \\ 3 & 1 \end{bmatrix}\right| = 14$, and $\det A = 2$. By Theorem 10, the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$ is

$$|\det A| \cdot \{\text{area of } S\} = 2 \cdot 14 = 28$$

CHAPTER 3 PROJECTS

Chapter 3 projects are available online at `bit.ly/30IM8gT`.

A. *Weighing Design*: This project develops the concept of weighing design and their corresponding matrices for use in weighing a few small, light objects.

B. *Jacobians*: This set of exercises examines how a particular determinant called the Jacobian may be used to allow us to change variables in double and triple integrals.

CHAPTER 3 SUPPLEMENTARY EXERCISES

In Exercises 1-15, mark each statement True or False **(T/F)**. Justify each answer. Assume that all matrices here are square.

1. **(T/F)** If A is a 2×2 matrix with a zero determinant, then one column of A is a multiple of the other.

2. **(T/F)** If two rows of a 3×3 matrix A are the same, then $\det A = 0$.

3. **(T/F)** If A is a 3×3 matrix, then $\det 5A = 5 \det A$.

4. **(T/F)** If A and B are $n \times n$ matrices, with $\det A = 2$ and $\det B = 3$, then $\det(A + B) = 5$.

5. **(T/F)** If A is $n \times n$ and $\det A = 2$, then $\det A^3 = 6$.

6. **(T/F)** If B is produced by interchanging two rows of A, then $\det B = \det A$.

7. **(T/F)** If B is produced by multiplying row 3 of A by 5, then $\det B = 5 \det A$.

8. **(T/F)** If B is formed by adding to one row of A a linear combination of the other rows, then $\det B = \det A$.

9. **(T/F)** $\det A^T = -\det A$.

10. **(T/F)** $\det(-A) = -\det A$.

11. **(T/F)** $\det A^TA \geq 0$.

12. **(T/F)** Any system of n linear equations in n variables can be solved by Cramer's rule.

13. **(T/F)** If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^2$ and $\det[\,\mathbf{u} \quad \mathbf{v}\,] = 10$, then the area of the triangle in the plane with vertices at $\mathbf{0}$, $\mathbf{u}$, and $\mathbf{v}$ is 10.

14. **(T/F)** If $A^3 = 0$, then $\det A = 0$.

15. **(T/F)** If A is invertible, then $\det A^{-1} = \det A$.

Use row operations to show that the determinants in Exercises 16–18 are all zero.

16. $\begin{vmatrix} 12 & 13 & 14 \\ 15 & 16 & 17 \\ 18 & 19 & 20 \end{vmatrix}$

17. $\begin{vmatrix} 1 & a & b+c \\ 1 & b & a+c \\ 1 & c & a+b \end{vmatrix}$

18. $\begin{vmatrix} a & b & c \\ a+x & b+x & c+x \\ a+y & b+y & c+y \end{vmatrix}$

Compute the determinants in Exercises 19 and 20.

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

20. $\begin{vmatrix} 4 & 8 & 8 & 8 & 5 \\ 0 & 1 & 0 & 0 & 0 \\ 6 & 8 & 8 & 8 & 7 \\ 0 & 8 & 8 & 3 & 0 \\ 0 & 8 & 2 & 0 & 0 \end{vmatrix}$

21. Show that the equation of the line in $\mathbb{R}^2$ through distinct points (x_1, y_1) and (x_2, y_2) can be written as

$$\det \begin{bmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{bmatrix} = 0$$

22. Find a 3×3 determinant equation similar to that in Exercise 21 that describes the equation of the line through (x_1, y_1) with slope m.

Exercises 23 and 24 concern determinants of the following *Vandermonde matrices*.

$$T = \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}, \quad V(t) = \begin{bmatrix} 1 & t & t^2 & t^3 \\ 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \end{bmatrix}$$

23. Use row operations to show that

$$\det T = (b - a)(c - a)(c - b)$$

24. Let $f(t) = \det V$, with x_1, x_2, and x_3 all distinct. Explain why $f(t)$ is a cubic polynomial, show that the coefficient of t^3 is nonzero, and find three points on the graph of f.

25. Find the area of the parallelogram determined by the points $(1, 4)$, $(-1, 5)$, $(3, 9)$, and $(5, 8)$. How can you tell that the quadrilateral determined by the points is actually a parallelogram?

26. Use the concept of area of a parallelogram to write a statement about a 2×2 matrix A that is true if and only if A is invertible.

27. Show that if A is invertible, then adj A is invertible, and

$$(\text{adj } A)^{-1} = \frac{1}{\det A} A$$

[*Hint:* Given matrices B and C, what calculation(s) would show that C is the inverse of B?]

28. Let A, B, C, D, and I be $n \times n$ matrices. Use the definition or properties of a determinant to justify the following formulas. Part (c) is useful in applications of eigenvalues (Chapter 5).

a. $\det \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} = \det A$ b. $\det \begin{bmatrix} I & 0 \\ C & D \end{bmatrix} = \det D$

c. $\det \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} = (\det A)(\det D) = \det \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$

29. Let A, B, C, and D be $n \times n$ matrices with A invertible.

a. Find matrices X and Y to produce the block LU factorization

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} A & B \\ 0 & Y \end{bmatrix}$$

and then show that

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = (\det A) \cdot \det(D - CA^{-1}B)$$

b. Show that if $AC = CA$, then

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(AD - CB)$$

30. Let J be the $n \times n$ matrix of all 1's, and consider $A = (a - b)I + bJ$; that is,

$$A = \begin{bmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{bmatrix}$$

Confirm that $\det A = (a - b)^{n-1}[a + (n - 1)b]$ as follows:

a. Subtract row 2 from row 1, row 3 from row 2, and so on, and explain why this does not change the determinant of the matrix.

b. With the resulting matrix from part (a), add column 1 to column 2, then add this new column 2 to column 3, and so on, and explain why this does not change the determinant.

c. Find the determinant of the resulting matrix from (b).

31. Let A be the original matrix given in Exercise 30, and let

$$B = \begin{bmatrix} a - b & b & b & \cdots & b \\ 0 & a & b & \cdots & b \\ 0 & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & b & b & \cdots & a \end{bmatrix},$$

$$C = \begin{bmatrix} b & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{bmatrix}$$

Notice that A, B, and C are nearly the same except that the first column of A equals the sum of the first columns of B and C. A *linearity property* of the determinant function, discussed in Section 3.2, says that $\det A = \det B + \det C$. Use this fact to prove the formula in Exercise 30 by induction on the size of matrix A.

T 32. Apply the result of Exercise 30 to find the determinants of the following matrices, and confirm your answers using a matrix program.

$$\begin{bmatrix} 3 & 8 & 8 & 8 \\ 8 & 3 & 8 & 8 \\ 8 & 8 & 3 & 8 \\ 8 & 8 & 8 & 3 \end{bmatrix} \quad \begin{bmatrix} 8 & 3 & 3 & 3 & 3 \\ 3 & 8 & 3 & 3 & 3 \\ 3 & 3 & 8 & 3 & 3 \\ 3 & 3 & 3 & 8 & 3 \\ 3 & 3 & 3 & 3 & 8 \end{bmatrix}$$

T **33.** Use a matrix program to compute the determinants of the following matrices.

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

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

Use the results to guess the determinant of the matrix M, and confirm your guess by using row operations to evaluate that determinant.

$$M = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & \cdots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}$$

T **34.** Use the method of Exercise 33 to guess the determinant of

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 3 & 3 & \cdots & 3 \\ 1 & 3 & 6 & \cdots & 6 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 3 & 6 & \cdots & 3(n-1) \end{bmatrix}$$

Justify your conjecture. [*Hint:* Use Exercise 28(c) and the result of Exercise 33.]

4 Vector Spaces

Introductory Example

DISCRETE-TIME SIGNALS AND DIGITAL SIGNAL PROCESSING

What is digital signal processing? Just ask Alexa, who uses signal processing to record your question and deliver the answer. In 2500 BC, the Egyptians created the first recorded discrete-time signal by carving information about the flooding of the Nile into a Palermo Stone. Despite the early beginnings of discrete-time signals, it was not until the 1940s that Claude Shannon set off the digital revolution with the ideas articulated in his paper "A Mathematical Theory of Communication."

When a person speaks into a digital processor like Alexa, it converts the sounds made by the voice into a discrete-time signal—basically a sequence of numbers $\{y_k\}$, where k represents the time at which the value y_k was recorded. Then using linear time invariant (LTI) transformations, the signal is processed to filter out unwanted noise, such as the sound of a fan running in the background. The processed signal is then compared to the signals produced by recordings of the individual sounds that make up the language of the speaker. Figure 1 shows a recording of the word "yes" and of the word "no" illustrating that the signals produced are quite distinct. Once the sounds spoken in the question are identified, machine learning is used to make a best guess at the intended question for a digital processor like Alexa. The digital processor then searches through digitized data to find the most appropriate response. Finally, the signal is processed further to produce the virtual sounds that replicate a spoken answer.

Digital signal processing (DSP) is the branch of engineering that, in the span of just a few decades, has revolutionized interpersonal communication and the entertainment industry. By reworking the principles of electronics, telecommunication, and computer science into a unifying paradigm, DSP is at the heart of the digital revolution. A smartphone fits easily in the palm of your hand, replacing numerous other devices such as cameras, video recorders, CD players, day planners, and calculators, and taking the fantasy component out of Borges's imagined Library of Babel.

The usefulness of discrete-time signals and DSP goes well beyond systems engineering. Technical analysis is employed in the investment sector. Trading opportunities are identified by applying DSP to the discrete-time signals created when the price or volume traded of a stock is recorded over time. In Example 11 of Section 4.2, price data is smoothed using a linear transformation. In the entertainment industry, audio and video are produced

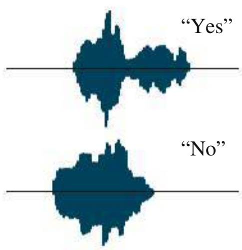


FIGURE 1

virtually and synthesized using DSP. In Example 3 of Section 4.7, we see how signal processing can be used to add richness to virtual sounds.

Discrete-time signals and DSP have become significant tools in many industries and areas of research. Mathematically speaking, discrete-time signals can be viewed as vectors that are processed using linear transformations. The operations of adding, scaling, and applying linear transformations to signals is completely analogous to the same operations for vectors in $\mathbb{R}^n$. For this reason, the set of all possible signals, $\mathbb{S}$, is treated as a *vector space*. In Sections 4.7 and 4.8, we look at the vector space of discrete-time signals in more detail.

The focus of Chapter 4 is to extend the theory of vectors in $\mathbb{R}^n$ to include signals and other mathematical structures that behave like the vectors you are already familiar with. Later on in the text, you will see how other vector spaces and their corresponding linear transformations arise in engineering, physics, biology, and statistics.

The mathematical seeds planted in Chapters 1 and 2 germinate and begin to blossom in this chapter. The beauty and power of linear algebra will be seen more clearly when you view $\mathbb{R}^n$ as only one of a variety of vector spaces that arise naturally in applied problems.

Beginning with basic definitions in Section 4.1, the general vector space framework develops gradually throughout the chapter. A goal of Sections 4.5 and 4.6 is to demonstrate how closely other vector spaces resemble $\mathbb{R}^n$. Sections 4.7 and 4.8 apply the theory of this chapter to discrete-time signals, DSP, and difference equations—the mathematics underlying the digital revolution.

4.1 Vector Spaces and Subspaces

Much of the theory in Chapters 1 and 2 rested on certain simple and obvious algebraic properties of $\mathbb{R}^n$, listed in Section 1.3. In fact, many other mathematical systems have the same properties. The specific properties of interest are listed in the following definition.

DEFINITION

A **vector space** is a nonempty set V of objects, called *vectors*, on which are defined two operations, called *addition* and *multiplication by scalars* (real numbers), subject to the ten axioms (or rules) listed below.[1] The axioms must hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and for all scalars c and d.

1. The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} + \mathbf{v}$, 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 is a **zero** vector $\mathbf{0}$ in V such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

[1] Technically, V is a *real vector space*. All of the theory in this chapter also holds for a *complex vector space* in which the scalars and matrix entries are complex numbers. We will look at this briefly in Chapter 5. Until then, all scalars and matrix entries are assumed to be real.

6. The scalar multiple of $\mathbf{u}$ by c, denoted by $c\mathbf{u}$, is in V.
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$.
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$.
10. $1\mathbf{u} = \mathbf{u}$.

Using only these axioms, one can show that the zero vector in Axiom 4 is unique, and the vector $-\mathbf{u}$, called the **negative** of $\mathbf{u}$, in Axiom 5 is unique for each $\mathbf{u}$ in V. See Exercises 33 and 34. Proofs of the following simple facts are also outlined in the exercises:

For each $\mathbf{u}$ in V and scalar c,

$$0\mathbf{u} = \mathbf{0} \tag{1}$$

$$c\mathbf{0} = \mathbf{0} \tag{2}$$

$$-\mathbf{u} = (-1)\mathbf{u} \tag{3}$$

EXAMPLE 1 The spaces $\mathbb{R}^n$, where $n \geq 1$, are the premier examples of vector spaces. The geometric intuition developed for $\mathbb{R}^3$ will help you understand and visualize many concepts throughout the chapter. ■

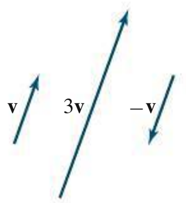


FIGURE 1

EXAMPLE 2 Let V be the set of all arrows (directed line segments) in three-dimensional space, with two arrows regarded as equal if they have the same length and point in the same direction. Define addition by the parallelogram rule (from Section 1.3), and for each $\mathbf{v}$ in V, define $c\mathbf{v}$ to be the arrow whose length is $|c|$ times the length of $\mathbf{v}$, pointing in the same direction as $\mathbf{v}$ if $c \geq 0$ and otherwise pointing in the opposite direction. (See Figure 1.) Show that V is a vector space. This space is a common model in physical problems for various forces.

SOLUTION The definition of V is geometric, using concepts of length and direction. No xyz-coordinate system is involved. An arrow of zero length is a single point and represents the zero vector. The negative of $\mathbf{v}$ is $(-1)\mathbf{v}$. So Axioms 1, 4, 5, 6, and 10 are evident. The rest are verified by geometry. For instance, see Figures 2 and 3. ■

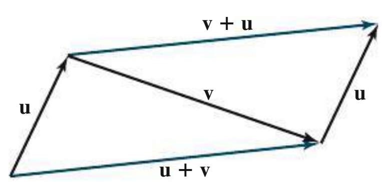


FIGURE 2 $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

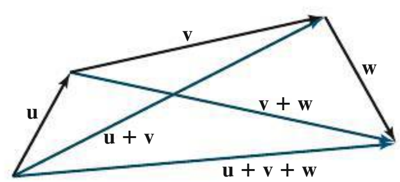


FIGURE 3 $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

EXAMPLE 3 Let $\mathbb{S}$ be the space of all doubly infinite sequences of numbers (usually written in a row rather than a column):

$$\{y_k\} = (\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots)$$

If $\{z_k\}$ is another element of $\mathbb{S}$, then the sum $\{y_k\} + \{z_k\}$ is the sequence $\{y_k + z_k\}$ formed by adding corresponding terms of $\{y_k\}$ and $\{z_k\}$. The scalar multiple $c\{y_k\}$ is the sequence $\{cy_k\}$. The vector space axioms are verified in the same way as for $\mathbb{R}^n$.

Elements of $\mathbb{S}$ arise in engineering, for example, whenever a signal is measured (or sampled) at discrete times. A signal might be electrical, mechanical, optical, biological, audio, and so on. The digital signal processors mentioned in the chapter introduction use discrete (or digital) signals. For convenience, we will call $\mathbb{S}$ the space of (discrete-time) **signals**. A signal may be visualized by a graph as in Figure 4.

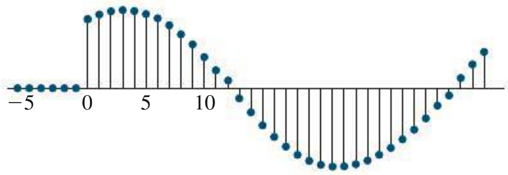


FIGURE 4 A discrete-time signal. ■

EXAMPLE 4 For $n \geq 0$, the set $\mathbb{P}_n$ of polynomials of degree at most n consists of all polynomials of the form

$$\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \tag{4}$$

where the coefficients $a_0, \ldots, a_n$ and the variable t are real numbers. The *degree* of $\mathbf{p}$ is the highest power of t in (4) whose coefficient is not zero. If $\mathbf{p}(t) = a_0 \neq 0$, the degree of $\mathbf{p}$ is zero. If all the coefficients are zero, $\mathbf{p}$ is called the *zero polynomial*. The zero polynomial is included in $\mathbb{P}_n$ even though its degree, for technical reasons, is not defined.

If $\mathbf{p}$ is given by (4) and if $\mathbf{q}(t) = b_0 + b_1 t + \cdots + b_n t^n$, then the sum $\mathbf{p} + \mathbf{q}$ is defined by

$$\begin{aligned}(\mathbf{p} + \mathbf{q})(t) &= \mathbf{p}(t) + \mathbf{q}(t) \\ &= (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n\end{aligned}$$

The scalar multiple $c\mathbf{p}$ is the polynomial defined by

$$(c\mathbf{p})(t) = c\mathbf{p}(t) = ca_0 + (ca_1)t + \cdots + (ca_n)t^n$$

These definitions satisfy Axioms 1 and 6 because $\mathbf{p} + \mathbf{q}$ and $c\mathbf{p}$ are polynomials of degree less than or equal to n. Axioms 2, 3, and 7–10 follow from properties of the real numbers. Clearly, the zero polynomial acts as the zero vector in Axiom 4. Finally, $(-1)\mathbf{p}$ acts as the negative of $\mathbf{p}$, so Axiom 5 is satisfied. Thus $\mathbb{P}_n$ is a vector space.

The vector spaces $\mathbb{P}_n$ for various n are used, for instance, in statistical trend analysis of data, discussed in Section 6.8. ■

EXAMPLE 5 Let V be the set of all real-valued functions defined on a set $\mathbb{D}$. (Typically, $\mathbb{D}$ is the set of real numbers or some interval on the real line.) Functions are added in the usual way: $\mathbf{f} + \mathbf{g}$ is the function whose value at t in the domain $\mathbb{D}$ is $\mathbf{f}(t) + \mathbf{g}(t)$. Likewise, for a scalar c and an $\mathbf{f}$ in V, the scalar multiple $c\mathbf{f}$ is the function whose value at t is $c\mathbf{f}(t)$. For instance, if $\mathbb{D} = \mathbb{R}$, $\mathbf{f}(t) = 1 + \sin 2t$, and $\mathbf{g}(t) = 2 + .5t$, then

$$(\mathbf{f} + \mathbf{g})(t) = 3 + \sin 2t + .5t \quad \text{and} \quad (2\mathbf{g})(t) = 4 + t$$

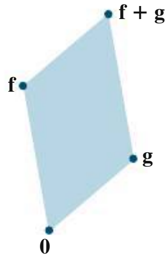


FIGURE 5 The sum of two vectors (functions).

Two functions in V are equal if and only if their values are equal for every t in $\mathbb{D}$. Hence the zero vector in V is the function that is identically zero, $\mathbf{f}(t) = 0$ for all t, and the negative of $\mathbf{f}$ is $(-1)\mathbf{f}$. Axioms 1 and 6 are obviously true, and the other axioms follow from properties of the real numbers, so V is a vector space. ■

It is important to think of each function in the vector space V of Example 5 as a single object, as just one "point" or vector in the vector space. The sum of two vectors $\mathbf{f}$ and $\mathbf{g}$ (functions in V, or elements of *any* vector space) can be visualized as in Figure 5, because this can help you carry over to a general vector space the geometric intuition you have developed while working with the vector space $\mathbb{R}^n$. See the *Study Guide* for help as you learn to adopt this more general point of view.

Subspaces

In many problems, a vector space consists of an appropriate subset of vectors from some larger vector space. In this case, only three of the ten vector space axioms need to be checked; the rest are automatically satisfied.

DEFINITION

A **subspace** of a vector space V is a subset H of V that has three properties:

a. The zero vector of V is in H.[2]

b. H is closed under vector addition. That is, for each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H.

c. H is closed under multiplication by scalars. That is, for each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is in H.

Properties (a), (b), and (c) guarantee that a subspace H of V is itself a *vector space*, under the vector space operations already defined in V. To verify this, note that properties (a), (b), and (c) are Axioms 1, 4, and 6. Axioms 2, 3, and 7–10 are automatically true in H because they apply to all elements of V, including those in H. Axiom 5 is also true in H, because if $\mathbf{u}$ is in H, then $(-1)\mathbf{u}$ is in H by property (c), and we know from equation (3) earlier in this section that $(-1)\mathbf{u}$ is the vector $-\mathbf{u}$ in Axiom 5.

So every subspace is a vector space. Conversely, every vector space is a subspace (of itself and possibly of other larger spaces). The term *subspace* is used when at least two vector spaces are in mind, with one inside the other, and the phrase *subspace of V* identifies V as the larger space. (See Figure 6.)

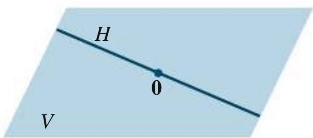


FIGURE 6 A subspace of V.

EXAMPLE 6 The set consisting of only the zero vector in a vector space V is a subspace of V, called the **zero subspace** and written as $\{\mathbf{0}\}$. ■

EXAMPLE 7 Let $\mathbb{P}$ be the set of all polynomials with real coefficients, with operations in $\mathbb{P}$ defined as for functions. Then $\mathbb{P}$ is a subspace of the space of all real-valued functions defined on $\mathbb{R}$. Also, for each $n \geq 0$, $\mathbb{P}_n$ is a subspace of $\mathbb{P}$, because $\mathbb{P}_n$ is a subset of $\mathbb{P}$ that contains the zero polynomial, the sum of two polynomials in $\mathbb{P}_n$ is also in $\mathbb{P}_n$, and a scalar multiple of a polynomial in $\mathbb{P}_n$ is also in $\mathbb{P}_n$. ■

[2] Some texts replace property (a) in this definition by the assumption that H is nonempty. Then (a) could be deduced from (c) and the fact that $0\mathbf{u} = \mathbf{0}$. But the best way to test for a subspace is to look first for the zero vector. If $\mathbf{0}$ is in H, then properties (b) and (c) must be checked. If $\mathbf{0}$ is *not* in H, then H cannot be a subspace and the other properties need not be checked.

EXAMPLE 8 The set of finitely supported signals $\mathbb{S}_f$ consists of the signals $\{y_k\}$, where only finitely many of the y_k are nonzero. Since the zero signal $\mathbf{0} = (\ldots, 0, 0, 0, \ldots)$ has no nonzero entries, it is clearly an element of $\mathbb{S}_f$. If two signals with finitely many nonzeros are added, the resulting signal will have finitely many nonzeros. Similarly if a signal with finitely many nonzeros is scaled, the result will still have finitely many nonzeros. Thus $\mathbb{S}_f$ is a subspace of $\mathbb{S}$, the discrete-time signals. See Figure 7.

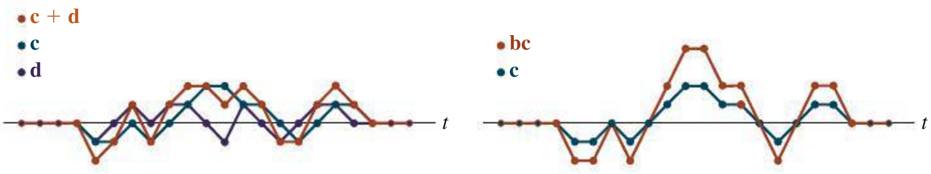


FIGURE 7 ■

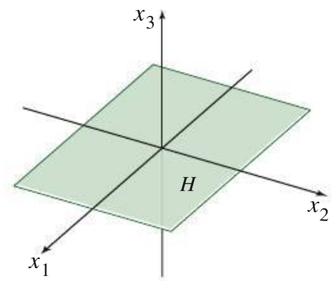


FIGURE 8
The x_1x_2-plane as a subspace of $\mathbb{R}^3$.

EXAMPLE 9 The vector space $\mathbb{R}^2$ is *not* a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$. (The vectors in $\mathbb{R}^3$ all have three entries, whereas the vectors in $\mathbb{R}^2$ have only two.) The set

$$H = \left\{ \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} : s \text{ and } t \text{ are real} \right\}$$

is a subset of $\mathbb{R}^3$ that "looks" and "acts" like $\mathbb{R}^2$, although it is logically distinct from $\mathbb{R}^2$. See Figure 8. Show that H is a subspace of $\mathbb{R}^3$.

SOLUTION The zero vector is in H, and H is closed under vector addition and scalar multiplication because these operations on vectors in H always produce vectors whose third entries are zero (and so belong to H). Thus H is a subspace of $\mathbb{R}^3$. ■

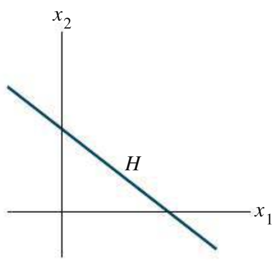


FIGURE 9
A line that is not a vector space.

EXAMPLE 10 A plane in $\mathbb{R}^3$ *not* through the origin is not a subspace of $\mathbb{R}^3$, because the plane does not contain the zero vector of $\mathbb{R}^3$. Similarly, a line in $\mathbb{R}^2$ *not* through the origin, such as in Figure 9, is *not* a subspace of $\mathbb{R}^2$. ■

A Subspace Spanned by a Set

The next example illustrates one of the most common ways of describing a subspace. As in Chapter 1, the term **linear combination** refers to any sum of scalar multiples of vectors, and Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ denotes the set of all vectors that can be written as linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

EXAMPLE 11 Given $\mathbf{v}_1$ and $\mathbf{v}_2$ in a vector space V, let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Show that H is a subspace of V.

SOLUTION The zero vector is in H, since $\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2$. To show that H is closed under vector addition, take two arbitrary vectors in H, say,

$$\mathbf{u} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 \quad \text{and} \quad \mathbf{w} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

By Axioms 2, 3, and 8 for the vector space V,

$$\begin{aligned} \mathbf{u} + \mathbf{w} &= (s_1\mathbf{v}_1 + s_2\mathbf{v}_2) + (t_1\mathbf{v}_1 + t_2\mathbf{v}_2) \\ &= (s_1 + t_1)\mathbf{v}_1 + (s_2 + t_2)\mathbf{v}_2 \end{aligned}$$

So $\mathbf{u}+\mathbf{w}$ is in H. Furthermore, if c is any scalar, then by Axioms 7 and 9,

$$c\mathbf{u} = c(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) = (cs_1)\mathbf{v}_1 + (cs_2)\mathbf{v}_2$$

which shows that $c\mathbf{u}$ is in H and H is closed under scalar multiplication. Thus H is a subspace of V. ■

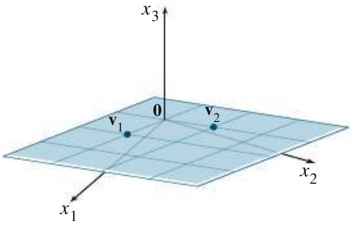


FIGURE 10
An example of a subspace.

In Section 4.5, you will see that every nonzero subspace of $\mathbb{R}^3$, other than $\mathbb{R}^3$ itself, is either Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for some linearly independent $\mathbf{v}_1$ and $\mathbf{v}_2$ or Span $\{\mathbf{v}\}$ for $\mathbf{v} \neq \mathbf{0}$. In the first case, the subspace is a plane through the origin; in the second case, it is a line through the origin. (See Figure 10.) It is helpful to keep these geometric pictures in mind, even for an abstract vector space.

The argument in Example 11 can easily be generalized to prove the following theorem.

THEOREM 1

If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in a vector space V, then Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a subspace of V.

We call Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ **the subspace spanned** (or **generated**) by $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$. Given any subspace H of V, a **spanning** (or **generating**) **set** for H is a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in H such that $H = \text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

The next example shows how to use Theorem 1.

EXAMPLE 12 Let H be the set of all vectors of the form $(a-3b, b-a, a, b)$, where a and b are arbitrary scalars. That is, let $H = \{(a-3b, b-a, a, b) : a \text{ and } b \text{ in } \mathbb{R}\}$. Show that H is a subspace of $\mathbb{R}^4$.

SOLUTION Write the vectors in H as column vectors. Then an arbitrary vector in H has the form

$$\begin{bmatrix} a-3b \\ b-a \\ a \\ b \end{bmatrix} = a\underbrace{\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_1} + b\underbrace{\begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_2}$$

This calculation shows that $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are the vectors indicated above. Thus H is a subspace of $\mathbb{R}^4$ by Theorem 1. ■

Example 12 illustrates a useful technique of expressing a subspace H as the set of linear combinations of some small collection of vectors. If $H = \text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$, we can think of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ in the spanning set as "handles" that allow us to hold on to the subspace H. Calculations with the infinitely many vectors in H are often reduced to operations with the finite number of vectors in the spanning set.

EXAMPLE 13 For what value(s) of h will $\mathbf{y}$ be in the subspace of $\mathbb{R}^3$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

SOLUTION This question is Practice Problem 2 in Section 1.3, written here with the term *subspace* rather than Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. The solution there shows that $\mathbf{y}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if $h = 5$. That solution is worth reviewing now, along with Exercises 11–16 and 19–21 in Section 1.3. ■

Although many vector spaces in this chapter will be subspaces of $\mathbb{R}^n$, it is important to keep in mind that the abstract theory applies to other vector spaces as well. Vector spaces of functions arise in many applications, and they will receive more attention later.

Practice Problems

1. Show that the set H of all points in $\mathbb{R}^2$ of the form $(3s, 2 + 5s)$ is not a vector space, by showing that it is not closed under scalar multiplication. (Find a specific vector $\mathbf{u}$ in H and a scalar c such that $c\mathbf{u}$ is not in H.)
2. Let $W = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$, where $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in a vector space V. Show that $\mathbf{v}_k$ is in W for $1 \le k \le p$. [*Hint:* First write an equation that shows that $\mathbf{v}_1$ is in W. Then adjust your notation for the general case.]
3. An $n \times n$ matrix A is said to be symmetric if $A^T = A$. Let S be the set of all 3×3 symmetric matrices. Show that S is a subspace of $M_{3\times3}$, the vector space of 3×3 matrices.

4.1 Exercises

1. Let V be the first quadrant in the xy-plane; that is, let
$$V = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \ge 0, y \ge 0 \right\}$$
   a. If $\mathbf{u}$ and $\mathbf{v}$ are in V, is $\mathbf{u} + \mathbf{v}$ in V? Why?
   b. Find a specific vector $\mathbf{u}$ in V and a specific scalar c such that $c\mathbf{u}$ is *not* in V. (This is enough to show that V is *not* a vector space.)

2. Let W be the union of the first and third quadrants in the xy-plane. That is, let $W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : xy \ge 0 \right\}$.
   a. If $\mathbf{u}$ is in W and c is any scalar, is $c\mathbf{u}$ in W? Why?
   b. Find specific vectors $\mathbf{u}$ and $\mathbf{v}$ in W such that $\mathbf{u} + \mathbf{v}$ is not in W. (This is enough to show that W is *not* a vector space.)

3. Let H be the set of points inside and on the unit circle in the xy-plane. That is, let $H = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x^2 + y^2 \le 1 \right\}$. Find a specific example—two vectors or a vector and a scalar—to show that H is not a subspace of $\mathbb{R}^2$.

4. Construct a geometric figure that illustrates why a line in $\mathbb{R}^2$ *not* through the origin is not closed under vector addition.

In Exercises 5–8, determine if the given set is a subspace of $\mathbb{P}_n$ for an appropriate value of n. Justify your answers.

5. All polynomials of the form $\mathbf{p}(t) = at^2$, where a is in $\mathbb{R}$.
6. All polynomials of the form $\mathbf{p}(t) = a + t^2$, where a is in $\mathbb{R}$.
7. All polynomials of degree at most 3, with integers as coefficients.
8. All polynomials in $\mathbb{P}_n$ such that $\mathbf{p}(0) = 0$.

9. Let H be the set of all vectors of the form $\begin{bmatrix} s \\ 3s \\ 2s \end{bmatrix}$. Find a vector $\mathbf{v}$ in $\mathbb{R}^3$ such that $H = \text{Span}\{\mathbf{v}\}$. Why does this show that H is a subspace of $\mathbb{R}^3$?

10. Let H be the set of all vectors of the form $\begin{bmatrix} 2t \\ 0 \\ -t \end{bmatrix}$. Show that H is a subspace of $\mathbb{R}^3$. (Use the method of Exercise 9.)

11. Let W be the set of all vectors of the form $\begin{bmatrix} 5b + 2c \\ b \\ c \end{bmatrix}$, where b and c are arbitrary. Find vectors $\mathbf{u}$ and $\mathbf{v}$ such that $W = \text{Span}\{\mathbf{u}, \mathbf{v}\}$. Why does this show that W is a subspace of $\mathbb{R}^3$?

12. Let W be the set of all vectors of the form $\begin{bmatrix} s + 3t \\ s - t \\ 2s - t \\ 4t \end{bmatrix}$. Show that W is a subspace of $\mathbb{R}^4$. (Use the method of Exercise 11.)

13. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$.

a. Is $\mathbf{w}$ in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? How many vectors are in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

b. How many vectors are in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

c. Is $\mathbf{w}$ in the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? Why?

14. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ be as in Exercise 13, and let $\mathbf{w} = \begin{bmatrix} 8 \\ 4 \\ 7 \end{bmatrix}$. Is $\mathbf{w}$ in the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? Why?

In Exercises 15–18, let W be the set of all vectors of the form shown, where a, b, and c represent arbitrary real numbers. In each case, either find a set S of vectors that spans W or give an example to show that W is *not* a vector space.

15. $\begin{bmatrix} 3a+b \\ 4 \\ a-5b \end{bmatrix}$

16. $\begin{bmatrix} -a+1 \\ a-6b \\ 2b+a \end{bmatrix}$

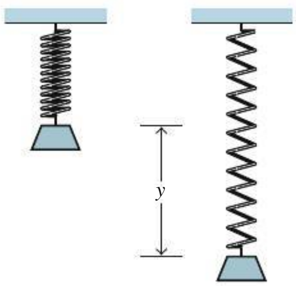
17. $\begin{bmatrix} a-b \\ b-c \\ c-a \\ b \end{bmatrix}$

18. $\begin{bmatrix} 4a+3b \\ 0 \\ a+b+c \\ c-2a \end{bmatrix}$

19. If a mass m is placed at the end of a spring, and if the mass is pulled downward and released, the mass–spring system will begin to oscillate. The displacement y of the mass from its resting position is given by a function of the form

$$y(t) = c_1 \cos \omega t + c_2 \sin \omega t \quad (5)$$

where ω is a constant that depends on the spring and the mass. (See the figure below.) Show that the set of all functions described in (5) (with ω fixed and c_1, c_2 arbitrary) is a vector space.



20. The set of all continuous real-valued functions defined on a closed interval $[a, b]$ in $\mathbb{R}$ is denoted by $C[a, b]$. This set is a subspace of the vector space of all real-valued functions defined on $[a, b]$.

a. What facts about continuous functions should be proved in order to demonstrate that $C[a, b]$ is indeed a subspace as claimed? (These facts are usually discussed in a calculus class.)

b. Show that $\{\mathbf{f} \text{ in } C[a, b] : \mathbf{f}(a) = \mathbf{f}(b)\}$ is a subspace of $C[a, b]$.

For fixed positive integers m and n, the set $M_{m\times n}$ of all $m \times n$ matrices is a vector space, under the usual operations of addition of matrices and multiplication by real scalars.

21. Determine if the set H of all matrices of the form $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ is a subspace of $M_{2\times 2}$.

22. Let F be a fixed 3×2 matrix, and let H be the set of all matrices A in $M_{2\times 4}$ with the property that $FA = 0$ (the zero matrix in $M_{3\times 4}$). Determine if H is a subspace of $M_{2\times 4}$.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** If $\mathbf{f}$ is a function in the vector space V of all real-valued functions on $\mathbb{R}$ and if $\mathbf{f}(t) = 0$ for some t, then $\mathbf{f}$ is the zero vector in V.

24. **(T/F)** A vector is any element of a vector space.

25. **(T/F)** An arrow in three-dimensional space can be considered to be a vector.

26. **(T/F)** If $\mathbf{u}$ is a vector in a vector space V, then $(-1)\,\mathbf{u}$ is the same as the negative of $\mathbf{u}$.

27. **(T/F)** A subset H of a vector space V is a subspace of V if the zero vector is in H.

28. **(T/F)** A vector space is also a subspace.

29. **(T/F)** A subspace is also a vector space.

30. **(T/F)** $\mathbb{R}^2$ is a subspace of $\mathbb{R}^3$.

31. **(T/F)** The polynomials of degree two or less are a subspace of the polynomials of degree three or less.

32. **(T/F)** A subset H of a vector space V is a subspace of V if the following conditions are satisfied: (i) the zero vector of V is in H, (ii) $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} + \mathbf{v}$ are in H, and (iii) c is a scalar and $c\mathbf{u}$ is in H.

Exercises 33–36 show how the axioms for a vector space V can be used to prove the elementary properties described after the definition of a vector space. Fill in the blanks with the appropriate axiom numbers. Because of Axiom 2, Axioms 4 and 5 imply, respectively, that $\mathbf{0} + \mathbf{u} = \mathbf{u}$ and $-\mathbf{u} + \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$.

33. Complete the following proof that the zero vector is unique. Suppose that $\mathbf{w}$ in V has the property that $\mathbf{u} + \mathbf{w} = \mathbf{w} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u}$ in V. In particular, $\mathbf{0} + \mathbf{w} = \mathbf{0}$. But $\mathbf{0} + \mathbf{w} = \mathbf{w}$, by Axiom ______. Hence $\mathbf{w} = \mathbf{0} + \mathbf{w} = \mathbf{0}$.

34. Complete the following proof that $-\mathbf{u}$ is the *unique vector* in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. Suppose that $\mathbf{w}$ satisfies $\mathbf{u} + \mathbf{w} = \mathbf{0}$. Adding $-\mathbf{u}$ to both sides, we have

$$(-\mathbf{u}) + [\mathbf{u} + \mathbf{w}] = (-\mathbf{u}) + \mathbf{0}$$

$$[(-\mathbf{u}) + \mathbf{u}] + \mathbf{w} = (-\mathbf{u}) + \mathbf{0} \qquad \text{by Axiom \_\_\_\_\_\_ (a)}$$

$$\mathbf{0} + \mathbf{w} = (-\mathbf{u}) + \mathbf{0} \qquad \text{by Axiom \_\_\_\_\_\_ (b)}$$

$$\mathbf{w} = -\mathbf{u} \qquad \text{by Axiom \_\_\_\_\_\_ (c)}$$

35. Fill in the missing axiom numbers in the following proof that $0\mathbf{u} = \mathbf{0}$ for every $\mathbf{u}$ in V.

$$0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u} \qquad \text{by Axiom \_\_\_\_\_\_ (a)}$$

Add the negative of $0\mathbf{u}$ to both sides:

$$0\mathbf{u} + (-0\mathbf{u}) = [0\mathbf{u} + 0\mathbf{u}] + (-0\mathbf{u})$$

$$0\mathbf{u} + (-0\mathbf{u}) = 0\mathbf{u} + [0\mathbf{u} + (-0\mathbf{u})] \quad \text{by Axiom \_\_\_\_ (b)}$$

$$\mathbf{0} = 0\mathbf{u} + \mathbf{0} \quad \text{by Axiom \_\_\_\_ (c)}$$

$$\mathbf{0} = 0\mathbf{u} \quad \text{by Axiom \_\_\_\_ (d)}$$

36. Fill in the missing axiom numbers in the following proof that $c\mathbf{0} = \mathbf{0}$ for every scalar c.

$$c\mathbf{0} = c(\mathbf{0} + \mathbf{0}) \quad \text{by Axiom \_\_\_\_ (a)}$$

$$= c\mathbf{0} + c\mathbf{0} \quad \text{by Axiom \_\_\_\_ (b)}$$

Add the negative of $c\mathbf{0}$ to both sides:

$$c\mathbf{0} + (-c\mathbf{0}) = [c\mathbf{0} + c\mathbf{0}] + (-c\mathbf{0})$$

$$c\mathbf{0} + (-c\mathbf{0}) = c\mathbf{0} + [c\mathbf{0} + (-c\mathbf{0})] \quad \text{by Axiom \_\_\_\_ (c)}$$

$$\mathbf{0} = c\mathbf{0} + \mathbf{0} \quad \text{by Axiom \_\_\_\_ (d)}$$

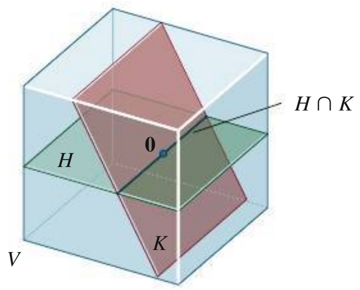
$$\mathbf{0} = c\mathbf{0} \quad \text{by Axiom \_\_\_\_ (e)}$$

37. Prove that $(-1)\mathbf{u} = -\mathbf{u}$. [*Hint:* Show that $\mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$. Use some axioms and the results of Exercises 34 and 35.]

38. Suppose $c\mathbf{u} = \mathbf{0}$ for some nonzero scalar c. Show that $\mathbf{u} = \mathbf{0}$. Mention the axioms or properties you use.

39. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in a vector space V, and let H be any subspace of V that contains both $\mathbf{u}$ and $\mathbf{v}$. Explain why H also contains Span $\{\mathbf{u}, \mathbf{v}\}$. This shows that Span $\{\mathbf{u}, \mathbf{v}\}$ is the smallest subspace of V that contains both $\mathbf{u}$ and $\mathbf{v}$.

40. Let H and K be subspaces of a vector space V. The **intersection** of H and K, written as $H \cap K$, is the set of $\mathbf{v}$ in V that belong to both H and K. Show that $H \cap K$ is a subspace of V. (See the figure.) Give an example in $\mathbb{R}^2$ to show that the union of two subspaces is not, in general, a subspace.



41. Given subspaces H and K of a vector space V, the **sum** of H and K, written as $H + K$, is the set of all vectors in V that can be written as the sum of two vectors, one in H and the other in K; that is,

$$H + K = \{\mathbf{w} : \mathbf{w} = \mathbf{u} + \mathbf{v} \text{ for some } \mathbf{u} \text{ in } H \text{ and some } \mathbf{v} \text{ in } K\}$$

a. Show that $H + K$ is a subspace of V.

b. Show that H is a subspace of $H + K$ and K is a subspace of $H + K$.

42. Suppose $\mathbf{u}_1, \dots, \mathbf{u}_p$ and $\mathbf{v}_1, \dots, \mathbf{v}_q$ are vectors in a vector space V, and let

$$H = \text{Span}\,\{\mathbf{u}_1, \dots, \mathbf{u}_p\} \text{ and } K = \text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_q\}$$

Show that $H + K = \text{Span}\,\{\mathbf{u}_1, \dots, \mathbf{u}_p, \mathbf{v}_1, \dots, \mathbf{v}_q\}$.

T 43. Show that $\mathbf{w}$ is in the subspace of $\mathbb{R}^4$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, where

$$\mathbf{w} = \begin{bmatrix} 9 \\ -4 \\ -4 \\ 7 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 8 \\ -4 \\ -3 \\ 9 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -4 \\ 3 \\ -2 \\ -8 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -7 \\ 6 \\ -5 \\ -18 \end{bmatrix}$$

T 44. Determine if $\mathbf{y}$ is in the subspace of $\mathbb{R}^4$ spanned by the columns of A, where

$$\mathbf{y} = \begin{bmatrix} -4 \\ -8 \\ 6 \\ -5 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & -5 & -9 \\ 8 & 7 & -6 \\ -5 & -8 & 3 \\ 2 & -2 & -9 \end{bmatrix}$$

T 45. The vector space $H = \text{Span}\,\{1, \cos^2 t, \cos^4 t, \cos^6 t\}$ contains at least two interesting functions that will be used in a later exercise:

$$\mathbf{f}(t) = 1 - 8\cos^2 t + 8\cos^4 t$$

$$\mathbf{g}(t) = -1 + 18\cos^2 t - 48\cos^4 t + 32\cos^6 t$$

Study the graph of $\mathbf{f}$ for $0 \le t \le 2\pi$, and guess a simple formula for $\mathbf{f}(t)$. Verify your conjecture by graphing the difference between $1 + \mathbf{f}(t)$ and your formula for $\mathbf{f}(t)$. (Hopefully, you will see the constant function 1.) Repeat for $\mathbf{g}$.

T 46. Repeat Exercise 45 for the functions

$$\mathbf{f}(t) = 3\sin t - 4\sin^3 t$$

$$\mathbf{g}(t) = 1 - 8\sin^2 t + 8\sin^4 t$$

$$\mathbf{h}(t) = 5\sin t - 20\sin^3 t + 16\sin^5 t$$

in the vector space Span $\{1, \sin t, \sin^2 t, \dots, \sin^5 t\}$.

Solutions to Practice Problems

1. Take any $\mathbf{u}$ in H—say, $\mathbf{u} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$—and take any $c \neq 1$—say, $c = 2$. Then $c\mathbf{u} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$. If this is in H, then there is some s such that

$$\begin{bmatrix} 3s \\ 2 + 5s \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

That is, $s = 2$ and $s = 12/5$, which is impossible. So $2\mathbf{u}$ is not in H and H is not a vector space.

2. $\mathbf{v}_1 = 1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p$. This expresses $\mathbf{v}_1$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$, so $\mathbf{v}_1$ is in W. In general, $\mathbf{v}_k$ is in W because

$$\mathbf{v}_k = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{k-1} + 1\mathbf{v}_k + 0\mathbf{v}_{k+1} + \cdots + 0\mathbf{v}_p$$

3. The subset S is a subspace of $M_{3\times 3}$ since it satisfies all three of the requirements listed in the definition of a subspace:
   a. Observe that the $\mathbf{0}$ in $M_{3\times 3}$ is the 3×3 zero matrix and since $\mathbf{0}^T = \mathbf{0}$, the matrix $\mathbf{0}$ is symmetric and hence $\mathbf{0}$ is in S.
   b. Let A and B in S. Notice that A and B are 3×3 symmetric matrices so $A^T = A$ and $B^T = B$. By the properties of transposes of matrices, $(A + B)^T = A^T + B^T = A + B$. Thus $A + B$ is symmetric and hence $A + B$ is in S.
   c. Let A be in S and let c be a scalar. Since A is symmetric, by the properties of symmetric matrices, $(cA)^T = c(A^T) = cA$. Thus cA is also a symmetric matrix and hence cA is in S.

4.2 Null Spaces, Column Spaces, Row Spaces, and Linear Transformations

In applications of linear algebra, subspaces of $\mathbb{R}^n$ usually arise in one of two ways: (1) as the set of all solutions to a system of homogeneous linear equations or (2) as the set of all linear combinations of certain specified vectors. In this section, we compare and contrast these two descriptions of subspaces, allowing us to practice using the concept of a subspace. Actually, as you will soon discover, we have been working with subspaces ever since Section 1.3. The main new feature here is the terminology. The section concludes with a discussion of the kernel and range of a linear transformation.

The Null Space of a Matrix

Consider the following system of homogeneous equations:

$$\begin{aligned} x_1 - 3x_2 - 2x_3 &= 0 \\ -5x_1 + 9x_2 + x_3 &= 0 \end{aligned} \tag{1}$$

In matrix form, this system is written as $A\mathbf{x} = \mathbf{0}$, where

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

Recall that the set of all $\mathbf{x}$ that satisfy (1) is called the **solution set** of the system (1). Often it is convenient to relate this set directly to the matrix A and the equation $A\mathbf{x} = \mathbf{0}$. We call the set of $\mathbf{x}$ that satisfy $A\mathbf{x} = \mathbf{0}$ the **null space** of the matrix A.

DEFINITION

The **null space** of an $m \times n$ matrix A, written as Nul A, is the set of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\text{Nul } A = \{\mathbf{x} : \mathbf{x} \text{ is in } \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}$$

A more dynamic description of Nul A is the set of all $\mathbf{x}$ in $\mathbb{R}^n$ that are mapped into the zero vector of $\mathbb{R}^m$ via the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. See Figure 1.

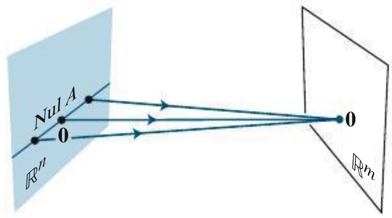


FIGURE 1

EXAMPLE 1 Let A be the matrix in (2), and let $\mathbf{u} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$. Determine if $\mathbf{u}$ belongs to the null space of A.

SOLUTION To test if $\mathbf{u}$ satisfies $A\mathbf{u} = \mathbf{0}$, simply compute

$$A\mathbf{u} = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 - 9 + 4 \\ -25 + 27 - 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Thus $\mathbf{u}$ is in Nul A. ■

The term *space* in *null space* is appropriate because the null space of a matrix is a vector space, as shown in the next theorem.

THEOREM 2

The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

PROOF Certainly Nul A is a subset of $\mathbb{R}^n$ because A has n columns. We must show that Nul A satisfies the three properties of a subspace. Of course, $\mathbf{0}$ is in Nul A. Next, let $\mathbf{u}$ and $\mathbf{v}$ represent any two vectors in Nul A. Then

$$A\mathbf{u} = \mathbf{0} \quad \text{and} \quad A\mathbf{v} = \mathbf{0}$$

To show that $\mathbf{u} + \mathbf{v}$ is in Nul A, we must show that $A(\mathbf{u} + \mathbf{v}) = \mathbf{0}$. Using a property of matrix multiplication, compute

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Thus $\mathbf{u} + \mathbf{v}$ is in Nul A, and Nul A is closed under vector addition. Finally, if c is any scalar, then

$$A(c\mathbf{u}) = c(A\mathbf{u}) = c(\mathbf{0}) = \mathbf{0}$$

which shows that $c\mathbf{u}$ is in Nul A. Thus Nul A is a subspace of $\mathbb{R}^n$. ■

EXAMPLE 2 Let H be the set of all vectors in $\mathbb{R}^4$ whose coordinates a, b, c, d satisfy the equations $a - 2b + 5c = d$ and $c - a = b$. Show that H is a subspace of $\mathbb{R}^4$.

SOLUTION Rearrange the equations that describe the elements of H, and note that H is the set of all solutions of the following system of homogeneous linear equations:

$$\begin{aligned} a - 2b + 5c - d &= 0 \\ -a - b + c \phantom{{}-d} &= 0 \end{aligned}$$

By Theorem 2, H is a subspace of $\mathbb{R}^4$. ■

It is important that the linear equations defining the set H are homogeneous. Otherwise, the set of solutions will definitely *not* be a subspace (because the zero vector is not a solution of a nonhomogeneous system). Also, in some cases, the set of solutions could be empty.

An Explicit Description of Nul *A*

There is no obvious relation between vectors in Nul A and the entries in A. We say that Nul A is defined *implicitly*, because it is defined by a condition that must be checked. No explicit list or description of the elements in Nul A is given. However, *solving* the equation $A\mathbf{x} = \mathbf{0}$ amounts to producing an *explicit* description of Nul A. The next example reviews the procedure from Section 1.5.

EXAMPLE 3 Find a spanning set for the null space of the matrix

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

SOLUTION The first step is to find the general solution of $A\mathbf{x} = \mathbf{0}$ in terms of free variables. Row reduce the augmented matrix $[\, A \quad \mathbf{0}\,]$ to *reduced* echelon form in order to write the basic variables in terms of the free variables:

$$\begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 - 2x_2 \phantom{{}+x_3} - x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The general solution is $x_1 = 2x_2 + x_4 - 3x_5$, $x_3 = -2x_4 + 2x_5$, with x_2, x_4, and x_5 free. Next, decompose the vector giving the general solution into a linear combination of vectors where *the weights are the free variables*. That is,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \underbrace{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\uparrow\ \mathbf{u}} + x_4 \underbrace{\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}}_{\uparrow\ \mathbf{v}} + x_5 \underbrace{\begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}}_{\uparrow\ \mathbf{w}}$$

$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w} \tag{3}$$

Every linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ is an element of Nul A and vice versa. Thus $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a spanning set for Nul A. ■

Two points should be made about the solution of Example 3 that apply to all problems of this type where Nul A contains nonzero vectors. We will use these facts later.

1. The spanning set produced by the method in Example 3 is automatically linearly independent because the free variables are the weights on the spanning vectors. For instance, look at the 2nd, 4th, and 5th entries in the solution vector in (3) and note that $x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$ can be $\mathbf{0}$ only if the weights x_2, x_4, and x_5 are all zero.
2. When Nul A contains nonzero vectors, the number of vectors in the spanning set for Nul A equals the number of free variables in the equation $A\mathbf{x} = \mathbf{0}$.

The Column Space of a Matrix

Another important subspace associated with a matrix is its column space. Unlike the null space, the column space is defined explicitly via linear combinations.

DEFINITION

The **column space** of an $m \times n$ matrix A, written as Col A, is the set of all linear combinations of the columns of A. If $A = [\,\mathbf{a}_1 \ \cdots \ \mathbf{a}_n\,]$, then

$$\text{Col } A = \text{Span}\,\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$$

Since Span $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a subspace, by Theorem 1, the next theorem follows from the definition of Col A and the fact that the columns of A are in $\mathbb{R}^m$.

THEOREM 3

The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.

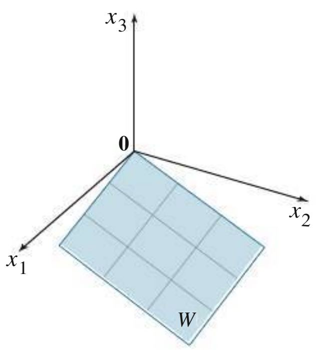
Note that a typical vector in Col A can be written as $A\mathbf{x}$ for some $\mathbf{x}$ because the notation $A\mathbf{x}$ stands for a linear combination of the columns of A. That is,

$$\boxed{\text{Col } A = \{\mathbf{b} : \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \text{ in } \mathbb{R}^n\}}$$

The notation $A\mathbf{x}$ for vectors in Col A also shows that Col A is the *range* of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. We will return to this point of view at the end of the section.

EXAMPLE 4 Find a matrix A such that $W = \text{Col } A$.

$$W = \left\{ \begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$$



SOLUTION First, write W as a set of linear combinations.

$$W = \left\{ a\begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix} + b\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\} = \text{Span}\left\{ \begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Second, use the vectors in the spanning set as the columns of A. Let $A = \begin{bmatrix} 6 & -1 \\ 1 & 1 \\ -7 & 0 \end{bmatrix}$. Then $W = \text{Col } A$, as desired. ■

Recall from Theorem 4 in Section 1.4 that the columns of A span $\mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$. We can restate this fact as follows:

The column space of an $m \times n$ matrix A is all of $\mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

The Row Space

If A is an $m \times n$ matrix, each row of A has n entries and thus can be identified with a vector in $\mathbb{R}^n$. The set of all linear combinations of the row vectors is called the **row space** of A and is denoted by Row A. Each row has n entries, so Row A is a subspace of $\mathbb{R}^n$. Since the rows of A are identified with the columns of A^T, we could also write Col A^T in place of Row A.

EXAMPLE 5 Let

$$A = \begin{bmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix} \quad \text{and} \quad \begin{aligned} \mathbf{r}_1 &= (-2, -5, 8, 0, -17) \\ \mathbf{r}_2 &= (1, 3, -5, 1, 5) \\ \mathbf{r}_3 &= (3, 11, -19, 7, 1) \\ \mathbf{r}_4 &= (1, 7, -13, 5, -3) \end{aligned}$$

The row space of A is the subspace of $\mathbb{R}^5$ spanned by $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$. That is, Row A = Span $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$. It is natural to write row vectors horizontally; however, they may also be written as column vectors if that is more convenient. ■

The Contrast Between Nul *A* and Col *A*

It is natural to wonder how the null space and column space of a matrix are related. In fact, the two spaces are quite dissimilar, as Examples 6–8 will show. Nevertheless, a surprising connection between the null space and column space will emerge in Section 4.5, after more theory is available.

EXAMPLE 6 Let

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

a. If the column space of A is a subspace of $\mathbb{R}^k$, what is k?

b. If the null space of A is a subspace of $\mathbb{R}^k$, what is k?

SOLUTION

a. The columns of A each have three entries, so Col A is a subspace of $\mathbb{R}^k$, where $k = 3$.

b. A vector $\mathbf{x}$ such that $A\mathbf{x}$ is defined must have four entries, so Nul A is a subspace of $\mathbb{R}^k$, where $k = 4$. ■

When a matrix is not square, as in Example 6, the vectors in Nul A and Col A live in entirely different "universes." For example, no linear combination of vectors in $\mathbb{R}^3$ can produce a vector in $\mathbb{R}^4$. When A is square, Nul A and Col A do have the zero vector in common, and in special cases it is possible that some nonzero vectors belong to both Nul A and Col A.

EXAMPLE 7 With A as in Example 6, find a nonzero vector in Col A and a nonzero vector in Nul A.

SOLUTION It is easy to find a vector in Col A. Any column of A will do, say, $\begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$.

To find a nonzero vector in Nul A, row reduce the augmented matrix $[\, A \quad \mathbf{0} \,]$ and obtain

$$[\, A \quad \mathbf{0} \,] \sim \begin{bmatrix} 1 & 0 & 9 & 0 & 0 \\ 0 & 1 & -5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Thus, if $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{0}$, then $x_1 = -9x_3$, $x_2 = 5x_3$, $x_4 = 0$, and x_3 is free. Assigning a nonzero value to x_3—say, $x_3 = 1$—we obtain a vector in Nul A, namely, $\mathbf{x} = (-9, 5, 1, 0)$. ■

EXAMPLE 8 With A as in Example 6, let $\mathbf{u} = \begin{bmatrix} 3 \\ -2 \\ -1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}$.

a. Determine if $\mathbf{u}$ is in Nul A. Could $\mathbf{u}$ be in Col A?

b. Determine if $\mathbf{v}$ is in Col A. Could $\mathbf{v}$ be in Nul A?

SOLUTION

a. An explicit description of Nul A is not needed here. Simply compute the product $A\mathbf{u}$.

$$A\mathbf{u} = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & 7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Obviously, $\mathbf{u}$ is *not* a solution of $A\mathbf{x} = \mathbf{0}$, so $\mathbf{u}$ is not in Nul A. Also, with four entries, $\mathbf{u}$ could not possibly be in Col A, since Col A is a subspace of $\mathbb{R}^3$.

b. Reduce $[\, A \quad \mathbf{v} \,]$ to an echelon form.

$$[\, A \quad \mathbf{v} \,] = \begin{bmatrix} 2 & 4 & -2 & 1 & 3 \\ -2 & -5 & 7 & 3 & -1 \\ 3 & 7 & -8 & 6 & 3 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -2 & 1 & 3 \\ 0 & 1 & -5 & -4 & -2 \\ 0 & 0 & 0 & 17 & 1 \end{bmatrix}$$

At this point, it is clear that the equation $A\mathbf{x} = \mathbf{v}$ is consistent, so $\mathbf{v}$ is in Col A. With only three entries, $\mathbf{v}$ could not possibly be in Nul A, since Nul A is a subspace of $\mathbb{R}^4$. ■

The table on page 217 summarizes what we have learned about Nul A and Col A. Item 8 is a restatement of Theorems 11 and 12(a) in Section 1.9.

Kernel and Range of a Linear Transformation

Subspaces of vector spaces other than $\mathbb{R}^n$ are often described in terms of a linear transformation instead of a matrix. To make this precise, we generalize the definition given in Section 1.8.

Contrast Between Nul A and Col A for an $m \times n$ Matrix A

Nul A	**Col A**
1. Nul A is a subspace of $\mathbb{R}^n$.	**1.** Col A is a subspace of $\mathbb{R}^m$.
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	**2.** Col A is explicitly defined; that is, you are told how to build vectors in Col A.
3. It takes time to find vectors in Nul A. Row operations on $[\,A \quad \mathbf{0}\,]$ are required.	**3.** It is easy to find vectors in Col A. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A.	**4.** There is an obvious relation between Col A and the entries in A, since each column of A is in Col A.
5. A typical vector $\mathbf{v}$ in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	**5.** A typical vector $\mathbf{v}$ in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector $\mathbf{v}$, it is easy to tell if $\mathbf{v}$ is in Nul A. Just compute $A\mathbf{v}$.	**6.** Given a specific vector $\mathbf{v}$, it may take time to tell if $\mathbf{v}$ is in Col A. Row operations on $[\,A \quad \mathbf{v}\,]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	**7.** Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	**8.** Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

DEFINITION

A **linear transformation** T from a vector space V into a vector space W is a rule that assigns to each vector $\mathbf{x}$ in V a unique vector $T(\mathbf{x})$ in W, such that

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$, $\mathbf{v}$ in V, and

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u}$ in V and all scalars c.

The **kernel** (or **null space**) of such a T is the set of all $\mathbf{u}$ in V such that $T(\mathbf{u}) = \mathbf{0}$ (the zero vector in W). The **range** of T is the set of all vectors in W of the form $T(\mathbf{x})$ for some $\mathbf{x}$ in V. If T happens to arise as a matrix transformation—say, $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A—then the kernel and the range of T are just the null space and the column space of A, as defined earlier.

It is not difficult to show that the kernel of T is a subspace of V. The proof is essentially the same as the one for Theorem 2. Also, the range of T is a subspace of W. See Figure 2 and Exercise 42.

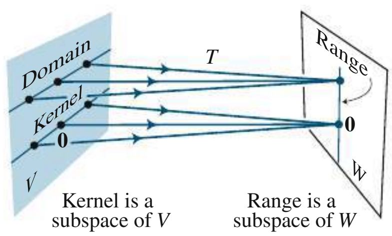


FIGURE 2 Subspaces associated with a linear transformation.

In applications, a subspace usually arises as either the kernel or the range of an appropriate linear transformation. For instance, the set of all solutions of a homogeneous linear differential equation turns out to be the kernel of a linear transformation. Typically, such a linear transformation is described in terms of one or more derivatives of a function. To explain this in any detail would take us too far afield at this point. So we consider only two examples. The first explains why the operation of differentiation is a linear transformation.

EXAMPLE 9 (Calculus required) Let V be the vector space of all real-valued functions f defined on an interval $[a, b]$ with the property that they are differentiable and their derivatives are continuous functions on $[a, b]$. Let W be the vector space $C[a, b]$ of all continuous functions on $[a, b]$, and let $D : V \to W$ be the transformation that changes f in V into its derivative f'. In calculus, two simple differentiation rules are

$$D(f + g) = D(f) + D(g) \quad \text{and} \quad D(cf) = cD(f)$$

That is, D is a linear transformation. It can be shown that the kernel of D is the set of constant functions on $[a, b]$ and the range of D is the set W of all continuous functions on $[a, b]$. ■

EXAMPLE 10 (Calculus required) The differential equation

$$y'' + \omega^2 y = 0 \tag{4}$$

where ω is a constant, is used to describe a variety of physical systems, such as the vibration of a weighted spring, the movement of a pendulum, and the voltage in an inductance-capacitance electrical circuit. The set of solutions of (4) is precisely the kernel of the linear transformation that maps a function $y = f(t)$ into the function $f''(t) + \omega^2 f(t)$. Finding an explicit description of this vector space is a problem in differential equations. The solution set turns out to be the space described in Exercise 19 in Section 4.1. ■

A common technique used in the stock market is technical analysis. Statistical trends gathered from stock-trading activity, such as price movement and volume, are analyzed. Technical analysts focus on patterns of stock-price movements, trading signals, and various other analytical charting tools to evaluate a security's strength or weakness. A moving average is a commonly used indicator in technical analysis. It smooths out price action by filtering out the effects from random price fluctuations. In the final example for this section, we examine the linear transformation that creates the two-day moving average from a "signal" of daily prices. We will look at moving average transformations that average over a longer period of time in Section 4.7.

EXAMPLE 11 Let $\{p_k\}$ in $\mathbb{S}$ represent the price of a stock that has been recorded daily over an extended period of time. Note that we can assume that $p_k = 0$ for k outside the time period under study. To create a two-day moving average, the mapping $M_2 : \mathbb{S} \to \mathbb{S}$ defined by $M_2(\{p_k\}) = \left\{ \dfrac{p_k + p_{k-1}}{2} \right\}$ is applied to the data. Show that M_2 is a linear transformation and find its kernel.

SOLUTION To see that M_2 is a linear transformation, observe that for two signals $\{p_k\}$ and $\{q_k\}$ in $\mathbb{S}$ and any scalar c,

$$M_2(\{p_k\} + \{q_k\}) = M_2(\{p_k + q_k\}) = \left\{\frac{p_k + q_k + p_{k-1} + q_{k-1}}{2}\right\}$$
$$= \left\{\frac{p_k + p_{k-1}}{2}\right\} + \left\{\frac{q_k + q_{k-1}}{2}\right\}$$
$$= M_2(\{p_k\}) + M_2(\{q_k\})$$

and

$$M_2(c\{p_k\}) = M_2(\{cp_k\}) = \left\{\frac{cp_k + cp_{k-1}}{2}\right\} = c\left\{\frac{p_k + p_{k-1}}{2}\right\} = cM_2(\{p_k\})$$

thus M_2 is a linear transformation.

To find the kernel of M_2, notice that $\{p_k\}$ is in the kernel if and only if $\frac{p_k + p_{k-1}}{2} = 0$ for all k, and hence $p_k = -p_{k-1}$. Since this relationship is true for all integers k, it can be applied recursively resulting in $p_k = -p_{k-1} = (-1)^2 p_{k-2} = (-1)^3 p_{k-3} \ldots$. Working out from $k = 0$, any signal in the kernel can be written as $p_k = p_0(-1)^k$, a multiple of the alternating signal described by $\{(-1)^k\}$. Since the kernel of the two-day moving average function consists of all multiples of the alternating sequence, it smooths out daily fluctuations, without leveling out overall trends. (See Figure 3.)

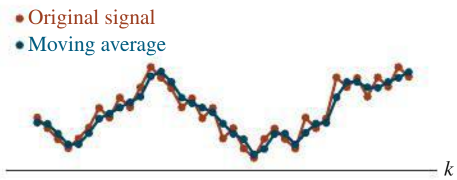


FIGURE 3

■

Practice Problems

1. Let $W = \left\{\begin{bmatrix} a \\ b \\ c \end{bmatrix} : a - 3b - c = 0\right\}$. Show in two different ways that W is a subspace of $\mathbb{R}^3$. (Use two theorems.)
2. Let $A = \begin{bmatrix} 7 & -3 & 5 \\ -4 & 1 & -5 \\ -5 & 2 & -4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 7 \\ 6 \\ -3 \end{bmatrix}$. Suppose you know that the equations $A\mathbf{x} = \mathbf{v}$ and $A\mathbf{x} = \mathbf{w}$ are both consistent. What can you say about the equation $A\mathbf{x} = \mathbf{v} + \mathbf{w}$?
3. Let A be an $n \times n$ matrix. If Col A = Nul A, show that Nul $A^2 = \mathbb{R}^n$.

4.2 Exercises

1. Determine if $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -4 \end{bmatrix}$ is in Nul A, where
$$A = \begin{bmatrix} 3 & -5 & -3 \\ 6 & -2 & 0 \\ -8 & 4 & 1 \end{bmatrix}.$$

2. Determine if $\mathbf{w} = \begin{bmatrix} 5 \\ -3 \\ 2 \end{bmatrix}$ is in Nul A, where
$$A = \begin{bmatrix} 5 & 21 & 19 \\ 13 & 23 & 2 \\ 8 & 14 & 1 \end{bmatrix}.$$

In Exercises 3–6, find an explicit description of Nul A by listing vectors that span the null space.

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

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

5. $A = \begin{bmatrix} 1 & -2 & 0 & 4 & 0 \\ 0 & 0 & 1 & -9 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

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

In Exercises 7–14, either use an appropriate theorem to show that the given set, W, is a vector space, or find a specific example to the contrary.

7. $\left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a + b + c = 2 \right\}$

8. $\left\{ \begin{bmatrix} r \\ s \\ t \end{bmatrix} : 5r - 1 = s + 2t \right\}$

9. $\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : \begin{matrix} a - 2b = 4c \\ 2a = c + 3d \end{matrix} \right\}$

10. $\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : \begin{matrix} a + 3b = c \\ b + c + a = d \end{matrix} \right\}$

11. $\left\{ \begin{bmatrix} b - 2d \\ 5 + d \\ b + 3d \\ d \end{bmatrix} : b, d \text{ real} \right\}$

12. $\left\{ \begin{bmatrix} b - 5d \\ 2b \\ 2d + 1 \\ d \end{bmatrix} : b, d \text{ real} \right\}$

13. $\left\{ \begin{bmatrix} c - 6d \\ d \\ c \end{bmatrix} : c, d \text{ real} \right\}$

14. $\left\{ \begin{bmatrix} -a + 2b \\ a - 2b \\ 3a - 6b \end{bmatrix} : a, b \text{ real} \right\}$

In Exercises 15 and 16, find A such that the given set is Col A.

15. $\left\{ \begin{bmatrix} 2s + 3t \\ r + s - 2t \\ 4r + s \\ 3r - s - t \end{bmatrix} : r, s, t \text{ real} \right\}$

16. $\left\{ \begin{bmatrix} b - c \\ 2b + c + d \\ 5c - 4d \\ d \end{bmatrix} : b, c, d \text{ real} \right\}$

For the matrices in Exercises 17–20, (a) find k such that Nul A is a subspace of $\mathbb{R}^k$, and (b) find k such that Col A is a subspace of $\mathbb{R}^k$.

17. $A = \begin{bmatrix} 2 & -6 \\ -1 & 3 \\ -4 & 12 \\ 3 & -9 \end{bmatrix}$

18. $A = \begin{bmatrix} 7 & -2 & 0 \\ -2 & 0 & -5 \\ 0 & -5 & 7 \\ -5 & 7 & -2 \end{bmatrix}$

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

20. $A = \begin{bmatrix} 1 & -3 & 9 & 0 & -5 \end{bmatrix}$

21. With A as in Exercise 17, find a nonzero vector in Nul A, a nonzero vector in Col A, and a nonzero vector in Row A.

22. With A as in Exercise 3, find a nonzero vector in Nul A, a nonzero vector in Col A, and a nonzero vector in Row A.

23. Let $A = \begin{bmatrix} -6 & 12 \\ -3 & 6 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Determine if $\mathbf{w}$ is in Col A. Is $\mathbf{w}$ in Nul A?

24. Let $A = \begin{bmatrix} -8 & -2 & -9 \\ 6 & 4 & 8 \\ 4 & 0 & 4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Determine if $\mathbf{w}$ is in Col A. Is $\mathbf{w}$ in Nul A?

In Exercises 25–38, A denotes an $m \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

25. **(T/F)** The null space of A is the solution set of the equation $A\mathbf{x} = \mathbf{0}$.

26. **(T/F)** A null space is a vector space.

27. **(T/F)** The null space of an $m \times n$ matrix is in $\mathbb{R}^m$.

28. **(T/F)** The column space of an $m \times n$ matrix is in $\mathbb{R}^m$.

29. **(T/F)** The column space of A is the range of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

30. **(T/F)** Col A is the set of all solutions of $A\mathbf{x} = \mathbf{b}$.

31. **(T/F)** If the equation $A\mathbf{x} = \mathbf{b}$ is consistent, then Col $A = \mathbb{R}^m$.

32. **(T/F)** Nul A is the kernel of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

33. **(T/F)** The kernel of a linear transformation is a vector space.

34. **(T/F)** The range of a linear transformation is a vector space.

35. **(T/F)** Col A is the set of all vectors that can be written as $A\mathbf{x}$ for some $\mathbf{x}$.

36. **(T/F)** The set of all solutions of a homogeneous linear differential equation is the kernel of a linear transformation.

37. **(T/F)** The row space of A is the same as the column space of A^T.

38. **(T/F)** The null space of A is the same as the row space of A^T.

39. It can be shown that a solution of the system below is $x_1 = 3$, $x_2 = 2$, and $x_3 = -1$. Use this fact and the theory from this section to explain why another solution is $x_1 = 30$, $x_2 = 20$, and $x_3 = -10$. (Observe how the solutions are related, but make no other calculations.)

$$\begin{aligned} x_1 - 3x_2 - 3x_3 &= 0 \\ -2x_1 + 4x_2 + 2x_3 &= 0 \\ -x_1 + 5x_2 + 7x_3 &= 0 \end{aligned}$$

40. Consider the following two systems of equations:

$$\begin{aligned} 5x_1 + x_2 - 3x_3 &= 0 \\ -9x_1 + 2x_2 + 5x_3 &= 1 \\ 4x_1 + x_2 - 6x_3 &= 9 \end{aligned} \qquad \begin{aligned} 5x_1 + x_2 - 3x_3 &= 0 \\ -9x_1 + 2x_2 + 5x_3 &= 5 \\ 4x_1 + x_2 - 6x_3 &= 45 \end{aligned}$$

It can be shown that the first system has a solution. Use this fact and the theory from this section to explain why the second system must also have a solution. (Make no row operations.)

41. Prove Theorem 3 as follows: Given an $m \times n$ matrix A, an element in Col A has the form $A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^n$. Let $A\mathbf{x}$ and $A\mathbf{w}$ represent any two vectors in Col A.

a. Explain why the zero vector is in Col A.

b. Show that the vector $A\mathbf{x} + A\mathbf{w}$ is in Col A.

c. Given a scalar c, show that $c(A\mathbf{x})$ is in Col A.

42. Let $T : V \to W$ be a linear transformation from a vector space V into a vector space W. Prove that the range of T is a subspace of W. [*Hint:* Typical elements of the range have the form $T(\mathbf{x})$ and $T(\mathbf{w})$ for some $\mathbf{x}$, $\mathbf{w}$ in V.]

43. Define $T : \mathbb{P}_2 \to \mathbb{R}^2$ by $T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(1) \end{bmatrix}$. For instance, if $\mathbf{p}(t) = 3 + 5t + 7t^2$, then $T(\mathbf{p}) = \begin{bmatrix} 3 \\ 15 \end{bmatrix}$.

a. Show that T is a linear transformation. [*Hint:* For arbitrary polynomials $\mathbf{p}$, $\mathbf{q}$ in $\mathbb{P}_2$, compute $T(\mathbf{p} + \mathbf{q})$ and $T(c\mathbf{p})$.]

b. Find a polynomial $\mathbf{p}$ in $\mathbb{P}_2$ that spans the kernel of T, and describe the range of T.

44. Define a linear transformation $T : \mathbb{P}_2 \to \mathbb{R}^2$ by $T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(0) \end{bmatrix}$. Find polynomials $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbb{P}_2$ that span the kernel of T, and describe the range of T.

45. Let $M_{2\times 2}$ be the vector space of all 2×2 matrices, and define $T : M_{2\times 2} \to M_{2\times 2}$ by $T(A) = A + A^T$, where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

a. Show that T is a linear transformation.

b. Let B be any element of $M_{2\times 2}$ such that $B^T = B$. Find an A in $M_{2\times 2}$ such that $T(A) = B$.

c. Show that the range of T is the set of B in $M_{2\times 2}$ with the property that $B^T = B$.

d. Describe the kernel of T.

46. (*Calculus required*) Define $T : C[0, 1] \to C[0, 1]$ as follows: For $\mathbf{f}$ in $C[0, 1]$, let $T(\mathbf{f})$ be the antiderivative $\mathbf{F}$ of $\mathbf{f}$ such that $\mathbf{F}(0) = 0$. Show that T is a linear transformation, and describe the kernel of T. (See the notation in Exercise 20 of Section 4.1.)

47. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. Given a subspace U of V, let $T(U)$ denote the set of all images of the form $T(\mathbf{x})$, where $\mathbf{x}$ is in U. Show that $T(U)$ is a subspace of W.

48. Given $T : V \to W$ as in Exercise 47, and given a subspace Z of W, let U be the set of all $\mathbf{x}$ in V such that $T(\mathbf{x})$ is in Z. Show that U is a subspace of V.

T **49.** Determine whether $\mathbf{w}$ is in the column space of A, the null space of A, or both, where

$$\mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -3 \end{bmatrix}, \quad A = \begin{bmatrix} 7 & 6 & -4 & 1 \\ -5 & -1 & 0 & -2 \\ 9 & -11 & 7 & -3 \\ 19 & -9 & 7 & 1 \end{bmatrix}$$

T **50.** Determine whether $\mathbf{w}$ is in the column space of A, the null space of A, or both, where

$$\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} -8 & 5 & -2 & 0 \\ -5 & 2 & 1 & -2 \\ 10 & -8 & 6 & -3 \\ 3 & -2 & 1 & 0 \end{bmatrix}$$

T **51.** Let $\mathbf{a}_1, \dots, \mathbf{a}_5$ denote the columns of the matrix A, where

$$A = \begin{bmatrix} 5 & 1 & 2 & 2 & 0 \\ 3 & 3 & 2 & -1 & -12 \\ 8 & 4 & 4 & -5 & 12 \\ 2 & 1 & 1 & 0 & -2 \end{bmatrix}, \quad B = [\, \mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_4 \,]$$

a. Explain why $\mathbf{a}_3$ and $\mathbf{a}_5$ are in the column space of B.

b. Find a set of vectors that spans Nul A.

c. Let $T : \mathbb{R}^5 \to \mathbb{R}^4$ be defined by $T(\mathbf{x}) = A\mathbf{x}$. Explain why T is neither one-to-one nor onto.

T **52.** Let $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$ and $K = \text{Span}\,\{\mathbf{v}_3, \mathbf{v}_4\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 3 \\ 8 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} 0 \\ -12 \\ -28 \end{bmatrix}.$$

Then H and K are subspaces of $\mathbb{R}^3$. In fact, H and K are planes in $\mathbb{R}^3$ through the origin, and they intersect in a line through $\mathbf{0}$. Find a nonzero vector $\mathbf{w}$ that generates that line. [*Hint:* $\mathbf{w}$ can be written as $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ and also as $c_3\mathbf{v}_3 + c_4\mathbf{v}_4$. To build $\mathbf{w}$, solve the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_3\mathbf{v}_3 + c_4\mathbf{v}_4$ for the unknown c_j's.]

STUDY GUIDE offers additional resources for mastering vector spaces, subspaces, and column row, and null spaces.

Solutions to Practice Problems

1. *First method:* W is a subspace of $\mathbb{R}^3$ by Theorem 2 because W is the set of all solutions to a system of homogeneous linear equations (where the system has only one equation). Equivalently, W is the null space of the 1×3 matrix $A = [\,1 \quad -3 \quad -1\,]$.
   Second method: Solve the equation $a - 3b - c = 0$ for the leading variable a in terms of the free variables b and c. Any solution has the form $\begin{bmatrix} 3b+c \\ b \\ c \end{bmatrix}$, where b and c are arbitrary, and

$$\begin{bmatrix} 3b+c \\ b \\ c \end{bmatrix} = b\underbrace{\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}}_{\uparrow\ \mathbf{v}_1} + c\underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\uparrow\ \mathbf{v}_2}$$

   This calculation shows that $W = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$. Thus W is a subspace of $\mathbb{R}^3$ by Theorem 1. We could also solve the equation $a - 3b - c = 0$ for b or c and get alternative descriptions of W as a set of linear combinations of two vectors.
2. Both $\mathbf{v}$ and $\mathbf{w}$ are in Col A. Since Col A is a vector space, $\mathbf{v} + \mathbf{w}$ must be in Col A. That is, the equation $A\mathbf{x} = \mathbf{v} + \mathbf{w}$ is consistent.
3. Let $\mathbf{x}$ be any vector in $\mathbb{R}^n$. Notice $A\mathbf{x}$ is in Col A, since it is a linear combination of the columns of A. Since Col A = Nul A, the vector $A\mathbf{x}$ is also in Nul A. Hence $A^2\mathbf{x} = A(A\mathbf{x}) = \mathbf{0}$ establishing that every vector $\mathbf{x}$ from $\mathbb{R}^n$ is in Nul A^2.

4.3 Linearly Independent Sets; Bases

In this section we identify and study the subsets that span a vector space V or a subspace H as "efficiently" as possible. The key idea is that of linear independence, defined as in $\mathbb{R}^n$.

An indexed set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in V is said to be **linearly independent** if the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0} \tag{1}$$

has *only* the trivial solution, $c_1 = 0, \dots, c_p = 0$.[1]

The set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is said to be **linearly dependent** if (1) has a nontrivial solution, that is, if there are some weights, $c_1, \dots, c_p$, *not all zero*, such that (1) holds. In such a case, (1) is called a **linear dependence relation** among $\mathbf{v}_1, \dots, \mathbf{v}_p$.

Just as in $\mathbb{R}^n$, a set containing a single vector $\mathbf{v}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$. Also, a set of two vectors is linearly dependent if and only if one of the vectors is a multiple of the other. And any set containing the zero vector is linearly dependent. The following theorem has the same proof as Theorem 7 in Section 1.7.

[1] It is convenient to use $c_1, \dots, c_p$ in (1) for the scalars instead of $x_1, \dots, x_p$, as we did previously.

THEOREM 4

An indexed set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ of two or more vectors, with $\mathbf{v}_1 \neq \mathbf{0}$, is linearly dependent if and only if some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\mathbf{v}_1, \dots, \mathbf{v}_{j-1}$.

The main difference between linear dependence in $\mathbb{R}^n$ and in a general vector space is that when the vectors are not n-tuples, the homogeneous equation (1) usually cannot be written as a system of n linear equations. That is, the vectors cannot be made into the columns of a matrix A in order to study the equation $A\mathbf{x} = \mathbf{0}$. We must rely instead on the definition of linear dependence and on Theorem 4.

EXAMPLE 1 Let $\mathbf{p}_1(t) = 1$, $\mathbf{p}_2(t) = t$, and $\mathbf{p}_3(t) = 4 - t$. Then $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ is linearly dependent in $\mathbb{P}$ because $\mathbf{p}_3 = 4\mathbf{p}_1 - \mathbf{p}_2$. ■

EXAMPLE 2 The set $\{\sin t, \cos t\}$ is linearly independent in $C[0, 1]$, the space of all continuous functions on $0 \le t \le 1$, because $\sin t$ and $\cos t$ are not multiples of one another *as vectors in* $C[0, 1]$. That is, there is no scalar c such that $\cos t = c \cdot \sin t$ for all t in $[0, 1]$. (Look at the graphs of $\sin t$ and $\cos t$.) However, $\{\sin t \cos t, \sin 2t\}$ is linearly dependent because of the identity $\sin 2t = 2 \sin t \cos t$, for all t. ■

DEFINITION

Let H be a subspace of a vector space V. A set of vectors $\mathcal{B}$ in V is a **basis** for H if

(i) $\mathcal{B}$ is a linearly independent set, and

(ii) the subspace spanned by $\mathcal{B}$ coincides with H; that is,

$$H = \text{Span}\,\mathcal{B}$$

The definition of a basis applies to the case when $H = V$, because any vector space is a subspace of itself. Thus a basis of V is a linearly independent set that spans V. Observe that when $H \neq V$, condition (ii) includes the requirement that each of the vectors $\mathbf{b}$ in $\mathcal{B}$ must belong to H, because Span $\mathcal{B}$ contains every element in $\mathcal{B}$, as shown in Section 4.1.

EXAMPLE 3 Let A be an invertible $n \times n$ matrix—say, $A = [\,\mathbf{a}_1 \ \cdots \ \mathbf{a}_n\,]$. Then the columns of A form a basis for $\mathbb{R}^n$ because they are linearly independent and they span $\mathbb{R}^n$, by the Invertible Matrix Theorem. ■

EXAMPLE 4 Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the columns of the $n \times n$ identity matrix, I_n. That is,

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$ (Figure 1). ■

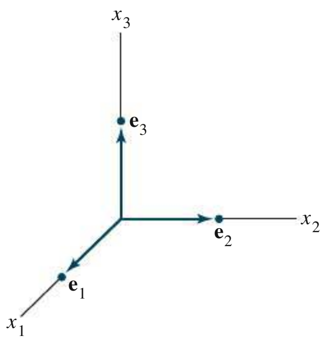


FIGURE 1
The standard basis for $\mathbb{R}^3$.

EXAMPLE 5 Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ -6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 7 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}$. Determine if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for $\mathbb{R}^3$.

SOLUTION Since there are exactly three vectors here in $\mathbb{R}^3$, we can use any of several methods to determine if the matrix $A = [\,\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \mathbf{v}_3\,]$ is invertible. For instance, two row replacements reveal that A has three pivot positions. Thus A is invertible. As in Example 3, the columns of A form a basis for $\mathbb{R}^3$. ■

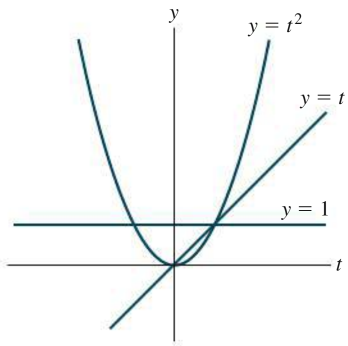


FIGURE 2
The standard basis for $\mathbb{P}_2$.

EXAMPLE 6 Let $S = \{1, t, t^2, \dots, t^n\}$. Verify that S is a basis for $\mathbb{P}_n$. This basis is called the **standard basis** for $\mathbb{P}_n$.

SOLUTION Certainly S spans $\mathbb{P}_n$. To show that S is linearly independent, suppose that $c_0, \dots, c_n$ satisfy

$$c_0 1 + c_1 t + c_2 t^2 + \cdots + c_n t^n = \mathbf{0}(t) \tag{2}$$

This equality means that the polynomial on the left has the same values as the zero polynomial on the right. A fundamental theorem in algebra says that the only polynomial in $\mathbb{P}_n$ with more than n zeros is the zero polynomial. That is, equation (2) holds for all t only if $c_0 = \cdots = c_n = 0$. This proves that S is linearly independent and hence is a basis for $\mathbb{P}_n$. See Figure 2. ■

Problems involving linear independence and spanning in $\mathbb{P}_n$ are handled best by a technique to be discussed in Section 4.4.

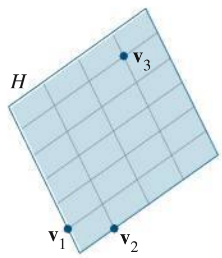
The Spanning Set Theorem

As we will see, a basis is an "efficient" spanning set that contains no unnecessary vectors. In fact, a basis can be constructed from a spanning set by discarding unneeded vectors.

EXAMPLE 7 Let

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 16 \\ -5 \end{bmatrix}, \quad \text{and} \quad H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}.$$

Note that $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, and show that Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ = Span $\{\mathbf{v}_1, \mathbf{v}_2\}$. Then find a basis for the subspace H.



SOLUTION Every vector in Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ belongs to H because

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + 0\mathbf{v}_3$$

Now let $\mathbf{x}$ be any vector in H—say, $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$. Since $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, we may substitute

$$\begin{aligned} \mathbf{x} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3(5\mathbf{v}_1 + 3\mathbf{v}_2) \\ &= (c_1 + 5c_3)\mathbf{v}_1 + (c_2 + 3c_3)\mathbf{v}_2 \end{aligned}$$

Thus $\mathbf{x}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2\}$, so every vector in H already belongs to Span $\{\mathbf{v}_1, \mathbf{v}_2\}$. We conclude that H and Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ are actually the same set of vectors. It follows that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of H since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is obviously linearly independent. ■

The next theorem generalizes Example 7.

THEOREM 5

The Spanning Set Theorem

Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a set in a vector space V, and let $H = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

a. If one of the vectors in S—say, $\mathbf{v}_k$—is a linear combination of the remaining vectors in S, then the set formed from S by removing $\mathbf{v}_k$ still spans H.

b. If $H \neq \{\mathbf{0}\}$, some subset of S is a basis for H.

PROOF

a. By rearranging the list of vectors in S, if necessary, we may suppose that $\mathbf{v}_p$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_{p-1}$—say,

$$\mathbf{v}_p = a_1\mathbf{v}_1 + \cdots + a_{p-1}\mathbf{v}_{p-1} \tag{3}$$

Given any $\mathbf{x}$ in H, we may write

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1} + c_p\mathbf{v}_p \tag{4}$$

for suitable scalars $c_1, \dots, c_p$. Substituting the expression for $\mathbf{v}_p$ from (3) into (4), it is easy to see that $\mathbf{x}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_{p-1}$. Thus $\{\mathbf{v}_1, \dots, \mathbf{v}_{p-1}\}$ spans H, because $\mathbf{x}$ was an arbitrary element of H.

b. If the original spanning set S is linearly independent, then it is already a basis for H. Otherwise, one of the vectors in S depends on the others and can be deleted, by part (a). So long as there are two or more vectors in the spanning set, we can repeat this process until the spanning set is linearly independent and hence is a basis for H. If the spanning set is eventually reduced to one vector, that vector will be nonzero (and hence linearly independent) because $H \neq \{\mathbf{0}\}$. ■

Bases for Nul *A*, Col *A*, and Row *A*

We already know how to find vectors that span the null space of a matrix A. The discussion in Section 4.2 pointed out that our method always produces a linearly independent set when Nul A contains nonzero vectors. So, in this case, that method produces a *basis* for Nul A.

The next two examples describe a simple algorithm for finding a basis for the column space.

EXAMPLE 8 Find a basis for Col B, where

$$B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

SOLUTION Each nonpivot column of B is a linear combination of the pivot columns. In fact, $\mathbf{b}_2 = 4\mathbf{b}_1$ and $\mathbf{b}_4 = 2\mathbf{b}_1 - \mathbf{b}_3$. By the Spanning Set Theorem, we may discard $\mathbf{b}_2$ and $\mathbf{b}_4$, and $\{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\}$ will still span Col B. Let

$$S = \{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Since $\mathbf{b}_1 \neq 0$ and no vector in S is a linear combination of the vectors that precede it, S is linearly independent (Theorem 4). Thus S is a basis for Col B. ■

What about a matrix A that is *not* in reduced echelon form? Recall that any linear dependence relationship among the columns of A can be expressed in the form $A\mathbf{x} = \mathbf{0}$, where $\mathbf{x}$ is a column of weights. (If some columns are not involved in a particular dependence relation, then their weights are zero.) When A is row reduced to a matrix B, the columns of B are often totally different from the columns of A. However, the equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have exactly the same set of solutions. If $A = [\,\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\,]$ and $B = [\,\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_n\,]$, then the vector equations

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0} \quad \text{and} \quad x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n = \mathbf{0}$$

also have the same set of solutions. That is, the columns of A have *exactly the same linear dependence relationships* as the columns of B.

EXAMPLE 9 It can be shown that the matrix

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix}$$

is row equivalent to the matrix B in Example 8. Find a basis for Col A.

SOLUTION In Example 8 we saw that

$$\mathbf{b}_2 = 4\mathbf{b}_1 \quad \text{and} \quad \mathbf{b}_4 = 2\mathbf{b}_1 - \mathbf{b}_3$$

so we can expect that

$$\mathbf{a}_2 = 4\mathbf{a}_1 \quad \text{and} \quad \mathbf{a}_4 = 2\mathbf{a}_1 - \mathbf{a}_3$$

Check that this is indeed the case! Thus we may discard $\mathbf{a}_2$ and $\mathbf{a}_4$ when selecting a minimal spanning set for Col A. In fact, $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ must be linearly independent because any linear dependence relationship among $\mathbf{a}_1$, $\mathbf{a}_3$, $\mathbf{a}_5$ would imply a linear dependence relationship among $\mathbf{b}_1$, $\mathbf{b}_3$, $\mathbf{b}_5$. But we know that $\{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\}$ is a linearly independent set. Thus $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ is a basis for Col A. The columns we have used for this basis are the pivot columns of A. ■

Examples 8 and 9 illustrate the following useful fact.

THEOREM 6 The pivot columns of a matrix A form a basis for Col A.

PROOF The general proof uses the arguments discussed above. Let B be the reduced echelon form of A. The set of pivot columns of B is linearly independent, for no vector in the set is a linear combination of the vectors that precede it. Since A is row equivalent to B, the pivot columns of A are linearly independent as well, because any linear dependence relation among the columns of A corresponds to a linear dependence relation among the columns of B. For this same reason, every nonpivot column of A is a linear combination of the pivot columns of A. Thus the nonpivot columns of A may be discarded from the spanning set for Col A, by the Spanning Set Theorem. This leaves the pivot columns of A as a basis for Col A. ■

Warning: The pivot columns of a matrix A are evident when A has been reduced only to *echelon* form. But, be careful to use the *pivot columns of A itself* for the basis of Col A. Row operations can change the column space of a matrix. The columns of an echelon form B of A are often not in the column space of A. For instance, the columns of matrix B in Example 8 all have zeros in their last entries, so they cannot span the column space of matrix A in Example 9.

In contrast, the following theorem establishes that row reduction does not change the row space of a matrix.

THEOREM 7 If two matrices A and B are row equivalent, then their row spaces are the same. If B is in echelon form, the nonzero rows of B form a basis for the row space of A as well as for that of B.

PROOF If B is obtained from A by row operations, the rows of B are linear combinations of the rows of A. It follows that any linear combination of the rows of B is automatically a linear combination of the rows of A. Thus the row space of B is contained in the row space of A. Since row operations are reversible, the same argument shows that the row space of A is a subset of the row space of B. So the two row spaces are the same. If B is in echelon form, its nonzero rows are linearly independent because no nonzero row is a linear combination of the nonzero rows below it. (Apply Theorem 4 to the nonzero rows of B in reverse order, with the first row last.) Thus the nonzero rows of B form a basis of the (common) row space of B and A. ■

EXAMPLE 10 Find a basis for the row space of the matrix A from Example 9.

SOLUTION To find a basis for the row space, recall that matrix A from Example 9 is row equivalent to matrix B from Example 8:

$$A = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix} \sim B = \begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem 7, the first three rows of B form a basis for the row space of A (as well as for the row space of B). Thus

$$\text{Basis for Row } A : \{(1, 4, 0, 2, 0), (0, 0, 1, -1, 0), (0, 0, 0, 0, 1)\}$$

Observe that, unlike the basis for Col A, the bases for Row A and Nul A have no simple connection with the entries in A itself.[2]

Two Views of a Basis

When the Spanning Set Theorem is used, the deletion of vectors from a spanning set must stop when the set becomes linearly independent. If an additional vector is deleted,

[2] It is possible to find a basis for the row space Row A that uses rows of A. First form A^T, and then row reduce until the pivot columns of A^T are found. These pivot columns of A^T are rows of A, and they form a basis for the row space of A.

it will not be a linear combination of the remaining vectors, and hence the smaller set will no longer span V. Thus a basis is a spanning set that is as small as possible.

A basis is also a linearly independent set that is as large as possible. If S is a basis for V, and if S is enlarged by one vector—say, $\mathbf{w}$—from V, then the new set cannot be linearly independent, because S spans V, and $\mathbf{w}$ is therefore a linear combination of the elements in S.

EXAMPLE 11 The following three sets in $\mathbb{R}^3$ show how a linearly independent set can be enlarged to a basis and how further enlargement destroys the linear independence of the set. Also, a spanning set can be shrunk to a basis, but further shrinking destroys the spanning property.

$$\underbrace{\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}2\\3\\0\end{bmatrix}\right\}}_{\substack{\text{Linearly independent}\\ \text{but does not span } \mathbb{R}^3}} \qquad \underbrace{\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}2\\3\\0\end{bmatrix},\begin{bmatrix}4\\5\\6\end{bmatrix}\right\}}_{\substack{\text{A basis}\\ \text{for } \mathbb{R}^3}} \qquad \underbrace{\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}2\\3\\0\end{bmatrix},\begin{bmatrix}4\\5\\6\end{bmatrix},\begin{bmatrix}7\\8\\9\end{bmatrix}\right\}}_{\substack{\text{Spans } \mathbb{R}^3 \text{ but is}\\ \text{linearly dependent}}}$$

Practice Problems

1. Let $\mathbf{v}_1 = \begin{bmatrix}1\\-2\\3\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}-2\\7\\-9\end{bmatrix}$. Determine if $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\mathbb{R}^3$. Is $\{\mathbf{v}_1, \mathbf{v}_2\}$ a basis for $\mathbb{R}^2$?
2. Let $\mathbf{v}_1 = \begin{bmatrix}1\\-3\\4\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}6\\2\\-1\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}2\\-2\\3\end{bmatrix}$, and $\mathbf{v}_4 = \begin{bmatrix}-4\\-8\\9\end{bmatrix}$. Find a basis for the subspace W spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.
3. Let $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$, and $H = \left\{\begin{bmatrix}s\\s\\0\end{bmatrix} : s \text{ in } \mathbb{R}\right\}$. Then every vector in H is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ because
$$\begin{bmatrix}s\\s\\0\end{bmatrix} = s\begin{bmatrix}1\\0\\0\end{bmatrix} + s\begin{bmatrix}0\\1\\0\end{bmatrix}$$
Is $\{\mathbf{v}_1, \mathbf{v}_2\}$ a basis for H?
4. Let V and W be vector spaces, let $T : V \to W$ and $U : V \to W$ be linear transformations, and let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a basis for V. If $T(\mathbf{v}_j) = U(\mathbf{v}_j)$ for every value of j between 1 and p, show that $T(\mathbf{x}) = U(\mathbf{x})$ for every vector $\mathbf{x}$ in V.

STUDY GUIDE offers additional resources for mastering the concept of basis.

4.3 Exercises

Determine which sets in Exercises 1–8 are bases for $\mathbb{R}^3$. Of the sets that are *not* bases, determine which ones are linearly independent and which ones span $\mathbb{R}^3$. Justify your answers.

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

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

8. $\begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$

Find bases for the null spaces of the matrices given in Exercises 9 and 10. Refer to the remarks that follow Example 3 in Section 4.2.

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

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

11. Find a basis for the set of vectors in $\mathbb{R}^3$ in the plane $x + 2y + z = 0$. [*Hint:* Think of the equation as a "system" of homogeneous equations.]

12. Find a basis for the set of vectors in $\mathbb{R}^2$ on the line $y = 5x$.

In Exercises 13 and 14, assume that A is row equivalent to B. Find bases for Nul A, Col A, and Row A.

13. $A = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 2 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

14. $A = \begin{bmatrix} 1 & 2 & -5 & 11 & -3 \\ 2 & 4 & -5 & 15 & 2 \\ 1 & 2 & 0 & 4 & 5 \\ 3 & 6 & -5 & 19 & -2 \end{bmatrix},$

$B = \begin{bmatrix} 1 & 2 & 0 & 4 & 5 \\ 0 & 0 & 5 & -7 & 8 \\ 0 & 0 & 0 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

In Exercises 15–18, find a basis for the space spanned by the given vectors, $\mathbf{v}_1, \dots, \mathbf{v}_5$.

15. $\begin{bmatrix} 1 \\ 0 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ -4 \\ 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ -8 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -6 \\ 9 \end{bmatrix}$

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

T 17. $\begin{bmatrix} 8 \\ 9 \\ -3 \\ -6 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 1 \\ -4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -9 \\ 6 \\ -7 \end{bmatrix}, \begin{bmatrix} 6 \\ 8 \\ 4 \\ -7 \\ 10 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 11 \\ -8 \\ -7 \end{bmatrix}$

T 18. $\begin{bmatrix} -8 \\ 7 \\ 6 \\ 5 \\ -7 \end{bmatrix}, \begin{bmatrix} 8 \\ -7 \\ -9 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} -8 \\ 7 \\ 4 \\ 5 \\ -7 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 9 \\ 6 \\ -7 \end{bmatrix}, \begin{bmatrix} -9 \\ 3 \\ -4 \\ -1 \\ 0 \end{bmatrix}$

19. Let $\mathbf{v}_1 = \begin{bmatrix} 4 \\ -3 \\ 7 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 9 \\ -2 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 7 \\ 11 \\ 6 \end{bmatrix}$, and $H =$ Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. It can be verified that $4\mathbf{v}_1 + 5\mathbf{v}_2 - 3\mathbf{v}_3 = \mathbf{0}$. Use this information to find a basis for H. There is more than one answer.

20. Let $\mathbf{v}_1 = \begin{bmatrix} 7 \\ 4 \\ -9 \\ -5 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ -7 \\ 2 \\ 5 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ -5 \\ 3 \\ 4 \end{bmatrix}$. It can be verified that $\mathbf{v}_1 - 3\mathbf{v}_2 + 5\mathbf{v}_3 = \mathbf{0}$. Use this information to find a basis for $H =$ Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

In Exercises 21–32, mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** A single vector by itself is linearly dependent.

22. **(T/F)** A linearly independent set in a subspace H is a basis for H.

23. **(T/F)** If $H =$ Span $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$, then $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for H.

24. **(T/F)** If a finite set S of nonzero vectors spans a vector space V, then some subset of S is a basis for V.

25. **(T/F)** The columns of an invertible $n \times n$ matrix form a basis for $\mathbb{R}^n$.

26. **(T/F)** A basis is a linearly independent set that is as large as possible.

27. **(T/F)** A basis is a spanning set that is as large as possible.

28. **(T/F)** The standard method for producing a spanning set for Nul A, described in Section 4.2, sometimes fails to produce a basis for Nul A.

29. **(T/F)** In some cases, the linear dependence relations among the columns of a matrix can be affected by certain elementary row operations on the matrix.

30. **(T/F)** If B is an echelon form of a matrix A, then the pivot columns of B form a basis for Col A.

31. **(T/F)** Row operations preserve the linear dependence relations among the rows of A.

32. **(T/F)** If A and B are row equivalent, then their row spaces are the same.

33. Suppose $\mathbb{R}^4 =$ Span $\{\mathbf{v}_1, \dots, \mathbf{v}_4\}$. Explain why $\{\mathbf{v}_1, \dots, \mathbf{v}_4\}$ is a basis for $\mathbb{R}^4$.

34. Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a linearly independent set in $\mathbb{R}^n$. Explain why $\mathcal{B}$ must be a basis for $\mathbb{R}^n$.

35. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and let H be the set of vectors in $\mathbb{R}^3$ whose second and third entries are equal. Then every vector in H has a unique expansion as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, because

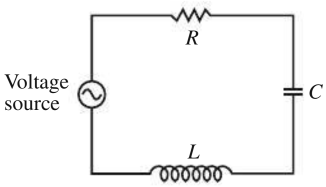
$$\begin{bmatrix} s \\ t \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + (t-s)\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

for any s and t. Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ a basis for H? Why or why not?

36. In the vector space of all real-valued functions, find a basis for the subspace spanned by $\{\sin t, \sin 2t, \sin t \cos t\}$.

37. Let V be the vector space of functions that describe the vibration of a mass–spring system. (Refer to Exercise 19 in Section 4.1.) Find a basis for V.

38. (*RLC circuit*) The circuit in the figure consists of a resistor (R ohms), an inductor (L henrys), a capacitor (C farads), and an initial voltage source. Let $b = R/(2L)$, and suppose R, L, and C have been selected so that b also equals $1/\sqrt{LC}$. (This is done, for instance, when the circuit is used in a voltmeter.) Let $v(t)$ be the voltage (in volts) at time t, measured across the capacitor. It can be shown that v is in the null space H of the linear transformation that maps $v(t)$ into $Lv''(t) + Rv'(t) + (1/C)v(t)$, and H consists of all functions of the form $v(t) = e^{-bt}(c_1 + c_2 t)$. Find a basis for H.



Exercises 39 and 40 show that every basis for $\mathbb{R}^n$ must contain exactly n vectors.

39. Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$, with $k < n$. Use a theorem from Section 1.4 to explain why S cannot be a basis for $\mathbb{R}^n$.

40. Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$, with $k > n$. Use a theorem from Chapter 1 to explain why S cannot be a basis for $\mathbb{R}^n$.

Exercises 41 and 42 reveal an important connection between linear independence and linear transformations and provide practice using the definition of linear dependence. Let V and W be vector spaces, let $T : V \to W$ be a linear transformation, and let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a subset of V.

41. Show that if $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly dependent in V, then the set of images, $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)\}$, is linearly dependent in W. This fact shows that if a linear transformation maps a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ onto a linearly *independent* set $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)\}$, then the original set is linearly independent, too (because it cannot be linearly dependent).

42. Suppose that T is a one-to-one transformation, so that an equation $T(\mathbf{u}) = T(\mathbf{v})$ always implies $\mathbf{u} = \mathbf{v}$. Show that if the set of images $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)\}$ is linearly dependent, then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly dependent. This fact shows that *a one-to-one linear transformation maps a linearly independent set onto a linearly independent set* (because in this case the set of images cannot be linearly dependent).

43. Consider the polynomials $\mathbf{p}_1(t) = 1 + t^2$ and $\mathbf{p}_2(t) = 1 - t^2$. Is $\{\mathbf{p}_1, \mathbf{p}_2\}$ a linearly independent set in $\mathbb{P}_3$? Why or why not?

44. Consider the polynomials $\mathbf{p}_1(t) = 1 + t$, $\mathbf{p}_2(t) = 1 - t$, and $\mathbf{p}_3(t) = 2$ (for all t). By inspection, write a linear dependence relation among $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$. Then find a basis for Span $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$.

45. Let V be a vector space that contains a linearly independent set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$. Describe how to construct a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ in V such that $\{\mathbf{v}_1, \mathbf{v}_3\}$ is a basis for Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

T **46.** Let $H =$ Span $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $K =$ Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 2 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 3 \\ 4 \\ 1 \\ -4 \end{bmatrix},$$

$$\mathbf{v}_1 = \begin{bmatrix} -2 \\ -2 \\ -1 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 2 \\ -6 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 4 \\ 6 \\ -2 \end{bmatrix}$$

Find bases for H, K, and $H + K$. (See Exercises 41 and 42 in Section 4.1.)

T **47.** Show that $\{t, \sin t, \cos 2t, \sin t \cos t\}$ is a linearly independent set of functions defined on $\mathbb{R}$. Start by assuming that

$$c_1 t + c_2 \sin t + c_3 \cos 2t + c_4 \sin t \cos t = 0 \qquad (5)$$

Equation (5) must hold for all real t, so choose several specific values of t (say, $t = 0, .1, .2$) until you get a system of enough equations to determine that all the c_j must be zero.

T **48.** Show that $\{1, \cos t, \cos^2 t, \dots, \cos^6 t\}$ is a linearly independent set of functions defined on $\mathbb{R}$. Use the method of Exercise 47. (This result will be needed in Exercise 54 in Section 4.5.)

Solutions to Practice Problems

1. Let $A = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]$. Row operations show that

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

Not every row of A contains a pivot position. So the columns of A do not span $\mathbb{R}^3$, by Theorem 4 in Section 1.4. Hence $\{\mathbf{v}_1, \mathbf{v}_2\}$ is not a basis for $\mathbb{R}^3$. Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are not in $\mathbb{R}^2$, they cannot possibly be a basis for $\mathbb{R}^2$. However, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are obviously linearly independent, they are a basis for a subspace of $\mathbb{R}^3$, namely Span $\{\mathbf{v}_1, \mathbf{v}_2\}$.

2. Set up a matrix A whose column space is the space spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$, and then row reduce A to find its pivot columns.

$$A = \begin{bmatrix} 1 & 6 & 2 & -4 \\ -3 & 2 & -2 & -8 \\ 4 & -1 & 3 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -4 \\ 0 & 20 & 4 & -20 \\ 0 & -25 & -5 & 25 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -4 \\ 0 & 5 & 1 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The first two columns of A are the pivot columns and hence form a basis of Col $A = W$. Hence $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. Note that the reduced echelon form of A is not needed in order to locate the pivot columns.

3. Neither $\mathbf{v}_1$ nor $\mathbf{v}_2$ is in H, so $\{\mathbf{v}_1, \mathbf{v}_2\}$ cannot be a basis for H. In fact, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for the *plane* of all vectors of the form $(c_1, c_2, 0)$, but H is only a *line*.

4. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis for V, for any vector $\mathbf{x}$ in V, there exist scalars $c_1, \dots, c_p$ such that $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$. Then since T and U are linear transformations

$$\begin{aligned} T(\mathbf{x}) &= T(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + \cdots + c_pT(\mathbf{v}_p) \\ &= c_1U(\mathbf{v}_1) + \cdots + c_pU(\mathbf{v}_p) = U(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) \\ &= U(\mathbf{x}) \end{aligned}$$

4.4 Coordinate Systems

An important reason for specifying a basis $\mathcal{B}$ for a vector space V is to impose a "coordinate system" on V. This section will show that if $\mathcal{B}$ contains n vectors, then the coordinate system will make V act like $\mathbb{R}^n$. If V is already $\mathbb{R}^n$ itself, then $\mathcal{B}$ will determine a coordinate system that gives a new "view" of V.

The existence of coordinate systems rests on the following fundamental result.

THEOREM 8

The Unique Representation Theorem

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V. Then for each $\mathbf{x}$ in V, there exists a unique set of scalars $c_1, \dots, c_n$ such that

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n \tag{1}$$

PROOF Since $\mathcal{B}$ spans V, there exist scalars such that (1) holds. Suppose $\mathbf{x}$ also has the representation

$$\mathbf{x} = d_1\mathbf{b}_1 + \cdots + d_n\mathbf{b}_n$$

for scalars $d_1, \ldots, d_n$. Then, subtracting, we have

$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (c_1 - d_1)\mathbf{b}_1 + \cdots + (c_n - d_n)\mathbf{b}_n \tag{2}$$

Since $\mathcal{B}$ is linearly independent, the weights in (2) must all be zero. That is, $c_j = d_j$ for $1 \leq j \leq n$. ■

DEFINITION

Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis for a vector space V and $\mathbf{x}$ is in V. The **coordinates of x relative to the basis $\mathcal{B}$** (or the **$\mathcal{B}$-coordinates of x**) are the weights $c_1, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$.

If $c_1, \ldots, c_n$ are the $\mathcal{B}$-coordinates of $\mathbf{x}$, then the vector in $\mathbb{R}^n$

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

is the **coordinate vector of x (relative to $\mathcal{B}$)**, or the **$\mathcal{B}$-coordinate vector of x**. The mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$ is the **coordinate mapping (determined by $\mathcal{B}$)**.[1]

EXAMPLE 1 Consider a basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ for $\mathbb{R}^2$, where $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Suppose an $\mathbf{x}$ in $\mathbb{R}^2$ has the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$. Find $\mathbf{x}$.

SOLUTION The $\mathcal{B}$-coordinates of $\mathbf{x}$ tell how to build $\mathbf{x}$ from the vectors in $\mathcal{B}$. That is,

$$\mathbf{x} = (-2)\mathbf{b}_1 + 3\mathbf{b}_2 = (-2)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$ ■

EXAMPLE 2 The entries in the vector $\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$ are the coordinates of $\mathbf{x}$ relative to the *standard basis* $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$, since

$$\begin{bmatrix} 1 \\ 6 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 6\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1\mathbf{e}_1 + 6\mathbf{e}_2$$

If $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$, then $[\,\mathbf{x}\,]_{\mathcal{E}} = \mathbf{x}$. ■

A Graphical Interpretation of Coordinates

A coordinate system on a set consists of a one-to-one mapping of the points in the set into $\mathbb{R}^n$. For example, ordinary graph paper provides a coordinate system for the plane

[1] The concept of a coordinate mapping assumes that the basis $\mathcal{B}$ is an indexed set whose vectors are listed in some fixed preassigned order. This property makes the definition of $[\,\mathbf{x}\,]_{\mathcal{B}}$ unambiguous.

when one selects perpendicular axes and a unit of measurement on each axis. Figure 1 shows the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$, the vectors $\mathbf{b}_1 (= \mathbf{e}_1)$ and $\mathbf{b}_2$ from Example 1, and the vector $\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$. The coordinates 1 and 6 give the location of $\mathbf{x}$ relative to the standard basis: 1 unit in the $\mathbf{e}_1$ direction and 6 units in the $\mathbf{e}_2$ direction.

Figure 2 shows the vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{x}$ from Figure 1. (Geometrically, the three vectors lie on a vertical line in both figures.) However, the standard coordinate grid was erased and replaced by a grid especially adapted to the basis $\mathcal{B}$ in Example 1. The coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ gives the location of $\mathbf{x}$ on this new coordinate system: -2 units in the $\mathbf{b}_1$ direction and 3 units in the $\mathbf{b}_2$ direction.

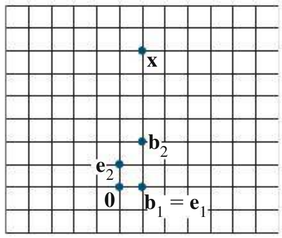


FIGURE 1 Standard graph paper.

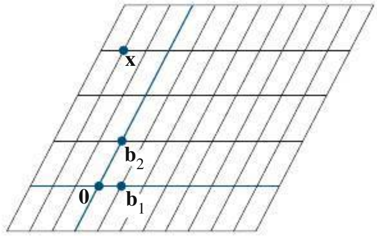


FIGURE 2 $\mathcal{B}$-graph paper.

EXAMPLE 3 In crystallography, the description of a crystal lattice is aided by choosing a basis $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ for $\mathbb{R}^3$ that corresponds to three adjacent edges of one "unit cell" of the crystal. An entire lattice is constructed by stacking together many copies of one cell. There are fourteen basic types of unit cells; three are displayed in Figure 3.[2]

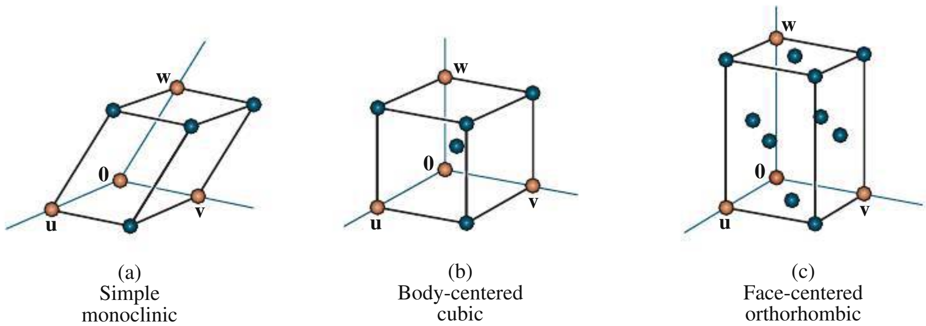


FIGURE 3 Examples of unit cells.

The coordinates of atoms within the crystal are given relative to the basis for the lattice. For instance,

$$\begin{bmatrix} 1/2 \\ 1/2 \\ 1 \end{bmatrix}$$

identifies the top face-centered atom in the cell in Figure 3(c). ■

[2] Adapted from *The Science and Engineering of Materials*, 4th Ed., by Donald R. Askeland (Boston: Prindle, Weber & Schmidt, © 2002), p. 36.

Coordinates in $\mathbb{R}^n$

When a basis $\mathcal{B}$ for $\mathbb{R}^n$ is fixed, the $\mathcal{B}$-coordinate vector of a specified $\mathbf{x}$ is easily found, as in the next example.

EXAMPLE 4 Let $\mathbf{b}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Find the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ of $\mathbf{x}$ relative to $\mathcal{B}$.

SOLUTION The $\mathcal{B}$-coordinates c_1, c_2 of $\mathbf{x}$ satisfy

$$c_1 \underset{\mathbf{b}_1}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}} + c_2 \underset{\mathbf{b}_2}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}} = \underset{\mathbf{x}}{\begin{bmatrix} 4 \\ 5 \end{bmatrix}}$$

or

$$\begin{array}{cc} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \\ \mathbf{b}_1 \quad \mathbf{b}_2 \end{array} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \underset{\mathbf{x}}{\begin{bmatrix} 4 \\ 5 \end{bmatrix}} \tag{3}$$

This equation can be solved by row operations on an augmented matrix or by multiplying the vector $\mathbf{x}$ by the inverse of the matrix. In any case, the solution is $c_1 = 3$, $c_2 = 2$. Thus $\mathbf{x} = 3\mathbf{b}_1 + 2\mathbf{b}_2$, and

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

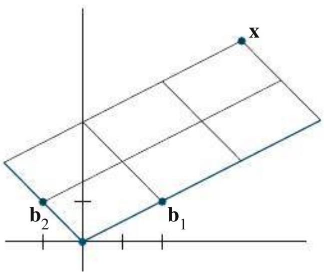


FIGURE 4 The $\mathcal{B}$-coordinate vector of $\mathbf{x}$ is $(3, 2)$.

See Figure 4. ■

The matrix in (3) changes the $\mathcal{B}$-coordinates of a vector $\mathbf{x}$ into the standard coordinates for $\mathbf{x}$. An analogous change of coordinates can be carried out in $\mathbb{R}^n$ for a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. Let

$$P_{\mathcal{B}} = [\,\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_n\,]$$

Then the vector equation

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$$

is equivalent to

$$\boxed{\mathbf{x} = P_{\mathcal{B}}[\,\mathbf{x}\,]_{\mathcal{B}}} \tag{4}$$

We call $P_{\mathcal{B}}$ the **change-of-coordinates matrix** from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$. Left-multiplication by $P_{\mathcal{B}}$ transforms the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ into $\mathbf{x}$. The change-of-coordinates equation (4) is important and will be needed at several points in Chapters 5 and 7.

Since the columns of $P_{\mathcal{B}}$ form a basis for $\mathbb{R}^n$, $P_{\mathcal{B}}$ is invertible (by the Invertible Matrix Theorem). Left-multiplication by $P_{\mathcal{B}}^{-1}$ converts $\mathbf{x}$ into its $\mathcal{B}$-coordinate vector:

$$P_{\mathcal{B}}^{-1}\mathbf{x} = [\,\mathbf{x}\,]_{\mathcal{B}}$$

The correspondence $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$, produced here by $P_{\mathcal{B}}^{-1}$, is the coordinate mapping mentioned earlier. Since $P_{\mathcal{B}}^{-1}$ is an invertible matrix, the coordinate mapping is a one-to-one linear transformation from $\mathbb{R}^n$ onto $\mathbb{R}^n$, by the Invertible Matrix Theorem. (See also Theorem 12 in Section 1.9.) This property of the coordinate mapping is also true in a general vector space that has a basis, as we shall see.

The Coordinate Mapping

Choosing a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ for a vector space V introduces a coordinate system in V. The coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$ connects the possibly unfamiliar space V to the familiar space $\mathbb{R}^n$. See Figure 5. Points in V can now be identified by their new "names."

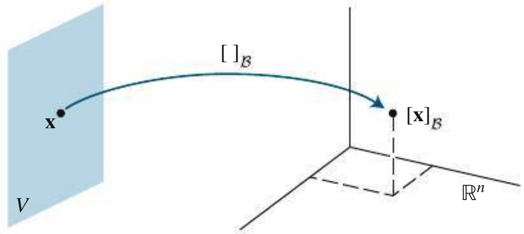


FIGURE 5 The coordinate mapping from V onto $\mathbb{R}^n$.

THEOREM 9

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V. Then the coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.

PROOF Take two typical vectors in V, say,

$$\begin{aligned}\mathbf{u} &= c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n \\ \mathbf{w} &= d_1\mathbf{b}_1 + \cdots + d_n\mathbf{b}_n\end{aligned}$$

Then, using vector operations,

$$\mathbf{u} + \mathbf{w} = (c_1 + d_1)\mathbf{b}_1 + \cdots + (c_n + d_n)\mathbf{b}_n$$

It follows that

$$[\,\mathbf{u} + \mathbf{w}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 + d_1 \\ \vdots \\ c_n + d_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = [\,\mathbf{u}\,]_{\mathcal{B}} + [\,\mathbf{w}\,]_{\mathcal{B}}$$

So the coordinate mapping preserves addition. If r is any scalar, then

$$r\mathbf{u} = r(c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n) = (rc_1)\mathbf{b}_1 + \cdots + (rc_n)\mathbf{b}_n$$

So

$$[\,r\mathbf{u}\,]_{\mathcal{B}} = \begin{bmatrix} rc_1 \\ \vdots \\ rc_n \end{bmatrix} = r\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = r[\,\mathbf{u}\,]_{\mathcal{B}}$$

Thus the coordinate mapping also preserves scalar multiplication and hence is a linear transformation. See Exercises 27 and 28 for verification that the coordinate mapping is one-to-one and maps V onto $\mathbb{R}^n$. ■

The linearity of the coordinate mapping extends to linear combinations, just as in Section 1.8. If $\mathbf{u}_1, \dots, \mathbf{u}_p$ are in V and if $c_1, \dots, c_p$ are scalars, then

$$[\,c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\,]_{\mathcal{B}} = c_1[\,\mathbf{u}_1\,]_{\mathcal{B}} + \cdots + c_p[\,\mathbf{u}_p\,]_{\mathcal{B}} \tag{5}$$

In words, (5) says that the $\mathcal{B}$-coordinate vector of a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_p$ is the *same* linear combination of their coordinate vectors.

The coordinate mapping in Theorem 9 is an important example of an *isomorphism* from V onto $\mathbb{R}^n$. In general, a one-to-one linear transformation from a vector space V onto a vector space W is called an **isomorphism** from V onto W (*iso* from the Greek for "the same," and *morph* from the Greek for "form" or "structure"). The notation and terminology for V and W may differ, but the two spaces are indistinguishable as vector spaces. *Every vector space calculation in V is accurately reproduced in W, and vice versa*. In particular, any real vector space with a basis of n vectors is indistinguishable from $\mathbb{R}^n$. See Exercises 29 and 30.

STUDY GUIDE offers additional resources about isomorphic vector spaces.

EXAMPLE 5 Let $\mathcal{B}$ be the standard basis of the space $\mathbb{P}_3$ of polynomials; that is, let $\mathcal{B} = \{1, t, t^2, t^3\}$. A typical element $\mathbf{p}$ of $\mathbb{P}_3$ has the form

$$\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

Since $\mathbf{p}$ is already displayed as a linear combination of the standard basis vectors, we conclude that

$$[\,\mathbf{p}\,]_{\mathcal{B}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Thus the coordinate mapping $\mathbf{p} \mapsto [\,\mathbf{p}\,]_{\mathcal{B}}$ is an isomorphism from $\mathbb{P}_3$ onto $\mathbb{R}^4$. All vector space operations in $\mathbb{P}_3$ correspond to operations in $\mathbb{R}^4$. ■

If we think of $\mathbb{P}_3$ and $\mathbb{R}^4$ as displays on two computer screens that are connected via the coordinate mapping, then every vector space operation in $\mathbb{P}_3$ on one screen is exactly duplicated by a corresponding vector operation in $\mathbb{R}^4$ on the other screen. The vectors on the $\mathbb{P}_3$ screen look different from those on the $\mathbb{R}^4$ screen, but they "act" as vectors in exactly the same way. See Figure 6.

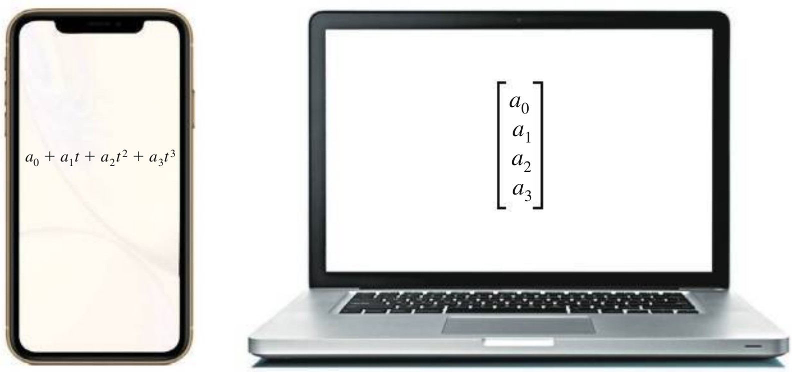


FIGURE 6 The space $\mathbb{P}_3$ is isomorphic to $\mathbb{R}^4$.

EXAMPLE 6 Use coordinate vectors to verify that the polynomials $1 + 2t^2$, $4 + t + 5t^2$, and $3 + 2t$ are linearly dependent in $\mathbb{P}_2$.

SOLUTION The coordinate mapping from Example 5 produces the coordinate vectors $(1, 0, 2)$, $(4, 1, 5)$, and $(3, 2, 0)$, respectively. Writing these vectors as the *columns* of a matrix A, we can determine their independence by row reducing the augmented matrix for $A\mathbf{x} = \mathbf{0}$:

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

The columns of A are linearly dependent, so the corresponding polynomials are linearly dependent. In fact, it is easy to check that column 3 of A is 2 times column 2 minus 5 times column 1. The corresponding relation for the polynomials is

$$3 + 2t = 2(4 + t + 5t^2) - 5(1 + 2t^2)$$ ■

The final example concerns a plane in $\mathbb{R}^3$ that is isomorphic to $\mathbb{R}^2$.

EXAMPLE 7 Let

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix},$$

and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Then $\mathcal{B}$ is a basis for $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Determine if $\mathbf{x}$ is in H, and if it is, find the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$.

SOLUTION If $\mathbf{x}$ is in H, then the following vector equation is consistent:

$$c_1 \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$$

The scalars c_1 and c_2, if they exist, are the $\mathcal{B}$-coordinates of $\mathbf{x}$. Using row operations, we obtain

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

Thus $c_1 = 2$, $c_2 = 3$, and $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. The coordinate system on H determined by $\mathcal{B}$ is shown in Figure 7. ■

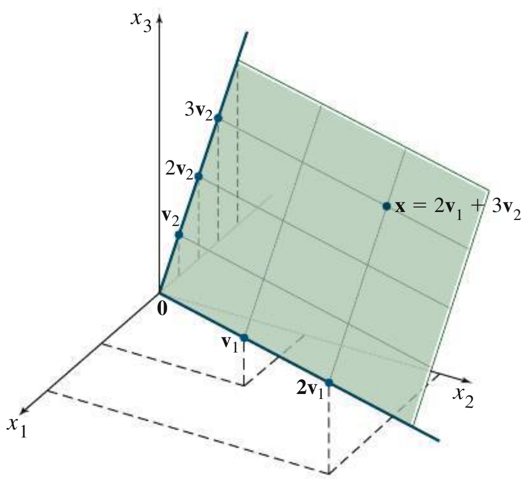


FIGURE 7 A coordinate system on a plane H in $\mathbb{R}^3$.

If a different basis for H were chosen, would the associated coordinate system also make H isomorphic to $\mathbb{R}^2$? Surely, this must be true. We shall prove it in the next section.

Practice Problems

1. Let $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 3 \\ -6 \\ 3 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} -8 \\ 2 \\ 3 \end{bmatrix}$.
   a. Show that the set $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is a basis of $\mathbb{R}^3$.
   b. Find the change-of-coordinates matrix from $\mathcal{B}$ to the standard basis.
   c. Write the equation that relates $\mathbf{x}$ in $\mathbb{R}^3$ to $[\,\mathbf{x}\,]_{\mathcal{B}}$.
   d. Find $[\,\mathbf{x}\,]_{\mathcal{B}}$, for the $\mathbf{x}$ given above.
2. The set $\mathcal{B} = \{1+t, 1+t^2, t+t^2\}$ is a basis for $\mathbb{P}_2$. Find the coordinate vector of $\mathbf{p}(t) = 6 + 3t - t^2$ relative to $\mathcal{B}$.

4.4 Exercises

In Exercises 1–4, find the vector $\mathbf{x}$ determined by the given coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ and the given basis $\mathcal{B}$.

1. $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ -5 \end{bmatrix}, \begin{bmatrix} -4 \\ 6 \end{bmatrix} \right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$

2. $\mathcal{B} = \left\{ \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 6 \\ 7 \end{bmatrix} \right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 8 \\ -5 \end{bmatrix}$

3. $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 4 \\ -7 \\ 0 \end{bmatrix} \right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$

4. $\mathcal{B} = \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ -7 \\ 3 \end{bmatrix} \right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -4 \\ 8 \\ -7 \end{bmatrix}$

In Exercises 5–8, find the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ of $\mathbf{x}$ relative to the given basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$.

5. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

6. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 5 \\ -6 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$

7. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -3 \\ 4 \\ 9 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 8 \\ -9 \\ 6 \end{bmatrix}$

8. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 3 \\ -5 \\ 4 \end{bmatrix}$

In Exercises 9 and 10, find the change-of-coordinates matrix from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$.

9. $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ -9 \end{bmatrix}, \begin{bmatrix} 1 \\ 8 \end{bmatrix} \right\}$

10. $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -5 \end{bmatrix}, \begin{bmatrix} 8 \\ -2 \\ 7 \end{bmatrix} \right\}$

In Exercises 11 and 12, use an inverse matrix to find $[\,\mathbf{x}\,]_{\mathcal{B}}$ for the given $\mathbf{x}$ and $\mathcal{B}$.

11. $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ -5 \end{bmatrix}, \begin{bmatrix} -4 \\ 6 \end{bmatrix} \right\}$, $\mathbf{x} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$

12. $\mathcal{B} = \left\{ \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 6 \\ 7 \end{bmatrix} \right\}$, $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

13. The set $\mathcal{B} = \{1+t^2, t+t^2, 1+2t+t^2\}$ is a basis for $\mathbb{P}_2$. Find the coordinate vector of $\mathbf{p}(t) = 1 + 4t + 7t^2$ relative to $\mathcal{B}$.

14. The set $\mathcal{B} = \{1-t^2, t-t^2, 2-2t+t^2\}$ is a basis for $\mathbb{P}_2$. Find the coordinate vector of $\mathbf{p}(t) = 3 + t - 6t^2$ relative to $\mathcal{B}$.

In Exercises 15–20, mark each statement True or False **(T/F)**. Justify each answer. Unless stated otherwise, $\mathcal{B}$ is a basis for a vector space V.

15. **(T/F)** If $\mathbf{x}$ is in V and if $\mathcal{B}$ contains n vectors, then the $\mathcal{B}$-coordinate vector of $\mathbf{x}$ is in $\mathbb{R}^n$.

16. **(T/F)** If $\mathcal{B}$ is the standard basis for $\mathbb{R}^n$, then the $\mathcal{B}$-coordinate vector of an $\mathbf{x}$ in $\mathbb{R}^n$ is $\mathbf{x}$ itself.

17. **(T/F)** If $P_{\mathcal{B}}$ is the change-of-coordinates matrix, then $[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}\,\mathbf{x}$, for $\mathbf{x}$ in V.

18. **(T/F)** The correspondence $[\,\mathbf{x}\,]_{\mathcal{B}} \mapsto \mathbf{x}$ is called the coordinate mapping.

19. **(T/F)** The vector spaces $\mathbb{P}_3$ and $\mathbb{R}^3$ are isomorphic.

20. **(T/F)** In some cases, a plane in $\mathbb{R}^3$ can be isomorphic to $\mathbb{R}^2$.

21. The vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ -8 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -3 \\ 7 \end{bmatrix}$ span $\mathbb{R}^2$ but do not form a basis. Find two different ways to express $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

22. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V. Explain why the $\mathcal{B}$-coordinate vectors of $\mathbf{b}_1, \dots, \mathbf{b}_n$ are the columns $\mathbf{e}_1, \dots, \mathbf{e}_n$ of the $n \times n$ identity matrix.

23. Let S be a finite set in a vector space V with the property that every $\mathbf{x}$ in V has a unique representation as a linear combination of elements of S. Show that S is a basis of V.

24. Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_4\}$ is a linearly dependent spanning set for a vector space V. Show that each $\mathbf{w}$ in V can be expressed in more than one way as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_4$. [*Hint:* Let $\mathbf{w} = k_1\mathbf{v}_1 + \cdots + k_4\mathbf{v}_4$ be an arbitrary vector in V. Use the linear dependence of $\{\mathbf{v}_1, \dots, \mathbf{v}_4\}$ to produce another representation of $\mathbf{w}$ as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_4$.]

25. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \begin{bmatrix} -2 \\ 9 \end{bmatrix} \right\}$. Since the coordinate mapping determined by $\mathcal{B}$ is a linear transformation from $\mathbb{R}^2$ into $\mathbb{R}^2$, this mapping must be implemented by some 2×2 matrix A. Find it. [*Hint:* Multiplication by A should transform a vector $\mathbf{x}$ into its coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$.]

26. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for $\mathbb{R}^n$. Produce a description of an $n \times n$ matrix A that implements the coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$. (See Exercise 25.)

Exercises 27–30 concern a vector space V, a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, and the coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$.

27. Show that the coordinate mapping is one-to-one. [*Hint:* Suppose $[\,\mathbf{u}\,]_{\mathcal{B}} = [\,\mathbf{w}\,]_{\mathcal{B}}$ for some $\mathbf{u}$ and $\mathbf{w}$ in V, and show that $\mathbf{u} = \mathbf{w}$.]

28. Show that the coordinate mapping is *onto* $\mathbb{R}^n$. That is, given any $\mathbf{y}$ in $\mathbb{R}^n$, with entries $y_1, \dots, y_n$, produce $\mathbf{u}$ in V such that $[\,\mathbf{u}\,]_{\mathcal{B}} = \mathbf{y}$.

29. Show that a subset $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ in V is linearly independent if and only if the set of coordinate vectors $\{[\,\mathbf{u}_1\,]_{\mathcal{B}}, \dots, [\,\mathbf{u}_p\,]_{\mathcal{B}}\}$ is linearly independent in $\mathbb{R}^n$. [*Hint:* Since the coordinate mapping is one-to-one, the following equations have the same solutions, $c_1, \dots, c_p$.]

$$\begin{aligned} c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p &= \mathbf{0} && \text{The zero vector in } V \\ [\,c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\,]_{\mathcal{B}} &= [\,\mathbf{0}\,]_{\mathcal{B}} && \text{The zero vector in } \mathbb{R}^n \end{aligned}$$

30. Given vectors $\mathbf{u}_1, \dots, \mathbf{u}_p$, and $\mathbf{w}$ in V, show that $\mathbf{w}$ is a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_p$ if and only if $[\,\mathbf{w}\,]_{\mathcal{B}}$ is a linear combination of the coordinate vectors $[\,\mathbf{u}_1\,]_{\mathcal{B}}, \dots, [\,\mathbf{u}_p\,]_{\mathcal{B}}$.

In Exercises 31–34, use coordinate vectors to test the linear independence of the sets of polynomials. Explain your work.

31. $\{1 + 2t^3,\ 2 + t - 3t^2,\ -t + 2t^2 - t^3\}$

32. $\{1 - 2t^2 - t^3,\ t + 2t^3,\ 1 + t - 2t^2\}$

33. $\{(1 - t)^2,\ t - 2t^2 + t^3,\ (1 - t)^3\}$

34. $\{(2 - t)^3,\ (3 - t)^2,\ 1 + 6t - 5t^2 + t^3\}$

35. Use coordinate vectors to test whether the following sets of polynomials span $\mathbb{P}_2$. Justify your conclusions.
   a. $\{1 - 3t + 5t^2, -3 + 5t - 7t^2, -4 + 5t - 6t^2, 1 - t^2\}$
   b. $\{5t + t^2, 1 - 8t - 2t^2, -3 + 4t + 2t^2, 2 - 3t\}$

36. Let $\mathbf{p}_1(t) = 1 + t^2$, $\mathbf{p}_2(t) = t - 3t^2$, $\mathbf{p}_3(t) = 1 + t - 3t^2$.
   a. Use coordinate vectors to show that these polynomials form a basis for $\mathbb{P}_2$.
   b. Consider the basis $\mathcal{B} = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ for $\mathbb{P}_2$. Find $\mathbf{q}$ in $\mathbb{P}_2$, given that $[\mathbf{q}]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$.

In Exercises 37 and 38, determine whether the sets of polynomials form a basis for $\mathbb{P}_3$. Justify your conclusions.

T 37. $3 + 7t, 5 + t - 2t^3, t - 2t^2, 1 + 16t - 6t^2 + 2t^3$

T 38. $5 - 3t + 4t^2 + 2t^3, 9 + t + 8t^2 - 6t^3, 6 - 2t + 5t^2, t^3$

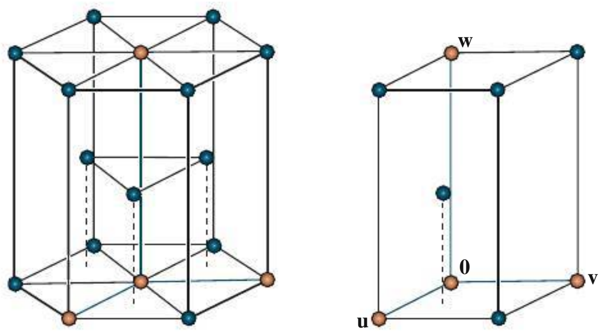
T 39. Let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Show that $\mathbf{x}$ is in H and find the $\mathcal{B}$-coordinate vector of $\mathbf{x}$, for

$$\mathbf{v}_1 = \begin{bmatrix} 11 \\ -5 \\ 10 \\ 7 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 14 \\ -8 \\ 13 \\ 10 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 19 \\ -13 \\ 18 \\ 15 \end{bmatrix}$$

T 40. Let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Show that $\mathcal{B}$ is a basis for H and $\mathbf{x}$ is in H, and find the $\mathcal{B}$-coordinate vector of $\mathbf{x}$, for

$$\mathbf{v}_1 = \begin{bmatrix} -6 \\ 4 \\ -9 \\ 4 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 8 \\ -3 \\ 7 \\ -3 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -9 \\ 5 \\ -8 \\ 3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 4 \\ 7 \\ -8 \\ 3 \end{bmatrix}$$

Exercises 41 and 42 concern the crystal lattice for titanium, which has the hexagonal structure shown on the left in the accompanying figure. The vectors $\begin{bmatrix} 2.6 \\ -1.5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 4.8 \end{bmatrix}$ in $\mathbb{R}^3$ form a basis for the unit cell shown on the right. The numbers here are Ångstrom units (1 Å $= 10^{-8}$ cm). In alloys of titanium, some additional atoms may be in the unit cell at the *octahedral* and *tetrahedral* sites (so named because of the geometric objects formed by atoms at these locations).



The hexagonal close-packed lattice and its unit cell.

41. One of the octahedral sites is $\begin{bmatrix} 1/2 \\ 1/4 \\ 1/6 \end{bmatrix}$, relative to the lattice basis. Determine the coordinates of this site relative to the standard basis of $\mathbb{R}^3$.

42. One of the tetrahedral sites is $\begin{bmatrix} 1/2 \\ 1/2 \\ 1/3 \end{bmatrix}$. Determine the coordinates of this site relative to the standard basis of $\mathbb{R}^3$.

Solutions to Practice Problems

1. a. It is evident that the matrix $P_{\mathcal{B}} = [\,\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3\,]$ is row-equivalent to the identity matrix. By the Invertible Matrix Theorem, $P_{\mathcal{B}}$ is invertible and its columns form a basis for $\mathbb{R}^3$.

b. From part (a), the change-of-coordinates matrix is $P_{\mathcal{B}} = \begin{bmatrix} 1 & -3 & 3 \\ 0 & 4 & -6 \\ 0 & 0 & 3 \end{bmatrix}$.

c. $\mathbf{x} = P_{\mathcal{B}}[\,\mathbf{x}\,]_{\mathcal{B}}$

d. To solve the equation in (c), it is probably easier to row reduce an augmented matrix than to compute $P_{\mathcal{B}}^{-1}$:

$$\begin{array}{c} \left[\begin{array}{rrrr} 1 & -3 & 3 & -8 \\ 0 & 4 & -6 & 2 \\ 0 & 0 & 3 & 3 \end{array}\right] \\ \quad P_{\mathcal{B}} \qquad\qquad \mathbf{x} \end{array} \sim \begin{array}{c} \left[\begin{array}{rrrr} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right] \\ \quad I \qquad\qquad [\,\mathbf{x}\,]_{\mathcal{B}} \end{array}$$

Hence

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}$$

2. The coordinates of $\mathbf{p}(t) = 6 + 3t - t^2$ with respect to $\mathcal{B}$ satisfy

$$c_1(1 + t) + c_2(1 + t^2) + c_3(t + t^2) = 6 + 3t - t^2$$

Equating coefficients of like powers of t, we have

$$\begin{aligned} c_1 + c_2 \qquad &= 6 \\ c_1 \qquad + c_3 &= 3 \\ c_2 + c_3 &= -1 \end{aligned}$$

Solving, we find that $c_1 = 5$, $c_2 = 1$, $c_3 = -2$, and $[\,\mathbf{p}\,]_{\mathcal{B}} = \begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}$.

4.5 The Dimension of a Vector Space

Theorem 9 in Section 4.4 implies that a vector space V with a basis $\mathcal{B}$ containing n vectors is isomorphic to $\mathbb{R}^n$. This section shows that this number n is an intrinsic property (called the dimension) of the space V that does not depend on the particular choice of basis. The discussion of dimension will give additional insight into properties of bases.

The first theorem generalizes a well-known result about the vector space $\mathbb{R}^n$.

THEOREM 10

If a vector space V has a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, then any set in V containing more than n vectors must be linearly dependent.

PROOF Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be a set in V with more than n vectors. The coordinate vectors $[\,\mathbf{u}_1\,]_\mathcal{B}, \dots, [\,\mathbf{u}_p\,]_\mathcal{B}$ form a linearly dependent set in $\mathbb{R}^n$, because there are more vectors (p) than entries (n) in each vector. So there exist scalars $c_1, \dots, c_p$, not all zero, such that

$$c_1[\,\mathbf{u}_1\,]_\mathcal{B} + \cdots + c_p[\,\mathbf{u}_p\,]_\mathcal{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector in $\mathbb{R}^n$

Since the coordinate mapping is a linear transformation,

$$\left[\,c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\,\right]_\mathcal{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector on the right displays the n weights needed to build the vector $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ from the basis vectors in $\mathcal{B}$. That is, $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = 0\mathbf{b}_1 + \cdots + 0\mathbf{b}_n = \mathbf{0}$. Since the c_i are not all zero, $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is linearly dependent.[1] ■

Theorem 10 implies that if a vector space V has a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, then each linearly independent set in V has no more than n vectors.

THEOREM 11

If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.

PROOF Let $\mathcal{B}_1$ be a basis of n vectors and $\mathcal{B}_2$ be any other basis (of V). Since $\mathcal{B}_1$ is a basis and $\mathcal{B}_2$ is linearly independent, $\mathcal{B}_2$ has no more than n vectors, by Theorem 10. Also, since $\mathcal{B}_2$ is a basis and $\mathcal{B}_1$ is linearly independent, $\mathcal{B}_2$ has at least n vectors. Thus $\mathcal{B}_2$ consists of exactly n vectors. ■

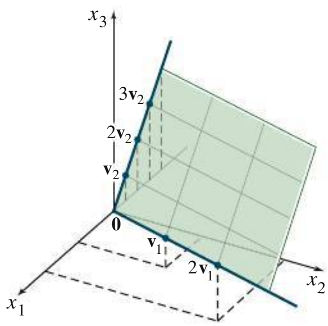
[1] Theorem 10 also applies to infinite sets in V. An infinite set is said to be linearly dependent if some finite subset is linearly dependent; otherwise, the set is linearly independent. If S is an infinite set in V, take any subset $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ of S, with $p > n$. The proof above shows that this subset is linearly dependent and hence so is S.

If a nonzero vector space V is spanned by a finite set S, then a subset of S is a basis for V, by the Spanning Set Theorem. In this case, Theorem 11 ensures that the following definition makes sense.

DEFINITION

If a vector space V is spanned by a finite set, then V is said to be **finite-dimensional**, and the **dimension** of V, written as dim V, is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

EXAMPLE 1 The standard basis for $\mathbb{R}^n$ contains n vectors, so $\dim \mathbb{R}^n = n$. The standard polynomial basis $\{1, t, t^2\}$ shows that $\dim \mathbb{P}_2 = 3$. In general, $\dim \mathbb{P}_n = n + 1$. The space $\mathbb{P}$ of all polynomials is infinite-dimensional. ■



EXAMPLE 2 Let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. Then H is the plane studied in Example 7 in Section 4.4. A basis for H is $\{\mathbf{v}_1, \mathbf{v}_2\}$, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are not multiples and hence are linearly independent. Thus $\dim H = 2$. ■

EXAMPLE 3 Find the dimension of the subspace

$$H = \left\{ \begin{bmatrix} a - 3b + 6c \\ 5a + 4d \\ b - 2c - d \\ 5d \end{bmatrix} : a, b, c, d \text{ in } \mathbb{R} \right\}$$

SOLUTION It is easy to see that H is the set of all linear combinations of the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}$$

Clearly, $\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2$ is not a multiple of $\mathbf{v}_1$, but $\mathbf{v}_3$ is a multiple of $\mathbf{v}_2$. By the Spanning Set Theorem, we may discard $\mathbf{v}_3$ and still have a set that spans H. Finally, $\mathbf{v}_4$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. So $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ is linearly independent (by Theorem 4 in Section 4.3) and hence is a basis for H. Thus $\dim H = 3$. ■

EXAMPLE 4 The subspaces of $\mathbb{R}^3$ can be classified by dimension. See Figure 1.

0-dimensional subspaces. Only the zero subspace.

1-dimensional subspaces. Any subspace spanned by a single nonzero vector. Such subspaces are lines through the origin.

2-dimensional subspaces. Any subspace spanned by two linearly independent vectors. Such subspaces are planes through the origin.

3-dimensional subspaces. Only $\mathbb{R}^3$ itself. Any three linearly independent vectors in $\mathbb{R}^3$ span all of $\mathbb{R}^3$, by the Invertible Matrix Theorem. ■

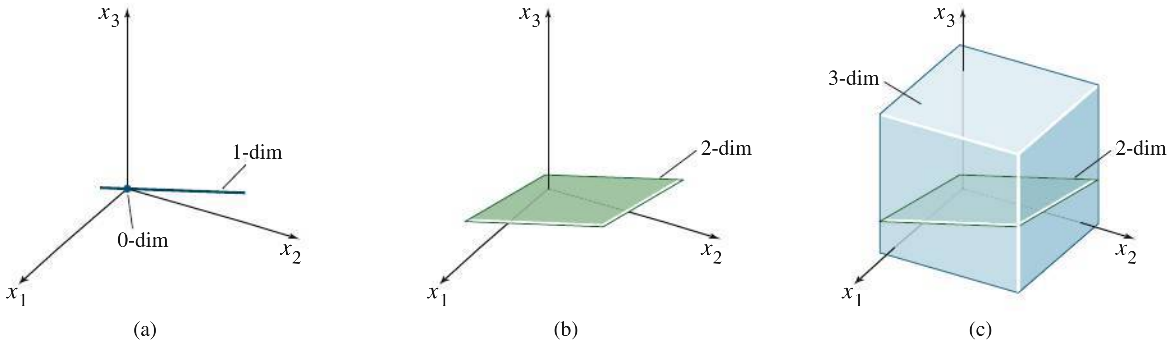


FIGURE 1 Sample subspaces of $\mathbb{R}^3$.

Subspaces of a Finite-Dimensional Space

The next theorem is a natural counterpart to the Spanning Set Theorem.

THEOREM 12

Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and

$$\dim H \leq \dim V$$

PROOF If $H = \{\mathbf{0}\}$, then certainly $\dim H = 0 \leq \dim V$. Otherwise, let $S = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ be any linearly independent set in H. If S spans H, then S is a basis for H. Otherwise, there is some $\mathbf{u}_{k+1}$ in H that is not in Span S. But then $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}\}$ will be linearly independent, because no vector in the set can be a linear combination of vectors that precede it (by Theorem 4).

So long as the new set does not span H, we can continue this process of expanding S to a larger linearly independent set in H. But the number of vectors in a linearly independent expansion of S can never exceed the dimension of V, by Theorem 10. So eventually the expansion of S will span H and hence will be a basis for H, and $\dim H \leq \dim V$. ∎

When the dimension of a vector space or subspace is known, the search for a basis is simplified by the next theorem. It says that if a set has the right number of elements, then one has only to show either that the set is linearly independent or that it spans the space. The theorem is of critical importance in numerous applied problems (involving differential equations or difference equations, for example) where linear independence is much easier to verify than spanning.

THEOREM 13

The Basis Theorem

Let V be a p-dimensional vector space, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.

PROOF By Theorem 12, a linearly independent set S of p elements can be extended to a basis for V. But that basis must contain exactly p elements, since $\dim V = p$. So S must already be a basis for V. Now suppose that S has p elements and spans V. Since V is nonzero, the Spanning Set Theorem implies that a subset S' of S is a basis of V. Since $\dim V = p$, S' must contain p vectors. Hence $S = S'$. ∎

The Dimensions of Nul A, Col A, and Row A

Since the dimensions of the null space and column space of an $m \times n$ matrix are referred to frequently, they have specific names:

DEFINITION

The **rank** of an $m \times n$ matrix A is the dimension of the column space and the **nullity** of A is the dimension of the null space.

The pivot columns of a matrix A form a basis for Col A, so the rank of A is just the number of pivot columns. Since a basis for Row A can be found by taking the pivot rows from the row reduced echelon form of A, the dimension of Row A is also equal to the rank of A.

The nullity of A might seem to require more work, since finding a basis for Nul A usually takes more time than finding a basis for Col A. There is a shortcut: Let A be an $m \times n$ matrix, and suppose the equation $A\mathbf{x} = \mathbf{0}$ has k free variables. From Section 4.2, we know that the standard method of finding a spanning set for Nul A will produce exactly k linearly independent vectors—say, $\mathbf{u}_1, \ldots, \mathbf{u}_k$ – one for each free variable. So $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a basis for Nul A, and the number of free variables determines the size of the basis.

To summarize these facts for future reference:

The rank of an $m \times n$ matrix A is the number of pivot columns and the nullity of A is the number of free variables. Since the dimension of the row space is the number of pivot rows, it is also equal to the rank of A.

Putting these observations together results in the rank theorem.

THEOREM 14

The Rank Theorem

The dimensions of the column space and the null space of an $m \times n$ matrix A satisfy the equation

$$\text{rank } A + \text{nullity } A = \text{number of columns in } A$$

PROOF By Theorem 6 in Section 4.3, rank A is the number of pivot columns in A. The nullity of A equals the number of free variables in the equation $A\mathbf{x} = \mathbf{0}$. Expressed another way, the nullity of A is the number of columns of A that are *not* pivot columns. (It is the number of these columns, not the columns themselves, that is related to Nul A.) Obviously,

$$\left\{\begin{array}{c}\text{number of} \\ \text{pivot columns}\end{array}\right\} + \left\{\begin{array}{c}\text{number of} \\ \text{nonpivot columns}\end{array}\right\} = \left\{\begin{array}{c}\text{number of} \\ \text{columns}\end{array}\right\}$$

This proves the theorem. ■

EXAMPLE 5 Find the nullity and rank of

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

SOLUTION Row reduce the augmented matrix $[A \;\; \mathbf{0}]$ to echelon form:

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

There are three free variables: x_2, x_4, and x_5. Hence the nullity of A is 3. Also, the rank of A is 2 because A has two pivot columns. ■

The ideas behind Theorem 14 are visible in the calculations in Example 5. The two pivot positions in B, an echelon form of A, determine the basic variables and identify the basis vectors for Col A and those for Row A.

EXAMPLE 6

a. If A is a 7×9 matrix with nullity 2, what is the rank of A?

b. Could a 6×9 matrix have nullity 2?

SOLUTION

a. Since A has 9 columns, $(\text{rank } A) + 2 = 9$, and hence rank $A = 7$.

b. No. If a 6×9 matrix, call it B, had a two-dimensional null space, it would have to have rank 7, by the Rank Theorem. But the columns of B are vectors in $\mathbb{R}^6$, and so the dimension of Col B cannot exceed 6; that is, rank B cannot exceed 6. ■

The next example provides a nice way to visualize the subspaces we have been studying. In Chapter 6, we will learn that Row A and Nul A have only the zero vector in common and are actually perpendicular to each other. The same fact applies to Row A^T ($=$ Col$A$) and Nul A^T. So Figure 2, which accompanies Example 7, creates a good mental image for the general case.

EXAMPLE 7 Let $A = \begin{bmatrix} 3 & 0 & -1 \\ 3 & 0 & -1 \\ 4 & 0 & 5 \end{bmatrix}$. It is readily checked that Nul A is the x_2-axis, Row A is the x_1x_3-plane, Col A is the plane whose equation is $x_1 - x_2 = 0$, and Nul A^T is the set of all multiples of $(1, -1, 0)$. Figure 2 shows Nul A and Row A in the domain of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$; the range of this mapping, Col A, is shown in a separate copy of $\mathbb{R}^3$, along with Nul A^T. ■

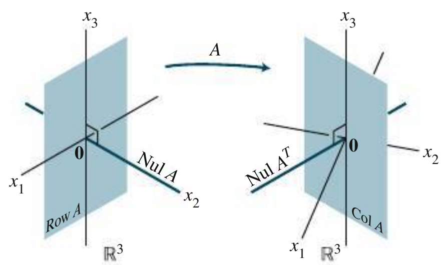


FIGURE 2 Subspaces determined by a matrix A.

Applications to Systems of Equations

The Rank Theorem is a powerful tool for processing information about systems of linear equations. The next example simulates the way a real-life problem using linear equations might be stated, without explicit mention of linear algebra terms such as matrix, subspace, and dimension.

EXAMPLE 8 A scientist has found two solutions to a homogeneous system of 40 equations in 42 variables. The two solutions are not multiples, and all other solutions can be constructed by adding together appropriate multiples of these two solutions. Can the scientist be *certain* that an associated nonhomogeneous system (with the same coefficients) has a solution?

SOLUTION Yes. Let A be the 40×42 coefficient matrix of the system. The given information implies that the two solutions are linearly independent and span Nul A. So nullity $A = 2$. By the Rank Theorem, rank $A = 42 - 2 = 40$. Since $\mathbb{R}^{40}$ is the only subspace of $\mathbb{R}^{40}$ whose dimension is 40, Col A must be all of $\mathbb{R}^{40}$. This means that every nonhomogeneous equation $A\mathbf{x} = \mathbf{b}$ has a solution. ■

Rank and the Invertible Matrix Theorem

The various vector space concepts associated with a matrix provide several more statements for the Invertible Matrix Theorem. The new statements listed here follow those in the original Invertible Matrix Theorem in Section 2.3 and other theorems in the text where statements have been added to it.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix.

m. The columns of A form a basis of $\mathbb{R}^n$.
n. Col $A = \mathbb{R}^n$
o. rank $A = n$
p. nullity $A = 0$
q. Nul $A = \{\mathbf{0}\}$

PROOF Statement (m) is logically equivalent to statements (e) and (h) regarding linear independence and spanning. The other five statements are linked to the earlier ones of the theorem by the following chain of almost trivial implications:

$$(g) \Rightarrow (n) \Rightarrow (o) \Rightarrow (p) \Rightarrow (q) \Rightarrow (d)$$

Statement (g), which says that the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$, implies (n), because Col A is precisely the set of all $\mathbf{b}$ such that the equation $A\mathbf{x} = \mathbf{b}$ is consistent. The implication (n) $\Rightarrow$ (o) follows from the definitions of dimension and rank. If the rank of A is n, the number of columns of A, then nullity $A = 0$, by the Rank Theorem, and so Nul $A = \{\mathbf{0}\}$. Thus (o) $\Rightarrow$ (p) $\Rightarrow$ (q). Also, (q) implies that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, which is statement (d). Since statements (d) and (g) are already known to be equivalent to the statement that A is invertible, the proof is complete. ■

We have refrained from adding to the Invertible Matrix Theorem obvious statements about the row space of A, because the row space is the column space of A^T. Recall from statement (l) of the Invertible Matrix Theorem that A is invertible if and only if A^T is invertible. Hence every statement in the Invertible Matrix Theorem can also be stated for A^T. To do so would double the length of the theorem and produce a list of more than 30 statements!

Numerical Notes

Many algorithms discussed in this text are useful for understanding concepts and making simple computations by hand. However, the algorithms are often unsuitable for large-scale problems in real life.

Rank determination is a good example. It would seem easy to reduce a matrix to echelon form and count the pivots. But unless exact arithmetic is performed on a matrix whose entries are specified exactly, row operations can change the apparent rank of a matrix. For instance, if the value of x in the matrix $\begin{bmatrix} 5 & 7 \\ 5 & x \end{bmatrix}$ is not stored exactly as 7 in a computer, then the rank may be 1 or 2, depending on whether the computer treats $x - 7$ as zero.

In practical applications, the effective rank of a matrix A is often determined from the singular value decomposition of A, to be discussed in Section 7.4. This decomposition is also a reliable source of bases for Col A, Row A, Nul A, and Nul A^T.

Practice Problems

1. Decide whether each statement is True or False, and give a reason for each answer. Here V is a nonzero finite-dimensional vector space.
   a. If $\dim V = p$ and if S is a linearly dependent subset of V, then S contains more than p vectors.
   b. If S spans V and if T is a subset of V that contains more vectors than S, then T is linearly dependent.
2. Let H and K be subspaces of a vector space V. In Section 4.1, Exercise 40, it is established that $H \cap K$ is also a subspace of V. Prove $\dim(H \cap K) \leq \dim H$.

4.5 Exercises

For each subspace in Exercises 1–8, (a) find a basis, and (b) state the dimension.

1. $\left\{ \begin{bmatrix} s-2t \\ s+t \\ 3t \end{bmatrix} : s, t \text{ in } \mathbb{R} \right\}$

2. $\left\{ \begin{bmatrix} 4s \\ -3s \\ -t \end{bmatrix} : s, t \text{ in } \mathbb{R} \right\}$

3. $\left\{ \begin{bmatrix} 2c \\ a-b \\ b-3c \\ a+2b \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

4. $\left\{ \begin{bmatrix} a+b \\ 2a \\ 3a-b \\ -b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$

5. $\left\{ \begin{bmatrix} a-4b-2c \\ 2a+5b-4c \\ -a+2c \\ -3a+7b+6c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

6. $\left\{ \begin{bmatrix} 3a+6b-c \\ 6a-2b-2c \\ -9a+5b+3c \\ -3a+b+c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

7. $\{(a, b, c) : a - 3b + c = 0, b - 2c = 0, 2b - c = 0\}$

8. $\{(a, b, c, d) : a - 3b + c = 0\}$

In Exercises 9 and 10, find the dimension of the subspace spanned by the given vectors.

9. $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 4 \\ -2 \end{bmatrix}, \begin{bmatrix} -7 \\ -3 \\ 1 \end{bmatrix}$

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

Determine the dimensions of Nul A, Col A, and Row A for the matrices shown in Exercises 11–16.

11. $A = \begin{bmatrix} 1 & -6 & 9 & 0 & -2 \\ 0 & 1 & 2 & -4 & 5 \\ 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

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

13. $A = \begin{bmatrix} 1 & 0 & 9 & 5 \\ 0 & 0 & 1 & -4 \end{bmatrix}$

14. $A = \begin{bmatrix} 3 & 4 \\ -6 & 10 \end{bmatrix}$

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

16. $A = \begin{bmatrix} 1 & 4 & -1 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

In Exercises 17–26, V is a vector space and A is an $m \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** The number of pivot columns of a matrix equals the dimension of its column space.

18. **(T/F)** The number of variables in the equation $A\mathbf{x} = \mathbf{0}$ equals the nullity A.

19. **(T/F)** A plane in $\mathbb{R}^3$ is a two-dimensional subspace of $\mathbb{R}^3$.

20. **(T/F)** The dimension of the vector space $\mathbb{P}_4$ is 4.

21. **(T/F)** The dimension of the vector space of signals, $\mathbb{S}$, is 10.

22. **(T/F)** The dimensions of the row space and the column space of A are the same, even if A is not square.

23. **(T/F)** If B is any echelon form of A, then the pivot columns of B form a basis for the column space of A.

24. **(T/F)** The nullity of A is the number of columns of A that are not pivot columns.

25. **(T/F)** If a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ spans a finite-dimensional vector space V and if T is a set of more than p vectors in V, then T is linearly dependent.

26. **(T/F)** A vector space is infinite-dimensional if it is spanned by an infinite set.

27. The first four Hermite polynomials are 1, $2t$, $-2 + 4t^2$, and $-12t + 8t^3$. These polynomials arise naturally in the study of certain important differential equations in mathematical physics.[2] Show that the first four Hermite polynomials form a basis of $\mathbb{P}_3$.

28. The first four Laguerre polynomials are 1, $1 - t$, $2 - 4t + t^2$, and $6 - 18t + 9t^2 - t^3$. Show that these polynomials form a basis of $\mathbb{P}_3$.

29. Let $\mathcal{B}$ be the basis of $\mathbb{P}_3$ consisting of the Hermite polynomials in Exercise 27, and let $\mathbf{p}(t) = 7 - 12t - 8t^2 + 12t^3$. Find the coordinate vector of $\mathbf{p}$ relative to $\mathcal{B}$.

30. Let $\mathcal{B}$ be the basis of $\mathbb{P}_2$ consisting of the first three Laguerre polynomials listed in Exercise 28, and let $\mathbf{p}(t) = 7 - 8t + 3t^2$. Find the coordinate vector of $\mathbf{p}$ relative to $\mathcal{B}$.

31. Let S be a subset of an n-dimensional vector space V, and suppose S contains fewer than n vectors. Explain why S cannot span V.

32. Let H be an n-dimensional subspace of an n-dimensional vector space V. Show that $H = V$.

33. If a 3×8 matrix A has rank 3, find nullity A, rank A, and rank A^T.

34. If a 6×3 matrix A has rank 3, find nullity A, rank A, and rank A^T.

35. Suppose a 4×7 matrix A has four pivot columns. Is Col $A = \mathbb{R}^4$? Is Nul $A = \mathbb{R}^3$? Explain your answers.

36. Suppose a 5×6 matrix A has four pivot columns. What is nullity A? Is Col $A = \mathbb{R}^4$? Why or why not?

37. If the nullity of a 5×6 matrix A is 4, what are the dimensions of the column and row spaces of A?

38. If the nullity of a 7×6 matrix A is 5, what are the dimensions of the column and row spaces of A?

39. If A is a 7×5 matrix, what is the largest possible rank of A? If A is a 5×7 matrix, what is the largest possible rank of A? Explain your answers.

40. If A is a 4×3 matrix, what is the largest possible dimension of the row space of A? If A is a 3×4 matrix, what is the largest possible dimension of the row space of A? Explain.

41. Explain why the space $\mathbb{P}$ of all polynomials is an infinite-dimensional space.

42. Show that the space $C(\mathbb{R})$ of all continuous functions defined on the real line is an infinite-dimensional space.

In Exercises 43–48, V is a nonzero finite-dimensional vector space, and the vectors listed belong to V. Mark each statement True or False **(T/F)**. Justify each answer. (These questions are more difficult than those in Exercises 17–26.)

---

[2] See *Introduction to Functional Analysis*, 2nd ed., by A. E. Taylor and David C. Lay (New York: John Wiley & Sons, 1980), pp. 92–93. Other sets of polynomials are discussed there, too.

43. **(T/F)** If there exists a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ that spans V, then $\dim V \le p$.

44. **(T/F)** If there exists a linearly dependent set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in V, then $\dim V \le p$.

45. **(T/F)** If there exists a linearly independent set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in V, then $\dim V \ge p$.

46. **(T/F)** If $\dim V = p$, then there exists a spanning set of $p+1$ vectors in V.

47. **(T/F)** If every set of p elements in V fails to span V, then $\dim V > p$.

48. **(T/F)** If $p \ge 2$ and $\dim V = p$, then every set of $p-1$ nonzero vectors is linearly independent.

49. Justify the following equality: $\dim \text{ Row } A + \text{nullity } A = n$, the number of columns of A

50. Justify the following equality: $\dim \text{ Row } A + \text{nullity } A^T = m$, the number of rows of A

Exercises 51 and 52 concern finite-dimensional vector spaces V and W and a linear transformation $T : V \to W$.

51. Let H be a nonzero subspace of V, and let $T(H)$ be the set of images of vectors in H. Then $T(H)$ is a subspace of W, by Exercise 47 in Section 4.2. Prove that $\dim T(H) \le \dim H$.

52. Let H be a nonzero subspace of V, and suppose T is a one-to-one (linear) mapping of V into W. Prove that $\dim T(H) = \dim H$. If T happens to be a one-to-one mapping of V *onto* W, then $\dim V = \dim W$. Isomorphic finite-dimensional vector spaces have the same dimension.

T 53. According to Theorem 12, a linearly independent set $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ can be expanded to a basis for $\mathbb{R}^n$. One way to do this is to create $A = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_k \ \ \mathbf{e}_1 \ \cdots \ \mathbf{e}_n\,]$, with $\mathbf{e}_1, \dots, \mathbf{e}_n$ the columns of the identity matrix; the pivot columns of A form a basis for $\mathbb{R}^n$.

a. Use the method described to extend the following vectors to a basis for $\mathbb{R}^5$:

$$\mathbf{v}_1 = \begin{bmatrix} -9 \\ -7 \\ 8 \\ -5 \\ 7 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 9 \\ 4 \\ 1 \\ 6 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 7 \\ -8 \\ 5 \\ -7 \end{bmatrix}$$

b. Explain why the method works in general: Why are the original vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ included in the basis found for Col A? Why is Col $A = \mathbb{R}^n$?

T 54. Let $\mathcal{B} = \{1, \cos t, \cos^2 t, \dots, \cos^6 t\}$ and $\mathcal{C} = \{1, \cos t, \cos 2t, \dots, \cos 6t\}$. Assume the following trigonometric identities (see Exercise 45 in Section 4.1).

$$\begin{aligned}
\cos 2t &= -1 + 2\cos^2 t \\
\cos 3t &= -3\cos t + 4\cos^3 t \\
\cos 4t &= 1 - 8\cos^2 t + 8\cos^4 t \\
\cos 5t &= 5\cos t - 20\cos^3 t + 16\cos^5 t \\
\cos 6t &= -1 + 18\cos^2 t - 48\cos^4 t + 32\cos^6 t
\end{aligned}$$

Let H be the subspace of functions spanned by the functions in $\mathcal{B}$. Then $\mathcal{B}$ is a basis for H, by Exercise 48 in Section 4.3.

a. Write the $\mathcal{B}$-coordinate vectors of the vectors in $\mathcal{C}$, and use them to show that $\mathcal{C}$ is a linearly independent set in H.

b. Explain why $\mathcal{C}$ is a basis for H.

Solutions to Practice Problems

1. a. False. Consider the set $\{\mathbf{0}\}$.

b. True. By the Spanning Set Theorem, S contains a basis for V; call that basis S'. Then T will contain more vectors than S'. By Theorem 10, T is linearly dependent.

2. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a basis for $H \cap K$. Notice $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a linearly independent subset of H, hence by Theorem 12, $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ can be expanded, if necessary, to a basis for H. Since the dimension of a subspace is just the number of vectors in a basis, it follows that $\dim (H \cap K) = p \le \dim H$.

4.6 Change of Basis

When a basis $\mathcal{B}$ is chosen for an n-dimensional vector space V, the associated coordinate mapping onto $\mathbb{R}^n$ provides a coordinate system for V. Each $\mathbf{x}$ in V is identified uniquely by its $\mathcal{B}$-coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$.[1]

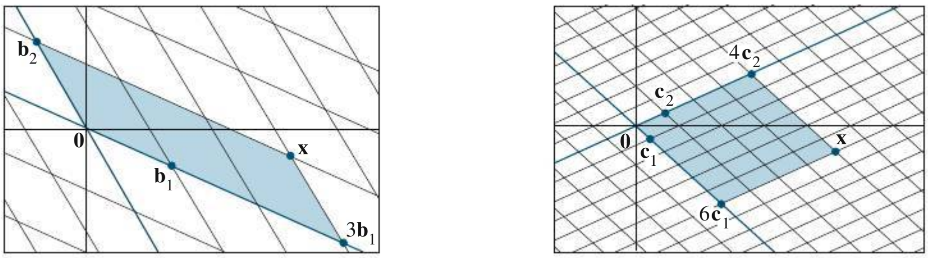
[1] Think of $[\,\mathbf{x}\,]_{\mathcal{B}}$ as a name for $\mathbf{x}$ that lists the weights used to build $\mathbf{x}$ as a linear combination of the basis vectors in $\mathcal{B}$.

In some applications, a problem is described initially using a basis $\mathcal{B}$, but the problem's solution is aided by changing $\mathcal{B}$ to a new basis $\mathcal{C}$. (Examples will be given in Chapters 5 and 7.) Each vector is assigned a new $\mathcal{C}$-coordinate vector. In this section, we study how $[\,\mathbf{x}\,]_{\mathcal{C}}$ and $[\,\mathbf{x}\,]_{\mathcal{B}}$ are related for each $\mathbf{x}$ in V.

To visualize the problem, consider the two coordinate systems in Figure 1. In Figure 1(a), $\mathbf{x} = 3\mathbf{b}_1 + \mathbf{b}_2$, while in Figure 1(b), the same $\mathbf{x}$ is shown as $\mathbf{x} = 6\mathbf{c}_1 + 4\mathbf{c}_2$. That is,

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \text{and} \quad [\,\mathbf{x}\,]_{\mathcal{C}} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

Our problem is to find the connection between the two coordinate vectors. Example 1 shows how to do this, provided we know how $\mathbf{b}_1$ and $\mathbf{b}_2$ are formed from $\mathbf{c}_1$ and $\mathbf{c}_2$.



(a) (b)

FIGURE 1 Two coordinate systems for the same vector space.

EXAMPLE 1 Consider two bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ for a vector space V, such that

$$\mathbf{b}_1 = 4\mathbf{c}_1 + \mathbf{c}_2 \quad \text{and} \quad \mathbf{b}_2 = -6\mathbf{c}_1 + \mathbf{c}_2 \tag{1}$$

Suppose

$$\mathbf{x} = 3\mathbf{b}_1 + \mathbf{b}_2 \tag{2}$$

That is, suppose $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Find $[\,\mathbf{x}\,]_{\mathcal{C}}$.

SOLUTION Apply the coordinate mapping determined by $\mathcal{C}$ to $\mathbf{x}$ in (2). Since the coordinate mapping is a linear transformation,

$$\begin{aligned} [\,\mathbf{x}\,]_{\mathcal{C}} &= [\,3\mathbf{b}_1 + \mathbf{b}_2\,]_{\mathcal{C}} \\ &= 3[\,\mathbf{b}_1\,]_{\mathcal{C}} + [\,\mathbf{b}_2\,]_{\mathcal{C}} \end{aligned}$$

We can write this vector equation as a matrix equation, using the vectors in the linear combination as the columns of a matrix:

$$[\,\mathbf{x}\,]_{\mathcal{C}} = \begin{bmatrix} [\,\mathbf{b}_1\,]_{\mathcal{C}} & [\,\mathbf{b}_2\,]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \tag{3}$$

This formula gives $[\,\mathbf{x}\,]_{\mathcal{C}}$, once we know the columns of the matrix. From (1),

$$[\,\mathbf{b}_1\,]_{\mathcal{C}} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \quad \text{and} \quad [\,\mathbf{b}_2\,]_{\mathcal{C}} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$$

Thus (3) provides the solution:

$$[\,\mathbf{x}\,]_{\mathcal{C}} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

The $\mathcal{C}$-coordinates of $\mathbf{x}$ match those of the $\mathbf{x}$ in Figure 1. ■

The argument used to derive formula (3) can be generalized to yield the following result. (See Exercises 17 and 18.)

THEOREM 15

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases of a vector space V. Then there is a unique $n \times n$ matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ such that

$$[\,\mathbf{x}\,]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\,\mathbf{x}\,]_{\mathcal{B}} \tag{4}$$

The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are the $\mathcal{C}$-coordinate vectors of the vectors in the basis $\mathcal{B}$. That is,

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_n]_{\mathcal{C}} \end{bmatrix} \tag{5}$$

The matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ in Theorem 15 is called the **change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$**. Multiplication by $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ converts $\mathcal{B}$-coordinates into $\mathcal{C}$-coordinates.[2] Figure 2 illustrates the change-of-coordinates equation (4).

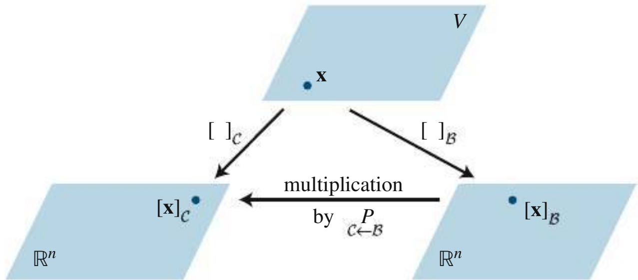


FIGURE 2 Two coordinate systems for V.

The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are linearly independent because they are the coordinate vectors of the linearly independent set $\mathcal{B}$. (See Exercise 29 in Section 4.4.) Since $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ is square, it must be invertible, by the Invertible Matrix Theorem. Left-multiplying both sides of equation (4) by $(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1}$ yields

$$(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1}[\,\mathbf{x}\,]_{\mathcal{C}} = [\,\mathbf{x}\,]_{\mathcal{B}}$$

Thus $(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1}$ is the matrix that converts $\mathcal{C}$-coordinates into $\mathcal{B}$-coordinates. That is,

$$(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P} \tag{6}$$

Change of Basis in $\mathbb{R}^n$

If $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{E}$ is the *standard basis* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ in $\mathbb{R}^n$, then $[\mathbf{b}_1]_{\mathcal{E}} = \mathbf{b}_1$, and likewise for the other vectors in $\mathcal{B}$. In this case, $\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}$ is the same as the change-of-coordinates matrix $P_{\mathcal{B}}$ introduced in Section 4.4, namely

$$P_{\mathcal{B}} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$$

[2] To remember how to construct the matrix, think of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\,\mathbf{x}\,]_{\mathcal{B}}$ as a linear combination of the columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$. The matrix-vector product is a $\mathcal{C}$-coordinate vector, so the columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ should be $\mathcal{C}$-coordinate vectors, too.

To change coordinates between two nonstandard bases in $\mathbb{R}^n$, we need Theorem 15. The theorem shows that to solve the change-of-basis problem, we need the coordinate vectors of the old basis relative to the new basis.

EXAMPLE 2 Let $\mathbf{b}_1 = \begin{bmatrix} -9 \\ 1 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -5 \\ -1 \end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$, $\mathbf{c}_2 = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$, and consider the bases for $\mathbb{R}^2$ given by $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.

SOLUTION The matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ involves the $\mathcal{C}$-coordinate vectors of $\mathbf{b}_1$ and $\mathbf{b}_2$. Let $[\,\mathbf{b}_1\,]_\mathcal{C} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $[\,\mathbf{b}_2\,]_\mathcal{C} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then, by definition,

$$\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{b}_1 \quad \text{and} \quad \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{b}_2$$

To solve both systems simultaneously, augment the coefficient matrix with $\mathbf{b}_1$ *and* $\mathbf{b}_2$, and row reduce:

$$\left[\begin{array}{cc:cc} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right] = \left[\begin{array}{cc:cc} 1 & 3 & -9 & -5 \\ -4 & -5 & 1 & -1 \end{array}\right] \sim \left[\begin{array}{cc:cc} 1 & 0 & 6 & 4 \\ 0 & 1 & -5 & -3 \end{array}\right] \tag{7}$$

Thus

$$[\,\mathbf{b}_1\,]_\mathcal{C} = \begin{bmatrix} 6 \\ -5 \end{bmatrix} \quad \text{and} \quad [\,\mathbf{b}_2\,]_\mathcal{C} = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$$

The desired change-of-coordinates matrix is therefore

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [\,\mathbf{b}_1\,]_\mathcal{C} & [\,\mathbf{b}_2\,]_\mathcal{C} \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ -5 & -3 \end{bmatrix}$$ ■

Observe that the matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ in Example 2 already appeared in (7). This is not surprising because the first column of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ results from row reducing $\left[\begin{array}{cc:c} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{b}_1 \end{array}\right]$ to $\left[\begin{array}{c:c} I & [\,\mathbf{b}_1\,]_\mathcal{C} \end{array}\right]$, and similarly for the second column of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$. Thus

$$\boxed{\left[\begin{array}{cc:cc} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right] \sim \left[\begin{array}{c:c} I & \underset{\mathcal{C}\leftarrow\mathcal{B}}{P} \end{array}\right]}$$

An analogous procedure works for finding the change-of-coordinates matrix between any two bases in $\mathbb{R}^n$.

EXAMPLE 3 Let $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix} -7 \\ 9 \end{bmatrix}$, $\mathbf{c}_2 = \begin{bmatrix} -5 \\ 7 \end{bmatrix}$, and consider the bases for $\mathbb{R}^2$ given by $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$.

a. Find the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$.

b. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.

SOLUTION

a. Notice that $\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}$ is needed rather than $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$, and compute

$$\left[\begin{array}{cc:cc} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{c}_1 & \mathbf{c}_2 \end{array}\right] = \left[\begin{array}{cc:cc} 1 & -2 & -7 & -5 \\ -3 & 4 & 9 & 7 \end{array}\right] \sim \left[\begin{array}{cc:cc} 1 & 0 & 5 & 3 \\ 0 & 1 & 6 & 4 \end{array}\right]$$

So

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} 5 & 3 \\ 6 & 4 \end{bmatrix}$$

b. By part (a) and property (6) (with $\mathcal{B}$ and $\mathcal{C}$ interchanged),

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = (\underset{\mathcal{B}\leftarrow\mathcal{C}}{P})^{-1} = \frac{1}{2}\begin{bmatrix} 4 & -3 \\ -6 & 5 \end{bmatrix} = \begin{bmatrix} 2 & -3/2 \\ -3 & 5/2 \end{bmatrix} \quad \blacksquare$$

Another description of the change-of-coordinates matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ uses the change-of-coordinate matrices $P_{\mathcal{B}}$ and $P_{\mathcal{C}}$ that convert $\mathcal{B}$-coordinates and $\mathcal{C}$-coordinates, respectively, into standard coordinates. Recall that for each $\mathbf{x}$ in $\mathbb{R}^n$,

$$P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}, \quad P_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}} = \mathbf{x}, \quad \text{and} \quad [\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}\mathbf{x}$$

Thus

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}\mathbf{x} = P_{\mathcal{C}}^{-1}P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

In $\mathbb{R}^n$, the change-of-coordinates matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ may be computed as $P_{\mathcal{C}}^{-1}P_{\mathcal{B}}$. Actually, for matrices larger than 2×2, an algorithm analogous to the one in Example 3 is faster than computing $P_{\mathcal{C}}^{-1}$ and then $P_{\mathcal{C}}^{-1}P_{\mathcal{B}}$. See Exercise 22 in Section 2.2.

Practice Problems

1. Let $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2\}$ and $\mathcal{G} = \{\mathbf{g}_1, \mathbf{g}_2\}$ be bases for a vector space V, and let P be a matrix whose columns are $[\,\mathbf{f}_1\,]_{\mathcal{G}}$ and $[\,\mathbf{f}_2\,]_{\mathcal{G}}$. Which of the following equations is satisfied by P for all $\mathbf{v}$ in V?

   (i) $[\,\mathbf{v}\,]_{\mathcal{F}} = P[\,\mathbf{v}\,]_{\mathcal{G}}$ (ii) $[\,\mathbf{v}\,]_{\mathcal{G}} = P[\,\mathbf{v}\,]_{\mathcal{F}}$
2. Let $\mathcal{B}$ and $\mathcal{C}$ be as in Example 1. Use the results of that example to find the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$.

4.6 Exercises

1. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ be bases for a vector space V, and suppose $\mathbf{b}_1 = 6\mathbf{c}_1 - 2\mathbf{c}_2$ and $\mathbf{b}_2 = 9\mathbf{c}_1 - 4\mathbf{c}_2$.
   a. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{C}}$ for $\mathbf{x} = -3\mathbf{b}_1 + 2\mathbf{b}_2$. Use part (a).

2. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ be bases for a vector space V, and suppose $\mathbf{b}_1 = -\mathbf{c}_1 + 4\mathbf{c}_2$ and $\mathbf{b}_2 = 5\mathbf{c}_1 - 3\mathbf{c}_2$.
   a. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{C}}$ for $\mathbf{x} = 5\mathbf{b}_1 + 3\mathbf{b}_2$.

3. Let $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $\mathcal{W} = \{\mathbf{w}_1, \mathbf{w}_2\}$ be bases for V, and let P be a matrix whose columns are $[\mathbf{u}_1]_{\mathcal{W}}$ and $[\mathbf{u}_2]_{\mathcal{W}}$. Which of the following equations is satisfied by P for all $\mathbf{x}$ in V?

   (i) $[\,\mathbf{x}\,]_{\mathcal{U}} = P[\,\mathbf{x}\,]_{\mathcal{W}}$ (ii) $[\,\mathbf{x}\,]_{\mathcal{W}} = P[\,\mathbf{x}\,]_{\mathcal{U}}$

4. Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ and $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ be bases for V, and let $P = [\,[\mathbf{d}_1]_{\mathcal{A}} \quad [\mathbf{d}_2]_{\mathcal{A}} \quad [\mathbf{d}_3]_{\mathcal{A}}\,]$. Which of the following equations is satisfied by P for all $\mathbf{x}$ in V?

   (i) $[\,\mathbf{x}\,]_{\mathcal{A}} = P[\,\mathbf{x}\,]_{\mathcal{D}}$ (ii) $[\,\mathbf{x}\,]_{\mathcal{D}} = P[\,\mathbf{x}\,]_{\mathcal{A}}$

5. Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be bases for a vector space V, and suppose $\mathbf{a}_1 = 4\mathbf{b}_1 - \mathbf{b}_2$, $\mathbf{a}_2 = -\mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3$, and $\mathbf{a}_3 = \mathbf{b}_2 - 2\mathbf{b}_3$.
   a. Find the change-of-coordinates matrix from $\mathcal{A}$ to $\mathcal{B}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{B}}$ for $\mathbf{x} = 3\mathbf{a}_1 + 4\mathbf{a}_2 + \mathbf{a}_3$.

6. Let $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ and $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ be bases for a vector space V, and suppose $\mathbf{f}_1 = 2\mathbf{d}_1 - \mathbf{d}_2 + \mathbf{d}_3$, $\mathbf{f}_2 = 3\mathbf{d}_2 + \mathbf{d}_3$, and $\mathbf{f}_3 = -3\mathbf{d}_1 + 2\mathbf{d}_3$.
   a. Find the change-of-coordinates matrix from $\mathcal{F}$ to $\mathcal{D}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{D}}$ for $\mathbf{x} = \mathbf{f}_1 - 2\mathbf{f}_2 + 2\mathbf{f}_3$.

In Exercises 7–10, let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ be bases for $\mathbb{R}^2$. In each exercise, find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$ and the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$.

7. $\mathbf{b}_1 = \begin{bmatrix} 7 \\ 5 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -3 \\ -1 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$

8. $\mathbf{b}_1 = \begin{bmatrix} -1 \\ 8 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

9. $\mathbf{b}_1 = \begin{bmatrix} -6 \\ -1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$

10. $\mathbf{b}_1 = \begin{bmatrix} 7 \\ -2 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

In Exercises 11–14, $\mathcal{B}$ and $\mathcal{C}$ are bases for a vector space V. Mark each statement True or False **(T/F)**. Justify each answer.

11. **(T/F)** The columns of the change-of-coordinates matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are $\mathcal{B}$-coordinate vectors of the vectors in $\mathcal{C}$.

12. **(T/F)** The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are linearly independent.

13. **(T/F)** If $V = \mathbb{R}^n$ and $\mathcal{C}$ is the *standard* basis for V, then $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ is the same as the change-of-coordinates matrix $P_{\mathcal{B}}$ introduced in Section 4.4.

14. **(T/F)** If $V = \mathbb{R}^2$, $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$, and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$, then row reduction of $[\,\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{b}_1 \quad \mathbf{b}_2\,]$ to $[\,I \quad P\,]$ produces a matrix P that satisfies $[\,\mathbf{x}\,]_{\mathcal{B}} = P[\,\mathbf{x}\,]_{\mathcal{C}}$ for all $\mathbf{x}$ in V.

15. In $\mathbb{P}_2$, find the change-of-coordinates matrix from the basis $\mathcal{B} = \{1 - 2t + t^2, 3 - 5t + 4t^2, 2t + 3t^2\}$ to the standard basis $\mathcal{C} = \{1, t, t^2\}$. Then find the $\mathcal{B}$-coordinate vector for $-1 + 2t$.

16. In $\mathbb{P}_2$, find the change-of-coordinates matrix from the basis $\mathcal{B} = \{1 - 3t^2, 2 + t - 5t^2, 1 + 2t\}$ to the standard basis. Then write t^2 as a linear combination of the polynomials in $\mathcal{B}$.

Exercises 17 and 18 provide a proof of Theorem 15. Fill in a justification for each step.

17. Given $\mathbf{v}$ in V, there exist scalars $x_1, \dots, x_n$, such that

$$\mathbf{v} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n$$

because (a) _____. Apply the coordinate mapping determined by the basis $\mathcal{C}$, and obtain

$$[\mathbf{v}]_{\mathcal{C}} = x_1[\mathbf{b}_1]_{\mathcal{C}} + x_2[\mathbf{b}_2]_{\mathcal{C}} + \cdots + x_n[\mathbf{b}_n]_{\mathcal{C}}$$

because (b) _____. This equation may be written in the form

$$[\,\mathbf{v}\,]_{\mathcal{C}} = \left[\,[\,\mathbf{b}_1\,]_{\mathcal{C}} \quad [\,\mathbf{b}_2\,]_{\mathcal{C}} \quad \cdots \quad [\,\mathbf{b}_n\,]_{\mathcal{C}}\,\right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{8}$$

by the definition of (c) _____. This shows that the matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ shown in (5) satisfies $[\mathbf{v}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\mathbf{v}]_{\mathcal{B}}$ for each $\mathbf{v}$ in V, because the vector on the right side of (8) is (d) _____.

18. Suppose Q is any matrix such that

$$[\mathbf{v}]_{\mathcal{C}} = Q[\mathbf{v}]_{\mathcal{B}} \quad \text{for each } \mathbf{v} \text{ in } V \tag{9}$$

Set $\mathbf{v} = \mathbf{b}_1$ in (9). Then (9) shows that $[\mathbf{b}_1]_{\mathcal{C}}$ is the first column of Q because (a) _____. Similarly, for $k = 2, \dots, n$, the kth column of Q is (b) _____ because (c) _____. This shows that the matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ defined by (5) in Theorem 15 is the only matrix that satisfies condition (4).

T 19. Let $\mathcal{B} = \{\mathbf{x}_0, \dots, \mathbf{x}_6\}$ and $C = \{\mathbf{y}_0, \dots, \mathbf{y}_6\}$, where $\mathbf{x}_k$ is the function $\cos^k t$ and $\mathbf{y}_k$ is the function $\cos kt$. Exercise 54 in Section 4.5 showed that both $\mathcal{B}$ and $\mathcal{C}$ are bases for the vector space $H = \text{Span}\,\{\mathbf{x}_0, \dots, \mathbf{x}_6\}$.

a. Set $P = \left[\,[\,\mathbf{y}_0\,]_{\mathcal{B}} \quad \cdots \quad [\,\mathbf{y}_6\,]_{\mathcal{B}}\,\right]$, and calculate P^{-1}.

b. Explain why the columns of P^{-1} are the $\mathcal{C}$-coordinate vectors of $\mathbf{x}_0, \dots, \mathbf{x}_6$. Then use these coordinate vectors to write trigonometric identities that express powers of $\cos t$ in terms of the functions in $\mathcal{C}$.

See the *Study Guide*.

T 20. (*Calculus required*)[3] Recall from calculus that integrals such as

$$\int (5\cos^3 t - 6\cos^4 t + 5\cos^5 t - 12\cos^6 t)\,dt \tag{10}$$

are tedious to compute. (The usual method is to apply integration by parts repeatedly and use the half-angle formula.) Use the matrix P or P^{-1} from Exercise 19 to transform (10); then compute the integral.

T 21. Let

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

$$\mathbf{v}_1 = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -8 \\ 5 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -7 \\ 2 \\ 6 \end{bmatrix}$$

a. Find a basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for $\mathbb{R}^3$ such that P is the change-of-coordinates matrix from $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ to the basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. [*Hint:* What do the columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ represent?]

b. Find a basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ for $\mathbb{R}^3$ such that P is the change-of-coordinates matrix from $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ to $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

T 22. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$, $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$, and $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2\}$ be bases for a two-dimensional vector space.

a. Write an equation that relates the matrices $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$, $\underset{\mathcal{D}\leftarrow\mathcal{C}}{P}$, and $\underset{\mathcal{D}\leftarrow\mathcal{B}}{P}$. Justify your result.

b. Use a matrix program either to help you find the equation or to check the equation you write. Work with three bases for $\mathbb{R}^2$. (See Exercises 7–10.)

[3] The idea for Exercises 19 and 20 and five related exercises in earlier sections came from a paper by Jack W. Rogers, Jr., of Auburn University, presented at a meeting of the International Linear Algebra Society, August 1995. See "Applications of Linear Algebra in Calculus," *American Mathematical Monthly* **104** (1), 1997.

Solutions to Practice Problems

1. Since the columns of P are $\mathcal{G}$-coordinate vectors, a vector of the form $P\mathbf{x}$ must be a $\mathcal{G}$-coordinate vector. Thus P satisfies equation (ii).
2. The coordinate vectors found in Example 1 show that

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \left[\,[\mathbf{b}_1]_{\mathcal{C}} \quad [\mathbf{b}_2]_{\mathcal{C}}\,\right] = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix}$$

Hence

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = (\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1} = \frac{1}{10}\begin{bmatrix} 1 & 6 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} .1 & .6 \\ -.1 & .4 \end{bmatrix}$$

4.7 Digital Signal Processing

Introduction

In the space of just a few decades, digital signal processing (DSP) has led to a dramatic shift in how data is collected, processed, and synthesized. DSP models unify the approach to dealing with data that was previously viewed as unrelated. From stock market analysis to telecommunications and computer science, the data collected over time can be viewed as discrete-time signals and DSP used to store and process the data for more efficient and effective use. Not only do digital signals arise in electrical and control systems engineering, but discrete-data sequences are also generated in biology, physics, economics, demography, and many other areas, wherever a process is measured, or *sampled*, at discrete time intervals. In this section, we will explore the properties of the discrete-time signal space, $\mathbb{S}$, and some of its subspaces, as well as how linear transformations can be used to process, filter, and synthesize the data contained in signals.

Discrete-Time Signals

The vector space $\mathbb{S}$ of discrete-time signals was introduced in Section 4.1. A **signal** in $\mathbb{S}$ is an infinite sequence of numbers, $\{y_k\}$, where the subscripts k range over all integers. Table 1 shows several examples of signals.

TABLE 1 Examples of Signals

Signals			
Name	**Symbol**	**Vector**	**Formal Description**
delta	δ	$(\ldots,0,0,0,1,0,0,0,\ldots)$	$\{d_k\}$, where $d_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$
unit step	υ	$(\ldots,0,0,0,1,1,1,1,\ldots)$	$\{u_k\}$, where $u_k = \begin{cases} 1 & \text{if } k \geq 0 \\ 0 & \text{if } k < 0 \end{cases}$
constant	χ	$(\ldots,1,1,1,1,1,1,1,\ldots)$	$\{c_k\}$, where $c_k = 1$
alternating	α	$(\ldots,-1,1,-1,1,-1,1,-1,\ldots)$	$\{a_k\}$, where $a_k = (-1)^k$
Fibonacci	F	$(\ldots,2,-1,1,0,1,1,2,\ldots)$	$\{f_k\}$, where $f_k = \begin{cases} 0 & \text{if } k = 0 \\ 1 & \text{if } k = 1 \\ f_{k-1}+f_{k-2} & \text{if } k > 1 \\ f_{k+2}-f_{k+1} & \text{if } k < 0 \end{cases}$
exponential	ϵ_c	$(\ldots,c^{-2},c^{-1},c^0,c^1,c^2,\ldots)$ $\uparrow$ $k=0$	$\{e_k\}$, where $e_k = c^k$

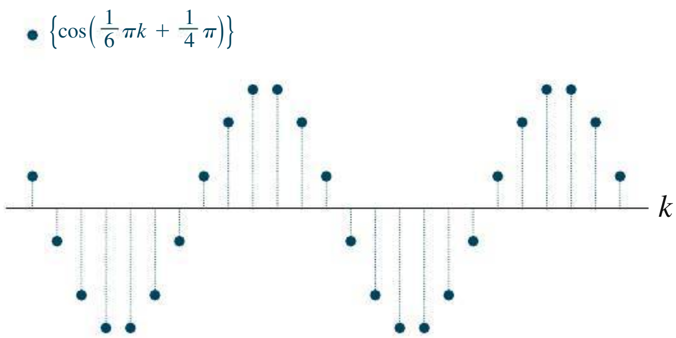


FIGURE 1

Another set of commonly used signals are the **periodic signals**—specifically signals $\{p_k\}$ for which there exists a positive integer q such that $p_k = p_{k+q}$ for all integers k. In particular, the sinusoidal signals, described by $\sigma_{f,\theta} = \{\cos(fk\pi + \theta\pi)\}$ where f and θ are fixed rational numbers, are periodic functions. (See Figure 1.)

Linear Time Invariant Transformations

Linear time invariant (LTI) transformations are used to process signals. One type of processing is to create signals as they are needed, rather than using valuable storage space to store the signals themselves.

To describe the standard basis for $\mathbb{R}^n$ given in Example 4 of Section 4.3, n vectors, $e_1, e_2, \dots, e_n$, are listed where e_j has a value of 1 in the j-th position and zeros elsewhere. In Example 1 that follows, the analogous signal to each e_j can be created by repeatedly applying a shift LTI transformation to just one signal, that of δ in Table 1.

EXAMPLE 1 Let S be the transformation that shifts each element in a signal to the right, specifically $S(\{x_k\}) = \{y_k\}$, where $y_k = x_{k-1}$. For ease of notation, write $S(\{x_k\}) = \{x_{k-1}\}$. To shift a signal to the left, consider $S^{-1}(\{x_k\}) = \{x_{k+1}\}$. Notice $S^{-1}S(\{x_k\}) = S^{-1}(\{x_{k-1}\}) = \{x_{(k-1)+1}\} = \{x_k\}$. It is easy to verify that $S^{-1}S = SS^{-1} = S^0 = I$, the identity transformation, and hence S is an example of an invertible transformation. Table 2 illustrates the effect of repeatedly applying S and S^{-1} to delta, and the resulting signals can be visualized using Figure 2.

TABLE 2 Applying a Shift Signal

$\vdots$	$\vdots$	$\vdots$
$S^{-2}(\delta)$	$(\dots, 1, 0, 0, 0, 0, \dots)$	$\{w_k\}$, where $w_k = \begin{cases} 1 & \text{if } k = -2 \\ 0 & \text{if } k \neq -2 \end{cases}$
$S^{-1}(\delta)$	$(\dots, 0, 1, 0, 0, 0, \dots)$	$\{x_k\}$, where $x_k = \begin{cases} 1 & \text{if } k = -1 \\ 0 & \text{if } k \neq -1 \end{cases}$
δ	$(\dots, 0, 0, 1, 0, 0, \dots)$	$\{d_k\}$, where $d_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$
$S^1(\delta)$	$(\dots, 0, 0, 0, 1, 0, \dots)$	$\{y_k\}$, where $y_k = \begin{cases} 1 & \text{if } k = 1 \\ 0 & \text{if } k \neq 1 \end{cases}$
$S^2(\delta)$	$(\dots, 0, 0, 0, 0, 1, \dots)$	$\{z_k\}$, where $z_k = \begin{cases} 1 & \text{if } k = 2 \\ 0 & \text{if } k \neq 2 \end{cases}$
$\vdots$	$\uparrow$ $k = 0$	$\vdots$

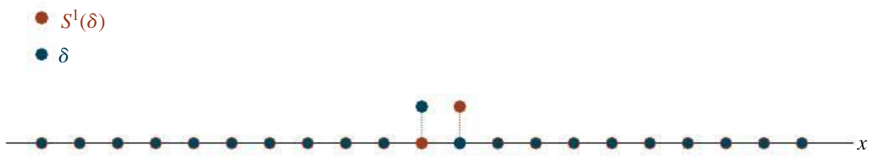


FIGURE 2

Notice that S satisfies the properties of a linear transformation. Specifically, for any scalar c and signals $\{x_k\}$ and $\{y_k\}$, applying S results in $S(\{x_k\} + \{y_k\}) = \{x_{k-1} + y_{k-1}\} = \{x_{k-1}\} + \{y_{k-1}\} = S(\{x_k\}) + S(\{y_k\})$ and $S(c\{x_k\}) = \{cx_{k-1}\} = cS(\{x_k\})$. The mapping S has an additional property. Notice that for any integer q, $S(\{x_{k+q}\}) = \{x_{k-1+q}\}$. One can think of this last property as the *time invariance* property. Transformations with the same properties as S are referred to as linear time invariant (LTI).

DEFINITION

Linear Time Invariant (LTI) Transformations

A transformation $T : \mathbb{S} \to \mathbb{S}$ is **linear time invariant** provided

(i) $T(\{x_k + y_k\}) = T(\{x_k\}) + T(\{y_k\})$ for all signals $\{x_k\}$ and $\{y_k\}$;

(ii) $T(c\{x_k\}) = cT(\{x_k\})$ for all scalars c and all signals $\{x_k\}$;

(iii) If $T(\{x_k\}) = \{y_k\}$, then $T(\{x_{k+q}\}) = \{y_{k+q}\}$ for all integers q and all signals $\{x_k\}$.

The first two properties in the definition of LTI transformations are the same as the two properties listed in the definition of a linear transformation resulting in the following theorem:

THEOREM 16

LTI Transformations are Linear Transformations

A linear time invariant transformation on the signal space $\mathbb{S}$ is a special type of linear transformation.

Digital Signal Processing

LTI transformations, like the shift transformation, can be used to create new signals from signals that are already stored in a system. Another type of LTI transformation is used for *smoothing* or *filtering* data. In Example 11 of Section 4.2, a two-day moving average LTI transformation is used to smooth out stock price fluctuations. In Example 2, this mapping is extended to encompass longer time periods. Smoothing out a signal can make it easier to spot trends in data. Filtering will be discussed in more detail in Section 4.8.

EXAMPLE 2 For any positive integer m, the **moving average** LTI transformation with time period m is given by

$$M_m(\{x_k\}) = \{y_k\} \text{ where } y_k = \frac{1}{m} \sum_{j=k-m+1}^{k} x_k$$

Figure 3 illustrates how M_3 smooths out a signal. Section 4.2, Figure 3 illustrates the smoothing that occurred when M_2 was applied to the same data. As m is increased, applying M_m smooths the signal even more.

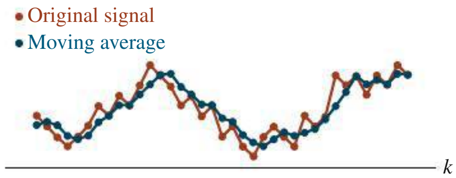


FIGURE 3

The kernel of M_2 is calculated in Example 11 of Section 4.2. It is the span of the alternating sequence α listed in Table 1. The kernel of the LTI transformation describes what is smoothed out of the original signal. Exercises 10, 12, and 14 explore properties of M_3 further. ■

Another type of DSP does the opposite of smoothing or filtering - it combines signals to increase their complexity. **Auralization** is a process used in the entertainment industry to give a more acoustic quality to virtually generated sounds. In Example 3, we illustrate how combining signals enhances the sound generated by the signal $\{\cos(440\pi k)\}$.

EXAMPLE 3 Combining several signals can be used to produce more realistic virtual sounds. In Figure 4, notice that the original cosine wave contains very little variation, whereas by enhancing the equation used, the waves created contain more variation by introducing echos or allowing a sound to fade out. ■

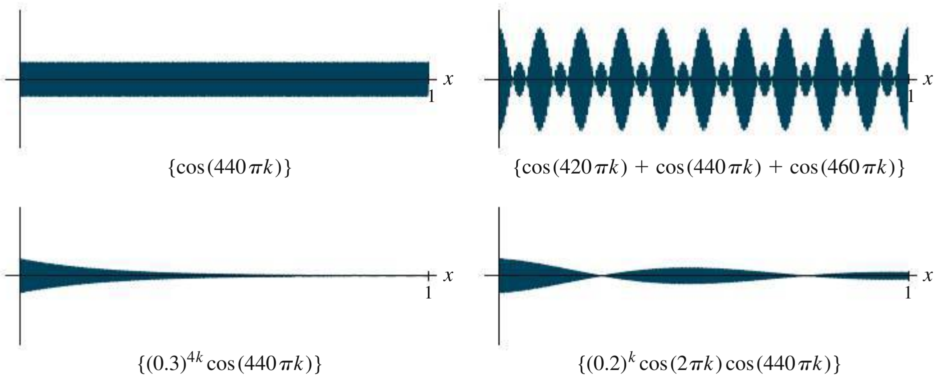


FIGURE 4

Generating Bases for Subspaces of $\mathbb{S}$

If several sets of data are being sampled over the same n time periods, it may be advantageous to view the signals created as part of $\mathbb{S}_n$. The set of **signals of length n**, $\mathbb{S}_n$, is defined to be the set of all signals $\{y_k\}$ such that $y_k = 0$ whenever $k < 0$ or $k > n$. Theorem 17 establishes that $\mathbb{S}_n$ is isomorphic to $\mathbb{R}^{n+1}$. A basis for $\mathbb{S}_n$ can be generated using the shift LTI transformation S from Example 1 and the signal δ as illustrated in Table 2.

THEOREM 17

The set $\mathbb{S}_n$ is a subspace of $\mathbb{S}$ isomorphic to $\mathbb{R}^{n+1}$, and the set of signals $\mathcal{B}_n = \{\delta, S(\delta), S^2(\delta), \ldots, S^n(\delta)\}$ forms a basis for $\mathbb{S}_n$.

PROOF Since the zero signal is in $\mathbb{S}_n$, and adding or scaling signals cannot create nonzeros in the positions that must contain zeros, the set $\mathbb{S}_n$ is a subspace of $\mathbb{S}$. Let $\{y_k\}$ be any signal in $\mathbb{S}_n$. Notice

$$\{y_k\} = \sum_{j=0}^{n} y_j S^j(\delta),$$

so $\mathcal{B}_n$ is a spanning set for $\mathbb{S}_n$. Conversely, if $c_0, \ldots, c_n$ are scalars such that

$$c_0\delta + c_1 S(\delta) + \ldots + c_n S^n(\delta) = \{0\},$$

specifically

$$(\ldots\ 0,\ 0,\ c_0,\ c_1,\ \ldots,\ c_n,\ 0,\ 0,\ \ldots) = (\ldots,\ 0,\ 0,\ 0,\ 0,\ \ldots,\ 0,\ 0,\ 0,\ \ldots),$$

then $c_0 = c_1 = \cdots = c_n = 0$, and thus the vectors in $\mathcal{B}_n$ form a linearly independent set. This establishes that $\mathcal{B}_n$ is a basis for $\mathbb{S}_n$ and hence it is an $n+1$ dimensional vector space isomorphic to $\mathbb{R}^{n+1}$.

Since $\mathbb{S}_n$ has a finite basis, any vector in $\mathbb{S}_n$ can be represented as a vector in $\mathbb{R}^{n+1}$. ■

EXAMPLE 4 Using the basis $\mathcal{B}_2 = \{\delta, S(\delta), S^2(\delta)\}$ for $\mathbb{S}_2$, represent the signal $\{y_k\}$, where

$$y_k = \begin{cases} 0 & \text{if } k < 0 \text{ or } k > 3 \\ 2 & \text{if } k = 0 \\ 3 & \text{if } k = 1 \\ -1 & \text{if } k = 2 \end{cases}$$

as a vector in $\mathbb{R}^3$.

SOLUTION First write $\{y_k\}$ as a linear combination of the basis vectors in $\mathcal{B}_2$.

$$\{y_k\} = 2\delta + 3S(\delta) + (-1)S^2(\delta)$$

The coefficients of this linear combination are precisely the entries in the coordinate vector. Thus $[\{y_k\}]_{\mathcal{B}_2} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$ ■

The set of **finitely supported signals**, $\mathbb{S}_f$, is the set of signals $\{y_k\}$, where only finitely many of the entries are nonzero. In Example 8 of Section 4.1, it is established that $\mathbb{S}_f$ is a subspace of $\mathbb{S}$. The signals created by recording the daily price of a stock increase in length each day, but remain finitely supported, and hence these signals belong to $\mathbb{S}_f$, but not to any particular $\mathbb{S}_n$. Conversely, if a signal is in $\mathbb{S}_n$ for some positive integer n, then it is also in $\mathbb{S}_f$. In Theorem 18, we see that $\mathbb{S}_f$ is an infinite dimensional subspace and so it is not isomorphic to $\mathbb{R}^n$ for any n.

THEOREM 18

The set $\mathcal{B}_f = \{S^j(\delta) : \text{where } j \in \mathbb{Z}\}$ is a basis for the infinite dimensional vector space $\mathbb{S}_f$.

PROOF Let $\{y_k\}$ be any signal in $\mathbb{S}_f$. Since only finitely many entries in $\{y_k\}$ are nonzero, there exist integers p and q such that $y_k = 0$ for all $k < p$ and $k > q$. Thus

$$\{y_k\} = \sum_{j=p}^{q} y_j S^j(\delta),$$

so $\mathcal{B}_f$ is a spanning set for $\mathbb{S}_n$. Moreover, if a linear combination of signals with scalars $c_p, c_{p+1}, \ldots, c_q$ add to zero,

$$\sum_{j=p}^{q} c_j S^j(\delta), = \{0\},$$

then $c_p = c_{p+1} = \cdots = c_q = 0$, and thus the vectors in $\mathcal{B}_f$ form a linearly independent set. This establishes that $\mathcal{B}_f$ is a basis for $\mathbb{S}_f$. Since $\mathcal{B}_f$ contains infinitely many signals, $\mathbb{S}_f$ is an infinite dimensional vector space. ■

The creative power of the shift LTI transformation falls short of being able to create a basis for $\mathbb{S}$ itself. The definition of linear combination requires that only finitely many vectors and scalars are used in a sum. Consider the unit step signal, υ, from Table 1. Although $\upsilon = \sum_{j=0}^{\infty} S^j(\delta)$, this is an infinite sum of vectors and hence not technically considered a *linear combination* of the basis elements from $\mathcal{B}_f$.

In calculus, sums with infinitely many terms are studied in detail. Although it can be shown that every vector space has a basis (using a finite number of terms in each linear combination), the proof relies on the Axiom of Choice and hence establishing that $\mathbb{S}$ has a basis is a topic you may see in higher level math classes. The sinusoidal and exponential signals, which have infinite support, are explored in detail in Section 4.8

Practice Problems

1. Find $\upsilon + \chi$ from Table 1. Express the answer as a vector and give its formal description.
2. Show that $T(\{x_k\}) = \{3x_k - 2x_{k-1}\}$ is a linear time invariant transformation.
3. Find a nonzero vector in the kernel of T for the linear time invariant transformation given in Practice Problem 2.

4.7 Exercises

For Exercises 1–4, find the indicated sums of the signals in Table 1.

1. $\chi + \alpha$
2. $\chi - \alpha$
3. $\upsilon + 2\alpha$
4. $\upsilon - 3\alpha$

For Exercises 5–8, recall that $I(\{x_k\}) = \{x_k\}$ and $S(\{x_k\}) = \{x_{k-1}\}$.

5. Which signals from Table 1 are in the kernel of $I + S$?
6. Which signals from Table 1 are in the kernel of $I - S$?

7. Which signals from Table 1 are in the kernel of $I - cS$ for a fixed nonzero scalar $c \neq 1$?

8. Which signals from Table 1 are in the kernel of $I - S - S^2$?

9. Show that $T(\{x_k\}) = \{x_k - x_{k-1}\}$ is a linear time invariant transformation.

10. Show that $M_3(\{x_k\}) = \left\{\frac{1}{3}(x_{k-2} + x_{k-1} + x_k)\right\}$ is a linear time invariant transformation.

11. Find a nonzero signal in the kernel of T from Exercise 9.

12. Find a nonzero signal in the kernel of M_3 from Exercise 10.

13. Find a nonzero signal in the range of T from Exercise 9.

14. Find a nonzero signal in the range of M_3 from Exercise 10.

In Exercises 15–22, V is a vector space and A is an $m \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

15. **(T/F)** The set of signals of length n, $\mathbb{S}_n$, has a basis with $n+1$ signals.

16. **(T/F)** The set of signals, $\mathbb{S}$, has a finite basis.

17. **(T/F)** Every subspace of the set of signals $\mathbb{S}$ is infinite dimensional.

18. **(T/F)** The vector space $\mathbb{R}^{n+1}$ is a subspace of $\mathbb{S}$.

19. **(T/F)** Every linear time invariant transformation is a linear transformation.

20. **(T/F)** The moving average function is a linear time invariant transformation.

21. **(T/F)** If you scale a signal by a fixed constant, the result is not a signal.

22. **(T/F)** If you scale a linear time invariant transformation by a fixed constant, the result is no longer a linear transformation.

Guess and check or working backwards through the solution to Practice Problem 3 are two good ways to find solutions to Exercises 23 and 24.

23. Construct a linear time invariant transformation that has the signal $\{x_k\} = \left\{\left(\frac{3}{4}\right)^k\right\}$ in its kernel.

24. Construct a linear time invariant transformation that has the signal $\{x_k\} = \left\{\left(\frac{-2}{3}\right)^k\right\}$ in its kernel.

25. Let $W = \left\{\{x_k\} \mid x_k = \begin{cases} 0 & \text{if } k \text{ is a multiple of } 2 \\ r & \text{if } k \text{ is not a multiple of } 2 \end{cases} \text{where } r \text{ can be any real number}\right\}$. A typical signal in W looks like

$$(\ldots,\ r,\ 0,\ r,\ \underset{\underset{k=0}{\uparrow}}{0},\ r,\ 0,\ r,\ \ldots)$$

Show that W is a subspace of $\mathbb{S}$.

26. Let $W = \left\{\{x_k\} \mid x_k = \begin{cases} 0 & \text{if } k < 0 \\ r & \text{if } k \geq 0 \end{cases} \text{where } r \text{ can be any real number.}\right\}$ A typical signal in W looks like

$$(\ldots,\ 0,\ 0,\ 0,\ \underset{\underset{k=0}{\uparrow}}{r},\ r,\ r,\ r,\ \ldots)$$

Show that W is a subspace of $\mathbb{S}$.

27. Find a basis for the subspace W in Exercise 25. What is the dimension of this subspace?

28. Find a basis for the subspace W in Exercise 26. What is the dimension of this subspace?

29. Let $W = \left\{\{x_k\} \mid x_k = \begin{cases} 0 & \text{if } k \text{ is a multiple of } 2 \\ r_k & \text{if } k \text{ is not a multiple of } 2 \end{cases} \text{where each } r_k \text{ can be any real number}\right\}$. A typical signal in W looks like

$$(\ldots,\ r_{-3},\ 0,\ r_{-1},\ \underset{\underset{k=0}{\uparrow}}{0},\ r_1,\ 0,\ r_3,\ \ldots)$$

Show that W is a subspace of $\mathbb{S}$.

30. Let $W = \left\{\{x_k\} \mid x_k = \begin{cases} 0 & \text{if } k < 0 \\ r_k & \text{if } k \geq 0 \end{cases} \text{where each } r_k \text{ can be any real number}\right\}$. A typical signal in W looks like

$$(\ldots,\ 0,\ 0,\ 0,\ \underset{\underset{k=0}{\uparrow}}{r_0},\ r_1,\ r_2,\ r_3,\ \ldots)$$

Show that W is a subspace of $\mathbb{S}$.

31. Describe an infinite linearly independent subset of the subspace W in Exercise 29. Does this establish that W is infinite dimensional? Justify your answer.

32. Describe an infinite linearly independent subset of the subspace W in Exercise 30. Does this establish that W is infinite dimensional? Justify your.

Solutions to Practice Problems

1. First add $\upsilon + \chi$ in vector form:

$$\begin{array}{rl} & (\ldots, 0, 0, 0, 1, 1, 1, 1, \ldots) \\ + & (\ldots, 1, 1, 1, 1, 1, 1, 1, \ldots) \\ = & (\ldots, 1, 1, 1, 2, 2, 2, 2, \ldots) \\ & \qquad\qquad\quad \uparrow \\ & \qquad\qquad\; k = 0 \end{array}$$

Then add the terms in the formal description to get a new formal description:

$$\upsilon + \chi = \{z_k\}, \text{where } z_k = u_k + c_k = \begin{cases} 1+1 & \text{if } k \geq 0 \\ 0+1 & \text{if } k < 0 \end{cases} = \begin{cases} 2 & \text{if } k \geq 0 \\ 1 & \text{if } k < 0 \end{cases}$$

2. Verify that the three conditions for a linear time invariant transformation hold. Specifically, for any two signals $\{x_k\}$ and $\{y_k\}$, and scalar c, observe that

a. $T(\{x_k + y_k\}) = \{3(x_k + y_k) - 2(x_{k-1} + y_{k-1})\} = \{3x_k - 2x_{k-1}\} + \{3y_k - 2y_{k-1}\} = T(\{x_k\}) + T(\{y_k\})$

b. $T(c\{x_k\}) = \{3cx_k - 2cx_{k-1}\} = c\{3x_k - 2x_{k-1}\} = cT(\{x_k\})$

c. $T(\{x_k\}) = \{3x_k - 2x_{k-1}\}$ and $T(\{x_{k+q}\}) = \{3x_{k+q} - 2x_{k+q-1}\} = \{3x_{k+q} - 2x_{k-1+q}\}$ for all integers q.

Thus T is a linear time invariant transformation.

3. To find a vector in the kernel of T, set $T(\{x_k\}) = \{3x_k - 2x_{k-1}\} = \{0\}$. Then for each k, notice $3x_k - 2x_{k-1} = 0$ and hence $x_k = \frac{2}{3}x_{k-1}$. Picking a nonzero value for x_0, say $x_0 = 1$, then $x_1 = \frac{2}{3}$, $x_2 = \left(\frac{2}{3}\right)^2$, and in general, $x_k = \left(\frac{2}{3}\right)^k$. To verify that this signal is indeed in the kernel of T observe that $T\left(\left\{\left(\frac{2}{3}\right)^k\right\}\right) = \left\{3\left(\frac{2}{3}\right)^k - 2\left(\frac{2}{3}\right)^{k-1}\right\} = \left\{\left(\frac{2}{3}\right)^{k-1}\left(3\left(\frac{2}{3}\right) - 2\right)\right\} = \{0\}$. Notice that $\left\{\left(\frac{2}{3}\right)^k\right\}$ is the exponential signal with $c = \frac{2}{3}$.

4.8 Applications to Difference Equations

Continuing our study of discrete-time signals, in this section we explore difference equations, a valuable tool used to filter the data contained in signals. Even when a differential equation is used to model a continuous process, a numerical solution is often produced from a related difference equation. This section highlights some fundamental properties of linear difference equations that are explained using linear algebra.

Linear Independence in the Space $\mathbb{S}$ of Signals

To simplify notation, we consider a set of only three signals in $\mathbb{S}$, say, $\{u_k\}$, $\{v_k\}$, and $\{w_k\}$. They are linearly independent precisely when the equation

$$c_1u_k + c_2v_k + c_3w_k = 0 \quad \text{for all } k \tag{1}$$

implies that $c_1 = c_2 = c_3 = 0$. The phrase "for all k" means for all integers—positive, negative, and zero. One could also consider signals that start with $k = 0$, for example, in which case, "for all k" would mean for all integers $k \geq 0$.

Suppose c_1, c_2, c_3 satisfy (1). Then equation (1) holds for any three consecutive values of k, say, k, $k+1$, and $k+2$. Thus (1) implies that

$$c_1 u_{k+1} + c_2 v_{k+1} + c_3 w_{k+1} = 0 \quad \text{for all } k$$

and

$$c_1 u_{k+2} + c_2 v_{k+2} + c_3 w_{k+2} = 0 \quad \text{for all } k$$

Hence c_1, c_2, c_3 satisfy

$$\begin{bmatrix} u_k & v_k & w_k \\ u_{k+1} & v_{k+1} & w_{k+1} \\ u_{k+2} & v_{k+2} & w_{k+2} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{for all } k \tag{2}$$

The coefficient matrix in this system is called the **Casorati matrix**, $C(k)$, of the signals, and the determinant of the matrix is called the **Casoratian** of $\{u_k\}$, $\{v_k\}$, and $\{w_k\}$. If the Casorati matrix is invertible for at least one value of k, then (2) will imply that $c_1 = c_2 = c_3 = 0$, which will prove that the three signals are linearly independent.

STUDY GUIDE offers additional resources on the Casorati Test.

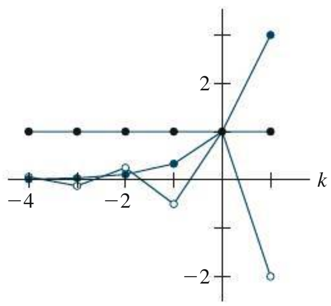
EXAMPLE 1 Verify that $\{1^k\}$, $\{(-2)^k\}$, and $\{3^k\}$ are linearly independent signals.

SOLUTION The Casorati matrix is

$$\begin{bmatrix} 1^k & (-2)^k & 3^k \\ 1^{k+1} & (-2)^{k+1} & 3^{k+1} \\ 1^{k+2} & (-2)^{k+2} & 3^{k+2} \end{bmatrix}$$

Row operations can show fairly easily that this matrix is always invertible. However, it is faster to substitute a value for k—say, $k = 0$—and row reduce the numerical matrix:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 3 \\ 1 & 4 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & -3 & 2 \\ 0 & 3 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & -3 & 2 \\ 0 & 0 & 10 \end{bmatrix}$$



The signals 1^k, $(-2)^k$, and 3^k.

The Casorati matrix is invertible for $k = 0$. So $\{1^k\}$, $\{(-2)^k\}$, and $\{3^k\}$ are linearly independent. ■

If a Casorati matrix is not invertible, the associated signals being tested may or may not be linearly dependent. (See Exercise 35.) However, it can be shown that if the signals are all solutions of the *same* homogeneous difference equation (described below), then either the Casorati matrix is invertible for all k and the signals are linearly independent, or else the Casorati matrix is not invertible for all k and the signals are linearly dependent. A nice proof using linear transformations is in the *Study Guide*.

Linear Difference Equations

Given scalars $a_0, \ldots, a_n$, with a_0 and a_n nonzero, and given a signal $\{z_k\}$, the equation

$$a_0 y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = z_k \quad \text{for all } k \tag{3}$$

is called a **linear difference equation** (or **linear recurrence relation**) **of order *n***. For simplicity, a_0 is often taken equal to 1. If $\{z_k\}$ is the zero sequence, the equation is **homogeneous**; otherwise, the equation is **nonhomogeneous**.

In digital signal processing (DSP), a difference equation such as (3) describes a **linear time invariant (LTI) filter**, and $a_0, \ldots, a_n$ are called the **filter coefficients**. The shift LTI transformations $S(\{y_k\}) = \{y_{k-1}\}$ and $S^{-1}(\{y_k\}) = \{y_{k+1}\}$ were introduced in Example 1 of Section 4.7 and are used here to describe the LTI filter associated with a linear difference equation. Define

$$T = a_0 S^{-n} + a_1 S^{-n+1} + \cdots + a_{n-1} S^{-1} + a_n S^0.$$

Notice if $\{z_k\} = T(\{y_k\})$, then for any k, Equation (3) describes the relationship between terms in the two signals.

EXAMPLE 2 Let us feed two different signals into the filter

$$.35y_{k+2} + .5y_{k+1} + .35y_k = z_k$$

Here .35 is an abbreviation for $\sqrt{2}/4$. The first signal is created by sampling the continuous signal $y = \cos(\pi t/4)$ at integer values of t, as in Figure 1(a). The discrete signal is

$$\{y_k\} = (\ldots, \cos(0), \cos(\pi/4), \cos(2\pi/4), \cos(3\pi/4), \ldots)$$

For simplicity, write $\pm.7$ in place of $\pm\sqrt{2}/2$, so that

$$\{y_k\} = (\ldots, \underset{\substack{\uparrow \\ k=0}}{1}, .7, 0, -.7, -1, -.7, 0, .7, 1, .7, 0, \ldots)$$

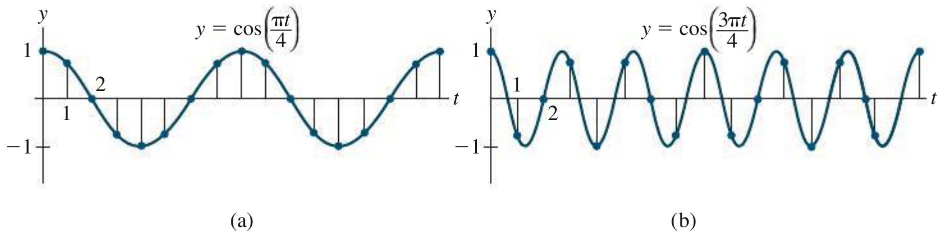


FIGURE 1 Discrete signals with different frequencies.

Table 1 shows a calculation of the output sequence $\{z_k\}$, where .35(.7) is an abbreviation for $(\sqrt{2}/4)(\sqrt{2}/2) = .25$. The output is $\{y_k\}$, shifted by one term.

TABLE I Computing the Output of a Filter

k	y_k	y_{k+1}	y_{k+2}	$.35y_k$	$+ .5y_{k+1}$	$+ .35y_{k+2}$	$=$	z_k
0	1	.7	0	.35(1)	+ .5(.7)	+ .35(0)	=	.7
1	.7	0	−.7	.35(.7)	+ .5(0)	+ .35(−.7)	=	0
2	0	−.7	−1	.35(0)	+ .5(−.7)	+ .35(−1)	=	−.7
3	−.7	−1	−.7	.35(−.7)	+ .5(−1)	+ .35(−.7)	=	−1
4	−1	−.7	0	.35(−1)	+ .5(−.7)	+ .35(0)	=	−.7
5	−.7	0	.7	.35(−.7)	+ .5(0)	+ .35(.7)	=	0
⋮	⋮							⋮

A different input signal is produced from the higher frequency signal $y = \cos(3\pi t/4)$, shown in Figure 1(b). Sampling at the same rate as before produces a new input sequence:

$$\{w_k\} = (\ldots, \underset{\substack{\uparrow \\ k=0}}{1}, -.7, 0, .7, -1, .7, 0, -.7, 1, -.7, 0, \ldots)$$

When $\{w_k\}$ is fed into the filter, the output is the zero sequence. The filter, called a *low-pass filter*, lets $\{y_k\}$ pass through, but stops the higher frequency $\{w_k\}$. ■

In many applications, a sequence $\{z_k\}$ is specified for the right side of a difference equation (3), and a $\{y_k\}$ that satisfies (3) is called a **solution** of the equation. The next example shows how to find solutions for a homogeneous equation.

EXAMPLE 3 Solutions of a homogeneous difference equation often have the form $\{y_k\} = \{r^k\}$ for some r. Find some solutions of the equation

$$y_{k+3} - 2y_{k+2} - 5y_{k+1} + 6y_k = 0 \quad \text{for all } k \tag{4}$$

SOLUTION Substitute r^k for y_k in the equation and factor the left side:

$$r^{k+3} - 2r^{k+2} - 5r^{k+1} + 6r^k = 0 \tag{5}$$

$$r^k(r^3 - 2r^2 - 5r + 6) = 0$$

$$r^k(r-1)(r+2)(r-3) = 0 \tag{6}$$

Since (5) is equivalent to (6), $\{r^k\}$ satisfies the difference equation (4) if and only if r satisfies (6). Thus $\{1^k\}$, $\{(-2)^k\}$, and $\{3^k\}$ are all solutions of (4). For instance, to verify that $\{3^k\}$ is a solution of (4), compute

$$3^{k+3} - 2 \cdot 3^{k+2} - 5 \cdot 3^{k+1} + 6 \cdot 3^k$$
$$= 3^k(27 - 18 - 15 + 6) = 0 \quad \text{for all } k$$ ■

In general, a nonzero signal $\{r^k\}$ satisfies the homogeneous difference equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0 \quad \text{for all } k$$

if and only if r is a root of the **auxiliary equation**

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0$$

We will not consider the case in which r is a repeated root of the auxiliary equation. When the auxiliary equation has a *complex root*, the difference equation has solutions of the form $\{s^k \cos k\omega\}$ and $\{s^k \sin k\omega\}$, for constants s and ω. This happened in Example 2.

Solution Sets of Linear Difference Equations

Given $a_1, \dots, a_n$, recall that the LTI transformation $T : \mathbb{S} \to \mathbb{S}$ given by

$$T = a_0 S^{-n} + a_1 S^{-n+1} + \cdots + a_{n-1} S^{-1} + a_n S^0$$

transforms a signal $\{y_k\}$ into the signal $\{w_k\}$ given by

$$w_k = y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k \text{ for all } k$$

This implies that the solution set of the homogeneous equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0 \text{ for all } k$$

is the kernel of T and describes the signals that are *filtered out* or transformed into the zero signal. Since the kernel of any linear transformation with domain $\mathbb{S}$ is a *subspace* of $\mathbb{S}$, so is the solution set of a homogeneous equation. Any linear combination of solutions is again a solution.

The next theorem, a simple but basic result, will lead to more information about the solution sets of difference equations.

THEOREM 19

If $a_n \neq 0$ and if $\{z_k\}$ is given, the equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = z_k \quad \text{for all } k \tag{7}$$

has a unique solution whenever $y_0, \ldots, y_{n-1}$ are specified.

PROOF If $y_0, \ldots, y_{n-1}$ are specified, use (7) to *define*

$$y_n = z_0 - [\, a_1 y_{n-1} + \cdots + a_{n-1} y_1 + a_n y_0 \,]$$

And now that $y_1, \ldots, y_n$ are specified, use (7) to define y_{n+1}. In general, use the recurrence relation

$$y_{n+k} = z_k - [\, a_1 y_{k+n-1} + \cdots + a_n y_k \,] \tag{8}$$

to define y_{n+k} for $k \geq 0$. To define y_k for $k < 0$, use the recurrence relation

$$y_k = \frac{1}{a_n} z_k - \frac{1}{a_n} [\, y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} \,] \tag{9}$$

This produces a signal that satisfies (7). Conversely, any signal that satisfies (7) for all k certainly satisfies (8) and (9), so the solution of (7) is unique. ■

THEOREM 20

The set H of all solutions of the nth-order homogeneous linear difference equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0 \quad \text{for all } k \tag{10}$$

is an n-dimensional vector space.

PROOF As was pointed out earlier, H is a subspace of $\mathbb{S}$ because H is the kernel of a linear transformation. For $\{y_k\}$ in H, let $F\{y_k\}$ be the vector in $\mathbb{R}^n$ given by $(y_0, y_1, \ldots, y_{n-1})$. It is readily verified that $F : H \to \mathbb{R}^n$ is a linear transformation. Given any vector $(y_0, y_1, \ldots, y_{n-1})$ in $\mathbb{R}^n$, Theorem 19 says that there is a unique signal $\{y_k\}$ in H such that $F\{y_k\} = (y_0, y_1, \ldots, y_{n-1})$. This means that F is a one-to-one linear transformation of H onto $\mathbb{R}^n$; that is, F is an isomorphism. Thus $\dim H = \dim \mathbb{R}^n = n$. (See Exercise 52 in Section 4.5.) ■

EXAMPLE 4 Find a basis for the set of all solutions to the difference equation

$$y_{k+3} - 2y_{k+2} - 5y_{k+1} + 6y_k = 0 \quad \text{for all } k$$

SOLUTION Our work in linear algebra really pays off now! We know from Examples 1 and 3 that $\{1^k\}$, $\{(-2)^k\}$, and $\{3^k\}$ are linearly independent solutions. In general, it can be difficult to verify directly that a set of signals *spans* the solution space. But that is no problem here because of two key theorems—Theorem 20, which shows that the solution space is exactly three-dimensional, and the Basis Theorem in Section 4.5, which says that a linearly independent set of n vectors in an n-dimensional space is automatically a basis. So $\{1^k\}$, $\{(-2)^k\}$, and $\{3^k\}$ form a basis for the solution space. ■

The standard way to describe the "general solution" of the difference equation (10) is to exhibit a basis for the subspace of all solutions. Such a basis is usually called a

fundamental set of solutions of (10). In practice, if you can find n linearly independent signals that satisfy (10), they will automatically span the n-dimensional solution space, as explained in Example 4.

Nonhomogeneous Equations

The general solution of the nonhomogeneous difference equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = z_k \quad \text{for all } k \tag{11}$$

can be written as one particular solution of (11) plus an arbitrary linear combination of a fundamental set of solutions of the corresponding homogeneous equation (10). This fact is analogous to the result in Section 1.5 showing that the solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$ are parallel. Both results have the same explanation: The mapping $\mathbf{x} \mapsto A\mathbf{x}$ is linear, and the mapping that transforms the signal $\{y_k\}$ into the signal $\{z_k\}$ in (11) is linear.

EXAMPLE 5 Verify that the signal $\{y_k\} = \{k^2\}$ satisfies the difference equation

$$y_{k+2} - 4y_{k+1} + 3y_k = -4k \quad \text{for all } k \tag{12}$$

Then find a description of all solutions of this equation.

SOLUTION Substitute k^2 for y_k on the left side of (12):

$$\begin{aligned}(k+2)^2 - 4(k+1)^2 + 3k^2 \\ &= (k^2 + 4k + 4) - 4(k^2 + 2k + 1) + 3k^2 \\ &= -4k\end{aligned}$$

So k^2 is indeed a solution of (12). The next step is to solve the homogeneous equation

$$y_{k+2} - 4y_{k+1} + 3y_k = 0 \quad \text{for all } k \tag{13}$$

The auxiliary equation is

$$r^2 - 4r + 3 = (r-1)(r-3) = 0$$

The roots are $r = 1, 3$. So two solutions of the homogeneous difference equation are $\{1^k\}$ and $\{3^k\}$. They are obviously not multiples of each other, so they are linearly independent signals. By Theorem 20, the solution space is two-dimensional, so $\{1^k\}$ and $\{3^k\}$ form a basis for the set of solutions of equation (13). Translating that set by a particular solution of the nonhomogeneous equation (12), we obtain the general solution of (12):

$$\{k^2\} + c_1\{1^k\} + c_2\{3^k\}, \quad \text{or} \quad \{k^2 + c_1 + c_2 3^k\}$$

Figure 2 gives a geometric visualization of the two solution sets. Each point in the figure corresponds to one signal in $\mathbb{S}$. ■

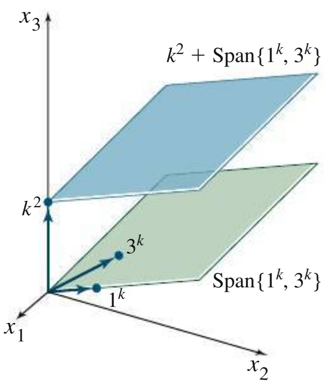


FIGURE 2
Solution sets of difference equations (12) and (13).

Reduction to Systems of First-Order Equations

A modern way to study a homogeneous nth-order linear difference equation is to replace it by an equivalent system of first-order difference equations, written in the form

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{for all } k$$

where the vectors $\mathbf{x}_k$ are in $\mathbb{R}^n$ and A is an $n \times n$ matrix.

A simple example of such a (vector-valued) difference equation was already studied in Section 1.10. Further examples will be covered in Sections 5.6 and 5.9.

EXAMPLE 6 Write the following difference equation as a first-order system:

$$y_{k+3} - 2y_{k+2} - 5y_{k+1} + 6y_k = 0 \quad \text{for all } k$$

SOLUTION For each k, set

$$\mathbf{x}_k = \begin{bmatrix} y_k \\ y_{k+1} \\ y_{k+2} \end{bmatrix}$$

The difference equation says that $y_{k+3} = -6y_k + 5y_{k+1} + 2y_{k+2}$, so

$$\mathbf{x}_{k+1} = \begin{bmatrix} y_{k+1} \\ y_{k+2} \\ y_{k+3} \end{bmatrix} = \begin{bmatrix} 0 & + & y_{k+1} & + 0 \\ 0 & + 0 & + & y_{k+2} \\ -6y_k & + 5y_{k+1} & + 2y_{k+2} & \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & 5 & 2 \end{bmatrix} \begin{bmatrix} y_k \\ y_{k+1} \\ y_{k+2} \end{bmatrix}$$

That is,

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{for all } k, \quad \text{where} \quad A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & 5 & 2 \end{bmatrix}$$

In general, the equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0 \quad \text{for all } k$$

can be rewritten as $\mathbf{x}_{k+1} = A\mathbf{x}_k$ for all k, where

$$\mathbf{x}_k = \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+n-1} \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$

Practice Problem

It can be shown that the signals 2^k, $3^k \sin \frac{k\pi}{2}$, and $3^k \cos \frac{k\pi}{2}$ are solutions of

$$y_{k+3} - 2y_{k+2} + 9y_{k+1} - 18y_k = 0$$

Show that these signals form a basis for the set of all solutions of the difference equation.

4.8 Exercises

Verify that the signals in Exercises 1 and 2 are solutions of the accompanying difference equation.

1. $2^k, (-4)^k$; $y_{k+2} + 2y_{k+1} - 8y_k = 0$

2. $3^k, (-3)^k$; $y_{k+2} - 9y_k = 0$

Show that the signals in Exercises 3–6 form a basis for the solution set of the accompanying difference equation.

3. The signals and equation in Exercise 1

4. The signals and equation in Exercise 2

5. $(-3)^k$, $k(-3)^k$; $y_{k+2} + 6y_{k+1} + 9y_k = 0$

6. $5^k \cos \frac{k\pi}{2}, 5^k \sin \frac{k\pi}{2}$; $y_{k+2} + 25y_k = 0$

In Exercises 7–12, assume the signals listed are solutions of the given difference equation. Determine if the signals form a basis for the solution space of the equation. Justify your answers using appropriate theorems.

7. $1^k, 2^k, (-2)^k$; $y_{k+3} - y_{k+2} - 4y_{k+1} + 4y_k = 0$

8. $2^k, 4^k, (-5)^k$; $y_{k+3} - y_{k+2} - 22y_{k+1} + 40y_k = 0$

9. $1^k, 3^k \cos \frac{k\pi}{2}, 3^k \sin \frac{k\pi}{2}$; $y_{k+3} - y_{k+2} + 9y_{k+1} - 9y_k = 0$

10. $(-1)^k, k(-1)^k, 5^k$; $y_{k+3} - 3y_{k+2} - 9y_{k+1} - 5y_k = 0$

11. $(-1)^k, 3^k$; $y_{k+3} + y_{k+2} - 9y_{k+1} - 9y_k = 0$

12. $1^k, (-1)^k$; $y_{k+4} - 2y_{k+2} + y_k = 0$

In Exercises 13–16, find a basis for the solution space of the difference equation. Prove that the solutions you find span the solution set.

13. $y_{k+2} - y_{k+1} + \frac{2}{9}y_k = 0$ **14.** $y_{k+2} - 7y_{k+1} + 12y_k = 0$

15. $y_{k+2} - 25y_k = 0$ **16.** $16y_{k+2} + 8y_{k+1} - 3y_k = 0$

17. The Fibonacci Sequence is listed in Table 1 of Section 4.7. It can be viewed as the sequence of numbers where each number is the sum of the two numbers before it. It can be described as the homogeneous difference equation

$$y_{k+2} - y_{k+1} - y_k = 0$$

with the initial conditions $y_0 = 0$ and $y_1 = 1$. Find the general solution of the Fibonacci sequence.

18. If the initial conditions are changed to $y_0 = 1$ and $y_1 = 2$ for the Fibonacci sequence in Exercise 17, list the terms of the sequence for $k = 2, 3, 4$ and 5. Find the solution to the difference equation from 17 with these new initial conditions.

Exercises 19 and 20 concern a simple model of the national economy described by the difference equation

$$Y_{k+2} - a(1+b)Y_{k+1} + abY_k = 1 \qquad (14)$$

Here Y_k is the total national income during year k, a is a constant less than 1, called the *marginal propensity to consume*, and b is a positive *constant of adjustment* that describes how changes in consumer spending affect the annual rate of private investment.[1]

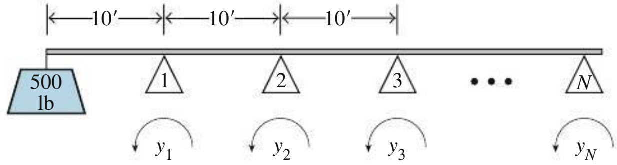
19. Find the general solution of equation (14) when $a = .9$ and $b = \frac{4}{9}$. What happens to Y_k as k increases? [*Hint:* First find a particular solution of the form $Y_k = T$, where T is a constant, called the equilibrium level of national income.]

20. Find the general solution of equation (14) when $a = .9$ and $b = .5$.

A lightweight cantilevered beam is supported at N points spaced 10 ft apart, and a weight of 500 lb is placed at the end of the beam, 10 ft from the first support, as in the figure. Let y_k be the bending moment at the kth support. Then $y_1 = 5000$ ft-lb. Suppose the beam is rigidly attached at the Nth support and the bending moment there is zero. In between, the moments satisfy the *three-moment equation*

$$y_{k+2} + 4y_{k+1} + y_k = 0 \quad \text{for } k = 1, 2, \ldots, N-2 \qquad (15)$$

[1] For example, see *Discrete Dynamical Systems*, by James T. Sandefur (Oxford: Clarendon Press, 1990), pp. 267–276. The original *accelerator-multiplier model* is attributed to the economist P. A. Samuelson.



Bending moments on a cantilevered beam.

21. Find the general solution of difference equation (15). Justify your answer.

22. Find the particular solution of (15) that satisfies the *boundary conditions* $y_1 = 5000$ and $y_N = 0$. (The answer involves N.)

23. When a signal is produced from a sequence of measurements made on a process (a chemical reaction, a flow of heat through a tube, a moving robot arm, etc.), the signal usually contains random *noise* produced by measurement errors. A standard method of preprocessing the data to reduce the noise is to smooth or filter the data. One simple filter is a *moving average* that replaces each y_k by its average with the two adjacent values:

$$\tfrac{1}{3}y_{k+1} + \tfrac{1}{3}y_k + \tfrac{1}{3}y_{k-1} = z_k \quad \text{for } k = 1, 2, \ldots$$

Suppose a signal y_k, for $k = 0, \ldots, 14$, is

9, 5, 7, 3, 2, 4, 6, 5, 7, 6, 8, 10, 9, 5, 7

Use the filter to compute $z_1, \ldots, z_{13}$. Make a broken-line graph that superimposes the original signal and the smoothed signal.

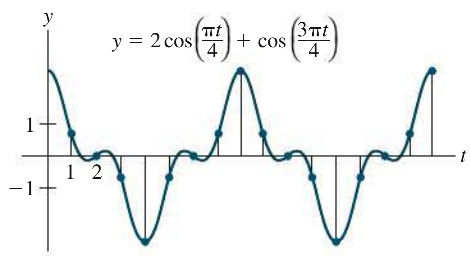
24. Let $\{y_k\}$ be the sequence produced by sampling the continuous signal $2\cos\frac{\pi t}{4} + \cos\frac{3\pi t}{4}$ at $t = 0, 1, 2, \ldots$, as shown in the figure. The values of y_k, beginning with $k = 0$, are

3, .7, 0, −.7, −3, −.7, 0, .7, 3, .7, 0, ...

where .7 is an abbreviation for $\sqrt{2}/2$.

a. Compute the output signal $\{z_k\}$ when $\{y_k\}$ is fed into the filter in Example 2.

b. Explain how and why the output in part (a) is related to the calculations in Example 2.



Sampled data from $2\cos\frac{\pi t}{4} + \cos\frac{3\pi t}{4}$.

Exercises 25 and 26 refer to a difference equation of the form $y_{k+1} - ay_k = b$, for suitable constants a and b.

25. A loan of \$10,000 has an interest rate of 1% per month and a monthly payment of \$450. The loan is made at month $k = 0$, and the first payment is made one month later, at $k = 1$. For

$k = 0, 1, 2, \ldots$, let y_k be the unpaid balance of the loan just after the kth monthly payment. Thus

$$\underset{\text{New balance}}{y_1} = \underset{\text{Balance due}}{10{,}000} + \underset{\text{Interest added}}{(.01)10{,}000} - \underset{\text{Payment}}{450}$$

a. Write a difference equation satisfied by $\{y_k\}$.

T b. Create a table showing k and the balance y_k at month k. List the program or the keystrokes you used to create the table.

T c. What will k be when the last payment is made? How much will the last payment be? How much money did the borrower pay in total?

26. At time $k = 0$, an initial investment of \$1000 is made into a savings account that pays 6% interest per year compounded monthly. (The interest rate per month is .005.) Each month after the initial investment, an additional \$200 is added to the account. For $k = 0, 1, 2, \ldots$, let y_k be the amount in the account at time k, just after a deposit has been made.

a. Write a difference equation satisfied by $\{y_k\}$.

T b. Create a table showing k and the total amount in the savings account at month k, for $k = 0$ through 60. List your program or the keystrokes you used to create the table.

T c. How much will be in the account after two years (that is, 24 months), four years, and five years? How much of the five-year total is interest?

In Exercises 27–30, show that the given signal is a solution of the difference equation. Then find the general solution of that difference equation.

27. $y_k = k^2$; $y_{k+2} + 3y_{k+1} - 4y_k = 7 + 10k$

28. $y_k = 1 + k$; $y_{k+2} - 8y_{k+1} + 15y_k = 2 + 8k$

29. $y_k = 2 - 2k$; $y_{k+2} - \frac{9}{2}y_{k+1} + 2y_k = 2 + 3k$

30. $y_k = 2k - 4$; $y_{k+2} + \frac{3}{2}y_{k+1} - y_k = 1 + 3k$

Write the difference equations in Exercises 31 and 32 as first-order systems, $\mathbf{x}_{k+1} = A\mathbf{x}_k$, for all k.

31. $y_{k+4} - 6y_{k+3} + 8y_{k+2} + 6y_{k+1} - 9y_k = 0$

32. $y_{k+3} - \frac{3}{4}y_{k+2} + \frac{1}{16}y_k = 0$

33. Is the following difference equation of order 3? Explain.

$$y_{k+3} + 5y_{k+2} + 6y_{k+1} = 0$$

34. What is the order of the following difference equation? Explain your answer.

$$y_{k+3} + a_1 y_{k+2} + a_2 y_{k+1} + a_3 y_k = 0$$

35. Let $y_k = k^2$ and $z_k = 2k|k|$. Are the signals $\{y_k\}$ and $\{z_k\}$ linearly independent? Evaluate the associated Casorati matrix $C(k)$ for $k = 0$, $k = -1$, and $k = -2$, and discuss your results.

36. Let f, g, and h be linearly independent functions defined for all real numbers, and construct three signals by sampling the values of the functions at the integers:

$$u_k = f(k), \qquad v_k = g(k), \qquad w_k = h(k)$$

Must the signals be linearly independent in $\mathbb{S}$? Discuss.

Solution to Practice Problem

Examine the Casorati matrix:

$$C(k) = \begin{bmatrix} 2^k & 3^k \sin\frac{k\pi}{2} & 3^k \cos\frac{k\pi}{2} \\ 2^{k+1} & 3^{k+1} \sin\frac{(k+1)\pi}{2} & 3^{k+1} \cos\frac{(k+1)\pi}{2} \\ 2^{k+2} & 3^{k+2} \sin\frac{(k+2)\pi}{2} & 3^{k+2} \cos\frac{(k+2)\pi}{2} \end{bmatrix}$$

Set $k = 0$ and row reduce the matrix to verify that it has three pivot positions and hence is invertible:

$$C(0) = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 0 \\ 4 & 0 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & -2 \\ 0 & 0 & -13 \end{bmatrix}$$

The Casorati matrix is invertible at $k = 0$, so the signals are linearly independent. Since there are three signals, and the solution space H of the difference equation has dimension 3 (Theorem 20), the signals form a basis for H, by the Basis Theorem.

CHAPTER 4 PROJECTS

Chapter 4 projects are available online at bit.ly/30IM8gT.

A. *Exploring Subspaces*: This project explores subspaces with a more hands-on approach.

B. *Hill Substitution Ciphers*: This project shows how to use matrices to encode and decode messages.

C. *Error Detecting and Error Correcting*: In this project, a method detecting and correcting errors made in the transmission of encoded messages is constructed. It will turn out that abstract vector spaces and the concepts of null space, rank, and dimension are needed for this construction.

D. *Signal Processing*: This project examines signal processing in more detail.

E. *Fibonacci Sequences*: The purpose of this project is to investigate further the Fibonacci sequence, which arises in number theory, applied mathematics, and biology.

CHAPTER 4 SUPPLEMENTARY EXERCISES

In Exercises 1–19, mark each statement True or False **(T/F)**. Justify each answer. (If true, cite appropriate facts or theorems. If false, explain why or give a counterexample that shows why the statement is not true in every case.) In Exercises 1–6, $\mathbf{v}_1, \dots, \mathbf{v}_p$ are vectors in a nonzero finite-dimensional vector space V, and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

1. **(T/F)** The set of all linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_p$ is a vector space.
2. **(T/F)** If $\{\mathbf{v}_1, \dots, \mathbf{v}_{p-1}\}$ spans V, then S spans V.
3. **(T/F)** If $\{\mathbf{v}_1, \dots, \mathbf{v}_{p-1}\}$ is linearly independent, then so is S.
4. **(T/F)** If S is linearly independent, then S is a basis for V.
5. **(T/F)** If Span $S = V$, then some subset of S is a basis for V.
6. **(T/F)** If $\dim V = p$ and Span $S = V$, then S cannot be linearly dependent.
7. **(T/F)** A plane in $\mathbb{R}^3$ is a two-dimensional subspace.
8. **(T/F)** The nonpivot columns of a matrix are always linearly dependent.
9. **(T/F)** Row operations on a matrix A can change the linear dependence relations among the rows of A.
10. **(T/F)** Row operations on a matrix can change the null space.
11. **(T/F)** The rank of a matrix equals the number of nonzero rows.
12. **(T/F)** If an $m \times n$ matrix A is row equivalent to an echelon matrix U and if U has k nonzero rows, then the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$ is $m - k$.
13. **(T/F)** If B is obtained from a matrix A by several elementary row operations, then rank B = rank A.
14. **(T/F)** The nonzero rows of a matrix A form a basis for Row A.
15. **(T/F)** If matrices A and B have the same reduced echelon form, then Row A = Row B.
16. **(T/F)** If H is a subspace of $\mathbb{R}^3$, then there is a 3×3 matrix A such that $H = \text{Col } A$.
17. **(T/F)** If A is $m \times n$ and rank $A = m$, then the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
18. **(T/F)** If A is $m \times n$ and the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is onto, then rank $A = m$.
19. **(T/F)** A change-of-coordinates matrix is always invertible.
20. Find a basis for the set of all vectors of the form
$$\begin{bmatrix} a - 2b + 5c \\ 2a + 5b - 8c \\ -a - 4b + 7c \\ 3a + b + c \end{bmatrix}. \quad \text{(Be careful.)}$$
21. Let $\mathbf{u}_1 = \begin{bmatrix} -2 \\ 4 \\ -6 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$, and $W = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$. Find an *implicit* description of W; that is, find a set of one or more homogeneous equations that characterize the points of W. [*Hint:* When is $\mathbf{b}$ in W?]
22. Explain what is wrong with the following discussion: Let $\mathbf{f}(t) = 3 + t$ and $\mathbf{g}(t) = 3t + t^2$, and note that $\mathbf{g}(t) = t\mathbf{f}(t)$. Then $\{\mathbf{f}, \mathbf{g}\}$ is linearly dependent because $\mathbf{g}$ is a multiple of $\mathbf{f}$.
23. Consider the polynomials $\mathbf{p}_1(t) = 1 + t$, $\mathbf{p}_2(t) = 1 - t$, $\mathbf{p}_3(t) = 4$, $\mathbf{p}_4(t) = t + t^2$, and $\mathbf{p}_5(t) = 1 + 2t + t^2$, and let H be the subspace of $\mathbb{P}_5$ spanned by the set $S = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_5\}$. Use the method described in the proof of the Spanning Set Theorem (Section 4.3) to produce a basis for H. (Explain how to select appropriate members of S.)
24. Suppose $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$, and $\mathbf{p}_4$ are specific polynomials that span a two-dimensional subspace H of $\mathbb{P}_5$. Describe how one can find a basis for H by examining the four polynomials and making almost no computations.
25. What would you have to know about the solution set of a homogeneous system of 18 linear equations in 20 variables in order to know that every associated nonhomogeneous equation has a solution? Discuss.

26. Let H be an n-dimensional subspace of an n-dimensional vector space V. Explain why $H = V$.

27. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.

a. What is the dimension of the range of T if T is a one-to-one mapping? Explain.

b. What is the dimension of the kernel of T (see Section 4.2) if T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$? Explain.

28. Let S be a maximal linearly independent subset of a vector space V. That is, S has the property that if a vector not in S is adjoined to S, then the new set will no longer be linearly independent. Prove that S must be a basis for V. [*Hint:* What if S were linearly independent but not a basis of V?]

29. Let S be a finite minimal spanning set of a vector space V. That is, S has the property that if a vector is removed from S, then the new set will no longer span V. Prove that S must be a basis for V.

Exercises 30–35 develop properties of rank that are sometimes needed in applications. Assume the matrix A is $m \times n$.

30. Show from parts (a) and (b) that rank AB cannot exceed the rank of A or the rank of B. (In general, the rank of a product of matrices cannot exceed the rank of any factor in the product.)

a. Show that if B is $n \times p$, then rank $AB \le$ rank A. [*Hint:* Explain why every vector in the column space of AB is in the column space of A.]

b. Show that if B is $n \times p$, then rank $AB \le$ rank B. [*Hint:* Use part (a) to study rank$(AB)^T$.]

31. Show that if P is an invertible $m \times m$ matrix, then rank $PA =$ rank A. [*Hint:* Apply Exercise 30 to PA and $P^{-1}(PA)$.]

32. Show that if Q is invertible, then rank $AQ =$ rank A. [*Hint:* Use Exercise 31 to study rank$(AQ)^T$.]

33. Let A be an $m \times n$ matrix, and let B be an $n \times p$ matrix such that $AB = 0$. Show that rank $A +$ rank $B \le n$. [*Hint:* One of the four subspaces Nul A, Col A, Nul B, and Col B is contained in one of the other three subspaces.]

34. If A is an $m \times n$ matrix of rank r, then a *rank factorization* of A is an equation of the form $A = CR$, where C is an $m \times r$ matrix of rank r and R is an $r \times n$ matrix of rank r. Show that such a factorization always exists. Then show that given any two $m \times n$ matrices A and B.

$$\text{rank}(A + B) \le \text{rank}\, A + \text{rank}\, B$$

[*Hint:* Write $A + B$ as the product of two partitioned matrices.]

35. A **submatrix** of a matrix A is any matrix that results from deleting some (or no) rows and/or columns of A. It can be shown that A has rank r if and only if A contains an invertible $r \times r$ submatrix and no larger square submatrix is invertible. Demonstrate part of this statement by explaining (a) why an $m \times n$ matrix A of rank r has an $m \times r$ submatrix A_1 of rank r, and (b) why A_1 has an invertible $r \times r$ submatrix A_2.

The concept of rank plays an important role in the design of engineering control systems. A *state-space model* of a control system includes a difference equation of the form

$$\mathbf{x}_{k+1} = A\mathbf{x}_k + B\mathbf{u}_k \quad \text{for } k = 0, 1, \ldots \tag{1}$$

where A is $n \times n$, B is $n \times m$, $\{\mathbf{x}_k\}$ is a sequence of "state vectors" in $\mathbb{R}^n$ that describe the state of the system at discrete times, and $\{\mathbf{u}_k\}$ is a *control*, or *input*, sequence. The pair (A, B) is said to be **controllable** if

$$\text{rank}\,[\, B \quad AB \quad A^2B \quad \cdots \quad A^{n-1}B \,] = n \tag{2}$$

The matrix that appears in (2) is called the **controllability matrix** for the system. If (A, B) is controllable, then the system can be controlled, or driven from the state $\mathbf{0}$ to any specified state $\mathbf{v}$ (in $\mathbb{R}^n$) in at most n steps, simply by choosing an appropriate control sequence in $\mathbb{R}^m$. This fact is illustrated in Exercise 36 for $n = 4$ and $m = 2$.

36. Suppose A is a 4×4 matrix and B is a 4×2 matrix, and let $\mathbf{u}_0, \ldots, \mathbf{u}_3$ represent a sequence of input vectors in $\mathbb{R}^2$.

a. Set $\mathbf{x}_0 = \mathbf{0}$, compute $\mathbf{x}_1, \ldots, \mathbf{x}_4$ from equation (1), and write a formula for $\mathbf{x}_4$ involving the controllability matrix M appearing in equation (2). (*Note:* The matrix M is constructed as a partitioned matrix. Its overall size here is 4×8.)

b. Suppose (A, B) is controllable and $\mathbf{v}$ is any vector in $\mathbb{R}^4$. Explain why there exists a control sequence $\mathbf{u}_0, \ldots, \mathbf{u}_3$ in $\mathbb{R}^2$ such that $\mathbf{x}_4 = \mathbf{v}$.

Determine if the matrix pairs in Exercises 37–40 are controllable.

37. $A = \begin{bmatrix} .9 & 1 & 0 \\ 0 & -.9 & 0 \\ 0 & 0 & .5 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

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

T **39.** $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -2 & -4.2 & -4.8 & -3.6 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$

T **40.** $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -13 & -12.2 & -1.5 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$

5 Eigenvalues and Eigenvectors

Introductory Example

DYNAMICAL SYSTEMS AND SPOTTED OWLS

In 1990, the northern spotted owl became the center of a nationwide controversy over the use and misuse of the majestic forests in the Pacific Northwest. Environmentalists convinced the federal government that the owl was threatened with extinction if logging continued in the old-growth forests (with trees more than 200 years old), where the owls prefer to live. The timber industry, anticipating the loss of 30,000–100,000 jobs as a result of new government restrictions on logging, argued that the owl should not be classified as a "threatened species" and cited a number of published scientific reports to support its case.[1]

The population of spotted owls continues to decline, and it remains a species caught in the crossfire between economic opportunities and conservation efforts. Mathematical ecologists help to analyze the effects on the spotted owl population of factors such as logging techniques, wildfires, and competition for habitat with the invasive barred owl. The life cycle of a spotted owl divides naturally into three stages: juvenile (up to 1 year old), subadult (1–2 years), and adult (older than 2 years). The owls mate for life during the subadult and adult stages, begin to breed as adults, and live for up to 20 years. Each owl pair requires about 1000 hectares (4 square miles) for its own home territory. A critical time in the life cycle is when the juveniles leave the nest. To survive and become a subadult, a juvenile must successfully find a new home range (and usually a mate).

[1] "The Great Spotted Owl War," *Reader's Digest*, November 1992, pp. 91–95.

A first step in studying the population dynamics is to model the population at yearly intervals, at times denoted by $k = 0, 1, 2, \ldots$. Usually, one assumes that there is a 1:1 ratio of males to females in each life stage and counts only the females. The population at year k can be described by a vector $\mathbf{x}_k = (j_k, s_k, a_k)$, where j_k, s_k, and a_k are the numbers of females in the juvenile, subadult, and adult stages, respectively.

Using actual field data from demographic studies, R. Lamberson and coworkers considered the following *stage-matrix model*:[2]

$$\begin{bmatrix} j_{k+1} \\ s_{k+1} \\ a_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & .33 \\ .18 & 0 & 0 \\ 0 & .71 & .94 \end{bmatrix} \begin{bmatrix} j_k \\ s_k \\ a_k \end{bmatrix}$$

Here the number of new juvenile females in year $k+1$ is .33 times the number of adult females in year k (based on the average birth rate per owl pair). Also, 18% of the juveniles survive to become subadults, and 71% of the subadults and 94% of the adults survive to be counted as adults.

The stage-matrix model is a difference equation of the form $\mathbf{x}_{k+1} = A\mathbf{x}_k$. Such an equation is often called a **dynamical system** (or a **discrete linear dynamical**

[2] R. H. Lamberson, R. McKelvey, B. R. Noon, and C. Voss, "A Dynamic Analysis of the Viability of the Northern Spotted Owl in a Fragmented Forest Environment," *Conservation Biology* **6** (1992), 505–512. Also, a private communication from Professor Lamberson, 1993.

system) because it describes the changes in a system as time passes.

The 18% juvenile survival rate in the Lamberson stage matrix is the entry affected most by the amount of old-growth forest available. Actually, 60% of the juveniles normally survive to leave the nest, but in the Willow Creek region of California studied by Lamberson and his colleagues, only 30% of the juveniles that left the nest were able to find new home ranges. The rest perished during the search process.

A significant reason for the failure of owls to find new home ranges is the increasing fragmentation of old-growth timber stands due to clear-cutting of scattered areas on the old-growth land. When an owl leaves the protective canopy of the forest and crosses a clear-cut area, the risk of attack by predators increases dramatically. Section 5.6 will show that the model described in the chapter introduction predicts the eventual demise of the spotted owl, but that if 50% of the juveniles who survive to leave the nest also find new home ranges, then the owl population will thrive.

The goal of this chapter is to dissect the action of a linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ into elements that are easily visualized. All matrices in the chapter are square. The main applications described here are to discrete dynamical systems, differential equations, and Markov chains. However, the basic concepts—eigenvectors and eigenvalues—are useful throughout pure and applied mathematics, and they appear in settings far more general than we consider here. Eigenvalues are also used to study differential equations and *continuous* dynamical systems, they provide critical information in engineering design, and they arise naturally in fields such as physics and chemistry.

5.1 Eigenvectors and Eigenvalues

Although a transformation $\mathbf{x} \mapsto A\mathbf{x}$ may move vectors in a variety of directions, it often happens that there are special vectors on which the action of A is quite simple.

EXAMPLE 1 Let $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. The images of $\mathbf{u}$ and $\mathbf{v}$ under multiplication by A are shown in Figure 1. In fact, $A\mathbf{v}$ is just $2\mathbf{v}$. So A only "stretches" or dilates $\mathbf{v}$. ■

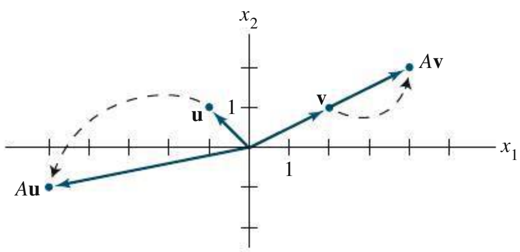


FIGURE 1 Effects of multiplication by A.

This section studies equations such as

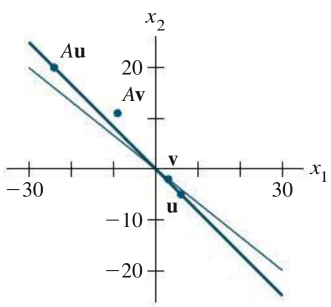
$$A\mathbf{x} = 2\mathbf{x} \quad \text{or} \quad A\mathbf{x} = -4\mathbf{x}$$

where special vectors are transformed by A into scalar multiples of themselves.

DEFINITION

An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ.[1]

It is easy to determine if a given vector is an eigenvector of a matrix. See Example 2. It is also easy to decide if a specified scalar is an eigenvalue. See Example 3.



$A\mathbf{u} = -4\mathbf{u}$, but $A\mathbf{v} \neq \lambda\mathbf{v}$.

EXAMPLE 2 Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Are $\mathbf{u}$ and $\mathbf{v}$ eigenvectors of A?

SOLUTION

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\mathbf{u}$$

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda\begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Thus $\mathbf{u}$ is an eigenvector corresponding to an eigenvalue (-4), but $\mathbf{v}$ is not an eigenvector of A, because $A\mathbf{v}$ is not a multiple of $\mathbf{v}$. ■

EXAMPLE 3 Show that 7 is an eigenvalue of matrix A in Example 2, and find the corresponding eigenvectors.

SOLUTION The scalar 7 is an eigenvalue of A if and only if the equation

$$A\mathbf{x} = 7\mathbf{x} \tag{1}$$

has a nontrivial solution. But (1) is equivalent to $A\mathbf{x} - 7\mathbf{x} = \mathbf{0}$, or

$$(A - 7I)\mathbf{x} = \mathbf{0} \tag{2}$$

To solve this homogeneous equation, form the matrix

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

The columns of $A - 7I$ are obviously linearly dependent, so (2) has nontrivial solutions. Thus 7 *is* an eigenvalue of A. To find the corresponding eigenvectors, use row operations:

$$\begin{bmatrix} -6 & 6 & 0 \\ 5 & -5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The general solution has the form $x_2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Each vector of this form with $x_2 \neq 0$ is an eigenvector corresponding to $\lambda = 7$. ■

[1] Note that an eigenvector must be *nonzero*, by definition, but an eigenvalue may be zero. The case in which the number 0 is an eigenvalue is discussed after Example 5.

Warning: Although row reduction was used in Example 3 to find eigen*vectors*, it cannot be used to find eigen*values*. An echelon form of a matrix A usually does *not* display the eigenvalues of A.

The equivalence of equations (1) and (2) obviously holds for any λ in place of $\lambda = 7$. Thus λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{3}$$

has a nontrivial solution. The set of *all* solutions of (3) is just the null space of the matrix $A - \lambda I$. So this set is a *subspace* of $\mathbb{R}^n$ and is called the **eigenspace** of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

Example 3 shows that for matrix A in Example 2, the eigenspace corresponding to $\lambda = 7$ consists of *all* multiples of $(1, 1)$, which is the line through $(1, 1)$ and the origin. From Example 2, you can check that the eigenspace corresponding to $\lambda = -4$ is the line through $(6, -5)$. These eigenspaces are shown in Figure 2, along with eigenvectors $(1, 1)$ and $(3/2, -5/4)$ and the geometric action of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ on each eigenspace.

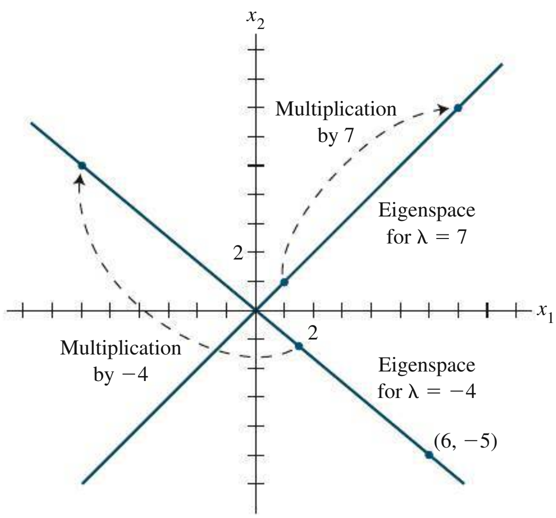


FIGURE 2 Eigenspaces for $\lambda = -4$ and $\lambda = 7$.

EXAMPLE 4 Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. An eigenvalue of A is 2. Find a basis for the corresponding eigenspace.

SOLUTION Form

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

and row reduce the augmented matrix for $(A - 2I)\mathbf{x} = \mathbf{0}$:

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

At this point, it is clear that 2 is indeed an eigenvalue of A because the equation $(A - 2I)\mathbf{x} = \mathbf{0}$ has free variables. The general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad x_2 \text{ and } x_3 \text{ free}$$

The eigenspace, shown in Figure 3, is a two-dimensional subspace of $\mathbb{R}^3$. A basis is

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

■

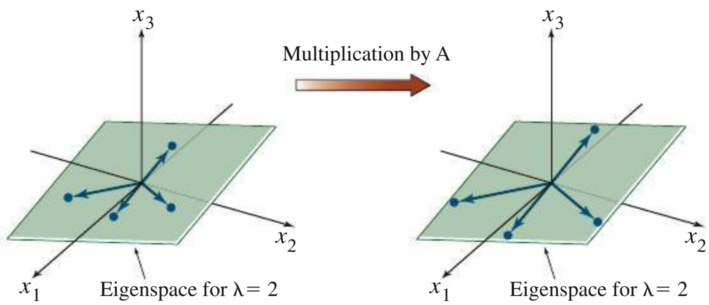


FIGURE 3 A acts as a dilation on the eigenspace.

Reasonable Answers

Remember that once you find a potential eigenvector $\mathbf{v}$, you can easily check your answer: just find $A\mathbf{v}$ and see if it is a multiple of $\mathbf{v}$. For example, to check whether $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}$, notice $A\mathbf{v} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$, which is not a multiple of $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, establishing that $\mathbf{v}$ is not an eigenvector. It turns out we had a sign error. The vector $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is a correct eigenvector for A since $A\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1\,\mathbf{u}$.

Numerical Notes

Example 4 shows a good method for manual computation of eigenvectors in simple cases when an eigenvalue is known. Using a matrix program and row reduction to find an eigenspace (for a specified eigenvalue) usually works, too, but this is not entirely reliable. Roundoff error can lead occasionally to a reduced echelon form with the wrong number of pivots. The best computer programs compute approximations for eigenvalues and eigenvectors simultaneously, to any desired degree of accuracy, for matrices that are not too large. The size of matrices that can be analyzed increases each year as computing power and software improve.

The following theorem describes one of the few special cases in which eigenvalues can be found precisely. Calculation of eigenvalues will also be discussed in Section 5.2.

THEOREM 1

The eigenvalues of a triangular matrix are the entries on its main diagonal.

PROOF For simplicity, consider the 3×3 case. If A is upper triangular, then $A - \lambda I$ has the form

$$\begin{aligned} A - \lambda I &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \\ &= \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix} \end{aligned}$$

The scalar λ is an eigenvalue of A if and only if the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution, that is, if and only if the equation has a free variable. Because of the zero entries in $A - \lambda I$, it is easy to see that $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a free variable if and only if at least one of the entries on the diagonal of $A - \lambda I$ is zero. This happens if and only if λ equals one of the entries a_{11}, a_{22}, a_{33} in A. For the case in which A is lower triangular, see Exercise 36. ■

EXAMPLE 5 Let $A = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & 3 & 4 \end{bmatrix}$. The eigenvalues of A are 3, 0, and 2. The eigenvalues of B are 4 and 1. ■

What does it mean for a matrix A to have an eigenvalue of 0, such as in Example 5? This happens if and only if the equation

$$A\mathbf{x} = 0\mathbf{x} \tag{4}$$

has a nontrivial solution. But (4) is equivalent to $A\mathbf{x} = \mathbf{0}$, which has a nontrivial solution if and only if A is not invertible. Thus 0 *is an eigenvalue of A if and only if A is not invertible*. This fact will be added to the Invertible Matrix Theorem in Section 5.2.

The following important theorem will be needed later. Its proof illustrates a typical calculation with eigenvectors. One way to prove the statement "If P then Q" is to show that P and the negation of Q leads to a contradiction. This strategy is used in the proof of the theorem.

THEOREM 2

If $\mathbf{v}_1, \dots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly independent.

PROOF Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly dependent. Since $\mathbf{v}_1$ is nonzero, Theorem 7 in Section 1.7 says that one of the vectors in the set is a linear combination of the preceding vectors. Let p be the least index such that $\mathbf{v}_{p+1}$ is a linear combination of the preceding (linearly independent) vectors. Then there exist scalars $c_1, \dots, c_p$ such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1} \tag{5}$$

Multiplying both sides of (5) by A and using the fact that $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$ for each k, we obtain

$$\begin{aligned} c_1 A\mathbf{v}_1 + \cdots + c_p A\mathbf{v}_p &= A\mathbf{v}_{p+1} \\ c_1 \lambda_1 \mathbf{v}_1 + \cdots + c_p \lambda_p \mathbf{v}_p &= \lambda_{p+1} \mathbf{v}_{p+1} \end{aligned} \tag{6}$$

Multiplying both sides of (5) by λ_{p+1} and subtracting the result from (6), we have

$$c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p = \mathbf{0} \tag{7}$$

Since $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly independent, the weights in (7) are all zero. But none of the factors $\lambda_i - \lambda_{p+1}$ are zero, because the eigenvalues are distinct. Hence $c_i = 0$ for $i = 1, \dots, p$. But then (5) says that $\mathbf{v}_{p+1} = \mathbf{0}$, which is impossible. Hence $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ cannot be linearly dependent and therefore must be linearly independent. ■

Eigenvectors and Difference Equations

This section concludes by showing how to construct solutions of the first-order difference equation discussed in the chapter introductory example:

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad (k = 0, 1, 2, \dots) \tag{8}$$

If A is an $n \times n$ matrix, then (8) is a *recursive* description of a sequence $\{\mathbf{x}_k\}$ in $\mathbb{R}^n$. A **solution** of (8) is an explicit description of $\{\mathbf{x}_k\}$ whose formula for each $\mathbf{x}_k$ does not depend directly on A or on the preceding terms in the sequence other than the initial term $\mathbf{x}_0$.

The simplest way to build a solution of (8) is to take an eigenvector $\mathbf{x}_0$ and its corresponding eigenvalue λ and let

$$\mathbf{x}_k = \lambda^k \mathbf{x}_0 \quad (k = 1, 2, \dots) \tag{9}$$

This sequence is a solution because

$$A\mathbf{x}_k = A(\lambda^k \mathbf{x}_0) = \lambda^k (A\mathbf{x}_0) = \lambda^k (\lambda \mathbf{x}_0) = \lambda^{k+1}\mathbf{x}_0 = \mathbf{x}_{k+1}$$

Linear combinations of solutions in the form of equation (9) are solutions, too! See Exercise 41.

Practice Problems

1. Is 5 an eigenvalue of $A = \begin{bmatrix} 6 & -3 & 1 \\ 3 & 0 & 5 \\ 2 & 2 & 6 \end{bmatrix}$?
2. If $\mathbf{x}$ is an eigenvector of A corresponding to λ, what is $A^3\mathbf{x}$?
3. Suppose that $\mathbf{b}_1$ and $\mathbf{b}_2$ are eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2, respectively, and suppose that $\mathbf{b}_3$ and $\mathbf{b}_4$ are linearly independent eigenvectors corresponding to a third distinct eigenvalue λ_3. Does it necessarily follow that $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set? [*Hint:* Consider the equation $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + (c_3\mathbf{b}_3 + c_4\mathbf{b}_4) = \mathbf{0}$.]
4. If A is an $n \times n$ matrix and λ is an eigenvalue of A, show that 2λ is an eigenvalue of $2A$.

5.1 Exercises

1. Is $\lambda = 2$ an eigenvalue of $\begin{bmatrix} 3 & 2 \\ 3 & 8 \end{bmatrix}$? Why or why not?

2. Is $\lambda = -2$ an eigenvalue of $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$? Why or why not?

3. Is $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} -3 & 1 \\ -3 & 8 \end{bmatrix}$? If so, find the eigenvalue.

4. Is $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$? If so, find the eigenvalue.

5. Is $\begin{bmatrix} 4 \\ -3 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 3 & 7 & 9 \\ -4 & -5 & 1 \\ 2 & 4 & 4 \end{bmatrix}$? If so, find the eigenvalue.

6. Is $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 2 & 6 & 7 \\ 3 & 2 & 7 \\ 5 & 6 & 4 \end{bmatrix}$? If so, find the eigenvalue.

7. Is $\lambda = 4$ an eigenvalue of $\begin{bmatrix} 3 & 0 & -1 \\ 2 & 3 & 1 \\ -3 & 4 & 5 \end{bmatrix}$? If so, find one corresponding eigenvector.

8. Is $\lambda = 3$ an eigenvalue of $\begin{bmatrix} 1 & 2 & 2 \\ 3 & -2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$? If so, find one corresponding eigenvector.

In Exercises 9–16, find a basis for the eigenspace corresponding to each listed eigenvalue.

9. $A = \begin{bmatrix} 5 & 0 \\ 2 & 1 \end{bmatrix}, \lambda = 1, 5$

10. $A = \begin{bmatrix} 10 & -9 \\ 4 & -2 \end{bmatrix}, \lambda = 4$

11. $A = \begin{bmatrix} 4 & -2 \\ -3 & 9 \end{bmatrix}, \lambda = 10$

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

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

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

15. $A = \begin{bmatrix} 4 & 2 & 3 \\ -1 & 1 & -3 \\ 2 & 4 & 9 \end{bmatrix}, \lambda = 3$

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

Find the eigenvalues of the matrices in Exercises 17 and 18.

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

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

19. For $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$, find one eigenvalue, with no calculation. Justify your answer.

20. Without calculation, find one eigenvalue and two linearly independent eigenvectors of $A = \begin{bmatrix} 5 & -5 & 5 \\ 5 & -5 & 5 \\ 5 & -5 & 5 \end{bmatrix}$. Justify your answer.

In Exercises 21–30, A is an $n \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If $A\mathbf{x} = \lambda\mathbf{x}$ for some vector $\mathbf{x}$, then λ is an eigenvalue of A.

22. **(T/F)** If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A.

23. **(T/F)** A matrix A is invertible if and only if 0 is an eigenvalue of A.

24. **(T/F)** A number c is an eigenvalue of A if and only if the equation $(A - cI)\mathbf{x} = 0$ has a nontrivial solution.

25. **(T/F)** Finding an eigenvector of A may be difficult, but checking whether a given vector is in fact an eigenvector is easy.

26. **(T/F)** To find the eigenvalues of A, reduce A to echelon form.

27. **(T/F)** If $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent eigenvectors, then they correspond to distinct eigenvalues.

28. **(T/F)** The eigenvalues of a matrix are on its main diagonal.

29. **(T/F)** If $\mathbf{v}$ is an eigenvector with eigenvalue 2, then $2\mathbf{v}$ is an eigenvector with eigenvalue 4.

30. **(T/F)** An eigenspace of A is a null space of a certain matrix.

31. Explain why a 2×2 matrix can have at most two distinct eigenvalues. Explain why an $n \times n$ matrix can have at most n distinct eigenvalues.

32. Construct an example of a 2×2 matrix with only one distinct eigenvalue.

33. Let λ be an eigenvalue of an invertible matrix A. Show that λ^{-1} is an eigenvalue of A^{-1}. [*Hint:* Suppose a nonzero $\mathbf{x}$ satisfies $A\mathbf{x} = \lambda\mathbf{x}$.]

34. Show that if A^2 is the zero matrix, then the only eigenvalue of A is 0.

35. Show that λ is an eigenvalue of A if and only if λ is an eigenvalue of A^T. [*Hint:* Find out how $A - \lambda I$ and $A^T - \lambda I$ are related.]

36. Use Exercise 35 to complete the proof of Theorem 1 for the case when A is lower triangular.

37. Consider an $n \times n$ matrix A with the property that the row sums all equal the same number s. Show that s is an eigenvalue of A. [*Hint:* Find an eigenvector.]

38. Consider an $n \times n$ matrix A with the property that the column sums all equal the same number s. Show that s is an eigenvalue of A. [*Hint:* Use Exercises 35 and 37.]

In Exercises 39 and 40, let A be the matrix of the linear transformation T. Without writing A, find an eigenvalue of A and describe the eigenspace.

39. T is the transformation on $\mathbb{R}^2$ that reflects points across some line through the origin.

40. T is the transformation on $\mathbb{R}^3$ that rotates points about some line through the origin.

41. Let $\mathbf{u}$ and $\mathbf{v}$ be eigenvectors of a matrix A, with corresponding eigenvalues λ and μ, and let c_1 and c_2 be scalars. Define

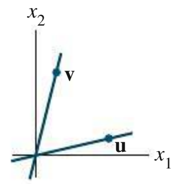
$$\mathbf{x}_k = c_1\lambda^k\mathbf{u} + c_2\mu^k\mathbf{v} \quad (k = 0, 1, 2, \ldots)$$

a. What is $\mathbf{x}_{k+1}$, by definition?

b. Compute $A\mathbf{x}_k$ from the formula for $\mathbf{x}_k$, and show that $A\mathbf{x}_k = \mathbf{x}_{k+1}$. This calculation will prove that the sequence $\{\mathbf{x}_k\}$ defined above satisfies the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$.

42. Describe how you might try to build a solution of a difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$ if you were given the initial $\mathbf{x}_0$ and this vector did not happen to be an eigenvector of A. [*Hint:* How might you relate $\mathbf{x}_0$ to eigenvectors of A?]

43. Let $\mathbf{u}$ and $\mathbf{v}$ be the vectors shown in the figure, and suppose $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of a 2×2 matrix A that correspond to eigenvalues 2 and 3, respectively. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $T(\mathbf{x}) = A\mathbf{x}$ for each $\mathbf{x}$ in $\mathbb{R}^2$, and let $\mathbf{w} = \mathbf{u} + \mathbf{v}$. Make a copy of the figure, and on the same coordinate system, carefully plot the vectors $T(\mathbf{u})$, $T(\mathbf{v})$, and $T(\mathbf{w})$.



44. Repeat Exercise 43, assuming $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of A that correspond to eigenvalues -1 and 3, respectively.

T In Exercises 45–48, use a matrix program to find the eigenvalues of the matrix. Then use the method of Example 4 with a row reduction routine to produce a basis for each eigenspace.

45. $\begin{bmatrix} 8 & -10 & -5 \\ 2 & 17 & 2 \\ -9 & -18 & 4 \end{bmatrix}$

46. $\begin{bmatrix} 9 & -4 & -2 & -4 \\ -56 & 32 & -28 & 44 \\ -14 & -14 & 6 & -14 \\ 42 & -33 & 21 & -45 \end{bmatrix}$

47. $\begin{bmatrix} 4 & -9 & -7 & 8 & 2 \\ -7 & -9 & 0 & 7 & 14 \\ 5 & 10 & 5 & -5 & -10 \\ -2 & 3 & 7 & 0 & 4 \\ -3 & -13 & -7 & 10 & 11 \end{bmatrix}$

48. $\begin{bmatrix} -4 & -4 & 20 & -8 & -1 \\ 14 & 12 & 46 & 18 & 2 \\ 6 & 4 & -18 & 8 & 1 \\ 11 & 7 & -37 & 17 & 2 \\ 18 & 12 & -60 & 24 & 5 \end{bmatrix}$

Solutions to Practice Problems

1. The number 5 is an eigenvalue of A if and only if the equation $(A - 5I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. Form

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

and row reduce the augmented matrix:

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

Solutions to Practice Problems (Continued)

At this point, it is clear that the homogeneous system has no free variables. Thus $A - 5I$ is an invertible matrix, which means that 5 is *not* an eigenvalue of A.

2. If $\mathbf{x}$ is an eigenvector of A corresponding to λ, then $A\mathbf{x} = \lambda\mathbf{x}$ and so

$$A^2\mathbf{x} = A(\lambda\mathbf{x}) = \lambda A\mathbf{x} = \lambda^2\mathbf{x}$$

Again, $A^3\mathbf{x} = A(A^2\mathbf{x}) = A(\lambda^2\mathbf{x}) = \lambda^2 A\mathbf{x} = \lambda^3\mathbf{x}$. The general pattern, $A^k\mathbf{x} = \lambda^k\mathbf{x}$, is proved by induction.

3. Yes. Suppose $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + (c_3\mathbf{b}_3 + c_4\mathbf{b}_4) = \mathbf{0}$. Since any linear combination of eigenvectors corresponding to the same eigenvalue is in the eigenspace for that eigenvalue, $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ is either $\mathbf{0}$ or an eigenvector for λ_3. If $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ were an eigenvector for λ_3, then by Theorem 2, $\{\mathbf{b}_1, \mathbf{b}_2, c_3\mathbf{b}_3 + c_4\mathbf{b}_4\}$ would be a linearly independent set, which would force $c_1 = c_2 = 0$ and $c_3\mathbf{b}_3 + c_4\mathbf{b}_4 = \mathbf{0}$, contradicting that $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ is an eigenvector. Thus $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ must be $\mathbf{0}$, implying that $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 = \mathbf{0}$ also. By Theorem 2, $\{\mathbf{b}_1, \mathbf{b}_2\}$ is a linearly independent set so $c_1 = c_2 = 0$. Moreover, $\{\mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set so $c_3 = c_4 = 0$. Since all of the coefficients c_1, c_2, c_3, and c_4 must be zero, it follows that $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set.

4. Since λ is an eigenvalue of A, there is a nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Multiplying both sides of this equation by 2 results in the equation $2(A\mathbf{x}) = 2(\lambda\mathbf{x})$. Thus $(2A)\mathbf{x} = (2\lambda)\mathbf{x}$ and hence 2λ is an eigenvalue of $2A$.

5.2 The Characteristic Equation

Useful information about the eigenvalues of a square matrix A is encoded in a special scalar equation called the characteristic equation of A. A simple example will lead to the general case.

EXAMPLE 1 Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

SOLUTION We must find all scalars λ such that the matrix equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

has a nontrivial solution. By the Invertible Matrix Theorem in Section 2.3, this problem is equivalent to finding all λ such that the matrix $A - \lambda I$ is *not* invertible, where

$$A - \lambda I = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}$$

By Theorem 4 in Section 2.2, this matrix fails to be invertible precisely when its determinant is zero. So the eigenvalues of A are the solutions of the equation

$$\det(A - \lambda I) = \det\begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = 0$$

Recall that

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

So

$$\begin{aligned}\det(A-\lambda I) &= (2-\lambda)(-6-\lambda)-(3)(3)\\ &= -12+6\lambda-2\lambda+\lambda^2-9\\ &= \lambda^2+4\lambda-21\\ &= (\lambda-3)(\lambda+7)\end{aligned}$$

If $\det(A-\lambda I)=0$, then $\lambda=3$ or $\lambda=-7$. So the eigenvalues of A are 3 and -7. ■

Determinants

The determinant in Example 1 transformed the matrix equation $(A-\lambda I)\mathbf{x}=\mathbf{0}$, which involves *two* unknowns λ and $\mathbf{x}$, into the scalar equation $\lambda^2+4\lambda-21=0$, which involves only *one* unknown. The same idea works for $n\times n$ matrices.

Before turning to larger matrices, recall from Section 3.1 that the matrix A_{ij} is obtained from A by deleting the ith row and jth column. The determinant of an $n\times n$ matrix A can be computed by an expansion across any row or down any column. The expansion across the ith row is given by

$$\det A = (-1)^{i+1}a_{i1}\det A_{i1} + (-1)^{i+2}a_{i2}\det A_{i2} + \cdots + (-1)^{i+n}a_{in}\det A_{in}$$

The expansion down the jth column is given by

$$\det A = (-1)^{1+j}a_{1j}\det A_{1j} + (-1)^{2+j}a_{2j}\det A_{2j} + \cdots + (-1)^{n+j}a_{nj}\det A_{nj}$$

EXAMPLE 2 Compute the determinant of

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

SOLUTION Any row or column can be chosen for the expansion. For example, expanding down the first column of A results in

$$\begin{aligned}\det A &= a_{11}\det A_{11} - a_{21}\det A_{21} + a_{31}\det A_{31}\\ &= 2\det\begin{bmatrix}0 & -1\\ 2 & 1\end{bmatrix} - 4\det\begin{bmatrix}3 & 1\\ 2 & 1\end{bmatrix} + 0\det\begin{bmatrix}3 & 1\\ 0 & -1\end{bmatrix}\\ &= 2(0-(-2)) - 4(3-2) + 0(-3-0) = 0\end{aligned}$$

The next theorem lists facts from Sections 3.1 and 3.2 and is included here for convenient reference. ■

THEOREM 3

Properties of Determinants

Let A and B be $n\times n$ matrices.

a. A is invertible if and only if $\det A \neq 0$.
b. $\det AB = (\det A)(\det B)$.
c. $\det A^T = \det A$.
d. If A is triangular, then $\det A$ is the product of the entries on the main diagonal of A.

e. A row replacement operation on A does not change the determinant. A row interchange changes the sign of the determinant. A row scaling also scales the determinant by the same scalar factor.

Recall that A is invertible if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Notice that the number 0 is an eigenvalue of A if and only if there *is* a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$, which happens if and only if $0 = \det(A - 0I) = \det A$. Hence A is invertible if and only if 0 is *not* an eigenvalue.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then A is invertible if and only if

r. The number 0 is *not* an eigenvalue of A.

The Characteristic Equation

Theorem 3(a) shows how to determine when a matrix of the form $A - \lambda I$ is *not* invertible. The scalar equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A, and the argument in Example 1 justifies the following fact.

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the characteristic equation

$$\det(A - \lambda I) = 0$$

EXAMPLE 3 Find the characteristic equation of

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

SOLUTION Form $A - \lambda I$, and use Theorem 3(d):

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix} \\ &= (5-\lambda)(3-\lambda)(5-\lambda)(1-\lambda) \end{aligned}$$

The characteristic equation is

$$(5-\lambda)^2(3-\lambda)(1-\lambda) = 0$$

or

$$(\lambda - 5)^2(\lambda - 3)(\lambda - 1) = 0$$

Expanding the product, we can also write

$$\lambda^4 - 14\lambda^3 + 68\lambda^2 - 130\lambda + 75 = 0 \qquad \blacksquare$$

Reasonable Answers

If you want to verify λ is an eigenvalue of A, row reduce $A - \lambda I$. If you get a pivot in every column, something is amiss—the scalar λ is not an eigenvalue of A. Looking back at Example 3, notice that $A - 5I$, $A - 3I$, and $A - I$ all have at least one column without a pivot; however, if λ is chosen to be any number other than 5, 3, or 1, the matrix $A - \lambda I$ has a pivot in every column.

In Examples 1 and 3, det $(A - \lambda I)$ is a polynomial in λ. It can be shown that if A is an $n \times n$ matrix, then det $(A - \lambda I)$ is a polynomial of degree n called the **characteristic polynomial** of A.

The eigenvalue 5 in Example 3 is said to have *multiplicity* 2 because $(\lambda - 5)$ occurs two times as a factor of the characteristic polynomial. In general, the (**algebraic**) **multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

EXAMPLE 4 The characteristic polynomial of a 6×6 matrix is $\lambda^6 - 4\lambda^5 - 12\lambda^4$. Find the eigenvalues and their multiplicities.

SOLUTION Factor the polynomial

$$\lambda^6 - 4\lambda^5 - 12\lambda^4 = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2)$$

The eigenvalues are 0 (multiplicity 4), 6 (multiplicity 1), and -2 (multiplicity 1). $\blacksquare$

We could also list the eigenvalues in Example 4 as 0, 0, 0, 0, 6, and -2, so that the eigenvalues are repeated according to their multiplicities.

Because the characteristic equation for an $n \times n$ matrix involves an nth-degree polynomial, the equation has exactly n roots, counting multiplicities, provided complex roots are allowed. Such complex roots, called *complex eigenvalues*, will be discussed in Section 5.5. Until then, we consider only real eigenvalues, and scalars will continue to be real numbers.

The characteristic equation is important for theoretical purposes. In practical work, however, eigenvalues of any matrix larger than 2×2 should be found by a computer, unless the matrix is triangular or has other special properties. Although a 3×3 characteristic polynomial is easy to compute by hand, factoring it can be difficult (unless the matrix is carefully chosen). See the Numerical Notes at the end of this section.

STUDY GUIDE has advice on how to factor a polynomial.

Similarity

The next theorem illustrates one use of the characteristic polynomial, and it provides the foundation for several iterative methods that *approximate* eigenvalues. If A and B are $n \times n$ matrices, then A **is similar to** B if there is an invertible matrix P such that $P^{-1}AP = B$, or, equivalently, $A = PBP^{-1}$. Writing Q for P^{-1}, we have $Q^{-1}BQ = A$. So B is also similar to A, and we say simply that A and B **are similar**. Changing A into $P^{-1}AP$ is called a **similarity transformation**.

THEOREM 4 If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).

PROOF If $B = P^{-1}AP$, then

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(AP - \lambda P) = P^{-1}(A - \lambda I)P$$

Using the multiplicative property (b) in Theorem 3, we compute

$$\begin{aligned} \det(B - \lambda I) &= \det[P^{-1}(A - \lambda I)P] \\ &= \det(P^{-1}) \cdot \det(A - \lambda I) \cdot \det(P) \end{aligned} \tag{1}$$

Since $\det(P^{-1}) \cdot \det(P) = \det(P^{-1}P) = \det I = 1$, we see from equation (1) that $\det(B - \lambda I) = \det(A - \lambda I)$. ■

***Warnings*:**

1. The matrices
$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$
are not similar even though they have the same eigenvalues.
2. Similarity is not the same as row equivalence. (If A is row equivalent to B, then $B = EA$ for some invertible matrix E.) Row operations on a matrix usually change its eigenvalues.

Application to Dynamical Systems

Eigenvalues and eigenvectors hold the key to the discrete evolution of a dynamical system, as mentioned in the chapter introduction.

EXAMPLE 5 Let $A = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}$. Analyze the long-term behavior (as k increases) of the dynamical system defined by $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$, with $\mathbf{x}_0 = \begin{bmatrix} .6 \\ .4 \end{bmatrix}$.

SOLUTION The first step is to find the eigenvalues of A and a basis for each eigenspace. The characteristic equation for A is

$$\begin{aligned} 0 &= \det\begin{bmatrix} .95 - \lambda & .03 \\ .05 & .97 - \lambda \end{bmatrix} = (.95 - \lambda)(.97 - \lambda) - (.03)(.05) \\ &= \lambda^2 - 1.92\lambda + .92 \end{aligned}$$

By the quadratic formula

$$\begin{aligned} \lambda &= \frac{1.92 \pm \sqrt{(-1.92)^2 - 4(.92)}}{2} = \frac{1.92 \pm \sqrt{.0064}}{2} \\ &= \frac{1.92 \pm .08}{2} = 1 \quad \text{or} \quad .92 \end{aligned}$$

It is readily checked that eigenvectors corresponding to $\lambda = 1$ and $\lambda = .92$ are multiples of

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

respectively.

The next step is to write the given $\mathbf{x}_0$ in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$. This can be done because $\{\mathbf{v}_1, \mathbf{v}_2\}$ is obviously a basis for $\mathbb{R}^2$. (Why?) So there exist weights c_1 and c_2 such that

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{2}$$

In fact,

$$\begin{aligned} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} &= [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]^{-1}\mathbf{x}_0 = \begin{bmatrix} 3 & 1 \\ 5 & -1 \end{bmatrix}^{-1}\begin{bmatrix} .60 \\ .40 \end{bmatrix} \\ &= \frac{1}{-8}\begin{bmatrix} -1 & -1 \\ -5 & 3 \end{bmatrix}\begin{bmatrix} .60 \\ .40 \end{bmatrix} = \begin{bmatrix} .125 \\ .225 \end{bmatrix} \end{aligned} \tag{3}$$

Because $\mathbf{v}_1$ and $\mathbf{v}_2$ in (3) are eigenvectors of A, with $A\mathbf{v}_1 = \mathbf{v}_1$ and $A\mathbf{v}_2 = .92\mathbf{v}_2$, we easily compute each $\mathbf{x}_k$:

$$\begin{aligned} \mathbf{x}_1 = A\mathbf{x}_0 &= c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 && \text{Using linearity of } \mathbf{x} \mapsto A\mathbf{x} \\ &= c_1\mathbf{v}_1 + c_2(.92)\mathbf{v}_2 && \mathbf{v}_1 \text{ and } \mathbf{v}_2 \text{ are eigenvectors.} \\ \mathbf{x}_2 = A\mathbf{x}_1 &= c_1A\mathbf{v}_1 + c_2(.92)A\mathbf{v}_2 \\ &= c_1\mathbf{v}_1 + c_2(.92)^2\mathbf{v}_2 \end{aligned}$$

and so on. In general,

$$\mathbf{x}_k = c_1\mathbf{v}_1 + c_2(.92)^k\mathbf{v}_2 \quad (k = 0, 1, 2, \ldots)$$

Using c_1 and c_2 from (4),

$$\mathbf{x}_k = .125\begin{bmatrix} 3 \\ 5 \end{bmatrix} + .225(.92)^k\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (k = 0, 1, 2, \ldots) \tag{4}$$

This explicit formula for $\mathbf{x}_k$ gives the solution of the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$. As $k \to \infty$, $(.92)^k$ tends to zero and $\mathbf{x}_k$ tends to $\begin{bmatrix} .375 \\ .625 \end{bmatrix} = .125\mathbf{v}_1$. ■

The calculations in Example 5 have an interesting application to a Markov chain discussed in Section 5.9. Those who read that section may recognize that matrix A in Example 5 above is the same as the migration matrix M in Section 5.9, $\mathbf{x}_0$ is the initial population distribution between city and suburbs, and $\mathbf{x}_k$ represents the population distribution after k years.

Numerical Notes

1. Computer software such as Mathematica and Maple can use symbolic calculations to find the characteristic polynomial of a moderate-sized matrix. But there is no formula or finite algorithm to solve the characteristic equation of a general $n \times n$ matrix for $n \geq 5$.

Numerical Notes (Continued)

2. The best numerical methods for finding eigenvalues avoid the characteristic polynomial entirely. In fact, MATLAB finds the characteristic polynomial of a matrix A by first computing the eigenvalues $\lambda_1, \dots, \lambda_n$ of A and then expanding the product $(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$.
3. Several common algorithms for estimating the eigenvalues of a matrix A are based on Theorem 4. The powerful *QR algorithm* is discussed in the exercises. Another technique, called *Jacobi's method*, works when $A = A^T$ and computes a sequence of matrices of the form

$$A_1 = A \quad \text{and} \quad A_{k+1} = P_k^{-1} A_k P_k \quad (k = 1, 2, \dots)$$

   Each matrix in the sequence is similar to A and so has the same eigenvalues as A. The nondiagonal entries of A_{k+1} tend to zero as k increases, and the diagonal entries tend to approach the eigenvalues of A.
4. Other methods of estimating eigenvalues are discussed in Section 5.8.

Practice Problem

Find the characteristic equation and eigenvalues of $A = \begin{bmatrix} 1 & -4 \\ 4 & 2 \end{bmatrix}$.

5.2 Exercises

Find the characteristic polynomial and the eigenvalues of the matrices in Exercises 1–8.

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

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

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

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

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

6. $\begin{bmatrix} 1 & -4 \\ 4 & 6 \end{bmatrix}$

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

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

Exercises 9–14 require techniques from Section 3.1. Find the characteristic polynomial of each matrix using expansion across a row or down a column. [*Note:* Finding the characteristic polynomial of a 3×3 matrix is not easy to do with just row operations, because the variable λ is involved.]

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

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

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

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

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

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

For the matrices in Exercises 15–17, list the eigenvalues, repeated according to their multiplicities.

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

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

17. $\begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ -5 & 1 & 0 & 0 & 0 \\ 3 & 8 & 0 & 0 & 0 \\ 0 & -7 & 2 & 1 & 0 \\ -4 & 1 & 9 & -2 & 3 \end{bmatrix}$

18. It can be shown that the algebraic multiplicity of an eigenvalue λ is always greater than or equal to the dimension of the eigenspace corresponding to λ. Find h in the matrix A below such that the eigenspace for $\lambda = 5$ is two-dimensional:

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

19. Let A be an $n \times n$ matrix, and suppose A has n real eigenvalues, $\lambda_1, \dots, \lambda_n$, repeated according to multiplicities, so that
$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda)$

Explain why $\det A$ is the product of the n eigenvalues of A. (This result is true for any square matrix when complex eigenvalues are considered.)

20. Use a property of determinants to show that A and A^T have the same characteristic polynomial.

In Exercises 21–30, A and B are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If 0 is an eigenvalue of A, then A is invertible.

22. **(T/F)** The zero vector is in the eigenspace of A associated with an eigenvalue λ.

23. **(T/F)** The matrix A and its transpose, A^T, have different sets of eigenvalues.

24. **(T/F)** The matrices A and $B^{-1}AB$ have the same sets of eigenvalues for every invertible matrix B.

25. **(T/F)** If 2 is an eigenvalue of A, then $A - 2I$ is not invertible.

26. **(T/F)** If two matrices have the same set of eigenvalues, then they are similar.

27. **(T/F)** If $\lambda + 5$ is a factor of the characteristic polynomial of A, then 5 is an eigenvalue of A.

28. **(T/F)** The multiplicity of a root r of the characteristic equation of A is called the algebraic multiplicity of r as an eigenvalue of A.

29. **(T/F)** The eigenvalue of the $n \times n$ identity matrix is 1 with algebraic multiplicity n.

30. **(T/F)** The matrix A can have more than n eigenvalues.

A widely used method for estimating eigenvalues of a general matrix A is the *QR algorithm*. Under suitable conditions, this algorithm produces a sequence of matrices, all similar to A, that become almost upper triangular, with diagonal entries that approach the eigenvalues of A. The main idea is to factor A (or another matrix similar to A) in the form $A = Q_1R_1$, where $Q_1^T = Q_1^{-1}$ and R_1 is upper triangular. The factors are interchanged to form $A_1 = R_1Q_1$, which is again factored as $A_1 = Q_2R_2$; then to form $A_2 = R_2Q_2$, and so on. The similarity of $A, A_1, \ldots$ follows from the more general result in Exercise 31.

31. Show that if $A = QR$ with Q invertible, then A is similar to $A_1 = RQ$.

32. Show that if A and B are similar, then $\det A = \det B$.

T 33. Construct a random integer-valued 4×4 matrix A, and verify that A and A^T have the same characteristic polynomial (the same eigenvalues with the same multiplicities). Do A and A^T have the same eigenvectors? Make the same analysis of a 5×5 matrix. Report the matrices and your conclusions.

T 34. Construct a random integer-valued 4×4 matrix A.

a. Reduce A to echelon form U with no row scaling, and compute $\det A$. (If A happens to be singular, start over with a new random matrix.)

b. Compute the eigenvalues of A and the product of these eigenvalues (as accurately as possible).

c. List the matrix A, and, to four decimal places, list the pivots in U and the eigenvalues of A. Compute $\det A$ with your matrix program, and compare it with the products you found in (a) and (b).

T 35. Let $A = \begin{bmatrix} -6 & 28 & 21 \\ 4 & -15 & -12 \\ -8 & a & 25 \end{bmatrix}$. For each value of a in the set $\{32, 31.9, 31.8, 32.1, 32.2\}$, compute the characteristic polynomial of A and the eigenvalues. In each case, create a graph of the characteristic polynomial $p(t) = \det(A - tI)$ for $0 \le t \le 3$. If possible, construct all graphs on one coordinate system. Describe how the graphs reveal the changes in the eigenvalues as a changes.

Solution to Practice Problem

The characteristic equation is

$$0 = \det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & -4 \\ 4 & 2-\lambda \end{bmatrix}$$
$$= (1-\lambda)(2-\lambda) - (-4)(4) = \lambda^2 - 3\lambda + 18$$

From the quadratic formula,

$$\lambda = \frac{3 \pm \sqrt{(-3)^2 - 4(18)}}{2} = \frac{3 \pm \sqrt{-63}}{2}$$

It is clear that the characteristic equation has no real solutions, so A has no real eigenvalues. The matrix A is acting on the real vector space $\mathbb{R}^2$, and there is no nonzero vector $\mathbf{v}$ in $\mathbb{R}^2$ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ.

5.3 Diagonalization

In many cases, the eigenvalue–eigenvector information contained within a matrix A can be displayed in a useful factorization of the form $A = PDP^{-1}$ where D is a diagonal matrix. In this section, the factorization enables us to compute A^k quickly for large values of k, a fundamental idea in several applications of linear algebra. Later, in Sections 5.6 and 5.7, the factorization will be used to analyze (and *decouple*) dynamical systems.

The following example illustrates that powers of a diagonal matrix are easy to compute.

EXAMPLE 1 If $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, then $D^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}$
and

$$D^3 = DD^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix} = \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix}$$

In general,

$$D^k = \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \quad \text{for } k \geq 1$$

■

If $A = PDP^{-1}$ for some invertible P and diagonal D, then A^k is also easy to compute, as the next example shows.

EXAMPLE 2 Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a formula for A^k, given that $A = PDP^{-1}$, where

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

SOLUTION The standard formula for the inverse of a 2×2 matrix yields

$$P^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Then, by associativity of matrix multiplication,

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD\underbrace{(P^{-1}P)}_{I}DP^{-1} = PDDP^{-1}$$

$$= PD^2P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Again,

$$A^3 = (PDP^{-1})A^2 = (PD\underbrace{P^{-1})P}_{I}D^2P^{-1} = PDD^2P^{-1} = PD^3P^{-1}$$

In general, for $k \geq 1$,

$$A^k = PD^kP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{bmatrix}$$

■

A square matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D. The next theorem gives a characterization of diagonalizable matrices and tells how to construct a suitable factorization.

THEOREM 5

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In fact, $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an **eigenvector basis** of $\mathbb{R}^n$.

PROOF First, observe that if P is any $n \times n$ matrix with columns $\mathbf{v}_1, \dots, \mathbf{v}_n$, and if D is any diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$, then

$$AP = A[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,] = [\,A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n\,] \tag{1}$$

while

$$PD = P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\,\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n\,] \tag{2}$$

Now suppose A is diagonalizable and $A = PDP^{-1}$. Then right-multiplying this relation by P, we have $AP = PD$. In this case, equations (1) and (2) imply that

$$[\,A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n\,] = [\,\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n\,] \tag{3}$$

Equating columns, we find that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \quad \dots, \quad A\mathbf{v}_n = \lambda_n\mathbf{v}_n \tag{4}$$

Since P is invertible, its columns $\mathbf{v}_1, \dots, \mathbf{v}_n$ must be linearly independent. Also, since these columns are nonzero, the equations in (4) show that $\lambda_1, \dots, \lambda_n$ are eigenvalues and $\mathbf{v}_1, \dots, \mathbf{v}_n$ are corresponding eigenvectors. This argument proves the "only if" parts of the first and second statements, along with the third statement, of the theorem.

Finally, given any n eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, use them to construct the columns of P and use corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ to construct D. By equations (1)–(3), $AP = PD$. This is true without any condition on the eigenvectors. If, in fact, the eigenvectors are linearly independent, then P is invertible (by the Invertible Matrix Theorem), and $AP = PD$ implies that $A = PDP^{-1}$. ■

Diagonalizing Matrices

EXAMPLE 3 Diagonalize the following matrix, if possible.

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

That is, find an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

SOLUTION There are four steps to implement the description in Theorem 5.

Step 1. Find the eigenvalues of A. As mentioned in Section 5.2, the mechanics of this step are appropriate for a computer when the matrix is larger than 2×2. To avoid unnecessary distractions, the text will usually supply information needed for this step. In the present case, the characteristic equation turns out to involve a cubic polynomial that can be factored:

$$\begin{aligned} 0 = \det(A - \lambda I) &= -\lambda^3 - 3\lambda^2 + 4 \\ &= -(\lambda - 1)(\lambda + 2)^2 \end{aligned}$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$.

Step 2. Find three linearly independent eigenvectors of A. *Three* vectors are needed because A is a 3×3 matrix. This is the critical step. If it fails, then Theorem 5 says that A cannot be diagonalized. The method in Section 5.1 produces a basis for each eigenspace:

$$\text{Basis for } \lambda = 1: \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

You can check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set.

Step 3. Construct P from the vectors in step 2. The vectors may be listed in any order. Using the order chosen in step 2, form

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Step 4. Construct D from the corresponding eigenvalues. In this step, it is essential that the order of the eigenvalues matches the order chosen for the columns of P. Use the eigenvalue $\lambda = -2$ twice, once for each of the eigenvectors corresponding to $\lambda = -2$:

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

It is a good idea to check that P and D really work. To avoid computing P^{-1}, simply verify that $AP = PD$. This is equivalent to $A = PDP^{-1}$ when P is invertible. (However, be sure that P is invertible!) Compute

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

$$PD = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix} \quad \blacksquare$$

EXAMPLE 4 Diagonalize the following matrix, if possible.

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

SOLUTION The characteristic equation of A turns out to be exactly the same as that in Example 3:

$$0 = \det(A - \lambda I) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$. However, it is easy to verify that each eigenspace is only one-dimensional:

$$\text{Basis for } \lambda = 1: \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \qquad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

There are no other eigenvalues, and every eigenvector of A is a multiple of either $\mathbf{v}_1$ or $\mathbf{v}_2$. Hence it is impossible to construct a basis of $\mathbb{R}^3$ using eigenvectors of A. By Theorem 5, A is *not* diagonalizable. ■

The following theorem provides a *sufficient* condition for a matrix to be diagonalizable.

THEOREM 6 An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

PROOF Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be eigenvectors corresponding to the n distinct eigenvalues of a matrix A. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent, by Theorem 2 in Section 5.1. Hence A is diagonalizable, by Theorem 5. ■

It is not *necessary* for an $n \times n$ matrix to have n distinct eigenvalues in order to be diagonalizable. The 3×3 matrix in Example 3 is diagonalizable even though it has only two distinct eigenvalues.

EXAMPLE 5 Determine if the following matrix is diagonalizable.

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

SOLUTION This is easy! Since the matrix is triangular, its eigenvalues are obviously 5, 0, and -2. Since A is a 3×3 matrix with three distinct eigenvalues, A is diagonalizable. ■

Matrices Whose Eigenvalues Are Not Distinct

If an $n \times n$ matrix A has n distinct eigenvalues, with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and if $P = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_n\,]$, then P is automatically invertible because its columns are linearly independent, by Theorem 2. When A is diagonalizable but has fewer than n distinct eigenvalues, it is still possible to build P in a way that makes P automatically invertible, as the next theorem shows.[1]

THEOREM 7

Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \dots, \lambda_p$.

a. For $1 \le k \le p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.

b. The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if (*i*) the characteristic polynomial factors completely into linear factors and (*ii*) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

c. If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \dots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

EXAMPLE 6 Diagonalize the following matrix, if possible.

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

SOLUTION Since A is a triangular matrix, the eigenvalues are 5 and -3, each with multiplicity 2. Using the method in Section 5.1, we find a basis for each eigenspace.

$$\text{Basis for } \lambda = 5: \quad \mathbf{v}_1 = \begin{bmatrix} -8 \\ 4 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -16 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -3: \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

The set $\{\mathbf{v}_1, \dots, \mathbf{v}_4\}$ is linearly independent, by Theorem 7. So the matrix $P = [\, \mathbf{v}_1 \ \cdots \ \mathbf{v}_4 \,]$ is invertible, and $A = PDP^{-1}$, where

$$P = \begin{bmatrix} -8 & -16 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

[1] The proof of Theorem 7 is somewhat lengthy but not difficult. For instance, see S. Friedberg, A. Insel, and L. Spence, *Linear Algebra*, 4th ed. (Englewood Cliffs, NJ: Prentice-Hall, 2002), Section 5.2.

Practice Problems

1. Compute A^8, where $A = \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$.
2. Let $A = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Suppose you are told that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. Use this information to diagonalize A.
3. Let A be a 4×4 matrix with eigenvalues 5, 3, and -2, and suppose you know that the eigenspace for $\lambda = 3$ is two-dimensional. Do you have enough information to determine if A is diagonalizable?

5.3 Exercises

In Exercises 1 and 2, let $A = PDP^{-1}$ and compute A^4.

1. $P = \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

2. $P = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

In Exercises 3 and 4, use the factorization $A = PDP^{-1}$ to compute A^k, where k represents an arbitrary positive integer.

3. $\begin{bmatrix} a & 0 \\ 3(a-b) & b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$

4. $\begin{bmatrix} -6 & 8 \\ -4 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$

In Exercises 5 and 6, the matrix A is factored in the form PDP^{-1}. Use the Diagonalization Theorem to find the eigenvalues of A and a basis for each eigenspace.

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

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

Diagonalize the matrices in Exercises 7–20, if possible. The eigenvalues for Exercises 11–16 are as follows: (11) $\lambda = 1, 2, 3$; (12) $\lambda = 1, 4$; (13) $\lambda = 5, 1$; (14) $\lambda = 3, 4$; (15) $\lambda = 3, 1$; (16) $\lambda = 2, 1$. For Exercise 18, one eigenvalue is $\lambda = 5$ and one eigenvector is $(-2, 1, 2)$.

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

8. $\begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}$

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

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

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

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

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

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

15. $\begin{bmatrix} 7 & 4 & 16 \\ 2 & 5 & 8 \\ -2 & -2 & -5 \end{bmatrix}$

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

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

18. $\begin{bmatrix} -7 & -16 & 4 \\ 6 & 13 & -2 \\ 12 & 16 & 1 \end{bmatrix}$

19. $\begin{bmatrix} 5 & -3 & 0 & 9 \\ 0 & 3 & 1 & -2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

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

In Exercises 21–28, A, P, and D are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer. (Study Theorems 5 and 6 and the examples in this section carefully before you try these exercises.)

21. **(T/F)** A is diagonalizable if $A = PDP^{-1}$ for some matrix D and some invertible matrix P.

22. **(T/F)** If $\mathbb{R}^n$ has a basis of eigenvectors of A, then A is diagonalizable.

23. **(T/F)** A is diagonalizable if and only if A has n eigenvalues, counting multiplicities.

24. **(T/F)** If A is diagonalizable, then A is invertible.

25. **(T/F)** A is diagonalizable if A has n eigenvectors.

26. **(T/F)** If A is diagonalizable, then A has n distinct eigenvalues.

27. **(T/F)** If $AP = PD$, with D diagonal, then the nonzero columns of P must be eigenvectors of A.

28. **(T/F)** If A is invertible, then A is diagonalizable.

29. A is a 5×5 matrix with two eigenvalues. One eigenspace is three-dimensional, and the other eigenspace is two-dimensional. Is A diagonalizable? Why?

30. A is a 3×3 matrix with two eigenvalues. Each eigenspace is one-dimensional. Is A diagonalizable? Why?

31. A is a 4×4 matrix with three eigenvalues. One eigenspace is one-dimensional, and one of the other eigenspaces is two-dimensional. Is it possible that A is *not* diagonalizable? Justify your answer.

32. A is a 7×7 matrix with three eigenvalues. One eigenspace is two-dimensional, and one of the other eigenspaces is three-dimensional. Is it possible that A is *not* diagonalizable? Justify your answer.

33. Show that if A is both diagonalizable and invertible, then so is A^{-1}.

34. Show that if A has n linearly independent eigenvectors, then so does A^T. [*Hint:* Use the Diagonalization Theorem.]

35. A factorization $A = PDP^{-1}$ is not unique. Demonstrate this for the matrix A in Example 2. With $D_1 = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$, use the information in Example 2 to find a matrix P_1 such that $A = P_1 D_1 P_1^{-1}$.

36. With A and D as in Example 2, find an invertible P_2 unequal to the P in Example 2, such that $A = P_2 D P_2^{-1}$.

37. Construct a nonzero 2×2 matrix that is invertible but not diagonalizable.

38. Construct a nondiagonal 2×2 matrix that is diagonalizable but not invertible.

Diagonalize the matrices in Exercises 39–42. Use your matrix program's eigenvalue command to find the eigenvalues, and then compute bases for the eigenspaces as in Section 5.1.

T **39.** $\begin{bmatrix} -6 & 4 & 0 & 9 \\ -3 & 0 & 1 & 6 \\ -1 & -2 & 1 & 0 \\ -4 & 4 & 0 & 7 \end{bmatrix}$

T **40.** $\begin{bmatrix} 0 & 13 & 8 & 4 \\ 4 & 9 & 8 & 4 \\ 8 & 6 & 12 & 8 \\ 0 & 5 & 0 & -4 \end{bmatrix}$

T **41.** $\begin{bmatrix} 11 & -6 & 4 & -10 & -4 \\ -3 & 5 & -2 & 4 & 1 \\ -8 & 12 & -3 & 12 & 4 \\ 1 & 6 & -2 & 3 & -1 \\ 8 & -18 & 8 & -14 & -1 \end{bmatrix}$

T **42.** $\begin{bmatrix} 4 & 4 & 2 & 3 & -2 \\ 0 & 1 & -2 & -2 & 2 \\ 6 & 12 & 11 & 2 & -4 \\ 9 & 20 & 10 & 10 & -6 \\ 15 & 28 & 14 & 5 & -3 \end{bmatrix}$

Solutions to Practice Problems

1. $\det(A - \lambda I) = \lambda^2 - 3\lambda + 2 = (\lambda - 2)(\lambda - 1)$. The eigenvalues are 2 and 1, and the corresponding eigenvectors are $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Next, form

$$P = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad P^{-1} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$$

Since $A = PDP^{-1}$,

$$\begin{aligned} A^8 = PD^8P^{-1} &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2^8 & 0 \\ 0 & 1^8 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 256 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 766 & -765 \\ 510 & -509 \end{bmatrix} \end{aligned}$$

2. Compute $A\mathbf{v}_1 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 1 \cdot \mathbf{v}_1$, and

$$A\mathbf{v}_2 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \cdot \mathbf{v}_2$$

So, $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors for the eigenvalues 1 and 3, respectively. Thus

$$A = PDP^{-1}, \quad \text{where} \quad P = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

3. Yes, A is diagonalizable. There is a basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the eigenspace corresponding to $\lambda = 3$. In addition, there will be at least one eigenvector for $\lambda = 5$ and one for $\lambda = -2$. Call them $\mathbf{v}_3$ and $\mathbf{v}_4$. Then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent by Theorem 2 and Practice Problem 3 in Section 5.1. There can be no additional eigenvectors that are linearly independent from $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$, because the vectors are all in $\mathbb{R}^4$. Hence the eigenspaces for $\lambda = 5$ and $\lambda = -2$ are both one-dimensional. It follows that A is diagonalizable by Theorem 7(b).

STUDY GUIDE has advice on mastering eigenvalues and eigenspaces.

5.4 Eigenvectors and Linear Transformations

In this section, we will look at eigenvalues and eigenvectors of linear transformations $T : V \to V$, where V is any vector space. In the case where V is a finite dimensional vector space and there is a basis for V consisting of eigenvectors of T, we will see how to represent the transformation T as left multiplication by a diagonal matrix.

Eigenvectors of Linear Transformations

Previously, we looked at a variety of vector spaces including the discrete-time signal space, $\mathbb{S}$, and the set of polynomials, $\mathbb{P}$. Eigenvalues and eigenvectors can be defined for linear transformations from any vector space to itself.

DEFINITION

Let V be a vector space. An **eigenvector** of a linear transformation $T : V \to V$ is a nonzero vector $\mathbf{x}$ in V such that $T(\mathbf{x}) = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of T if there is a nontrivial solution $\mathbf{x}$ of $T(\mathbf{x}) = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an **eigenvector** corresponding to λ.

EXAMPLE 1 The sinusoidal signals were studied in detail in Sections 4.7 and 4.8. Consider the signal defined by $\{s_k\} = \left\{\cos\left(\frac{k\pi}{2}\right)\right\}$, where k ranges over all integers. The left double-shift linear transformation D is defined by $D(\{x_k\}) = \{x_{k+2}\}$. Show that $\{s_k\}$ is an eigenvector of D and determine the associated eigenvalue.

SOLUTION The trigonometric formula $\cos(\theta + \pi) = -\cos(\theta)$ is useful here. Set $\{y_k\} = D(\{s_k\})$ and observe that

$$y_k = s_{k+2} = \cos\left(\frac{(k+2)\pi}{2}\right) = \cos\left(\frac{k\pi}{2} + \pi\right) = -\cos\left(\frac{k\pi}{2}\right) = -s_k$$

and so $D(\{s_k\}) = \{-s_k\} = -\{s_k\}$. This establishes that $\{s_k\}$ is an eigenvector of D with eigenvalue -1. ■

In Figure 1, different values for the frequency, f, are chosen to graph a section of the sinusoidal signals $\left\{\cos\left(\frac{fk\pi}{4}\right)\right\}$ and $D\left(\left\{\cos\left(\frac{fk\pi}{4}\right)\right\}\right)$. Setting $f = 2$ illustrates the eigenvector for D established in Example 1. What is the relationship in the patterns of the dots that signifies an eigenvector relationship between the original signal and the transformed signal? Which other choices of the frequency, f, create a signal that is an eigenvector for D? What are the associated eigenvalues? In Figure 1, the graph on the left illustrates the sinusoidal signal with $f = 1$ and the graph on the right illustrates the sinusoidal signal with $f = 2$.

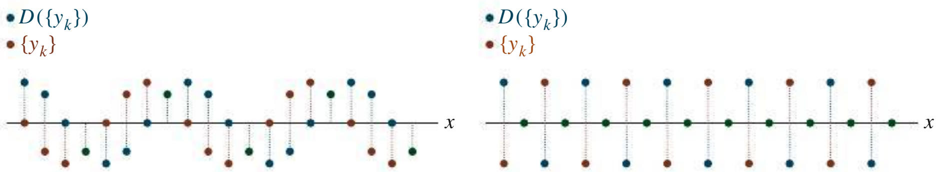


FIGURE 1

The Matrix of a Linear Transformation

There are branches of linear algebra that use infinite dimensional matrices to transform infinite dimensional vector spaces; however, in the remainder of this chapter we will restrict our study to linear transformations and matrices associated with finite dimensional vector spaces.

Let V be an n-dimensional vector space and let T be any linear transformation from V to V. To associate a matrix with T, choose any basis $\mathcal{B}$ for V. Given any $\mathbf{x}$ in V, the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ is in $\mathbb{R}^n$, as is the coordinate vector of its image, $[T(\mathbf{x})]_{\mathcal{B}}$.

The connection between $[\mathbf{x}]_{\mathcal{B}}$ and $[T(\mathbf{x})]_{\mathcal{B}}$ is easy to find. Let $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be the basis $\mathcal{B}$ for V. If $\mathbf{x} = r_1\mathbf{b}_1 + \cdots + r_n\mathbf{b}_n$, then

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

and

$$T(\mathbf{x}) = T(r_1\mathbf{b}_1 + \cdots + r_n\mathbf{b}_n) = r_1T(\mathbf{b}_1) + \cdots + r_nT(\mathbf{b}_n) \tag{1}$$

because T is linear. Now, since the coordinate mapping from V to $\mathbb{R}^n$ is linear (Theorem 8 in Section 4.4), equation (1) leads to

$$[T(\mathbf{x})]_{\mathcal{B}} = r_1[T(\mathbf{b}_1)]_{\mathcal{B}} + \cdots + r_n[T(\mathbf{b}_n)]_{\mathcal{B}} \tag{2}$$

Since $\mathcal{B}$-coordinate vectors are in $\mathbb{R}^n$, the vector equation (2) can be written as a matrix equation, namely

$$[T(\mathbf{x})]_{\mathcal{B}} = M[\mathbf{x}]_{\mathcal{B}} \tag{3}$$

where

$$M = [[T(\mathbf{b}_1)]_{\mathcal{B}} \ [T(\mathbf{b}_2)]_{\mathcal{B}} \ \cdots \ [T(\mathbf{b}_n)]_{\mathcal{B}}] \tag{4}$$

The matrix M is a matrix representation of T, called the **matrix for T relative to the basis** $\mathcal{B}$ and denoted by $[T]_{\mathcal{B}}$. See Figure 2.

Equation (3) says that, so far as coordinate vectors are concerned, the action of T on $\mathbf{x}$ may be viewed as left-multiplication by M.

$$\begin{tikzcd}
\mathbf{x} \arrow[r, "T"] \arrow[d] & T(\mathbf{x}) \arrow[d] \\
{[\mathbf{x}]_{\mathcal{B}}} \arrow[r, "\text{Multiplication}", "\text{by } M = [T]_{\mathcal{B}}"'] & {[T(\mathbf{x})]_{\mathcal{B}}}
\end{tikzcd}$$

FIGURE 2

EXAMPLE 2 Suppose $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ is a basis for V. Let $T : V \to V$ be a linear transformation with the property that

$$T(\mathbf{b}_1) = 3\mathbf{b}_1 - 2\mathbf{b}_2 \quad \text{and} \quad T(\mathbf{b}_2) = 4\mathbf{b}_1 + 7\mathbf{b}_2$$

Find the matrix M for T relative to $\mathcal{B}$.

SOLUTION The $\mathcal{B}$-coordinate vectors of the *images* of $\mathbf{b}_1$ and $\mathbf{b}_2$ are

$$[T(\mathbf{b}_1)]_{\mathcal{B}} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \text{and } [T(\mathbf{b}_2)]_{\mathcal{B}} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

Hence

$$M = \begin{bmatrix} 3 & 4 \\ -2 & 7 \end{bmatrix}$$

■

EXAMPLE 3 The mapping $T : \mathbb{P}_2 \to \mathbb{P}_2$ defined by

$$T\left(a_0 + a_1 t + a_2 t^2\right) = a_1 + 2a_2 t$$

is a linear transformation. (Calculus students will recognize T as the differentiation operator.)

a. Find the $\mathcal{B}$-matrix for T, when $\mathcal{B}$ is the basis $\{1, t, t^2\}$.

b. Verify that $[T(\mathbf{p})]_{\mathcal{B}} = [T]_{\mathcal{B}}[\mathbf{p}]_{\mathcal{B}}$ for each $\mathbf{p}$ in $\mathbb{P}_2$.

SOLUTION

a. Compute the images of the basis vectors:

$$\begin{aligned} T(1) &= 0 && \text{The zero polynomial} \\ T(t) &= 1 && \text{The polynomial whose value is always 1} \\ T\left(t^2\right) &= 2t && \end{aligned}$$

Then write the $\mathcal{B}$-coordinate vectors of $T(1)$, $T(t)$, and $T\left(t^2\right)$ (which are found by inspection in this example) and place them together as the $\mathcal{B}$-matrix for T:

$$[T(1)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, [T(t)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \left[T\left(t^2\right)\right]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

$$[T]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

b. For a general $\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2$

$$\begin{aligned} [T(\mathbf{p})]_{\mathcal{B}} &= [a_1 + 2a_2 t]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ 2a_2 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = [T]_{\mathcal{B}}[\mathbf{p}]_{\mathcal{B}} \end{aligned}$$

See Figure 3. ■

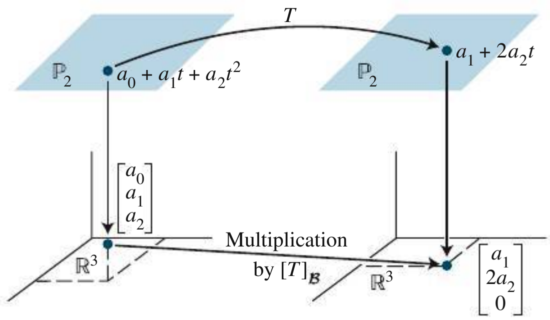


FIGURE 3 Matrix representation of a linear transformation.

Linear Transformations on $\mathbb{R}^n$

In an applied problem involving $\mathbb{R}^n$, a linear transformation T usually appears first as a matrix transformation, $\mathbf{x} \mapsto A\mathbf{x}$. If A is diagonalizable, then there is a basis $\mathcal{B}$ for $\mathbb{R}^n$ consisting of eigenvectors of A. Theorem 8 below shows that, in this case, the $\mathcal{B}$-matrix for T is diagonal. Diagonalizing A amounts to finding a diagonal matrix representation of $\mathbf{x} \mapsto A\mathbf{x}$.

THEOREM 8

Diagonal Matrix Representation

Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

PROOF Denote the columns of P by $\mathbf{b}_1, \dots, \mathbf{b}_n$, so that $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $P = [\,\mathbf{b}_1 \ \cdots \ \mathbf{b}_n\,]$. In this case, P is the change-of-coordinates matrix $P_{\mathcal{B}}$ discussed in Section 4.4, where

$$P[\,\mathbf{x}\,]_{\mathcal{B}} = \mathbf{x} \quad \text{and} \quad [\,\mathbf{x}\,]_{\mathcal{B}} = P^{-1}\mathbf{x}$$

If $T(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x}$ in $\mathbb{R}^n$, then

$$\begin{aligned}
[\,T\,]_{\mathcal{B}} &= \big[\,[\,T(\mathbf{b}_1)\,]_{\mathcal{B}} \ \ \cdots \ \ [\,T(\mathbf{b}_n)\,]_{\mathcal{B}}\,\big] && \text{Definition of } [\,T\,]_{\mathcal{B}} \\
&= \big[\,[\,A\mathbf{b}_1\,]_{\mathcal{B}} \ \ \cdots \ \ [\,A\mathbf{b}_n\,]_{\mathcal{B}}\,\big] && \text{Since } T(\mathbf{x}) = A\mathbf{x} \\
&= [\,P^{-1}A\mathbf{b}_1 \ \ \cdots \ \ P^{-1}A\mathbf{b}_n\,] && \text{Change of coordinates} \\
&= P^{-1}A[\,\mathbf{b}_1 \ \ \cdots \ \ \mathbf{b}_n\,] && \text{Matrix multiplication} \\
&= P^{-1}AP && (6)
\end{aligned}$$

Since $A = PDP^{-1}$, we have $[\,T\,]_{\mathcal{B}} = P^{-1}AP = D$. ■

EXAMPLE 4 Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$, where $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a basis $\mathcal{B}$ for $\mathbb{R}^2$ with the property that the $\mathcal{B}$-matrix for T is a diagonal matrix.

SOLUTION From Example 2 in Section 5.3, we know that $A = PDP^{-1}$, where

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

The columns of P, call them $\mathbf{b}_1$ and $\mathbf{b}_2$, are eigenvectors of A. By Theorem 8, D is the $\mathcal{B}$-matrix for T when $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. The mappings $\mathbf{x} \mapsto A\mathbf{x}$ and $\mathbf{u} \mapsto D\mathbf{u}$ describe the same linear transformation, relative to different bases. ■

Similarity of Matrix Representations

The proof of Theorem 8 did not use the information that D was diagonal. Hence, if A is similar to a matrix C, with $A = PCP^{-1}$, then C is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$ when the basis $\mathcal{B}$ is formed from the columns of P. The factorization $A = PCP^{-1}$ is shown in Figure 4.

$$\begin{tikzcd}
\mathbf{x} \arrow[r, "\text{Multiplication}", "\text{by } A"'] \arrow[d, "\text{Multiplication}"', "\text{by } P^{-1}"] & A\mathbf{x} \\
{[\mathbf{x}]_{\mathcal{B}}} \arrow[r, "\text{Multiplication}", "\text{by } C"'] & {[A\mathbf{x}]_{\mathcal{B}}} \arrow[u, "\text{Multiplication}"', "\text{by } P"]
\end{tikzcd}$$

FIGURE 4 Similarity of two matrix representations: $A = PCP^{-1}$.

Conversely, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $T(\mathbf{x}) = A\mathbf{x}$, and if $\mathcal{B}$ is any basis for $\mathbb{R}^n$, then the $\mathcal{B}$-matrix for T is similar to A. In fact, the calculations in the proof of Theorem 8 show that if P is the matrix whose columns come from the vectors in $\mathcal{B}$, then $[T]_{\mathcal{B}} = P^{-1}AP$. This important connection between the matrix of a linear transformation and similar matrices is highlighted here.

> The set of all matrices similar to a matrix A coincides with the set of all matrix representations of the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

EXAMPLE 5 Let $A = \begin{bmatrix} 4 & -9 \\ 4 & -8 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, and $\mathbf{b}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. The characteristic polynomial of A is $(\lambda + 2)^2$, but the eigenspace for the eigenvalue -2 is only one-dimensional; so A is not diagonalizable. However, the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ has the property that the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is a triangular matrix called the *Jordan form* of A.[1] Find this $\mathcal{B}$-matrix.

SOLUTION If $P = [\,\mathbf{b}_1 \quad \mathbf{b}_2\,]$, then the $\mathcal{B}$-matrix is $P^{-1}AP$. Compute

$$AP = \begin{bmatrix} 4 & -9 \\ 4 & -8 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} -6 & -1 \\ -4 & 0 \end{bmatrix}$$

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

Notice that the eigenvalue of A is on the diagonal. ■

[1] Every square matrix A is similar to a matrix in Jordan form. The basis used to produce a Jordan form consists of eigenvectors and so-called "generalized eigenvectors" of A. See Chapter 9 of *Applied Linear Algebra*, 3rd ed. (Englewood Cliffs, NJ: Prentice-Hall, 1988), by B. Noble and J. W. Daniel.

Numerical Notes

An efficient way to compute a $\mathcal{B}$-matrix $P^{-1}AP$ is to compute AP and then to row reduce the augmented matrix $[\,P \quad AP\,]$ to $[\,I \quad P^{-1}AP\,]$. A separate computation of P^{-1} is unnecessary. See Exercise 22 in Section 2.2.

Practice Problems

1. Find $T(a_0 + a_1t + a_2t^2)$, if T is the linear transformation from $\mathbb{P}_2$ to $\mathbb{P}_2$ whose matrix relative to $\mathcal{B} = \{1, t, t^2\}$ is

$$[\,T\,]_{\mathcal{B}} = \begin{bmatrix} 3 & 4 & 0 \\ 0 & 5 & -1 \\ 1 & -2 & 7 \end{bmatrix}$$

2. Let A, B, and C be $n \times n$ matrices. The text has shown that if A is similar to B, then B is similar to A. This property, together with the statements below, shows that "similar to" is an *equivalence relation*. (Row equivalence is another example of an equivalence relation.) Verify parts (a) and (b).
   a. A is similar to A.
   b. If A is similar to B and B is similar to C, then A is similar to C.

5.4 Exercises

1. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be a basis for the vector space V. Let $T : V \to V$ be a linear transformation with the property that

$$T(\mathbf{b}_1) = 3\mathbf{b}_1 - 5\mathbf{b}_2,\ T(\mathbf{b}_2) = -\mathbf{b}_1 + 6\mathbf{b}_2,\ T(\mathbf{b}_3) = 4\mathbf{b}_2$$

Find $[T]_{\mathcal{B}}$, the matrix for T relative to $\mathcal{B}$.

2. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ be a basis for vector space V. Let $T : V \to V$ be a linear transformation with the property that

$$T(\mathbf{b}_1) = 2\mathbf{b}_1 - 3\mathbf{b}_2,\ T(\mathbf{b}_2) = -4\mathbf{b}_1 + 5\mathbf{b}_2$$

Find $[T]_{\mathcal{B}}$, the matrix for T relative to $\mathcal{B}$.

3. Assume the mapping $T : \mathbb{P}_2 \to \mathbb{P}_2$ defined by

$$T\left(a_0 + a_1t + a_2t^2\right) = 3a_0 + (5a_0 - 2a_1)t + (4a_1 + a_2)t^2$$

is linear. Find the matrix representation of T relative to the basis $\mathcal{B} = \{1, t, t^2\}$.

4. Define $T : \mathbb{P}_2 \to \mathbb{P}_2$ by $T(\mathbf{p}) = \mathbf{p}(0) - \mathbf{p}(1)t + \mathbf{p}(2)t^2$.
   a. Show that T is a linear transformation.
   b. Find $T(\mathbf{p})$ when $\mathbf{p}(t) = -2 + t$. Is $\mathbf{p}$ an eigenvector of T?
   c. Find the matrix for T relative to the basis $\{1, t, t^2\}$ for $\mathbb{P}_2$.

5. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be a basis for a vector space V. Find $T(3\mathbf{b}_1 - 4\mathbf{b}_2)$ when T is a linear transformation from V to V whose matrix relative to $\mathcal{B}$ is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 0 & -6 & 1 \\ 0 & 5 & -1 \\ 1 & -2 & 7 \end{bmatrix}$$

6. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be a basis for a vector space V. Find $T(2\mathbf{b}_1 - \mathbf{b}_2 + 4\mathbf{b}_3)$ when T is a linear transformation from V to V whose matrix relative to $\mathcal{B}$ is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 0 & -6 & 1 \\ 0 & 5 & -1 \\ 1 & -2 & 7 \end{bmatrix}$$

In Exercises 7 and 8, find the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$, when $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$.

7. $A = \begin{bmatrix} 3 & 4 \\ -1 & -1 \end{bmatrix}, \mathbf{b}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

8. $A = \begin{bmatrix} -1 & 4 \\ -2 & 3 \end{bmatrix}, \mathbf{b}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

In Exercises 9–12, define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$. Find a basis $\mathcal{B}$ for $\mathbb{R}^2$ with the property that $[T]_{\mathcal{B}}$ is diagonal.

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

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

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

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

13. Let $A = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}$ and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$, for $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$.

a. Verify that $\mathbf{b}_1$ is an eigenvector of A but A is not diagonalizable.

b. Find the $\mathcal{B}$ matrix for T.

14. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$, where A is a 3×3 matrix with eigenvalues 5 and -2. Does there exist a basis $\mathcal{B}$ for $\mathbb{R}^3$ such that the $\mathcal{B}$-matrix for T is a diagonal matrix? Discuss.

15. Define $T : \mathbb{P}_2 \to \mathbb{P}_2$ by $T(\mathbf{p}) = \mathbf{p}(1) + \mathbf{p}(1)t + \mathbf{p}(1)t^2$.

a. Find $T(\mathbf{p})$ when $\mathbf{p}(t) = 1 + t + t^2$. Is $\mathbf{p}$ an eigenvector of T? If $\mathbf{p}$ is an eigenvector, what is its eigenvalue?

b. Find $T(\mathbf{p})$ when $\mathbf{p}(t) = -2 + t$. Is $\mathbf{p}$ an eigenvector of T? If $\mathbf{p}$ is an eigenvector, what is its eigenvalue?

16. Define $T : \mathbb{P}_3 \to \mathbb{P}_3$ by $T(\mathbf{p}) = \mathbf{p}(0) - \mathbf{p}(1)t - \mathbf{p}(1)t^2 + \mathbf{p}(0)t^3$.

a. Find $T(\mathbf{p})$ when $\mathbf{p}(t) = 1 + t + t^2 + t^3$. Is $\mathbf{p}$ an eigenvector of T? If $\mathbf{p}$ is an eigenvector, what is its eigenvalue?

b. Find $T(\mathbf{p})$ when $\mathbf{p}(t) = t + t^2$. Is $\mathbf{p}$ an eigenvector of T? If $\mathbf{p}$ is an eigenvector, what is its eigenvalue?

In Exercises 17 through 20, mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** Similar matrices have the same eigenvalues.

18. **(T/F)** Similar matrices have the same eigenvectors.

19. **(T/F)** Only linear transformations on finite vectors spaces have eigenvectors.

20. **(T/F)** If there is a nonzero vector in the kernel of a linear transformation T, then 0 is an eigenvalue of T.

Verify the statements in Exercises 21–28 by providing justification for each statement. In each case, the matrices are square.

21. If A is invertible and similar to B, then B is invertible and A^{-1} is similar to B^{-1}. [*Hint:* $P^{-1}AP = B$ for some invertible P. Explain why B is invertible. Then find an invertible Q such that $Q^{-1}A^{-1}Q = B^{-1}$.]

22. If A is similar to B, then A^2 is similar to B^2.

23. If B is similar to A and C is similar to A, then B is similar to C.

24. If A is diagonalizable and B is similar to A, then B is also diagonalizable.

25. If $B = P^{-1}AP$ and $\mathbf{x}$ is an eigenvector of A corresponding to an eigenvalue λ, then $P^{-1}\mathbf{x}$ is an eigenvector of B corresponding also to λ.

26. If A and B are similar, then they have the same rank. [*Hint:* Refer to Supplementary Exercises 31 and 32 for Chapter 4.]

27. The *trace* of a square matrix A is the sum of the diagonal entries in A and is denoted by tr A. It can be verified that $\text{tr}(FG) = \text{tr}(GF)$ for any two $n \times n$ matrices F and G. Show that if A and B are similar, then tr A = tr B.

28. It can be shown that the trace of a matrix A equals the sum of the eigenvalues of A. Verify this statement for the case when A is diagonalizable.

Exercises 29–32 refer to the vector space of signals, $\mathbb{S}$, from Section 4.7. The shift transformation, $S(\{y_k\}) = \{y_{k-1}\}$, shifts each entry in the signal one position to the right. The moving average transformation, $M_2(\{y_k\}) = \left\{\dfrac{y_k + y_{k-1}}{2}\right\}$, creates a new signal by averaging two consecutive terms in the given signal. The constant signal of all ones is given by $\chi = \{1^k\}$ and the alternating signal by $\alpha = \{(-1)^k\}$.

29. Show that χ is an eigenvector of the shift transformation S. What is the associated eigenvalue?

30. Show that α is an eigenvector of the shift transformation S. What is the associated eigenvalue?

31. Show that α is an eigenvector of the moving average transformation M_2. What is the associated eigenvalue?

32. Show that χ is an eigenvector of the moving average transformation M_2. What is the associated eigenvalue?

In Exercises 33 and 34, find the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$ when $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$.

T **33.** $A = \begin{bmatrix} -14 & 4 & -14 \\ -33 & 9 & -31 \\ 11 & -4 & 11 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$,
$\mathbf{b}_2 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} -1 \\ -2 \\ 0 \end{bmatrix}$

T **34.** $A = \begin{bmatrix} -7 & -48 & -16 \\ 1 & 14 & 6 \\ -3 & -45 & -19 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 1 \\ -3 \end{bmatrix}$,
$\mathbf{b}_2 = \begin{bmatrix} -2 \\ 1 \\ -3 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$

T **35.** Let T be the transformation whose standard matrix is given below. Find a basis for $\mathbb{R}^4$ with the property that $[T]_{\mathcal{B}}$ is diagonal.

$$A = \begin{bmatrix} 15 & -66 & -44 & -33 \\ 0 & 13 & 21 & -15 \\ 1 & -15 & -21 & 12 \\ 2 & -18 & -22 & 8 \end{bmatrix}$$

Solutions to Practice Problems

1. Let $\mathbf{p}(t) = a_0 + a_1t + a_2t^2$ and compute

$$[T(\mathbf{p})]_{\mathcal{B}} = [T]_{\mathcal{B}}[\mathbf{p}]_{\mathcal{B}} = \begin{bmatrix} 3 & 4 & 0 \\ 0 & 5 & -1 \\ 1 & -2 & 7 \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 3a_0 + 4a_1 \\ 5a_1 - a_2 \\ a_0 - 2a_1 + 7a_2 \end{bmatrix}$$

So $T(\mathbf{p}) = (3a_0 + 4a_1) + (5a_1 - a_2)t + (a_0 - 2a_1 + 7a_2)t^2$.

2. a. $A = (I)^{-1}AI$, so A is similar to A.

b. By hypothesis, there exist invertible matrices P and Q with the property that $B = P^{-1}AP$ and $C = Q^{-1}BQ$. Substitute the formula for B into the formula for C, and use a fact about the inverse of a product:

$$C = Q^{-1}BQ = Q^{-1}(P^{-1}AP)Q = (PQ)^{-1}A(PQ)$$

This equation has the proper form to show that A is similar to C.

5.5 Complex Eigenvalues

Since the characteristic equation of an $n \times n$ matrix involves a polynomial of degree n, the equation always has exactly n roots, counting multiplicities, *provided that possibly complex roots are included*. This section shows that if the characteristic equation of a real matrix A has some complex roots, then these roots provide critical information about A. The key is to let A act on the space $\mathbb{C}^n$ of n-tuples of complex numbers.[1]

Our interest in $\mathbb{C}^n$ does not arise from a desire to "generalize" the results of the earlier chapters, although that would in fact open up significant new applications of linear algebra.[2] Rather, this study of complex eigenvalues is essential in order to uncover "hidden" information about certain matrices with real entries that arise in a variety of real-life problems. Such problems include many real dynamical systems that involve periodic motion, vibration, or some type of rotation in space.

The matrix eigenvalue–eigenvector theory already developed for $\mathbb{R}^n$ applies equally well to $\mathbb{C}^n$. So a complex scalar λ satisfies $\det(A - \lambda I) = 0$ if and only if there is a nonzero vector $\mathbf{x}$ in $\mathbb{C}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. We call λ a (**complex**) **eigenvalue** and $\mathbf{x}$ a (**complex**) **eigenvector** corresponding to λ.

EXAMPLE 1 If $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ on $\mathbb{R}^2$ rotates the plane counterclockwise through a quarter-turn. The action of A is periodic, since after four quarter-turns, a vector is back where it started. Obviously, no nonzero vector is mapped into a multiple of itself, so A has no eigenvectors in $\mathbb{R}^2$ and hence no real eigenvalues. In fact, the characteristic equation of A is

$$\lambda^2 + 1 = 0$$

[1] Refer to Appendix B for a brief discussion of complex numbers. Matrix algebra and concepts about real vector spaces carry over to the case with complex entries and scalars. In particular, $A(c\mathbf{x} + d\mathbf{y}) = cA\mathbf{x} + dA\mathbf{y}$, for A an $m \times n$ matrix with complex entries, $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{C}^n$, and c, d in $\mathbb{C}$.

[2] A second course in linear algebra often discusses such topics. They are of particular importance in electrical engineering.

The only roots are complex: $\lambda = i$ and $\lambda = -i$. However, if we permit A to act on $\mathbb{C}^2$, then

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

Thus i and $-i$ are eigenvalues, with $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ and $\begin{bmatrix} 1 \\ i \end{bmatrix}$ as corresponding eigenvectors. (A method for *finding* complex eigenvectors is discussed in Example 2.) ■

The main focus of this section will be on the matrix in the next example.

EXAMPLE 2 Let $A = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}$. Find the eigenvalues of A, and find a basis for each eigenspace.

SOLUTION The characteristic equation of A is

$$0 = \det\begin{bmatrix} .5-\lambda & -.6 \\ .75 & 1.1-\lambda \end{bmatrix} = (.5-\lambda)(1.1-\lambda) - (-.6)(.75)$$
$$= \lambda^2 - 1.6\lambda + 1$$

From the quadratic formula, $\lambda = \frac{1}{2}[1.6 \pm \sqrt{(-1.6)^2 - 4}] = .8 \pm .6i$. For the eigenvalue $\lambda = .8 - .6i$, construct

$$A - (.8 - .6i)I = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix} - \begin{bmatrix} .8-.6i & 0 \\ 0 & .8-.6i \end{bmatrix}$$
$$= \begin{bmatrix} -.3+.6i & -.6 \\ .75 & .3+.6i \end{bmatrix} \qquad (1)$$

Row reduction of the usual augmented matrix is quite unpleasant by hand because of the complex arithmetic. However, here is a nice observation that really simplifies matters: Since $.8 - .6i$ is an eigenvalue, the system

$$\begin{aligned} (-.3+.6i)x_1 - \quad\quad .6x_2 &= 0 \\ .75x_1 + (.3+.6i)x_2 &= 0 \end{aligned} \qquad (2)$$

has a nontrivial solution (with x_1 and x_2 possibly complex numbers). Therefore, *both equations in (2) determine the same relationship between x_1 and x_2, and either equation can be used to express one variable in terms of the other.*[3]

The second equation in (2) leads to

$$\begin{aligned} .75x_1 &= (-.3-.6i)x_2 \\ x_1 &= (-.4-.8i)x_2 \end{aligned}$$

Choose $x_2 = 5$ to eliminate the decimals, and obtain $x_1 = -2 - 4i$. A basis for the eigenspace corresponding to $\lambda = .8 - .6i$ is

$$\mathbf{v}_1 = \begin{bmatrix} -2-4i \\ 5 \end{bmatrix}$$

[3] Another way to see this is to realize that the matrix in equation (1) is not invertible, so its rows are linearly dependent (as vectors in $\mathbb{C}^2$), and hence one row is a (complex) multiple of the other.

Analogous calculations for $\lambda = .8 + .6i$ produce the eigenvector

$$\mathbf{v}_2 = \begin{bmatrix} -2 + 4i \\ 5 \end{bmatrix}$$

As a check on the work, compute

$$A\mathbf{v}_2 = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix} \begin{bmatrix} -2 + 4i \\ 5 \end{bmatrix} = \begin{bmatrix} -4 + 2i \\ 4 + 3i \end{bmatrix} = (.8 + .6i)\mathbf{v}_2 \qquad \blacksquare$$

Surprisingly, the matrix A in Example 2 determines a transformation $\mathbf{x} \mapsto A\mathbf{x}$ that is essentially a rotation. This fact becomes evident when appropriate points are plotted, as illustrated in Figure 1.

EXAMPLE 3 One way to see how multiplication by the matrix A in Example 2 affects points is to plot an arbitrary initial point—say, $\mathbf{x}_0 = (2, 0)$—and then to plot successive images of this point under repeated multiplications by A. That is, plot

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} \\ \mathbf{x}_2 &= A\mathbf{x}_1 = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix} \begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} = \begin{bmatrix} -.4 \\ 2.4 \end{bmatrix} \\ \mathbf{x}_3 &= A\mathbf{x}_2, \ldots \end{aligned}$$

Figure 1 shows $\mathbf{x}_0, \ldots, \mathbf{x}_8$ as larger dots. The smaller dots are the locations of $\mathbf{x}_9, \ldots, \mathbf{x}_{100}$. The sequence lies along an elliptical orbit. $\blacksquare$

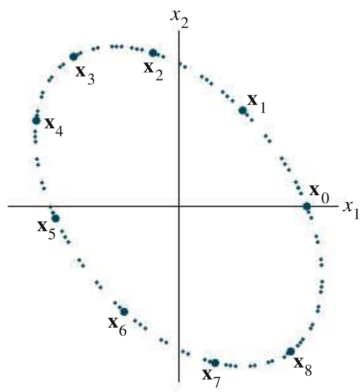


FIGURE 1 Iterates of a point $\mathbf{x}_0$ under the action of a matrix with a complex eigenvalue.

Of course, Figure 1 does not explain *why* the rotation occurs. The secret to the rotation is hidden in the real and imaginary parts of a complex eigenvector.

Real and Imaginary Parts of Vectors

The complex conjugate of a complex vector $\mathbf{x}$ in $\mathbb{C}^n$ is the vector $\overline{\mathbf{x}}$ in $\mathbb{C}^n$ whose entries are the complex conjugates of the entries in $\mathbf{x}$. The **real** and **imaginary parts** of a complex vector $\mathbf{x}$ are the vectors Re $\mathbf{x}$ and Im $\mathbf{x}$ in $\mathbb{R}^n$ formed from the real and imaginary parts of the entries of $\mathbf{x}$. Thus,

$$\mathbf{x} = \text{Re}\,\mathbf{x} + i\text{Im}\,\mathbf{x} \tag{3}$$

EXAMPLE 4 If $\mathbf{x} = \begin{bmatrix} 3-i \\ i \\ 2+5i \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + i\begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix}$, then

$$\text{Re}\,\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}, \quad \text{Im}\,\mathbf{x} = \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix}, \quad \text{and} \quad \overline{\mathbf{x}} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} - i\begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 3+i \\ -i \\ 2-5i \end{bmatrix} \qquad \blacksquare$$

If B is an $m \times n$ matrix with possibly complex entries, then $\overline{B}$ denotes the matrix whose entries are the complex conjugates of the entries in B. Let r be a complex number and $\mathbf{x}$ any vector. Properties of conjugates for complex numbers carry over to complex matrix algebra:

$$\overline{r\mathbf{x}} = \overline{r}\,\overline{\mathbf{x}}, \quad \overline{B\mathbf{x}} = \overline{B}\,\overline{\mathbf{x}}, \quad \overline{BC} = \overline{B}\,\overline{C}, \quad \text{and} \quad \overline{rB} = \overline{r}\,\overline{B}$$

Eigenvalues and Eigenvectors of a Real Matrix That Acts on $\mathbb{C}^n$

Let A be an $n \times n$ matrix whose entries are real. Then $\overline{A\mathbf{x}} = \overline{A}\overline{\mathbf{x}} = A\overline{\mathbf{x}}$. If λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector in $\mathbb{C}^n$, then

$$A\overline{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

Hence $\overline{\lambda}$ is also an eigenvalue of A, with $\overline{\mathbf{x}}$ a corresponding eigenvector. This shows that *when A is real, its complex eigenvalues occur in conjugate pairs*. (Here and elsewhere, we use the term *complex eigenvalue* to refer to an eigenvalue $\lambda = a + bi$, with $b \neq 0$.)

EXAMPLE 5 The eigenvalues of the real matrix in Example 2 are complex conjugates, namely $.8 - .6i$ and $.8 + .6i$. The corresponding eigenvectors found in Example 2 are also conjugates:

$$\mathbf{v}_1 = \begin{bmatrix} -2-4i \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -2+4i \\ 5 \end{bmatrix} = \overline{\mathbf{v}}_1 \qquad \blacksquare$$

The next example provides the basic "building block" for all real 2×2 matrices with complex eigenvalues.

EXAMPLE 6 If $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, where a and b are real and not both zero, then the eigenvalues of C are $\lambda = a \pm bi$. (See the Practice Problem at the end of this section.) Also, if $r = |\lambda| = \sqrt{a^2+b^2}$, then

$$C = r\begin{bmatrix} a/r & -b/r \\ b/r & a/r \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

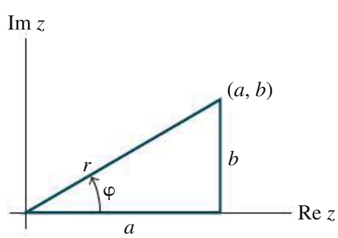


FIGURE 2

where φ is the angle between the positive x-axis and the ray from $(0,0)$ through (a,b). See Figure 2 and Appendix B. The angle φ is called the *argument* of $\lambda = a + bi$. Thus the transformation $\mathbf{x} \mapsto C\mathbf{x}$ may be viewed as the composition of a rotation through the angle φ and a scaling by $|\lambda|$ (see Figure 3). ■

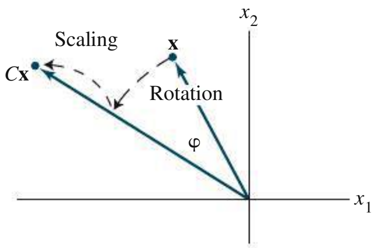


FIGURE 3 A rotation followed by a scaling.

Finally, we are ready to uncover the rotation that is hidden within a real matrix having a complex eigenvalue.

EXAMPLE 7 Let $A = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}$, $\lambda = .8 - .6i$, and $\mathbf{v}_1 = \begin{bmatrix} -2-4i \\ 5 \end{bmatrix}$, as in Example 2. Also, let P be the 2×2 real matrix, described in Theorem 9,

$$P = \begin{bmatrix} \operatorname{Re} \mathbf{v}_1 & \operatorname{Im} \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} -2 & -4 \\ 5 & 0 \end{bmatrix}$$

and let

$$C = P^{-1}AP = \frac{1}{20}\begin{bmatrix} 0 & 4 \\ -5 & -2 \end{bmatrix}\begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}\begin{bmatrix} -2 & -4 \\ 5 & 0 \end{bmatrix} = \begin{bmatrix} .8 & -.6 \\ .6 & .8 \end{bmatrix}$$

By Example 6, C is a pure rotation because $|\lambda|^2 = (.8)^2 + (.6)^2 = 1$. From $C = P^{-1}AP$, we obtain

$$A = PCP^{-1} = P\begin{bmatrix} .8 & -.6 \\ .6 & .8 \end{bmatrix}P^{-1}$$

Here is the rotation "inside" A! The matrix P provides a change of variable, say, $\mathbf{x} = P\mathbf{u}$. The action of A amounts to a change of variable from $\mathbf{x}$ to $\mathbf{u}$, followed by a rotation, and then a return to the original variable. See Figure 4. The rotation produces an ellipse, as in Figure 1, instead of a circle, because the coordinate system determined by the columns of P is not rectangular and does not have equal unit lengths on the two axes. ■

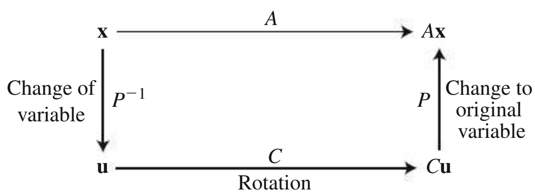


FIGURE 4 Rotation due to a complex eigenvalue.

The next theorem shows that the calculations in Example 7 can be carried out for any 2×2 real matrix A having a complex eigenvalue λ. The proof uses the fact that if the entries in A are real, then $A(\operatorname{Re} \mathbf{x}) = \operatorname{Re}(A\mathbf{x})$ and $A(\operatorname{Im} \mathbf{x}) = \operatorname{Im}(A\mathbf{x})$, and if $\mathbf{x}$ is an eigenvector for a complex eigenvalue, then $\operatorname{Re} \mathbf{x}$ and $\operatorname{Im} \mathbf{x}$ are linearly independent in $\mathbb{R}^2$. (See Exercises 29 and 30.) The details are omitted.

THEOREM 9

Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ $(b \neq 0)$ and an associated eigenvector $\mathbf{v}$ in $\mathbb{C}^2$. Then

$$A = PCP^{-1}, \quad \text{where} \quad P = [\,\text{Re}\,\mathbf{v} \quad \text{Im}\,\mathbf{v}\,] \quad \text{and} \quad C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

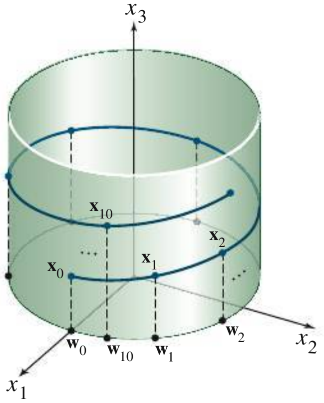


FIGURE 5

Iterates of two points under the action of a 3×3 matrix with a complex eigenvalue.

The phenomenon displayed in Example 7 persists in higher dimensions. For instance, if A is a 3×3 matrix with a complex eigenvalue, then there is a plane in $\mathbb{R}^3$ on which A acts as a rotation (possibly combined with scaling). Every vector in that plane is rotated into another point on the same plane. We say that the plane is **invariant** under A.

EXAMPLE 8 The matrix $A = \begin{bmatrix} .8 & -.6 & 0 \\ .6 & .8 & 0 \\ 0 & 0 & 1.07 \end{bmatrix}$ has eigenvalues $.8 \pm .6i$ and 1.07. Any vector $\mathbf{w}_0$ in the x_1x_2-plane (with third coordinate 0) is rotated by A into another point in the plane. Any vector $\mathbf{x}_0$ not in the plane has its x_3-coordinate multiplied by 1.07. The iterates of the points $\mathbf{w}_0 = (2, 0, 0)$ and $\mathbf{x}_0 = (2, 0, 1)$ under multiplication by A are shown in Figure 5. ■

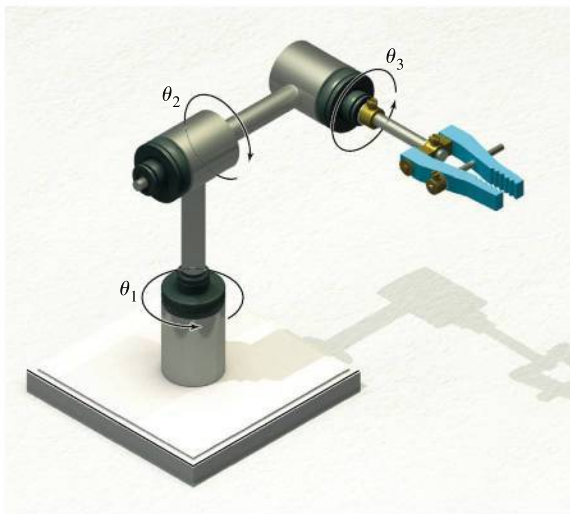


FIGURE 6

EXAMPLE 9 Many robots work by rotating at various joints, just as matrices with complex eigenvalues rotate points in space. Figure 6 illustrates a robot arm made using linear transformations, each with a pair of complex eigenvalues. In Project C, at the end of the chapter, you will be asked to find videos of robots on the web that use rotations as a key element of their functioning.

Practice Problem

Show that if a and b are real, then the eigenvalues of $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ are $a \pm bi$, with corresponding eigenvectors $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ and $\begin{bmatrix} 1 \\ i \end{bmatrix}$.

5.5 Exercises

Let each matrix in Exercises 1–6 act on $\mathbb{C}^2$. Find the eigenvalues and a basis for each eigenspace in $\mathbb{C}^2$.

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

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

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

In Exercises 7–12, use Example 6 to list the eigenvalues of A. In each case, the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is the composition of a rotation and a scaling. Give the angle φ of the rotation, where $-\pi < \varphi \le \pi$, and give the scale factor r.

7. $\begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$ **8.** $\begin{bmatrix} \sqrt{3} & 3 \\ -3 & \sqrt{3} \end{bmatrix}$

9. $\begin{bmatrix} -\sqrt{3}/2 & 1/2 \\ -1/2 & -\sqrt{3}/2 \end{bmatrix}$ **10.** $\begin{bmatrix} 3 & 3 \\ -3 & 3 \end{bmatrix}$

11. $\begin{bmatrix} .1 & .1 \\ -.1 & .1 \end{bmatrix}$ **12.** $\begin{bmatrix} 0 & 4 \\ -4 & 0 \end{bmatrix}$

In Exercises 13–20, find an invertible matrix P and a matrix C of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ such that the given matrix has the form $A = PCP^{-1}$. For Exercises 13–16, use information from Exercises 1–4.

13. $\begin{bmatrix} 1 & -2 \\ 1 & 3 \end{bmatrix}$ **14.** $\begin{bmatrix} -1 & -1 \\ 5 & -5 \end{bmatrix}$

15. $\begin{bmatrix} 1 & 5 \\ -2 & 3 \end{bmatrix}$ **16.** $\begin{bmatrix} -3 & -1 \\ 2 & -5 \end{bmatrix}$

17. $\begin{bmatrix} 1 & -.8 \\ 4 & -2.2 \end{bmatrix}$ **18.** $\begin{bmatrix} 1 & -1 \\ .4 & .6 \end{bmatrix}$

19. $\begin{bmatrix} 1.52 & -.7 \\ .56 & .4 \end{bmatrix}$ **20.** $\begin{bmatrix} -1.64 & -2.4 \\ 1.92 & 2.2 \end{bmatrix}$

21. In Example 2, solve the first equation in (2) for x_2 in terms of x_1, and from that produce the eigenvector $\mathbf{y} = \begin{bmatrix} 2 \\ -1+2i \end{bmatrix}$ for the matrix A. Show that this $\mathbf{y}$ is a (complex) multiple of the vector $\mathbf{v}_1$ used in Example 2.

22. Let A be a complex (or real) $n \times n$ matrix, and let $\mathbf{x}$ in $\mathbb{C}^n$ be an eigenvector corresponding to an eigenvalue λ in $\mathbb{C}$. Show that for each nonzero complex scalar μ, the vector $\mu\mathbf{x}$ is an eigenvector of A.

In Exercises 23–26, A is a 2×2 matrix with real entries, and $\mathbf{x}$ is a vector in $\mathbb{R}^2$. Mark each statement True or False **(T/F)**. Justify each answer.

23. (T/F) The matrix A can have one real and one complex eigenvalue.

24. (T/F) The points $A\mathbf{x}$, $A^2\mathbf{x}$, $A^3\mathbf{x}, \ldots$ always lie on the same circle.

25. (T/F) The matrix A always has two eigenvalues, but sometimes they have algebraic multiplicity 2 or are complex numbers.

26. (T/F) If the matrix A has two complex eigenvalues, then it also has two linearly independent real eigenvectors.

Chapter 7 will focus on matrices A with the property that $A^T = A$. Exercises 27 and 28 show that every eigenvalue of such a matrix is necessarily real.

27. Let A be an $n \times n$ real matrix with the property that $A^T = A$, let $\mathbf{x}$ be any vector in $\mathbb{C}^n$, and let $q = \bar{\mathbf{x}}^T A\mathbf{x}$. The equalities below show that q is a real number by verifying that $\bar{q} = q$. Give a reason for each step.

$$\bar{q} \underset{\text{(a)}}{=} \overline{\bar{\mathbf{x}}^T A\mathbf{x}} \underset{\text{(b)}}{=} \mathbf{x}^T \overline{A\mathbf{x}} \underset{\text{(c)}}{=} \mathbf{x}^T A\bar{\mathbf{x}} \underset{\text{(d)}}{=} (\mathbf{x}^T A\bar{\mathbf{x}})^T \underset{\text{(e)}}{=} \bar{\mathbf{x}}^T A^T \mathbf{x} = q$$

28. Let A be an $n \times n$ real matrix with the property that $A^T = A$. Show that if $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero vector $\mathbf{x}$ in $\mathbb{C}^n$, then, in fact, λ is real and the real part of $\mathbf{x}$ is an eigenvector of A. [*Hint:* Compute $\bar{\mathbf{x}}^T A\mathbf{x}$, and use Exercise 27. Also, examine the real and imaginary parts of $A\mathbf{x}$.]

29. Let A be a real $n \times n$ matrix, and let $\mathbf{x}$ be a vector in $\mathbb{C}^n$. Show that $\text{Re}(A\mathbf{x}) = A(\text{Re}\,\mathbf{x})$ and $\text{Im}(A\mathbf{x}) = A(\text{Im}\,\mathbf{x})$.

30. Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ $(b \ne 0)$ and an associated eigenvector $\mathbf{v}$ in $\mathbb{C}^2$.

a. Show that $A(\text{Re}\,\mathbf{v}) = a\,\text{Re}\,\mathbf{v} + b\,\text{Im}\,\mathbf{v}$ and $A(\text{Im}\,\mathbf{v}) = -b\,\text{Re}\,\mathbf{v} + a\,\text{Im}\,\mathbf{v}$. [*Hint:* Write $\mathbf{v} = \text{Re}\,\mathbf{v} + i\,\text{Im}\,\mathbf{v}$, and compute $A\mathbf{v}$.]

b. Verify that if P and C are given as in Theorem 9, then $AP = PC$.

T In Exercises 31 and 32, find a factorization of the given matrix A in the form $A = PCP^{-1}$, where C is a block-diagonal matrix with 2×2 blocks of the form shown in Example 6. (For each conjugate pair of eigenvalues, use the real and imaginary parts of one eigenvector in $\mathbb{C}^4$ to create two columns of P.)

31. $\begin{bmatrix} .7 & 1.1 & 2.0 & 1.7 \\ -2.0 & -4.0 & -8.6 & -7.4 \\ 0 & -.5 & -1.0 & -1.0 \\ 1.0 & 2.8 & 6.0 & 5.3 \end{bmatrix}$

32. $\begin{bmatrix} -1.4 & -2.0 & -2.0 & -2.0 \\ -1.3 & -.8 & -.1 & -.6 \\ .3 & -1.9 & -1.6 & -1.4 \\ 2.0 & 3.3 & 2.3 & 2.6 \end{bmatrix}$

Solution to Practice Problem

Remember that it is easy to test whether a vector is an eigenvector. There is no need to examine the characteristic equation. Compute

$$A\mathbf{x} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} a+bi \\ b-ai \end{bmatrix} = (a+bi)\begin{bmatrix} 1 \\ -i \end{bmatrix}$$

Thus $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ is an eigenvector corresponding to $\lambda = a + bi$. From the discussion in this section, $\begin{bmatrix} 1 \\ i \end{bmatrix}$ must be an eigenvector corresponding to $\overline{\lambda} = a - bi$.

5.6 Discrete Dynamical Systems

Eigenvalues and eigenvectors provide the key to understanding the long-term behavior, or *evolution*, of a dynamical system described by a difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$. Such an equation was used to model population movement in Section 1.10, and will be used in various Markov chains in Section 5.9 and the spotted owl population in the introductory example for this chapter. The vectors $\mathbf{x}_k$ give information about the system as time (denoted by k) passes, where k is a nonnegative integer. In the spotted owl example, for instance, $\mathbf{x}_k$ listed the numbers of owls in three age classes at time k.

The applications in this section focus on ecological problems because they are easier to state and explain than, say, problems in physics or engineering. However, dynamical systems arise in many scientific fields. For instance, standard undergraduate courses in control systems discuss several aspects of dynamical systems. The modern *state-space* design method in such courses relies heavily on matrix algebra.[1] The *steady-state response* of a control system is the engineering equivalent of what we call here the "long-term behavior" of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$.

Until Example 6, we assume that A is diagonalizable, with n linearly independent eigenvectors, $\mathbf{v}_1, \dots, \mathbf{v}_n$, and corresponding eigenvalues, $\lambda_1, \dots, \lambda_n$. For convenience, assume the eigenvectors are arranged so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$, any initial vector $\mathbf{x}_0$ can be written uniquely as

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \tag{1}$$

This *eigenvector decomposition* of $\mathbf{x}_0$ determines what happens to the sequence $\{\mathbf{x}_k\}$. The next calculation generalizes the simple case examined in Example 5 of Section 5.2. Since the $\mathbf{v}_i$ are eigenvectors,

$$\begin{aligned} \mathbf{x}_1 = A\mathbf{x}_0 &= c_1 A\mathbf{v}_1 + \cdots + c_n A\mathbf{v}_n \\ &= c_1\lambda_1\mathbf{v}_1 + \cdots + c_n\lambda_n\mathbf{v}_n \end{aligned}$$

and

$$\begin{aligned} \mathbf{x}_2 = A\mathbf{x}_1 &= c_1\lambda_1 A\mathbf{v}_1 + \cdots + c_n\lambda_n A\mathbf{v}_n \\ &= c_1(\lambda_1)^2\mathbf{v}_1 + \cdots + c_n(\lambda_n)^2\mathbf{v}_n \end{aligned}$$

[1] See G. F. Franklin, J. D. Powell, and A. Emami-Naeimi, *Feedback Control of Dynamic Systems*, 5th ed. (Upper Saddle River, NJ: Prentice-Hall, 2006). This undergraduate text has a nice introduction to dynamic models (Chapter 2). State-space design is covered in Chapters 7 and 8.

In general,

$$\mathbf{x}_k = c_1(\lambda_1)^k \mathbf{v}_1 + \cdots + c_n(\lambda_n)^k \mathbf{v}_n \qquad (k = 0, 1, 2, \ldots) \tag{2}$$

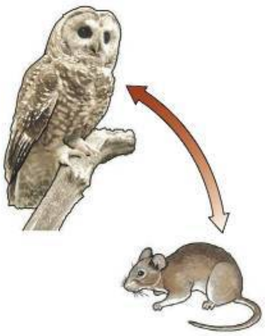
The examples that follow illustrate what can happen in (2) as $k \to \infty$.

A Predator–Prey System

Deep in the redwood forests of California, dusky-footed wood rats provide up to 80% of the diet for the spotted owl, the main predator of the wood rat. Example 1 uses a linear dynamical system to model the physical system of the owls and the rats. (Admittedly, the model is unrealistic in several respects, but it can provide a starting point for the study of more complicated nonlinear models used by environmental scientists.)

EXAMPLE 1 Denote the owl and wood rat populations at time k by $\mathbf{x}_k = \begin{bmatrix} O_k \\ R_k \end{bmatrix}$, where k is the time in months, O_k is the number of owls in the region studied, and R_k is the number of rats (measured in thousands). Suppose

$$\begin{aligned} O_{k+1} &= (.5)O_k + (.4)R_k \\ R_{k+1} &= -p \cdot O_k + (1.1)R_k \end{aligned} \tag{3}$$



where p is a positive parameter to be specified. The $(.5)O_k$ in the first equation says that with no wood rats for food, only half of the owls will survive each month, while the $(1.1)R_k$ in the second equation says that with no owls as predators, the rat population will grow by 10% per month. If rats are plentiful, the $(.4)R_k$ will tend to make the owl population rise, while the negative term $-p \cdot O_k$ measures the deaths of rats due to predation by owls. (In fact, $1000p$ is the average number of rats eaten by one owl in one month.) Determine the evolution of this system when the predation parameter p is .104.

SOLUTION When $p = .104$, the eigenvalues of the coefficient matrix $A = \begin{bmatrix} .5 & .4 \\ -p & 1.1 \end{bmatrix}$ for the equations in (3) turn out to be $\lambda_1 = 1.02$ and $\lambda_2 = .58$. Corresponding eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} 10 \\ 13 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

An initial $\mathbf{x}_0$ can be written as $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. Then, for $k \geq 0$,

$$\begin{aligned} \mathbf{x}_k &= c_1(1.02)^k \mathbf{v}_1 + c_2(.58)^k \mathbf{v}_2 \\ &= c_1(1.02)^k \begin{bmatrix} 10 \\ 13 \end{bmatrix} + c_2(.58)^k \begin{bmatrix} 5 \\ 1 \end{bmatrix} \end{aligned}$$

As $k \to \infty$, $(.58)^k$ rapidly approaches zero. Assume $c_1 > 0$. Then, for all sufficiently large k, $\mathbf{x}_k$ is approximately the same as $c_1(1.02)^k \mathbf{v}_1$, and we write

$$\mathbf{x}_k \approx c_1(1.02)^k \begin{bmatrix} 10 \\ 13 \end{bmatrix} \tag{4}$$

The approximation in (4) improves as k increases, and so for large k,

$$\mathbf{x}_{k+1} \approx c_1(1.02)^{k+1} \begin{bmatrix} 10 \\ 13 \end{bmatrix} = (1.02)c_1(1.02)^k \begin{bmatrix} 10 \\ 13 \end{bmatrix} \approx 1.02\mathbf{x}_k \tag{5}$$

The approximation in (5) says that eventually both entries of $\mathbf{x}_k$ (the numbers of owls and rats) grow by a factor of almost 1.02 each month, a 2% monthly growth rate. By (4), $\mathbf{x}_k$ is approximately a multiple of (10, 13), so the entries in $\mathbf{x}_k$ are nearly in the same ratio as 10 to 13. That is, for every 10 owls there are about 13 thousand rats. ■

Example 1 illustrates two general facts about a dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ in which A is $n \times n$, its eigenvalues satisfy $|\lambda_1| \geq 1$ and $1 > |\lambda_j|$ for $j = 2, \dots, n$, and $\mathbf{v}_1$ is an eigenvector corresponding to λ_1. If $\mathbf{x}_0$ is given by equation (1), with $c_1 \neq 0$, then for all sufficiently large k,

$$\mathbf{x}_{k+1} \approx \lambda_1 \mathbf{x}_k \tag{6}$$

and

$$\mathbf{x}_k \approx c_1(\lambda_1)^k \mathbf{v}_1 \tag{7}$$

The approximations in (6) and (7) can be made as close as desired by taking k sufficiently large. By (6), the $\mathbf{x}_k$ eventually grow almost by a factor of λ_1 each time, so λ_1 determines the eventual growth rate of the system. Also, by (7), the ratio of any two entries in $\mathbf{x}_k$ (for large k) is nearly the same as the ratio of the corresponding entries in $\mathbf{v}_1$. The case in which $\lambda_1 = 1$ is illustrated in Example 5 in Section 5.2.

Graphical Description of Solutions

When A is 2×2, algebraic calculations can be supplemented by a geometric description of a system's evolution. We can view the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ as a description of what happens to an initial point $\mathbf{x}_0$ in $\mathbb{R}^2$ as it is transformed repeatedly by the mapping $\mathbf{x} \mapsto A\mathbf{x}$. The graph of $\mathbf{x}_0, \mathbf{x}_1, \dots$ is called a **trajectory** of the dynamical system.

EXAMPLE 2 Plot several trajectories of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$, when

$$A = \begin{bmatrix} .80 & 0 \\ 0 & .64 \end{bmatrix}$$

SOLUTION The eigenvalues of A are .8 and .64, with eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. If $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, then

$$\mathbf{x}_k = c_1(.8)^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2(.64)^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Of course, $\mathbf{x}_k$ tends to $\mathbf{0}$ because $(.8)^k$ and $(.64)^k$ both approach 0 as $k \to \infty$. But *the way* $\mathbf{x}_k$ goes toward $\mathbf{0}$ is interesting. Figure 1 shows the first few terms of several trajectories

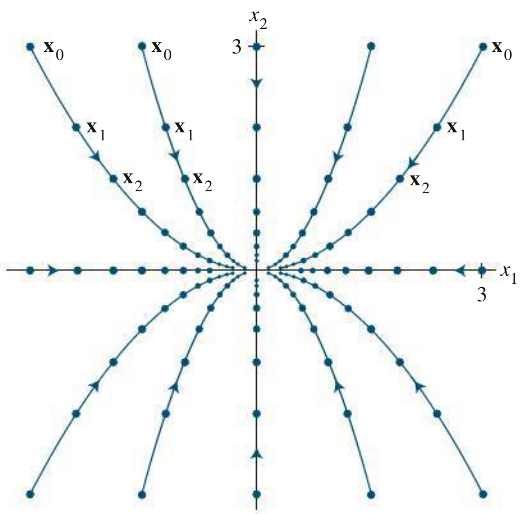


FIGURE 1 The origin as an attractor.

that begin at points on the boundary of the box with corners at $(\pm 3, \pm 3)$. The points on each trajectory are connected by a thin curve, to make the trajectory easier to see. ■

In Example 2, the origin is called an **attractor** of the dynamical system because all trajectories tend toward $\mathbf{0}$. This occurs whenever both eigenvalues are less than 1 in magnitude. The direction of greatest attraction is along the line through $\mathbf{0}$ and the eigenvector $\mathbf{v}_2$ for the eigenvalue of smaller magnitude.

In the next example, both eigenvalues of A are larger than 1 in magnitude, and $\mathbf{0}$ is called a **repeller** of the dynamical system. All solutions of $\mathbf{x}_{k+1} = A\mathbf{x}_k$ except the (constant) zero solution are unbounded and tend away from the origin.[2]

EXAMPLE 3 Plot several typical solutions of the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$, where

$$A = \begin{bmatrix} 1.44 & 0 \\ 0 & 1.2 \end{bmatrix}$$

SOLUTION The eigenvalues of A are 1.44 and 1.2. If $\mathbf{x}_0 = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$, then

$$\mathbf{x}_k = c_1(1.44)^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2(1.2)^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Both terms grow in size, but the first term grows faster. So the direction of greatest repulsion is the line through $\mathbf{0}$ and the eigenvector for the eigenvalue of larger magnitude. Figure 2 shows several trajectories that begin at points quite close to $\mathbf{0}$. ■

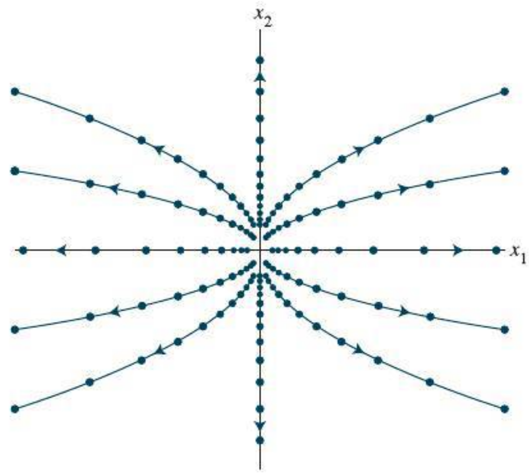


FIGURE 2 The origin as a repeller.

In the next example, $\mathbf{0}$ is called a **saddle point** because the origin attracts solutions from some directions and repels them in other directions. This occurs whenever one eigenvalue is greater than 1 in magnitude and the other is less than 1 in magnitude. The direction of greatest attraction is determined by an eigenvector for the eigenvalue of smaller magnitude. The direction of greatest repulsion is determined by an eigenvector for the eigenvalue of greater magnitude.

[2] The origin is the only possible attractor or repeller in a *linear* dynamical system, but there can be multiple attractors and repellers in a more general dynamical system for which the mapping $\mathbf{x}_k \mapsto \mathbf{x}_{k+1}$ is not linear. In such a system, attractors and repellers are defined in terms of the eigenvalues of a special matrix (with variable entries) called the *Jacobian matrix* of the system.

EXAMPLE 4 Plot several typical solutions of the equation $\mathbf{y}_{k+1} = D\mathbf{y}_k$, where

$$D = \begin{bmatrix} 2.0 & 0 \\ 0 & 0.5 \end{bmatrix}$$

(We write D and $\mathbf{y}$ here instead of A and $\mathbf{x}$ because this example will be used later.) Show that a solution $\{\mathbf{y}_k\}$ is unbounded if its initial point is not on the x_2-axis.

SOLUTION The eigenvalues of D are 2 and .5. If $\mathbf{y}_0 = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$, then

$$\mathbf{y}_k = c_1 2^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2(.5)^k \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{8}$$

If $\mathbf{y}_0$ is on the x_2-axis, then $c_1 = 0$ and $\mathbf{y}_k \to \mathbf{0}$ as $k \to \infty$. But if $\mathbf{y}_0$ is not on the x_2-axis, then the first term in the sum for $\mathbf{y}_k$ becomes arbitrarily large, and so $\{\mathbf{y}_k\}$ is unbounded. Figure 3 shows ten trajectories that begin near or on the x_2-axis. ■

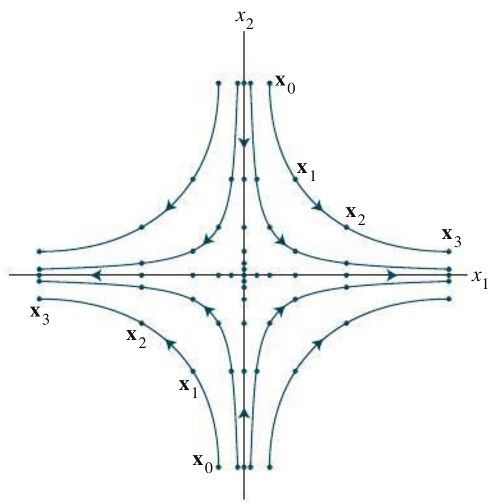


FIGURE 3 The origin as a saddle point.

Change of Variable

The preceding three examples involved diagonal matrices. To handle the nondiagonal case, we return for a moment to the $n \times n$ case in which eigenvectors of A form a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for $\mathbb{R}^n$. Let $P = [\, \mathbf{v}_1 \ \cdots \ \mathbf{v}_n \,]$, and let D be the diagonal matrix with the corresponding eigenvalues on the diagonal. Given a sequence $\{\mathbf{x}_k\}$ satisfying $\mathbf{x}_{k+1} = A\mathbf{x}_k$, define a new sequence $\{\mathbf{y}_k\}$ by

$$\mathbf{y}_k = P^{-1}\mathbf{x}_k, \quad \text{or equivalently,} \quad \mathbf{x}_k = P\mathbf{y}_k$$

Substituting these relations into the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ and using the fact that $A = PDP^{-1}$, we find that

$$P\mathbf{y}_{k+1} = AP\mathbf{y}_k = (PDP^{-1})P\mathbf{y}_k = PD\mathbf{y}_k$$

Left-multiplying both sides by P^{-1}, we obtain

$$\mathbf{y}_{k+1} = D\mathbf{y}_k$$

If we write $\mathbf{y}_k$ as $\mathbf{y}(k)$ and denote the entries in $\mathbf{y}(k)$ by $y_1(k), \dots, y_n(k)$, then

$$\begin{bmatrix} y_1(k+1) \\ y_2(k+1) \\ \vdots \\ y_n(k+1) \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_n(k) \end{bmatrix}$$

The change of variable from $\mathbf{x}_k$ to $\mathbf{y}_k$ has *decoupled* the system of difference equations. The evolution of $y_1(k)$, for example, is unaffected by what happens to $y_2(k), \dots, y_n(k)$, because $y_1(k+1) = \lambda_1 \cdot y_1(k)$ for each k.

The equation $\mathbf{x}_k = P\mathbf{y}_k$ says that $\mathbf{y}_k$ is the coordinate vector of $\mathbf{x}_k$ with respect to the eigenvector basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. We can decouple the system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ by making calculations in the new eigenvector coordinate system. When $n = 2$, this amounts to using graph paper with axes in the directions of the two eigenvectors.

EXAMPLE 5 Show that the origin is a saddle point for solutions of $\mathbf{x}_{k+1} = A\mathbf{x}_k$, where

$$A = \begin{bmatrix} 1.25 & -.75 \\ -.75 & 1.25 \end{bmatrix}$$

Find the directions of greatest attraction and greatest repulsion.

SOLUTION Using standard techniques, we find that A has eigenvalues 2 and .5, with corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, respectively. Since $|2| > 1$ and $|.5| < 1$, the origin is a saddle point of the dynamical system. If $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, then

$$\mathbf{x}_k = c_1 2^k \mathbf{v}_1 + c_2 (.5)^k \mathbf{v}_2 \tag{9}$$

This equation looks just like equation (8) in Example 4, with $\mathbf{v}_1$ and $\mathbf{v}_2$ in place of the standard basis.

On graph paper, draw axes through $\mathbf{0}$ and the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$. See Figure 4. Movement along these axes corresponds to movement along the standard axes in Figure 3. In Figure 4, the direction of greatest *repulsion* is the line through $\mathbf{0}$ and the

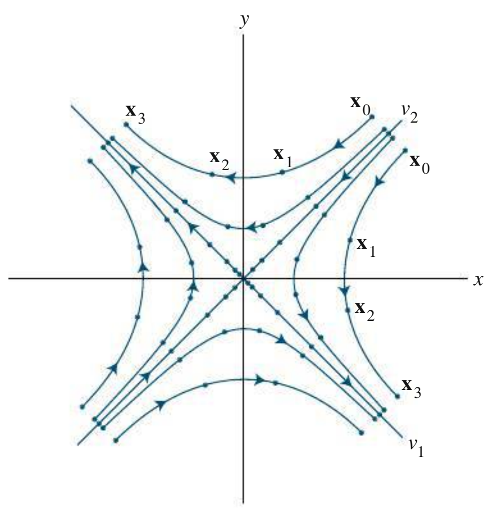


FIGURE 4 The origin as a saddle point.

eigenvector $\mathbf{v}_1$ whose eigenvalue is greater than 1 in magnitude. If $\mathbf{x}_0$ is on this line, the c_2 in (9) is zero and $\mathbf{x}_k$ moves quickly away from $\mathbf{0}$. The direction of greatest *attraction* is determined by the eigenvector $\mathbf{v}_2$ whose eigenvalue is less than 1 in magnitude.

A number of trajectories are shown in Figure 4. When this graph is viewed in terms of the eigenvector axes, the picture "looks" essentially the same as the picture in Figure 3. ■

Complex Eigenvalues

When a real 2×2 matrix A has complex eigenvalues, A is not diagonalizable (when acting on $\mathbb{R}^2$), but the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ is easy to describe. Example 3 of Section 5.5 illustrated the case in which the eigenvalues have absolute value 1. The iterates of a point $\mathbf{x}_0$ spiral around the origin along an elliptical trajectory.

If A has two complex eigenvalues whose absolute value is greater than 1, then $\mathbf{0}$ is a repeller and iterates of $\mathbf{x}_0$ will spiral outward around the origin. If the absolute values of the complex eigenvalues are less than 1, then the origin is an attractor and the iterates of $\mathbf{x}_0$ spiral inward toward the origin, as in the following example.

EXAMPLE 6 It can be verified that the matrix

$$A = \begin{bmatrix} .8 & .5 \\ -.1 & 1.0 \end{bmatrix}$$

has eigenvalues $.9 \pm .2i$, with eigenvectors $\begin{bmatrix} 1 \mp 2i \\ 1 \end{bmatrix}$. Figure 5 shows three trajectories of the system $\mathbf{x}_{k+1} = A\mathbf{x}_k$, with initial vectors $\begin{bmatrix} 0 \\ 2.5 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$, and $\begin{bmatrix} 0 \\ -2.5 \end{bmatrix}$. ■

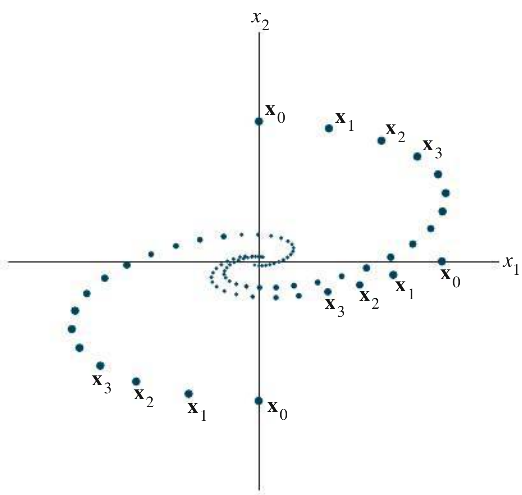


FIGURE 5 Rotation associated with complex eigenvalues.

Survival of the Spotted Owls

Recall from this chapter's introductory example that the spotted owl population in the Willow Creek area of California was modeled by a dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ in which the entries in $\mathbf{x}_k = (j_k, s_k, a_k)$ listed the numbers of females (at time k) in the juvenile, subadult, and adult life stages, respectively, and A is the stage-matrix

$$A = \begin{bmatrix} 0 & 0 & .33 \\ .18 & 0 & 0 \\ 0 & .71 & .94 \end{bmatrix} \tag{10}$$

MATLAB shows that the eigenvalues of A are approximately $\lambda_1 = .98$, $\lambda_2 = -.02 + .21i$, and $\lambda_3 = -.02 - .21i$. Observe that all three eigenvalues are less than 1 in magnitude, because $|\lambda_2|^2 = |\lambda_3|^2 = (-.02)^2 + (.21)^2 = .0445$.

For the moment, let A act on the complex vector space $\mathbb{C}^3$. Then, because A has three distinct eigenvalues, the three corresponding eigenvectors are linearly independent and form a basis for $\mathbb{C}^3$. Denote the eigenvectors by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Then the general solution of $\mathbf{x}_{k+1} = A\mathbf{x}_k$ (using vectors in $\mathbb{C}^3$) has the form

$$\mathbf{x}_k = c_1(\lambda_1)^k\mathbf{v}_1 + c_2(\lambda_2)^k\mathbf{v}_2 + c_3(\lambda_3)^k\mathbf{v}_3 \tag{11}$$

If $\mathbf{x}_0$ is a real initial vector, then $\mathbf{x}_1 = A\mathbf{x}_0$ is real because A is real. Similarly, the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ shows that each $\mathbf{x}_k$ on the left side of (11) is real, even though it is expressed as a sum of complex vectors. However, each term on the right side of (11) is approaching the zero vector, because the eigenvalues are all less than 1 in magnitude. Therefore the real sequence $\mathbf{x}_k$ approaches the zero vector, too. Sadly, this model predicts that the spotted owls will eventually all perish.

Is there hope for the spotted owl? Recall from the introductory example that the 18% entry in the matrix A in (10) comes from the fact that although 60% of the juvenile owls live long enough to leave the nest and search for new home territories, only 30% of that group survive the search and find new home ranges. Search survival is strongly influenced by the number of clear-cut areas in the forest, which make the search more difficult and dangerous.

Some owl populations live in areas with few or no clear-cut areas. It may be that a larger percentage of the juvenile owls there survive and find new home ranges. Of course, the problem of the spotted owl is more complex than we have described, but the final example provides a happy ending to the story.

EXAMPLE 7 Suppose the search survival rate of the juvenile owls is 50%, so the $(2, 1)$-entry in the stage-matrix A in (10) is .3 instead of .18. What does the stage-matrix model predict about this spotted owl population?

SOLUTION Now the eigenvalues of A turn out to be approximately $\lambda_1 = 1.01$, $\lambda_2 = -.03 + .26i$, and $\lambda_3 = -.03 - .26i$. An eigenvector for λ_1 is approximately $\mathbf{v}_1 = (10, 3, 31)$. Let $\mathbf{v}_2$ and $\mathbf{v}_3$ be (complex) eigenvectors for λ_2 and λ_3. In this case, equation (11) becomes

$$\mathbf{x}_k = c_1(1.01)^k\mathbf{v}_1 + c_2(-.03 + .26i)^k\mathbf{v}_2 + c_3(-.03 - .26i)^k\mathbf{v}_3$$

As $k \to \infty$, the second two vectors tend to zero. So $\mathbf{x}_k$ becomes more and more like the (real) vector $c_1(1.01)^k\mathbf{v}_1$. The approximations in equations (6) and (7), following Example 1, apply here. Also, it can be shown that the constant c_1 in the initial

Further Reading: Franklin, G. F., J. D. Powell, and M. L. Workman. *Digital Control of Dynamic Systems*, 3rd ed. Reading, MA: Addison-Wesley, 1998; Sandefur, James T. *Discrete Dynamical Systems—Theory and Applications*. Oxford: Oxford University Press, 1990; Tuchinsky, Philip. *Management of a Buffalo Herd*, UMAP Module 207. Lexington, MA: COMAP, 1980.

decomposition of $\mathbf{x}_0$ is positive when the entries in $\mathbf{x}_0$ are nonnegative. Thus the owl population will grow slowly, with a long-term growth rate of 1.01. The eigenvector $\mathbf{v}_1$ describes the eventual distribution of the owls by life stages: for every 31 adults, there will be about 10 juveniles and 3 subadults. ■

Practice Problems

1. The matrix A below has eigenvalues 1, $\frac{2}{3}$, and $\frac{1}{3}$, with corresponding eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$:

$$A = \frac{1}{9}\begin{bmatrix} 7 & -2 & 0 \\ -2 & 6 & 2 \\ 0 & 2 & 5 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

Find the general solution of the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ if $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 11 \\ -2 \end{bmatrix}$.

2. What happens to the sequence $\{\mathbf{x}_k\}$ in Practice Problem 1 as $k \to \infty$?

5.6 Exercises

1. Let A be a 2×2 matrix with eigenvalues 3 and 1/3 and corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Let $\{\mathbf{x}_k\}$ be a solution of the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$, $\mathbf{x}_0 = \begin{bmatrix} 9 \\ 1 \end{bmatrix}$.
   a. Compute $\mathbf{x}_1 = A\mathbf{x}_0$. [*Hint:* You do not need to know A itself.]
   b. Find a formula for $\mathbf{x}_k$ involving k and the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

2. Suppose the eigenvalues of a 3×3 matrix A are 3, 4/5, and 3/5, with corresponding eigenvectors $\begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \\ -5 \end{bmatrix}$, and $\begin{bmatrix} -3 \\ -3 \\ 7 \end{bmatrix}$. Let $\mathbf{x}_0 = \begin{bmatrix} -2 \\ -5 \\ 3 \end{bmatrix}$. Find the solution of the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ for the specified $\mathbf{x}_0$, and describe what happens as $k \to \infty$.

In Exercises 3–6, assume that any initial vector $\mathbf{x}_0$ has an eigenvector decomposition such that the coefficient c_1 in equation (1) of this section is positive.[3]

3. Determine the evolution of the dynamical system in Example 1 when the predation parameter p is .2 in equation (3). (Give a formula for $\mathbf{x}_k$.) Does the owl population grow or decline? What about the wood rat population?

4. Determine the evolution of the dynamical system in Example 1 when the predation parameter p is .125. (Give a formula for $\mathbf{x}_k$.) As time passes, what happens to the sizes of the owl and wood rat populations? The system tends toward what is sometimes called an unstable equilibrium. What do you think might happen to the system if some aspect of the model (such as birth rates or the predation rate) were to change slightly?

5. In old-growth forests of Douglas fir, the spotted owl dines mainly on flying squirrels. Suppose the predator–prey matrix for these two populations is $A = \begin{bmatrix} .4 & .3 \\ -p & 1.2 \end{bmatrix}$. Show that if the predation parameter p is .325, both populations grow. Estimate the long-term growth rate and the eventual ratio of owls to flying squirrels.

6. Show that if the predation parameter p in Exercise 5 is .5, both the owls and the squirrels will eventually perish. Find a value of p for which populations of both owls and squirrels tend toward constant levels. What are the relative population sizes in this case?

7. Let A have the properties described in Exercise 1.
   a. Is the origin an attractor, a repeller, or a saddle point of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$?
   b. Find the directions of greatest attraction and/or repulsion for this dynamical system.
   c. Make a graphical description of the system, showing the directions of greatest attraction or repulsion. Include a rough sketch of several typical trajectories (without computing specific points).

8. Determine the nature of the origin (attractor, repeller, or saddle point) for the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ if A has

[3] One of the limitations of the model in Example 1 is that there always exist initial population vectors $\mathbf{x}_0$ with positive entries such that the coefficient c_1 is negative. The approximation (7) is still valid, but the entries in $\mathbf{x}_k$ eventually become negative.

the properties described in Exercise 2. Find the directions of greatest attraction or repulsion.

In Exercises 9–14, classify the origin as an attractor, repeller, or saddle point of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$. Find the directions of greatest attraction and/or repulsion.

9. $A = \begin{bmatrix} 1.7 & -.3 \\ -1.2 & .8 \end{bmatrix}$

10. $A = \begin{bmatrix} .3 & .4 \\ -.3 & 1.1 \end{bmatrix}$

11. $A = \begin{bmatrix} .4 & .5 \\ -.4 & 1.3 \end{bmatrix}$

12. $A = \begin{bmatrix} .5 & .6 \\ -.3 & 1.4 \end{bmatrix}$

13. $A = \begin{bmatrix} .8 & .3 \\ -.4 & 1.5 \end{bmatrix}$

14. $A = \begin{bmatrix} 1.7 & .6 \\ -.4 & .7 \end{bmatrix}$

15. Let $A = \begin{bmatrix} .4 & 0 & .2 \\ .3 & .8 & .3 \\ .3 & .2 & .5 \end{bmatrix}$. The vector $\mathbf{v}_1 = \begin{bmatrix} .1 \\ .6 \\ .3 \end{bmatrix}$ is an eigenvector for A, and two eigenvalues are .5 and .2. Construct the solution of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ that satisfies $\mathbf{x}_0 = (0, .3, .7)$. What happens to $\mathbf{x}_k$ as $k \to \infty$?

T 16. Produce the general solution of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ when $A = \begin{bmatrix} .90 & .01 & .09 \\ .01 & .90 & .01 \\ .09 & .09 & .90 \end{bmatrix}$.

17. Construct a stage-matrix model for an animal species that has two life stages: juvenile (up to 1 year old) and adult. Suppose the female adults give birth each year to an average of 1.6 female juveniles. Each year, 30% of the juveniles survive to become adults and 80% of the adults survive. For $k \geq 0$, let $\mathbf{x}_k = (j_k, a_k)$, where the entries in $\mathbf{x}_k$ are the numbers of female juveniles and female adults in year k.

a. Construct the stage-matrix A such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$ for $k \geq 0$.

b. Show that the population is growing, compute the eventual growth rate of the population, and give the eventual ratio of juveniles to adults.

T c. Suppose that initially there are 15 juveniles and 10 adults in the population. Produce four graphs that show how the population changes over eight years: (a) the number of juveniles, (b) the number of adults, (c) the total population, and (d) the ratio of juveniles to adults (each year). When does the ratio in (d) seem to stabilize? Include a listing of the program or keystrokes used to produce the graphs for (c) and (d).

18. A herd of American buffalo (bison) can be modeled by a stage matrix similar to that for the spotted owls. The females can be divided into calves (up to 1 year old), yearlings (1 to 2 years), and adults. Suppose an average of 42 female calves are born each year per 100 adult females. (Only adults produce offspring.) Each year, about 60% of the calves survive, 75% of the yearlings survive, and 95% of the adults survive. For $k \geq 0$, let $\mathbf{x}_k = (c_k, y_k, a_k)$, where the entries in $\mathbf{x}_k$ are the numbers of females in each life stage at year k.

a. Construct the stage-matrix A for the buffalo herd, such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$ for $k \geq 0$.

T b. Show that the buffalo herd is growing, determine the expected growth rate after many years, and give the expected numbers of calves and yearlings present per 100 adults.

Solutions to Practice Problems

1. The first step is to write $\mathbf{x}_0$ as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Row reduction of $[\,\mathbf{v}_1 \ \ \mathbf{v}_2 \ \ \mathbf{v}_3 \ \ \mathbf{x}_0\,]$ produces the weights $c_1 = 2$, $c_2 = 1$, and $c_3 = 3$, so that

$$\mathbf{x}_0 = 2\mathbf{v}_1 + 1\mathbf{v}_2 + 3\mathbf{v}_3$$

Since the eigenvalues are 1, $\frac{2}{3}$, and $\frac{1}{3}$, the general solution is

$$\begin{aligned} \mathbf{x}_k &= 2 \cdot 1^k \mathbf{v}_1 + 1 \cdot \left(\frac{2}{3}\right)^k \mathbf{v}_2 + 3 \cdot \left(\frac{1}{3}\right)^k \mathbf{v}_3 \\ &= 2\begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} + \left(\frac{2}{3}\right)^k \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} + 3 \cdot \left(\frac{1}{3}\right)^k \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} \end{aligned} \tag{12}$$

2. As $k \to \infty$, the second and third terms in (12) tend to the zero vector, and

$$\mathbf{x}_k = 2\mathbf{v}_1 + \left(\frac{2}{3}\right)^k \mathbf{v}_2 + 3\left(\frac{1}{3}\right)^k \mathbf{v}_3 \to 2\mathbf{v}_1 = \begin{bmatrix} -4 \\ 4 \\ 2 \end{bmatrix}$$

5.7 Applications to Differential Equations

This section describes continuous analogues of the difference equations studied in Section 5.6. In many applied problems, several quantities are varying continuously in time, and they are related by a system of differential equations:

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n \\ &\vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n \end{aligned}$$

Here $x_1, \ldots, x_n$ are differentiable functions of t, with derivatives $x_1', \ldots, x_n'$, and the a_{ij} are constants. The crucial feature of this system is that it is *linear*. To see this, write the system as a matrix differential equation

$$\mathbf{x}'(t) = A\mathbf{x}(t) \tag{1}$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

A **solution** of equation (1) is a vector-valued function that satisfies (1) for all t in some interval of real numbers, such as $t \geq 0$.

Equation (1) is *linear* because both differentiation of functions and multiplication of vectors by a matrix are linear transformations. Thus, if $\mathbf{u}$ and $\mathbf{v}$ are solutions of $\mathbf{x}' = A\mathbf{x}$, then $c\mathbf{u} + d\mathbf{v}$ is also a solution, because

$$\begin{aligned} (c\mathbf{u} + d\mathbf{v})' &= c\mathbf{u}' + d\mathbf{v}' \\ &= cA\mathbf{u} + dA\mathbf{v} = A(c\mathbf{u} + d\mathbf{v}) \end{aligned}$$

(Engineers call this property *superposition* of solutions.) Also, the identically zero function is a (trivial) solution of (1). In the terminology of Chapter 4, the set of all solutions of (1) is a *subspace* of the set of all continuous functions with values in $\mathbb{R}^n$.

Standard texts on differential equations show that there always exists what is called a **fundamental set of solutions** to (1). If A is $n \times n$, then there are n linearly independent functions in a fundamental set, and each solution of (1) is a unique linear combination of these n functions. That is, a fundamental set of solutions is a *basis* for the set of all solutions of (1), and the solution set is an n-dimensional vector space of functions. If a vector $\mathbf{x}_0$ is specified, then the **initial value problem** is to construct the (unique) function $\mathbf{x}$ such that $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{x}(0) = \mathbf{x}_0$.

When A is a diagonal matrix, the solutions of (1) can be produced by elementary calculus. For instance, consider

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \tag{2}$$

that is,

$$\begin{aligned} x_1'(t) &= 3x_1(t) \\ x_2'(t) &= -5x_2(t) \end{aligned} \tag{3}$$

The system (2) is said to be *decoupled* because each derivative of a function depends only on the function itself, not on some combination or "coupling" of both $x_1(t)$ and $x_2(t)$. From calculus, the solutions of (3) are $x_1(t) = c_1e^{3t}$ and $x_2(t) = c_2e^{-5t}$, for any constants c_1 and c_2. Each solution of equation (2) can be written in the form

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} c_1e^{3t} \\ c_2e^{-5t} \end{bmatrix} = c_1\begin{bmatrix} 1 \\ 0 \end{bmatrix}e^{3t} + c_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}e^{-5t}$$

This example suggests that for the general equation $\mathbf{x}' = A\mathbf{x}$, a solution might be a linear combination of functions of the form

$$\mathbf{x}(t) = \mathbf{v}e^{\lambda t} \tag{4}$$

for some scalar λ and some fixed nonzero vector $\mathbf{v}$. [If $\mathbf{v} = \mathbf{0}$, the function $\mathbf{x}(t)$ is identically zero and hence satisfies $\mathbf{x}' = A\mathbf{x}$.] Observe that

$$\begin{aligned} \mathbf{x}'(t) &= \lambda\mathbf{v}e^{\lambda t} && \text{By calculus, since } \mathbf{v} \text{ is a constant vector} \\ A\mathbf{x}(t) &= A\mathbf{v}e^{\lambda t} && \text{Multiplying both sides of (4) by } A \end{aligned}$$

Since $e^{\lambda t}$ is never zero, $\mathbf{x}'(t)$ will equal $A\mathbf{x}(t)$ if and only if $\lambda\mathbf{v} = A\mathbf{v}$, that is, if and only if λ is an eigenvalue of A and $\mathbf{v}$ is a corresponding eigenvector. Thus each eigenvalue–eigenvector pair provides a solution (4) of $\mathbf{x}' = A\mathbf{x}$. Such solutions are sometimes called *eigenfunctions* of the differential equation. Eigenfunctions provide the key to solving systems of differential equations.

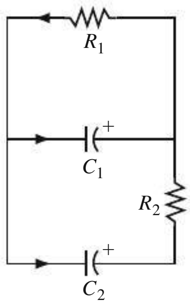


FIGURE 1

EXAMPLE 1 The circuit in Figure 1 can be described by the differential equation

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = \begin{bmatrix} -(1/R_1 + 1/R_2)/C_1 & 1/(R_2C_1) \\ 1/(R_2C_2) & -1/(R_2C_2) \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

where $x_1(t)$ and $x_2(t)$ are the voltages across the two capacitors at time t. Suppose resistor R_1 is 1 ohm, R_2 is 2 ohms, capacitor C_1 is 1 farad, and C_2 is .5 farad, and suppose there is an initial charge of 5 volts on capacitor C_1 and 4 volts on capacitor C_2. Find formulas for $x_1(t)$ and $x_2(t)$ that describe how the voltages change over time.

SOLUTION Let A denote the matrix displayed above, and let $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$. For the data given, $A = \begin{bmatrix} -1.5 & .5 \\ 1 & -1 \end{bmatrix}$, and $\mathbf{x}(0) = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$. The eigenvalues of A are $\lambda_1 = -.5$ and $\lambda_2 = -2$, with corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

The eigenfunctions $\mathbf{x}_1(t) = \mathbf{v}_1e^{\lambda_1 t}$ and $\mathbf{x}_2(t) = \mathbf{v}_2e^{\lambda_2 t}$ both satisfy $\mathbf{x}' = A\mathbf{x}$, and so does any linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$. Set

$$\mathbf{x}(t) = c_1\mathbf{v}_1e^{\lambda_1 t} + c_2\mathbf{v}_2e^{\lambda_2 t} = c_1\begin{bmatrix} 1 \\ 2 \end{bmatrix}e^{-.5t} + c_2\begin{bmatrix} -1 \\ 1 \end{bmatrix}e^{-2t}$$

and note that $\mathbf{x}(0) = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are obviously linearly independent and hence span $\mathbb{R}^2$, c_1 and c_2 can be found to make $\mathbf{x}(0)$ equal to $\mathbf{x}_0$. In fact, the equation

$$c_1\underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}_{\mathbf{v}_1} + c_2\underbrace{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}_{\mathbf{v}_2} = \underbrace{\begin{bmatrix} 5 \\ 4 \end{bmatrix}}_{\mathbf{x}_0}$$

leads easily to $c_1 = 3$ and $c_2 = -2$. Thus the desired solution of the differential equation $\mathbf{x}' = A\mathbf{x}$ is

$$\mathbf{x}(t) = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} - 2\begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}$$

or

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 3e^{-.5t} + 2e^{-2t} \\ 6e^{-.5t} - 2e^{-2t} \end{bmatrix}$$

Figure 2 shows the graph, or *trajectory*, of $\mathbf{x}(t)$, for $t \geq 0$, along with trajectories for some other initial points. The trajectories of the two eigenfunctions $\mathbf{x}_1$ and $\mathbf{x}_2$ lie in the eigenspaces of A.

The functions $\mathbf{x}_1$ and $\mathbf{x}_2$ both decay to zero as $t \to \infty$, but the values of $\mathbf{x}_2$ decay faster because its exponent is more negative. The entries in the corresponding eigenvector $\mathbf{v}_2$ show that the voltages across the capacitors will decay to zero as rapidly as possible if the initial voltages are equal in magnitude but opposite in sign. ■

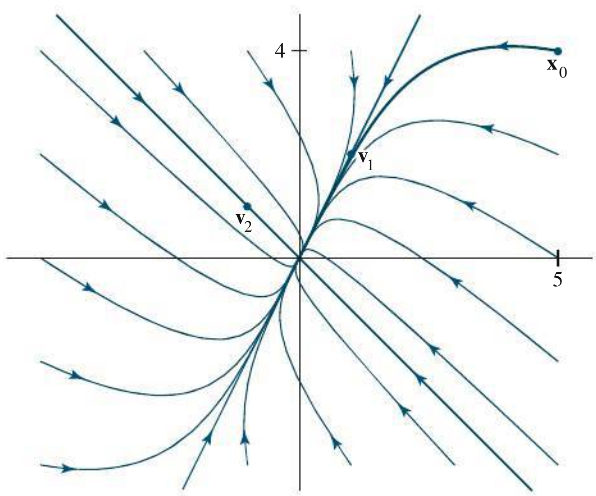


FIGURE 2 The origin as an attractor.

In Figure 2, the origin is called an **attractor**, or **sink**, of the dynamical system because all trajectories are drawn into the origin. The direction of greatest attraction is along the trajectory of the eigenfunction $\mathbf{x}_2$ (along the line through $\mathbf{0}$ and $\mathbf{v}_2$) corresponding to the more negative eigenvalue, $\lambda = -2$. Trajectories that begin at points not on this line become asymptotic to the line through $\mathbf{0}$ and $\mathbf{v}_1$ because their components in the $\mathbf{v}_2$ direction decay so rapidly.

If the eigenvalues in Example 1 were positive instead of negative, the corresponding trajectories would be similar in shape, but the trajectories would be traversed *away* from the origin. In such a case, the origin is called a **repeller**, or **source**, of the dynamical system, and the direction of greatest repulsion is the line containing the trajectory of the eigenfunction corresponding to the more positive eigenvalue.

EXAMPLE 2 Suppose a particle is moving in a planar force field and its position vector $\mathbf{x}$ satisfies $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{x}(0) = \mathbf{x}_0$, where

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 1 \end{bmatrix}, \qquad \mathbf{x}_0 = \begin{bmatrix} 2.9 \\ 2.6 \end{bmatrix}$$

Solve this initial value problem for $t \geq 0$, and sketch the trajectory of the particle.

SOLUTION The eigenvalues of A turn out to be $\lambda_1 = 6$ and $\lambda_2 = -1$, with corresponding eigenvectors $\mathbf{v}_1 = (-5, 2)$ and $\mathbf{v}_2 = (1, 1)$. For any constants c_1 and c_2, the function

$$\mathbf{x}(t) = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t} = c_1\begin{bmatrix} -5 \\ 2 \end{bmatrix} e^{6t} + c_2\begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{-t}$$

is a solution of $\mathbf{x}' = A\mathbf{x}$. We want c_1 and c_2 to satisfy $\mathbf{x}(0) = \mathbf{x}_0$, that is,

$$c_1\begin{bmatrix} -5 \\ 2 \end{bmatrix} + c_2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.9 \\ 2.6 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} -5 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2.9 \\ 2.6 \end{bmatrix}$$

Calculations show that $c_1 = -3/70$ and $c_2 = 188/70$, and so the desired function is

$$\mathbf{x}(t) = \frac{-3}{70}\begin{bmatrix} -5 \\ 2 \end{bmatrix} e^{6t} + \frac{188}{70}\begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{-t}$$

Trajectories of $\mathbf{x}$ and other solutions are shown in Figure 3. ■

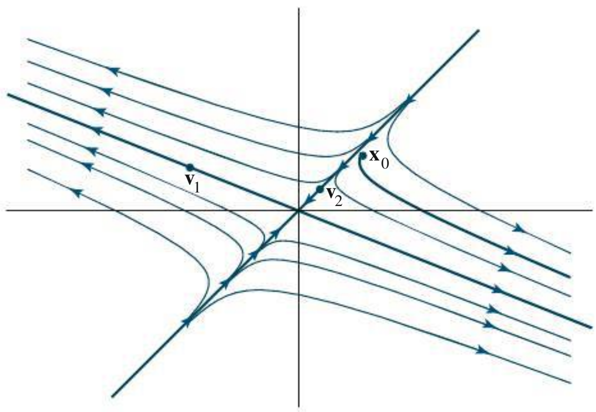


FIGURE 3 The origin as a saddle point.

In Figure 3, the origin is called a **saddle point** of the dynamical system because some trajectories approach the origin at first and then change direction and move away from the origin. A saddle point arises whenever the matrix A has both positive and negative eigenvalues. The direction of greatest repulsion is the line through $\mathbf{v}_1$ and $\mathbf{0}$, corresponding to the positive eigenvalue. The direction of greatest attraction is the line through $\mathbf{v}_2$ and $\mathbf{0}$, corresponding to the negative eigenvalue.

Decoupling a Dynamical System

The following discussion shows that the method of Examples 1 and 2 produces a fundamental set of solutions for any dynamical system described by $\mathbf{x}' = A\mathbf{x}$ when A is $n \times n$ and has n linearly independent eigenvectors, that is, when A is diagonalizable. Suppose the eigenfunctions for A are

$$\mathbf{v}_1 e^{\lambda_1 t}, \ldots, \mathbf{v}_n e^{\lambda_n t}$$

with $\mathbf{v}_1, \ldots, \mathbf{v}_n$ linearly independent eigenvectors. Let $P = [\,\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n\,]$, and let D be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$, so that $A = PDP^{-1}$. Now make a *change of variable*, defining a new function $\mathbf{y}$ by

$$\mathbf{y}(t) = P^{-1}\mathbf{x}(t) \quad \text{or, equivalently,} \quad \mathbf{x}(t) = P\mathbf{y}(t)$$

The equation $\mathbf{x}(t) = P\mathbf{y}(t)$ says that $\mathbf{y}(t)$ is the coordinate vector of $\mathbf{x}(t)$ relative to the eigenvector basis. Substitution of $P\mathbf{y}$ for $\mathbf{x}$ in the equation $\mathbf{x}' = A\mathbf{x}$ gives

$$\frac{d}{dt}(P\mathbf{y}) = A(P\mathbf{y}) = (PDP^{-1})P\mathbf{y} = PD\mathbf{y} \tag{5}$$

Since P is a constant matrix, the left side of (5) is $P\mathbf{y}'$. Left-multiply both sides of (5) by P^{-1} and obtain $\mathbf{y}' = D\mathbf{y}$, or

$$\begin{bmatrix} y_1'(t) \\ y_2'(t) \\ \vdots \\ y_n'(t) \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix}$$

The change of variable from $\mathbf{x}$ to $\mathbf{y}$ has *decoupled* the system of differential equations, because the derivative of each scalar function y_k depends only on y_k. (Review the analogous change of variables in Section 5.6.) Since $y_1' = \lambda_1 y_1$, we have $y_1(t) = c_1 e^{\lambda_1 t}$, with similar formulas for $y_2, \dots, y_n$. Thus

$$\mathbf{y}(t) = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix}, \quad \text{where} \quad \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{y}(0) = P^{-1}\mathbf{x}(0) = P^{-1}\mathbf{x}_0$$

To obtain the general solution $\mathbf{x}$ of the original system, compute

$$\begin{aligned} \mathbf{x}(t) = P\mathbf{y}(t) &= [\,\mathbf{v}_1 \;\; \cdots \;\; \mathbf{v}_n\,]\,\mathbf{y}(t) \\ &= c_1\mathbf{v}_1 e^{\lambda_1 t} + \cdots + c_n \mathbf{v}_n e^{\lambda_n t} \end{aligned}$$

This is the eigenfunction expansion constructed as in Example 1.

Complex Eigenvalues

In the next example, a real matrix A has a pair of complex eigenvalues λ and $\overline{\lambda}$, with associated complex eigenvectors $\mathbf{v}$ and $\overline{\mathbf{v}}$. (Recall from Section 5.5 that for a real matrix, complex eigenvalues and associated eigenvectors come in conjugate pairs.) So two solutions of $\mathbf{x}' = A\mathbf{x}$ are

$$\mathbf{x}_1(t) = \mathbf{v}e^{\lambda t} \quad \text{and} \quad \mathbf{x}_2(t) = \overline{\mathbf{v}}e^{\overline{\lambda} t} \tag{6}$$

It can be shown that $\mathbf{x}_2(t) = \overline{\mathbf{x}_1(t)}$ by using a power series representation for the complex exponential function. Although the complex eigenfunctions $\mathbf{x}_1$ and $\mathbf{x}_2$ are convenient for some calculations (particularly in electrical engineering), real functions are more appropriate for many purposes. Fortunately, the real and imaginary parts of $\mathbf{x}_1$ are (real) solutions of $\mathbf{x}' = A\mathbf{x}$, because they are linear combinations of the solutions in (6):

$$\operatorname{Re}(\mathbf{v}e^{\lambda t}) = \frac{1}{2}[\,\mathbf{x}_1(t) + \overline{\mathbf{x}_1(t)}\,], \qquad \operatorname{Im}(\mathbf{v}e^{\lambda t}) = \frac{1}{2i}[\,\mathbf{x}_1(t) - \overline{\mathbf{x}_1(t)}\,]$$

To understand the nature of $\operatorname{Re}(\mathbf{v}e^{\lambda t})$, recall from calculus that for any number x, the exponential function e^x can be computed from the power series:

$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + \cdots$$

This series can be used to define $e^{\lambda t}$ when λ is complex:

$$e^{\lambda t} = 1 + (\lambda t) + \frac{1}{2!}(\lambda t)^2 + \cdots + \frac{1}{n!}(\lambda t)^n + \cdots$$

By writing $\lambda = a + bi$ (with a and b real), and using similar power series for the cosine and sine functions, one can show that

$$e^{(a+bi)t} = e^{at}e^{ibt} = e^{at}(\cos bt + i \sin bt) \tag{7}$$

Hence

$$\begin{aligned} \mathbf{v}e^{\lambda t} &= (\operatorname{Re}\mathbf{v} + i \operatorname{Im}\mathbf{v})\,(e^{at})(\cos bt + i \sin bt) \\ &= [\,(\operatorname{Re}\mathbf{v})\cos bt - (\operatorname{Im}\mathbf{v})\sin bt\,]e^{at} \\ &\quad + i[\,(\operatorname{Re}\mathbf{v})\sin bt + (\operatorname{Im}\mathbf{v})\cos bt\,]e^{at} \end{aligned}$$

So two real solutions of $\mathbf{x}' = A\mathbf{x}$ are

$$\begin{aligned} \mathbf{y}_1(t) &= \operatorname{Re}\mathbf{x}_1(t) = [\,(\operatorname{Re}\mathbf{v})\cos bt - (\operatorname{Im}\mathbf{v})\sin bt\,]\,e^{at} \\ \mathbf{y}_2(t) &= \operatorname{Im}\mathbf{x}_1(t) = [\,(\operatorname{Re}\mathbf{v})\sin bt + (\operatorname{Im}\mathbf{v})\cos bt\,]\,e^{at} \end{aligned}$$

It can be shown that $\mathbf{y}_1$ and $\mathbf{y}_2$ are linearly independent functions (when $b \neq 0$).[1]

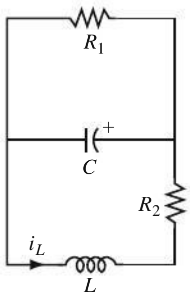


FIGURE 4

EXAMPLE 3 The circuit in Figure 4 can be described by the equation

$$\begin{bmatrix} i_L' \\ v_C' \end{bmatrix} = \begin{bmatrix} -R_2/L & -1/L \\ 1/C & -1/(R_1C) \end{bmatrix} \begin{bmatrix} i_L \\ v_C \end{bmatrix}$$

where i_L is the current passing through the inductor L and v_C is the voltage drop across the capacitor C. Suppose R_1 is 5 ohms, R_2 is .8 ohm, C is .1 farad, and L is .4 henry. Find formulas for i_L and v_C, if the initial current through the inductor is 3 amperes and the initial voltage across the capacitor is 3 volts.

SOLUTION For the data given, $A = \begin{bmatrix} -2 & -2.5 \\ 10 & -2 \end{bmatrix}$ and $\mathbf{x}_0 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$. The method discussed in Section 5.5 produces the eigenvalue $\lambda = -2 + 5i$ and the corresponding eigenvector $\mathbf{v}_1 = \begin{bmatrix} i \\ 2 \end{bmatrix}$. The complex solutions of $\mathbf{x}' = A\mathbf{x}$ are complex linear combinations of

$$\mathbf{x}_1(t) = \begin{bmatrix} i \\ 2 \end{bmatrix} e^{(-2+5i)t} \quad \text{and} \quad \mathbf{x}_2(t) = \begin{bmatrix} -i \\ 2 \end{bmatrix} e^{(-2-5i)t}$$

Next, use equation (7) to write

$$\mathbf{x}_1(t) = \begin{bmatrix} i \\ 2 \end{bmatrix} e^{-2t}(\cos 5t + i \sin 5t)$$

The real and imaginary parts of $\mathbf{x}_1$ provide real solutions:

$$\mathbf{y}_1(t) = \begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix} e^{-2t}, \qquad \mathbf{y}_2(t) = \begin{bmatrix} \cos 5t \\ 2\sin 5t \end{bmatrix} e^{-2t}$$

[1] Since $\mathbf{x}_2(t)$ is the complex conjugate of $\mathbf{x}_1(t)$, the real and imaginary parts of $\mathbf{x}_2(t)$ are $\mathbf{y}_1(t)$ and $-\mathbf{y}_2(t)$, respectively. Thus one can use either $\mathbf{x}_1(t)$ or $\mathbf{x}_2(t)$, but not both, to produce two real linearly independent solutions of $\mathbf{x}' = A\mathbf{x}$.

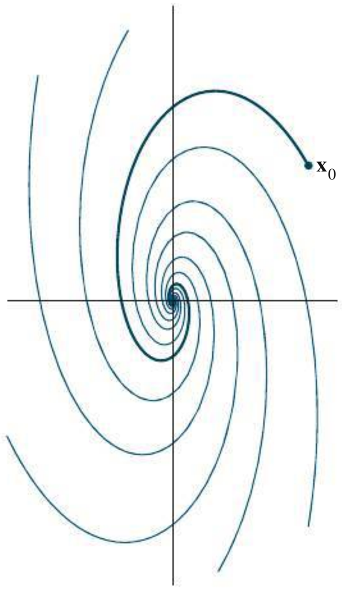


FIGURE 5
The origin as a spiral point.

Since $\mathbf{y}_1$ and $\mathbf{y}_2$ are linearly independent functions, they form a basis for the two-dimensional real vector space of solutions of $\mathbf{x}' = A\mathbf{x}$. Thus the general solution is

$$\mathbf{x}(t) = c_1\begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix}e^{-2t} + c_2\begin{bmatrix} \cos 5t \\ 2\sin 5t \end{bmatrix}e^{-2t}$$

To satisfy $\mathbf{x}(0) = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, we need $c_1\begin{bmatrix} 0 \\ 2 \end{bmatrix} + c_2\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, which leads to $c_1 = 1.5$ and $c_2 = 3$. Thus

$$\mathbf{x}(t) = 1.5\begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix}e^{-2t} + 3\begin{bmatrix} \cos 5t \\ 2\sin 5t \end{bmatrix}e^{-2t}$$

or

$$\begin{bmatrix} i_L(t) \\ v_C(t) \end{bmatrix} = \begin{bmatrix} -1.5\sin 5t + 3\cos 5t \\ 3\cos 5t + 6\sin 5t \end{bmatrix}e^{-2t}$$

See Figure 5. ■

In Figure 5, the origin is called a **spiral point** of the dynamical system. The rotation is caused by the sine and cosine functions that arise from a complex eigenvalue. The trajectories spiral inward because the factor e^{-2t} tends to zero. Recall that -2 is the real part of the eigenvalue in Example 3. When A has a complex eigenvalue with positive real part, the trajectories spiral outward. If the real part of the eigenvalue is zero, the trajectories form ellipses around the origin.

Practice Problems

A real 3×3 matrix A has eigenvalues $-.5$, $.2 + .3i$, and $.2 - .3i$, with corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1+2i \\ 4i \\ 2 \end{bmatrix}, \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} 1-2i \\ -4i \\ 2 \end{bmatrix}$$

1. Is A diagonalizable as $A = PDP^{-1}$, using complex matrices?
2. Write the general solution of $\mathbf{x}' = A\mathbf{x}$ using complex eigenfunctions, and then find the general real solution.
3. Describe the shapes of typical trajectories.

5.7 Exercises

1. A particle moving in a planar force field has a position vector $\mathbf{x}$ that satisfies $\mathbf{x}' = A\mathbf{x}$. The 2×2 matrix A has eigenvalues 4 and 2, with corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Find the position of the particle at time t, assuming that $\mathbf{x}(0) = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$.

2. Let A be a 2×2 matrix with eigenvalues -3 and -1 and corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Let $\mathbf{x}(t)$ be the position of a particle at time t. Solve the initial value problem $\mathbf{x}' = A\mathbf{x}$, $\mathbf{x}(0) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

In Exercises 3–6, solve the initial value problem $\mathbf{x}'(t) = A\mathbf{x}(t)$ for $t \geq 0$, with $\mathbf{x}(0) = (3, 2)$. Classify the nature of the origin as an attractor, repeller, or saddle point of the dynamical system described by $\mathbf{x}' = A\mathbf{x}$. Find the directions of greatest attraction and/or repulsion. When the origin is a saddle point, sketch typical trajectories.

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

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

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

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

In Exercises 7 and 8, make a change of variable that decouples the equation $\mathbf{x}' = A\mathbf{x}$. Write the equation $\mathbf{x}(t) = P\mathbf{y}(t)$ and show the calculation that leads to the uncoupled system $\mathbf{y}' = D\mathbf{y}$, specifying P and D.

7. A as in Exercise 5

8. A as in Exercise 6

In Exercises 9–18, construct the general solution of $\mathbf{x}' = A\mathbf{x}$ involving complex eigenfunctions and then obtain the general real solution. Describe the shapes of typical trajectories.

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

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

11. $A = \begin{bmatrix} -3 & -9 \\ 2 & 3 \end{bmatrix}$

12. $A = \begin{bmatrix} -7 & 10 \\ -4 & 5 \end{bmatrix}$

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

14. $A = \begin{bmatrix} -2 & 1 \\ -8 & 2 \end{bmatrix}$

15. $A = \begin{bmatrix} -8 & -12 & -6 \\ 2 & 1 & 2 \\ 7 & 12 & 5 \end{bmatrix}$

16. $A = \begin{bmatrix} -6 & -11 & 16 \\ 2 & 5 & -4 \\ -4 & -5 & 10 \end{bmatrix}$

17. $A = \begin{bmatrix} 30 & 64 & 23 \\ -11 & -23 & -9 \\ 6 & 15 & 4 \end{bmatrix}$

18. $A = \begin{bmatrix} 53 & -30 & -2 \\ 90 & -52 & -3 \\ 20 & -10 & 2 \end{bmatrix}$

19. Find formulas for the voltages v_1 and v_2 (as functions of time t) for the circuit in Example 1, assuming that $R_1 = 1/5$ ohm, $R_2 = 1/3$ ohm, $C_1 = 4$ farads, $C_2 = 3$ farads, and the initial charge on each capacitor is 4 volts.

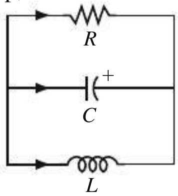
20. Find formulas for the voltages v_1 and v_2 for the circuit in Example 1, assuming that $R_1 = 1/15$ ohm, $R_2 = 1/3$ ohm, $C_1 = 9$ farads, $C_2 = 2$ farads, and the initial charge on each capacitor is 3 volts.

21. Find formulas for the current i_L and the voltage v_C for the circuit in Example 3, assuming that $R_1 = 1$ ohm, $R_2 = .125$ ohm, $C = .2$ farad, $L = .125$ henry, the initial current is 0 amp, and the initial voltage is 15 volts.

22. The circuit in the figure is described by the equation

$$\begin{bmatrix} i_L' \\ v_C' \end{bmatrix} = \begin{bmatrix} 0 & 1/L \\ -1/C & -1/(RC) \end{bmatrix} \begin{bmatrix} i_L \\ v_C \end{bmatrix}$$

where i_L is the current through the inductor L and v_C is the voltage drop across the capacitor C. Find formulas for i_L and v_C when $R = .5$ ohm, $C = 2.5$ farads, $L = .5$ henry, the initial current is 0 amp, and the initial voltage is 12 volts.



Solutions to Practice Problems

1. Yes, the 3×3 matrix is diagonalizable because it has three distinct eigenvalues. Theorem 2 in Section 5.1 and Theorem 6 in Section 5.3 are valid when complex scalars are used. (The proofs are essentially the same as for real scalars.)

2. The general solution has the form

$$\mathbf{x}(t) = c_1 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} 1+2i \\ 4i \\ 2 \end{bmatrix} e^{(.2+.3i)t} + c_3 \begin{bmatrix} 1-2i \\ -4i \\ 2 \end{bmatrix} e^{(.2-.3i)t}$$

The scalars c_1, c_2, and c_3 here can be any complex numbers. The first term in $\mathbf{x}(t)$ is real, provided c_1 is real. Two more real solutions can be produced using the real and imaginary parts of the second term in $\mathbf{x}(t)$:

$$\begin{bmatrix} 1+2i \\ 4i \\ 2 \end{bmatrix} e^{.2t}(\cos .3t + i \sin .3t)$$

The general real solution has the following form, with *real* scalars c_1, c_2, and c_3:

$$c_1 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} \cos .3t - 2\sin .3t \\ -4\sin .3t \\ 2\cos .3t \end{bmatrix} e^{.2t} + c_3 \begin{bmatrix} \sin .3t + 2\cos .3t \\ 4\cos .3t \\ 2\sin .3t \end{bmatrix} e^{.2t}$$

3. Any solution with $c_2 = c_3 = 0$ is attracted to the origin because of the negative exponential factor. Other solutions have components that grow without bound, and the trajectories spiral outward.

Be careful not to mistake this problem for one in Section 5.6. There the condition for attraction toward $\mathbf{0}$ was that an eigenvalue be less than 1 in magnitude, to make $|\lambda|^k \to 0$. Here the real part of the eigenvalue must be negative, to make $e^{\lambda t} \to 0$.

5.8 Iterative Estimates for Eigenvalues

In scientific applications of linear algebra, eigenvalues are seldom known precisely. Fortunately, a close numerical approximation is usually quite satisfactory. In fact, some applications require only a rough approximation to the largest eigenvalue. The first algorithm described below can work well for this case. Also, it provides a foundation for a more powerful method that can give fast estimates for other eigenvalues as well.

The Power Method

The power method applies to an $n \times n$ matrix A with a **strictly dominant eigenvalue** λ_1, which means that λ_1 must be larger in absolute value than all the other eigenvalues. In this case, the power method produces a scalar sequence that approaches λ_1 and a vector sequence that approaches a corresponding eigenvector. The background for the method rests on the eigenvector decomposition used at the beginning of Section 5.6.

Assume for simplicity that A is diagonalizable and $\mathbb{R}^n$ has a basis of eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, arranged so their corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ decrease in size, with the strictly dominant eigenvalue first. That is,

$$\underset{\uparrow \text{ Strictly larger}}{|\lambda_1| > |\lambda_2|} \geq |\lambda_3| \geq \cdots \geq |\lambda_n| \tag{1}$$

As we saw in equation (2) of Section 5.6, if $\mathbf{x}$ in $\mathbb{R}^n$ is written as $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, then

$$A^k\mathbf{x} = c_1(\lambda_1)^k\mathbf{v}_1 + c_2(\lambda_2)^k\mathbf{v}_2 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n \quad (k = 1, 2, \dots)$$

Assume $c_1 \neq 0$. Then, dividing by $(\lambda_1)^k$,

$$\frac{1}{(\lambda_1)^k}A^k\mathbf{x} = c_1\mathbf{v}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\mathbf{v}_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\mathbf{v}_n \quad (k = 1, 2, \dots) \tag{2}$$

From inequality (1), the fractions $\lambda_2/\lambda_1, \dots, \lambda_n/\lambda_1$ are all less than 1 in magnitude and so their powers go to zero. Hence

$$(\lambda_1)^{-k}A^k\mathbf{x} \to c_1\mathbf{v}_1 \quad \text{as } k \to \infty \tag{3}$$

Thus, for large k, a scalar multiple of $A^k\mathbf{x}$ determines almost the same *direction* as the eigenvector $c_1\mathbf{v}_1$. Since positive scalar multiples do not change the direction of a vector, $A^k\mathbf{x}$ itself points almost in the same direction as $\mathbf{v}_1$ or $-\mathbf{v}_1$, provided $c_1 \neq 0$.

EXAMPLE 1 Let $A = \begin{bmatrix} 1.8 & .8 \\ .2 & 1.2 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} -.5 \\ 1 \end{bmatrix}$. Then A has eigenvalues 2 and 1, and the eigenspace for $\lambda_1 = 2$ is the line through $\mathbf{0}$ and $\mathbf{v}_1$. For $k = 0, \dots, 8$, compute $A^k\mathbf{x}$ and construct the line through $\mathbf{0}$ and $A^k\mathbf{x}$. What happens as k increases?

SOLUTION The first three calculations are

$$A\mathbf{x} = \begin{bmatrix} 1.8 & .8 \\ .2 & 1.2 \end{bmatrix}\begin{bmatrix} -.5 \\ 1 \end{bmatrix} = \begin{bmatrix} -.1 \\ 1.1 \end{bmatrix}$$

$$A^2\mathbf{x} = A(A\mathbf{x}) = \begin{bmatrix} 1.8 & .8 \\ .2 & 1.2 \end{bmatrix}\begin{bmatrix} -.1 \\ 1.1 \end{bmatrix} = \begin{bmatrix} .7 \\ 1.3 \end{bmatrix}$$

$$A^3\mathbf{x} = A(A^2\mathbf{x}) = \begin{bmatrix} 1.8 & .8 \\ .2 & 1.2 \end{bmatrix}\begin{bmatrix} .7 \\ 1.3 \end{bmatrix} = \begin{bmatrix} 2.3 \\ 1.7 \end{bmatrix}$$

Analogous calculations complete Table 1.

TABLE 1 Iterates of a Vector

k	0	1	2	3	4	5	6	7	8
$A^k\mathbf{x}$	$\begin{bmatrix} -.5 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -.1 \\ 1.1 \end{bmatrix}$	$\begin{bmatrix} .7 \\ 1.3 \end{bmatrix}$	$\begin{bmatrix} 2.3 \\ 1.7 \end{bmatrix}$	$\begin{bmatrix} 5.5 \\ 2.5 \end{bmatrix}$	$\begin{bmatrix} 11.9 \\ 4.1 \end{bmatrix}$	$\begin{bmatrix} 24.7 \\ 7.3 \end{bmatrix}$	$\begin{bmatrix} 50.3 \\ 13.7 \end{bmatrix}$	$\begin{bmatrix} 101.5 \\ 26.5 \end{bmatrix}$

The vectors $\mathbf{x}$, $A\mathbf{x}, \dots, A^4\mathbf{x}$ are shown in Figure 1. The other vectors are growing too long to display. However, line segments are drawn showing the directions of those vectors. In fact, the directions of the vectors are what we really want to see, not the vectors themselves. The lines seem to be approaching the line representing the eigenspace spanned by $\mathbf{v}_1$. More precisely, the angle between the line (subspace) determined by $A^k\mathbf{x}$ and the line (eigenspace) determined by $\mathbf{v}_1$ goes to zero as $k \to \infty$. ■

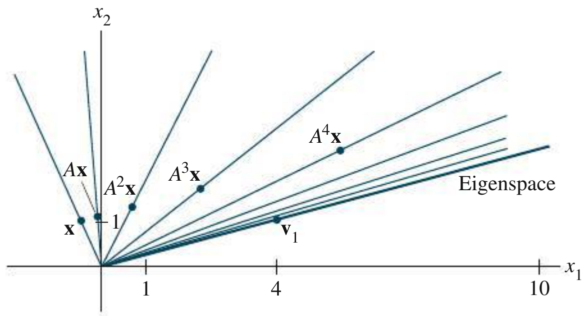


FIGURE 1 Directions determined by $\mathbf{x}$, $A\mathbf{x}$, $A^2\mathbf{x}, \dots, A^7\mathbf{x}$.

The vectors $(\lambda_1)^{-k}A^k\mathbf{x}$ in (3) are scaled to make them converge to $c_1\mathbf{v}_1$, provided $c_1 \neq 0$. We cannot scale $A^k\mathbf{x}$ in this way because we do not know λ_1. But we can scale each $A^k\mathbf{x}$ to make its largest entry a 1. It turns out that the resulting sequence $\{\mathbf{x}_k\}$ will converge to a multiple of $\mathbf{v}_1$ whose largest entry is 1. Figure 2 shows the scaled sequence for Example 1. The eigenvalue λ_1 can be estimated from the sequence $\{\mathbf{x}_k\}$, too. When $\mathbf{x}_k$

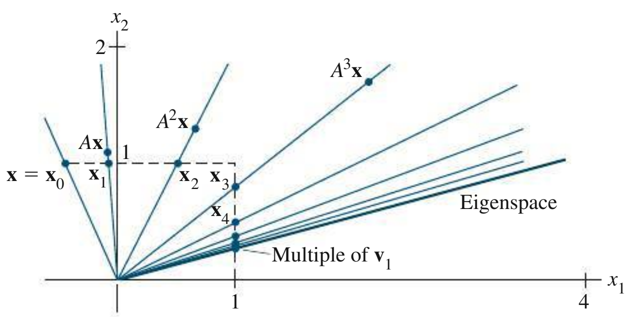


FIGURE 2 Scaled multiples of $\mathbf{x}$, $A\mathbf{x}$, $A^2\mathbf{x}, \dots, A^7\mathbf{x}$.

is close to an eigenvector for λ_1, the vector $A\mathbf{x}_k$ is close to $\lambda_1\mathbf{x}_k$, with each entry in $A\mathbf{x}_k$ approximately λ_1 times the corresponding entry in $\mathbf{x}_k$. Because the largest entry in $\mathbf{x}_k$ is 1, the largest entry in $A\mathbf{x}_k$ is close to λ_1. (Careful proofs of these statements are omitted.)

THE POWER METHOD FOR ESTIMATING A STRICTLY DOMINANT EIGENVALUE

1. Select an initial vector $\mathbf{x}_0$ whose largest entry is 1.

2. For $k = 0, 1, \dots,$

a. Compute $A\mathbf{x}_k$.

b. Let μ_k be an entry in $A\mathbf{x}_k$ whose absolute value is as large as possible.

c. Compute $\mathbf{x}_{k+1} = (1/\mu_k)A\mathbf{x}_k$.

3. For almost all choices of $\mathbf{x}_0$, the sequence $\{\mu_k\}$ approaches the dominant eigenvalue, and the sequence $\{\mathbf{x}_k\}$ approaches a corresponding eigenvector.

EXAMPLE 2 Apply the power method to $A = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}$ with $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Stop when $k = 5$, and estimate the dominant eigenvalue and a corresponding eigenvector of A.

SOLUTION Calculations in this example and the next were made with MATLAB, which computes with 16-digit accuracy, although we show only a few significant figures here. To begin, compute $A\mathbf{x}_0$ and identify the largest entry μ_0 in $A\mathbf{x}_0$:

$$A\mathbf{x}_0 = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \qquad \mu_0 = 5$$

Scale $A\mathbf{x}_0$ by $1/\mu_0$ to get $\mathbf{x}_1$, compute $A\mathbf{x}_1$, and identify the largest entry in $A\mathbf{x}_1$:

$$\mathbf{x}_1 = \frac{1}{\mu_0}A\mathbf{x}_0 = \frac{1}{5}\begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ .4 \end{bmatrix}$$

$$A\mathbf{x}_1 = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ .4 \end{bmatrix} = \begin{bmatrix} 8 \\ 1.8 \end{bmatrix}, \qquad \mu_1 = 8$$

Scale $A\mathbf{x}_1$ by $1/\mu_1$ to get $\mathbf{x}_2$, compute $A\mathbf{x}_2$, and identify the largest entry in $A\mathbf{x}_2$:

$$\mathbf{x}_2 = \frac{1}{\mu_1}A\mathbf{x}_1 = \frac{1}{8}\begin{bmatrix} 8 \\ 1.8 \end{bmatrix} = \begin{bmatrix} 1 \\ .225 \end{bmatrix}$$

$$A\mathbf{x}_2 = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ .225 \end{bmatrix} = \begin{bmatrix} 7.125 \\ 1.450 \end{bmatrix}, \qquad \mu_2 = 7.125$$

Scale $A\mathbf{x}_2$ by $1/\mu_2$ to get $\mathbf{x}_3$, and so on. The results of MATLAB calculations for the first five iterations are arranged in Table 2.

TABLE 2 The Power Method for Example 2

k	**0**	**1**	**2**	**3**	**4**	**5**
$\mathbf{x}_k$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ .4 \end{bmatrix}$	$\begin{bmatrix} 1 \\ .225 \end{bmatrix}$	$\begin{bmatrix} 1 \\ .2035 \end{bmatrix}$	$\begin{bmatrix} 1 \\ .2005 \end{bmatrix}$	$\begin{bmatrix} 1 \\ .20007 \end{bmatrix}$
$A\mathbf{x}_k$	$\begin{bmatrix} 5 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 1.8 \end{bmatrix}$	$\begin{bmatrix} 7.125 \\ 1.450 \end{bmatrix}$	$\begin{bmatrix} 7.0175 \\ 1.4070 \end{bmatrix}$	$\begin{bmatrix} 7.0025 \\ 1.4010 \end{bmatrix}$	$\begin{bmatrix} 7.00036 \\ 1.40014 \end{bmatrix}$
μ_k	5	8	7.125	7.0175	7.0025	7.00036

The evidence from Table 2 strongly suggests that $\{\mathbf{x}_k\}$ approaches $(1, .2)$ and $\{\mu_k\}$ approaches 7. If so, then $(1, .2)$ is an eigenvector and 7 is the dominant eigenvalue. This is easily verified by computing

$$A\begin{bmatrix} 1 \\ .2 \end{bmatrix} = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ .2 \end{bmatrix} = \begin{bmatrix} 7 \\ 1.4 \end{bmatrix} = 7\begin{bmatrix} 1 \\ .2 \end{bmatrix} \qquad \blacksquare$$

The sequence $\{\mu_k\}$ in Example 2 converged quickly to $\lambda_1 = 7$ because the second eigenvalue of A was much smaller. (In fact, $\lambda_2 = 1$.) In general, the rate of convergence depends on the ratio $|\lambda_2/\lambda_1|$, because the vector $c_2(\lambda_2/\lambda_1)^k\mathbf{v}_2$ in equation (2) is the main source of error when using a scaled version of $A^k\mathbf{x}$ as an estimate of $c_1\mathbf{v}_1$. (The other fractions λ_j/λ_1 are likely to be smaller.) If $|\lambda_2/\lambda_1|$ is close to 1, then $\{\mu_k\}$ and $\{\mathbf{x}_k\}$ can converge very slowly, and other approximation methods may be preferred.

With the power method, there is a slight chance that the chosen initial vector $\mathbf{x}$ will have no component in the $\mathbf{v}_1$ direction (when $c_1 = 0$). But computer rounding errors during the calculations of the $\mathbf{x}_k$ are likely to create a vector with at least a small component in the direction of $\mathbf{v}_1$. If that occurs, the $\mathbf{x}_k$ will start to converge to a multiple of $\mathbf{v}_1$.

The Inverse Power Method

This method provides an approximation for *any* eigenvalue, provided a good initial estimate α of the eigenvalue λ is known. In this case, we let $B = (A - \alpha I)^{-1}$ and apply the power method to B. It can be shown that if the eigenvalues of A are $\lambda_1, \dots, \lambda_n$, then the eigenvalues of B are

$$\frac{1}{\lambda_1 - \alpha}, \quad \frac{1}{\lambda_2 - \alpha}, \quad \dots, \quad \frac{1}{\lambda_n - \alpha}$$

and the corresponding eigenvectors are the same as those for A. (See Exercises 15 and 16.)

Suppose, for example, that α is closer to λ_2 than to the other eigenvalues of A. Then $1/(\lambda_2 - \alpha)$ will be a strictly dominant eigenvalue of B. If α is really close to λ_2, then $1/(\lambda_2 - \alpha)$ is *much* larger than the other eigenvalues of B, and the inverse power method produces a very rapid approximation to λ_2 for almost all choices of $\mathbf{x}_0$. The following algorithm gives the details.

THE INVERSE POWER METHOD FOR ESTIMATING AN EIGENVALUE λ OF A

1. Select an initial estimate α sufficiently close to λ.
2. Select an initial vector $\mathbf{x}_0$ whose largest entry is 1.
3. For $k = 0, 1, \dots,$
   a. Solve $(A - \alpha I)\mathbf{y}_k = \mathbf{x}_k$ for $\mathbf{y}_k$.
   b. Let μ_k be an entry in $\mathbf{y}_k$ with the largest absolute value.
   c. Compute $\nu_k = \alpha + (1/\mu_k)$.
   d. Compute $\mathbf{x}_{k+1} = (1/\mu_k)\mathbf{y}_k$.
4. For almost all choices of $\mathbf{x}_0$, the sequence $\{\nu_k\}$ approaches the eigenvalue λ of A, and the sequence $\{\mathbf{x}_k\}$ approaches a corresponding eigenvector.

Notice that B, or rather $(A - \alpha I)^{-1}$, does not appear in the algorithm. Instead of computing $(A - \alpha I)^{-1}\mathbf{x}_k$ to get the next vector in the sequence, it is better to *solve* the equation $(A - \alpha I)\mathbf{y}_k = \mathbf{x}_k$ for $\mathbf{y}_k$ (and then scale $\mathbf{y}_k$ to produce $\mathbf{x}_{k+1}$). Since this equation for $\mathbf{y}_k$ must be solved for each k, an LU factorization of $A - \alpha I$ will speed up the process.

EXAMPLE 3 It is not uncommon in some applications to need to know the smallest eigenvalue of a matrix A and to have at hand rough estimates of the eigenvalues. Suppose 21, 3.3, and 1.9 are estimates for the eigenvalues of the matrix A below. Find the smallest eigenvalue, accurate to six decimal places.

$$A = \begin{bmatrix} 10 & -8 & -4 \\ -8 & 13 & 4 \\ -4 & 5 & 4 \end{bmatrix}$$

SOLUTION The two smallest eigenvalues seem close together, so we use the inverse power method for $A - 1.9I$. Results of a MATLAB calculation are shown in Table 3. Here $\mathbf{x}_0$ was chosen arbitrarily, $\mathbf{y}_k = (A - 1.9I)^{-1}\mathbf{x}_k$, μ_k is the largest entry in $\mathbf{y}_k$, $\nu_k = 1.9 + 1/\mu_k$, and $\mathbf{x}_{k+1} = (1/\mu_k)\mathbf{y}_k$. As it turns out, the initial eigenvalue estimate was fairly good, and the inverse power sequence converged quickly. The smallest eigenvalue is exactly 2. ■

TABLE 3 The Inverse Power Method

k	0	1	2	3	4
$\mathbf{x}_k$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} .5736 \\ .0646 \\ 1 \end{bmatrix}$	$\begin{bmatrix} .5054 \\ .0045 \\ 1 \end{bmatrix}$	$\begin{bmatrix} .5004 \\ .0003 \\ 1 \end{bmatrix}$	$\begin{bmatrix} .50003 \\ .00002 \\ 1 \end{bmatrix}$
$\mathbf{y}_k$	$\begin{bmatrix} 4.45 \\ .50 \\ 7.76 \end{bmatrix}$	$\begin{bmatrix} 5.0131 \\ .0442 \\ 9.9197 \end{bmatrix}$	$\begin{bmatrix} 5.0012 \\ .0031 \\ 9.9949 \end{bmatrix}$	$\begin{bmatrix} 5.0001 \\ .0002 \\ 9.9996 \end{bmatrix}$	$\begin{bmatrix} 5.000006 \\ .000015 \\ 9.999975 \end{bmatrix}$
μ_k	7.76	9.9197	9.9949	9.9996	9.999975
ν_k	2.03	2.0008	2.00005	2.000004	2.0000002

If an estimate for the smallest eigenvalue of a matrix is not available, one can simply take $\alpha = 0$ in the inverse power method. This choice of α works reasonably well if the smallest eigenvalue is much closer to zero than to the other eigenvalues.

The two algorithms presented in this section are practical tools for many simple situations, and they provide an introduction to the problem of eigenvalue estimation. A more robust and widely used iterative method is the QR algorithm. For instance, it is the heart of the MATLAB command `eig(A)`, which rapidly computes eigenvalues and eigenvectors of A. A brief description of the QR algorithm was given in the exercises for Section 5.2. Further details are presented in most modern numerical analysis texts.

Practice Problem

How can you tell if a given vector $\mathbf{x}$ is a good approximation to an eigenvector of a matrix A? If it is, how would you estimate the corresponding eigenvalue? Experiment with

$$A = \begin{bmatrix} 5 & 8 & 4 \\ 8 & 3 & -1 \\ 4 & -1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1.0 \\ -4.3 \\ 8.1 \end{bmatrix}$$

5.8 Exercises

In Exercises 1–4, the matrix A is followed by a sequence $\{\mathbf{x}_k\}$ produced by the power method. Use these data to estimate the largest eigenvalue of A, and give a corresponding eigenvector.

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

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ .25 \end{bmatrix}, \begin{bmatrix} 1 \\ .3158 \end{bmatrix}, \begin{bmatrix} 1 \\ .3298 \end{bmatrix}, \begin{bmatrix} 1 \\ .3326 \end{bmatrix}$

2. $A = \begin{bmatrix} 1.8 & -.8 \\ -3.2 & 4.2 \end{bmatrix}$;

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -.5625 \\ 1 \end{bmatrix}, \begin{bmatrix} -.3021 \\ 1 \end{bmatrix}, \begin{bmatrix} -.2601 \\ 1 \end{bmatrix}, \begin{bmatrix} -.2520 \\ 1 \end{bmatrix}$

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

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ .8 \end{bmatrix}, \begin{bmatrix} .6875 \\ 1 \end{bmatrix}, \begin{bmatrix} .5577 \\ 1 \end{bmatrix}, \begin{bmatrix} .5188 \\ 1 \end{bmatrix}$

4. $A = \begin{bmatrix} 4.1 & -6 \\ 3 & -4.4 \end{bmatrix}$;

$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ .7368 \end{bmatrix}, \begin{bmatrix} 1 \\ .7541 \end{bmatrix}, \begin{bmatrix} 1 \\ .7490 \end{bmatrix}, \begin{bmatrix} 1 \\ .7502 \end{bmatrix}$

5. Let $A = \begin{bmatrix} 15 & 16 \\ -20 & -21 \end{bmatrix}$. The vectors $\mathbf{x}, \ldots, A^5\mathbf{x}$ are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$,

$\begin{bmatrix} 31 \\ -41 \end{bmatrix}, \begin{bmatrix} -191 \\ 241 \end{bmatrix}, \begin{bmatrix} 991 \\ -1241 \end{bmatrix}, \begin{bmatrix} -4991 \\ 6241 \end{bmatrix}, \begin{bmatrix} 24991 \\ -31241 \end{bmatrix}.$

Find a vector with a 1 in the second entry that is close to an eigenvector of A. Use four decimal places. Check your estimate, and give an estimate for the dominant eigenvalue of A.

6. Let $A = \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix}$. Repeat Exercise 5, using the following sequence $\mathbf{x}, A\mathbf{x}, \ldots, A^5\mathbf{x}$.

$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -5 \\ 13 \end{bmatrix}, \begin{bmatrix} -29 \\ 61 \end{bmatrix}, \begin{bmatrix} -125 \\ 253 \end{bmatrix}, \begin{bmatrix} -509 \\ 1021 \end{bmatrix}, \begin{bmatrix} -2045 \\ 4093 \end{bmatrix}$

Exercises 7–12 require MATLAB or other computational aid. In Exercises 7 and 8, use the power method with the $\mathbf{x}_0$ given. List $\{\mathbf{x}_k\}$ and $\{\mu_k\}$ for $k = 1, \ldots, 5$. In Exercises 9 and 10, list μ_5 and μ_6.

T 7. $A = \begin{bmatrix} 6 & 7 \\ 8 & 5 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

T 8. $A = \begin{bmatrix} 2 & 1 \\ 4 & 5 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

T 9. $A = \begin{bmatrix} 8 & 0 & 12 \\ 1 & -2 & 1 \\ 0 & 3 & 0 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

T 10. $A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 9 \\ 0 & 1 & 9 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

Another estimate can be made for an eigenvalue when an approximate eigenvector is available. Observe that if $A\mathbf{x} = \lambda\mathbf{x}$, then $\mathbf{x}^TA\mathbf{x} = \mathbf{x}^T(\lambda\mathbf{x}) = \lambda(\mathbf{x}^T\mathbf{x})$, and the **Rayleigh quotient**

$$R(\mathbf{x}) = \frac{\mathbf{x}^TA\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

equals λ. If $\mathbf{x}$ is close to an eigenvector for λ, then this quotient is close to λ. When A is a symmetric matrix ($A^T = A$), the Rayleigh quotient $R(\mathbf{x}_k) = (\mathbf{x}_k^TA\mathbf{x}_k)/(\mathbf{x}_k^T\mathbf{x}_k)$ will have roughly twice as many digits of accuracy as the scaling factor μ_k in the power method. Verify this increased accuracy in Exercises 11 and 12 by computing μ_k and $R(\mathbf{x}_k)$ for $k = 1, \ldots, 4$.

T 11. $A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

T 12. $A = \begin{bmatrix} -3 & 2 \\ 2 & 0 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Exercises 13 and 14 apply to a 3×3 matrix A whose eigenvalues are estimated to be 4, -4, and 3.

13. If the eigenvalues close to 4 and -4 are known to have different absolute values, will the power method work? Is it likely to be useful?

14. Suppose the eigenvalues close to 4 and -4 are known to have exactly the same absolute value. Describe how one might obtain a sequence that estimates the eigenvalue close to 4.

15. Suppose $A\mathbf{x} = \lambda\mathbf{x}$ with $\mathbf{x} \neq \mathbf{0}$. Let α be a scalar different from the eigenvalues of A, and let $B = (A - \alpha I)^{-1}$. Subtract $\alpha\mathbf{x}$ from both sides of the equation $A\mathbf{x} = \lambda\mathbf{x}$, and use algebra to show that $1/(\lambda - \alpha)$ is an eigenvalue of B, with $\mathbf{x}$ a corresponding eigenvector.

16. Suppose μ is an eigenvalue of the B in Exercise 15, and that $\mathbf{x}$ is a corresponding eigenvector, so that $(A - \alpha I)^{-1}\mathbf{x} = \mu\mathbf{x}$. Use this equation to find an eigenvalue of A in terms of μ and α. [*Note:* $\mu \neq 0$ because B is invertible.]

T 17. Use the inverse power method to estimate the middle eigenvalue of the A in Example 3, with accuracy to four decimal places. Set $\mathbf{x}_0 = (1, 0, 0)$.

T 18. Let A be as in Exercise 9. Use the inverse power method with $\mathbf{x}_0 = (1, 0, 0)$ to estimate the eigenvalue of A near $\alpha = -1.4$, with an accuracy to four decimal places.

In Exercises 19 and 20, find (a) the largest eigenvalue and (b) the eigenvalue closest to zero. In each case, set $\mathbf{x}_0 = (1, 0, 0, 0)$ and carry out approximations until the approximating sequence seems accurate to four decimal places. Include the approximate eigenvector.

19. $A = \begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix}$

20. $A = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & 12 & 13 & 11 \\ -2 & 3 & 0 & 2 \\ 4 & 5 & 7 & 2 \end{bmatrix}$

21. A common misconception is that if A has a strictly dominant eigenvalue, then, for any sufficiently large value of k, the vector $A^k\mathbf{x}$ is approximately equal to an eigenvector of A. For the three matrices below, study what happens to $A^k\mathbf{x}$ when $\mathbf{x} = (.5, .5)$, and try to draw general conclusions (for a 2×2 matrix).

a. $A = \begin{bmatrix} .8 & 0 \\ 0 & .2 \end{bmatrix}$ b. $A = \begin{bmatrix} 1 & 0 \\ 0 & .8 \end{bmatrix}$

c. $A = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}$

Solution to Practice Problem

For the given A and $\mathbf{x}$,

$$A\mathbf{x} = \begin{bmatrix} 5 & 8 & 4 \\ 8 & 3 & -1 \\ 4 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1.00 \\ -4.30 \\ 8.10 \end{bmatrix} = \begin{bmatrix} 3.00 \\ -13.00 \\ 24.50 \end{bmatrix}$$

If $A\mathbf{x}$ is nearly a multiple of $\mathbf{x}$, then the ratios of corresponding entries in the two vectors should be nearly constant. So compute:

$$\begin{array}{ccccc} \{\text{entry in } A\mathbf{x}\} & \div & \{\text{entry in } \mathbf{x}\} & = & \{\text{ratio}\} \\ 3.00 & & 1.00 & & 3.000 \\ -13.00 & & -4.30 & & 3.023 \\ 24.50 & & 8.10 & & 3.025 \end{array}$$

Each entry in $A\mathbf{x}$ is about 3 times the corresponding entry in $\mathbf{x}$, so $\mathbf{x}$ is close to an eigenvector. Any of the ratios above is an estimate for the eigenvalue. (To five decimal places, the eigenvalue is 3.02409.)

5.9 Applications to Markov Chains

The Markov chains described in this section are used as mathematical models of a wide variety of situations in biology, business, chemistry, engineering, physics, and elsewhere. In each case, the model is used to describe an experiment or measurement that is performed many times in the same way, where the outcome of each trial of the experiment will be one of several specified possible outcomes, and where the outcome of one trial depends only on the immediately preceding trial.

For example, if the population of a city and its suburbs were measured each year, then a vector such as

$$\mathbf{x}_0 = \begin{bmatrix} .60 \\ .40 \end{bmatrix} \tag{1}$$

could indicate that 60% of the population lives in the city and 40% in the suburbs. The decimals in $\mathbf{x}_0$ add up to 1 because they account for the entire population of the region. Percentages are more convenient for our purposes here than population totals.

DEFINITION

A vector with nonnegative entries that add up to 1 is called a **probability vector**. A **stochastic matrix** is a square matrix whose columns are probability vectors.

A **Markov chain** is a sequence of probability vectors $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$, together with a stochastic matrix P, such that

$$\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P\mathbf{x}_1, \quad \mathbf{x}_3 = P\mathbf{x}_2, \quad \ldots$$

Thus the Markov chain is described by the first-order difference equation

$$\mathbf{x}_{k+1} = P\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \ldots$$

When a Markov chain of vectors in $\mathbb{R}^n$ describes a system or a sequence of experiments, the entries in $\mathbf{x}_k$ list, respectively, the probabilities that the system is in each of n possible states, or the probabilities that the outcome of the experiment is one of n possible outcomes. For this reason, $\mathbf{x}_k$ is often called a **state vector**.

EXAMPLE 1 Section 1.10 examined a model for population movement between a city and its suburbs. See Figure 1. The annual migration between these two parts of the metropolitan region was governed by the *migration matrix* M:

$$\begin{array}{cc} & \text{From} \\ & \begin{array}{cc} \text{City} & \text{Suburbs} \end{array} \quad \text{To} \\ M = & \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{array}{l} \text{City} \\ \text{Suburbs} \end{array} \end{array}$$

That is, each year 5% of the city population moves to the suburbs, and 3% of the suburban population moves to the city. The columns of M are probability vectors, so M is a stochastic matrix. Suppose the 2020 population of the region is 600,000 in the city and 400,000 in the suburbs. Then the initial distribution of the population in the region is given previously by $\mathbf{x}_0$ in (1). What is the distribution of the population in 2021? In 2022?

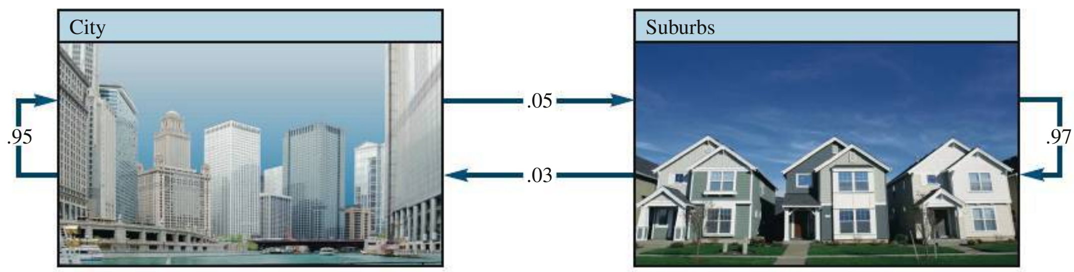


FIGURE 1 Annual percentage migration between city and suburbs.

SOLUTION In Example 3 of Section 1.10, we saw that after one year, the population vector $\begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$ changed to

$$\begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix} = \begin{bmatrix} 582{,}000 \\ 418{,}000 \end{bmatrix}$$

If we divide both sides of this equation by the total population of 1 million, and use the fact that $kM\mathbf{x} = M(k\mathbf{x})$, we find that

$$\begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}\begin{bmatrix} .600 \\ .400 \end{bmatrix} = \begin{bmatrix} .582 \\ .418 \end{bmatrix}$$

The vector $\mathbf{x}_1 = \begin{bmatrix} .582 \\ .418 \end{bmatrix}$ gives the population distribution in 2021. That is, 58.2% of the region lived in the city and 41.8% lived in the suburbs. Similarly, the population distribution in 2022 is described by a vector $\mathbf{x}_2$, where

$$\mathbf{x}_2 = M\mathbf{x}_1 = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}\begin{bmatrix} .582 \\ .418 \end{bmatrix} = \begin{bmatrix} .565 \\ .435 \end{bmatrix}$$ ■

EXAMPLE 2 Suppose the voting results of a congressional election at a certain voting precinct are represented by a vector $\mathbf{x}$ in $\mathbb{R}^3$:

$$\mathbf{x} = \begin{bmatrix} \text{\% voting Democratic (D)} \\ \text{\% voting Republican (R)} \\ \text{\% voting Other (O)} \end{bmatrix}$$

Suppose we record the outcome of the congressional election every two years by a vector of this type and the outcome of one election depends only on the results of the preceding election. Then the sequence of vectors that describe the votes every two years may be a Markov chain. As an example of a stochastic matrix P for this chain, we take

$$\begin{array}{c} \text{From} \\ \begin{array}{ccc} \text{D} & \text{R} & \text{O} \end{array} \end{array}$$

$$P = \begin{array}{c} \\ \left[\begin{array}{ccc} .70 & .10 & .30 \\ .20 & .80 & .30 \\ .10 & .10 & .40 \end{array}\right] \end{array} \begin{array}{c} \text{To} \\ \text{D} \\ \text{R} \\ \text{O} \end{array}$$

The entries in the first column, labeled D, describe what the persons voting Democratic in one election will do in the next election. Here we have supposed that 70% will vote D again in the next election, 20% will vote R, and 10% will vote O. Similar interpretations hold for the other columns of P. A diagram for this matrix is shown in Figure 2.

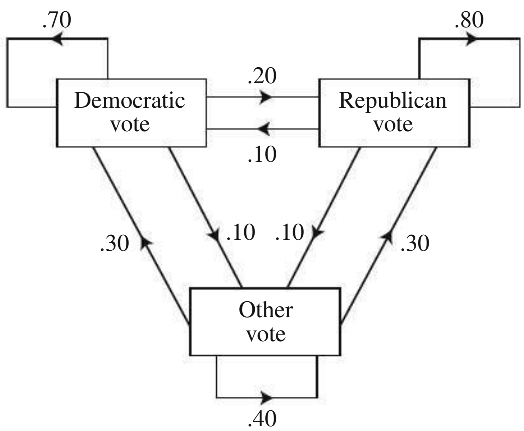


FIGURE 2 Voting changes from one election to the next.

If the "transition" percentages remain constant over many years from one election to the next, then the sequence of vectors that give the voting outcomes forms a Markov chain. Suppose the outcome of one election is given by

$$\mathbf{x}_0 = \begin{bmatrix} .55 \\ .40 \\ .05 \end{bmatrix}$$

Determine the likely outcome of the next election and the likely outcome of the election after that.

SOLUTION The outcome of the next election is described by the state vector $\mathbf{x}_1$ and that of the election after that by $\mathbf{x}_2$, where

$$\mathbf{x}_1 = P\mathbf{x}_0 = \begin{bmatrix} .70 & .10 & .30 \\ .20 & .80 & .30 \\ .10 & .10 & .40 \end{bmatrix} \begin{bmatrix} .55 \\ .40 \\ .05 \end{bmatrix} = \begin{bmatrix} .440 \\ .445 \\ .115 \end{bmatrix} \quad \begin{array}{l} \text{44\% will vote D.} \\ \text{44.5\% will vote R.} \\ \text{11.5\% will vote O.} \end{array}$$

$$\mathbf{x}_2 = P\mathbf{x}_1 = \begin{bmatrix} .70 & .10 & .30 \\ .20 & .80 & .30 \\ .10 & .10 & .40 \end{bmatrix} \begin{bmatrix} .440 \\ .445 \\ .115 \end{bmatrix} = \begin{bmatrix} .3870 \\ .4785 \\ .1345 \end{bmatrix} \quad \begin{array}{l} \text{38.7\% will vote D.} \\ \text{47.9\% will vote R.} \\ \text{13.5\% will vote O.} \end{array}$$

To understand why $\mathbf{x}_1$ does indeed give the outcome of the next election, suppose 1000 persons voted in the "first" election, with 550 voting D, 400 voting R, and 50 voting O. (See the percentages in $\mathbf{x}_0$.) In the next election, 70% of the 550 will vote D again, 10% of the 400 will switch from R to D, and 30% of the 50 will switch from O to D. Thus the total D vote will be

$$.70(550) + .10(400) + .30(50) = 385 + 40 + 15 = 440 \tag{2}$$

Thus 44% of the vote next time will be for the D candidate. The calculation in (2) is essentially the same as that used to compute the first entry in $\mathbf{x}_1$. Analogous calculations could be made for the other entries in $\mathbf{x}_1$, for the entries in $\mathbf{x}_2$, and so on. ■

Predicting the Distant Future

The most interesting aspect of Markov chains is the study of a chain's long-term behavior. For instance, what can be said in Example 2 about the voting after many elections have passed (assuming that the given stochastic matrix continues to describe the transition percentages from one election to the next)? Or, what happens to the population distribution in Example 1 "in the long run"? Here our work on eigenvalues and eigenvectors becomes helpful.

THEOREM 10

Stochastic Matrices

If P is a stochastic matrix, then 1 is an eigenvalue of P.

PROOF Since the columns of P sum to 1, the rows of P^T will also sum to 1. Let $\mathbf{e}$ represent the vector for which every entry is 1. Notice that multiplying P^T by $\mathbf{e}$ has the effect of adding up the values in each row, hence $P^T\mathbf{e} = \mathbf{e}$, establishing that $\mathbf{e}$ is an eigenvector of P^T with eigenvalue 1. Since P and P^T have the same eigenvalues (Exercise 20 in Section 5.2), 1 is also an eigenvalue of P. ■

In the next example, we see that the vectors generated in a Markov chain are almost the same as the vectors generated using the power method outlined in Section 5.8 – the only difference is that in a Markov chain, the vectors are not scaled at each step. Based on our experience from Section 5.8, as k increases we expect $\mathbf{x}_k \to \mathbf{q}$, where $\mathbf{q}$ is an eigenvector of P.

EXAMPLE 3 Let $P = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix}$ and $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Consider a system whose state is described by the Markov chain $\mathbf{x}_{k+1} = P\mathbf{x}_k$, for $k = 0, 1, \ldots$ What happens to the system as time passes? Compute the state vectors $\mathbf{x}_1, \ldots, \mathbf{x}_{15}$ to find out.

SOLUTION

$$\mathbf{x}_1 = P\mathbf{x}_0 = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} .5 \\ .3 \\ .2 \end{bmatrix}$$

$$\mathbf{x}_2 = P\mathbf{x}_1 = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix}\begin{bmatrix} .5 \\ .3 \\ .2 \end{bmatrix} = \begin{bmatrix} .37 \\ .45 \\ .18 \end{bmatrix}$$

$$\mathbf{x}_3 = P\mathbf{x}_2 = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix}\begin{bmatrix} .37 \\ .45 \\ .18 \end{bmatrix} = \begin{bmatrix} .329 \\ .525 \\ .146 \end{bmatrix}$$

The results of further calculations are shown below, with entries rounded to four or five significant figures.

$$\mathbf{x}_4 = \begin{bmatrix} .3133 \\ .5625 \\ .1242 \end{bmatrix}, \quad \mathbf{x}_5 = \begin{bmatrix} .3064 \\ .5813 \\ .1123 \end{bmatrix}, \quad \mathbf{x}_6 = \begin{bmatrix} .3032 \\ .5906 \\ .1062 \end{bmatrix}, \quad \mathbf{x}_7 = \begin{bmatrix} .3016 \\ .5953 \\ .1031 \end{bmatrix}$$

$$\mathbf{x}_8 = \begin{bmatrix} .3008 \\ .5977 \\ .1016 \end{bmatrix}, \quad \mathbf{x}_9 = \begin{bmatrix} .3004 \\ .5988 \\ .1008 \end{bmatrix}, \quad \mathbf{x}_{10} = \begin{bmatrix} .3002 \\ .5994 \\ .1004 \end{bmatrix}, \quad \mathbf{x}_{11} = \begin{bmatrix} .3001 \\ .5997 \\ .1002 \end{bmatrix}$$

$$\mathbf{x}_{12} = \begin{bmatrix} .30005 \\ .59985 \\ .10010 \end{bmatrix}, \quad \mathbf{x}_{13} = \begin{bmatrix} .30002 \\ .59993 \\ .10005 \end{bmatrix}, \quad \mathbf{x}_{14} = \begin{bmatrix} .30001 \\ .59996 \\ .10002 \end{bmatrix}, \quad \mathbf{x}_{15} = \begin{bmatrix} .30001 \\ .59998 \\ .10001 \end{bmatrix}$$

These vectors seem to be approaching $\mathbf{q} = \begin{bmatrix} .3 \\ .6 \\ .1 \end{bmatrix}$. The probabilities are hardly changing from one value of k to the next. Observe that the following calculation is exact (with no rounding error):

$$P\mathbf{q} = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix}\begin{bmatrix} .3 \\ .6 \\ .1 \end{bmatrix} = \begin{bmatrix} .15 + .12 + .03 \\ .09 + .48 + .03 \\ .06 + 0 + .04 \end{bmatrix} = \begin{bmatrix} .30 \\ .60 \\ .10 \end{bmatrix} = \mathbf{q}$$

When the system is in state $\mathbf{q}$, there is no change in the system from one measurement to the next. ■

Steady-State Vectors

If P is a stochastic matrix, then a **steady-state vector** (or **equilibrium vector**) for P is a probability vector $\mathbf{q}$ such that

$$P\mathbf{q} = \mathbf{q}$$

In Theorem 10, it is established that 1 is an eigenvalue of any stochastic matrix. It can be shown that 1 is actually the largest eigenvalue of a stochastic matrix and the associated eigenvector can be chosen to be a steady-state vector. In Example 3, $\mathbf{q}$ is a steady-state vector for P.

EXAMPLE 4 The probability vector $\mathbf{q} = \begin{bmatrix} .375 \\ .625 \end{bmatrix}$ is a steady-state vector for the population migration matrix M in Example 1, because

$$M\mathbf{q} = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}\begin{bmatrix} .375 \\ .625 \end{bmatrix} = \begin{bmatrix} .35625 + .01875 \\ .01875 + .60625 \end{bmatrix} = \begin{bmatrix} .375 \\ .625 \end{bmatrix} = \mathbf{q} \qquad \blacksquare$$

If the total population of the metropolitan region in Example 1 is 1 million, then $\mathbf{q}$ from Example 4 would correspond to having 375,000 persons in the city and 625,000 in the suburbs. At the end of one year, the migration *out of* the city would be $(.05)(375{,}000) = 18{,}750$ persons, and the migration *into* the city from the suburbs would be $(.03)(625{,}000) = 18{,}750$ persons. As a result, the population in the city would remain the same. Similarly, the suburban population would be stable.

The next example shows how to *find* a steady-state vector. Notice that we are just finding an eigenvector associated with the eigenvalue 1 and then scaling it to create a probability vector.

EXAMPLE 5 Let $P = \begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix}$. Find a steady-state vector for P.

SOLUTION First, solve the equation $P\mathbf{x} = \mathbf{x}$.

$$\begin{aligned} P\mathbf{x} - \mathbf{x} &= \mathbf{0} \\ P\mathbf{x} - I\mathbf{x} &= \mathbf{0} \qquad \text{Recall from Section 1.4 that } I\mathbf{x} = \mathbf{x}. \\ (P - I)\mathbf{x} &= \mathbf{0} \end{aligned}$$

For P as above,

$$P - I = \begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -.4 & .3 \\ .4 & -.3 \end{bmatrix}$$

To find all solutions of $(P - I)\mathbf{x} = \mathbf{0}$, row reduce the augmented matrix:

$$\begin{bmatrix} -.4 & .3 & 0 \\ .4 & -.3 & 0 \end{bmatrix} \sim \begin{bmatrix} -.4 & .3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3/4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Then $x_1 = \frac{3}{4}x_2$ and x_2 is free. The general solution is $x_2\begin{bmatrix} 3/4 \\ 1 \end{bmatrix}$.

Next, choose a simple basis for the solution space. One obvious choice is $\begin{bmatrix} 3/4 \\ 1 \end{bmatrix}$ but a better choice with no fractions is $\mathbf{w} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ (corresponding to $x_2 = 4$).

Finally, find a probability vector in the set of all solutions of $P\mathbf{x} = \mathbf{x}$. This process is easy, since every solution is a multiple of the solution $\mathbf{w}$. Divide $\mathbf{w}$ by the sum of its entries and obtain

$$\mathbf{q} = \begin{bmatrix} 3/7 \\ 4/7 \end{bmatrix}$$

As a check, compute

$$P\mathbf{q} = \begin{bmatrix} 6/10 & 3/10 \\ 4/10 & 7/10 \end{bmatrix} \begin{bmatrix} 3/7 \\ 4/7 \end{bmatrix} = \begin{bmatrix} 18/70 + 12/70 \\ 12/70 + 28/70 \end{bmatrix} = \begin{bmatrix} 30/70 \\ 40/70 \end{bmatrix} = \mathbf{q} \qquad \blacksquare$$

The next theorem shows that what happened in Example 3 is typical of many stochastic matrices. We say that a stochastic matrix is **regular** if some matrix power P^k contains only strictly positive entries. For P in Example 3,

$$P^2 = \begin{bmatrix} .37 & .26 & .33 \\ .45 & .70 & .45 \\ .18 & .04 & .22 \end{bmatrix}$$

Since every entry in P^2 is strictly positive, P is a regular stochastic matrix.

Also, we say that a sequence of vectors, $\mathbf{x}_k$ for $k = 1, 2, \ldots,$ **converges** to a vector $\mathbf{q}$ as $k \to \infty$, if the entries in $\mathbf{x}_k$ can be made as close as desired to the corresponding entries in $\mathbf{q}$ by taking k sufficiently large.

THEOREM 11

If P is an $n \times n$ regular stochastic matrix, then P has a unique steady-state vector $\mathbf{q}$. Further, if $\mathbf{x}_0$ is any initial state and $\mathbf{x}_{k+1} = P\mathbf{x}_k$ for $k = 0, 1, 2, \ldots,$ then the Markov chain $\{\mathbf{x}_k\}$ converges to $\mathbf{q}$ as $k \to \infty$.

This theorem is proved in standard texts on Markov chains. The amazing part of the theorem is that the initial state has no effect on the long-term behavior of the Markov chain.

EXAMPLE 6 In Example 2, what percentage of the voters are likely to vote for the Republican candidate in some election many years from now, assuming that the election outcomes form a Markov chain?

SOLUTION If you want to compute the precise entries of the steady-state vector by hand, it is better to recognize that it is an eigenvector with eigenvalue 1 rather than to pick some initial vector $\mathbf{x}_0$ and compute $\mathbf{x}_1, \ldots, \mathbf{x}_k$ for some large value of k. You have no way of knowing how many vectors to compute, and you cannot be sure of the limiting values of the entries in $\mathbf{x}_k$.

A better approach is to compute the steady-state vector and then appeal to Theorem 11. Given P as in Example 2, form $P - I$ by subtracting 1 from each diagonal entry in P. Then row reduce the augmented matrix:

$$[\,(P - I) \quad \mathbf{0}\,] = \begin{bmatrix} -.3 & .1 & .3 & 0 \\ .2 & -.2 & .3 & 0 \\ .1 & .1 & -.6 & 0 \end{bmatrix}$$

Recall from earlier work with decimals that the arithmetic is simplified by multiplying each row by 10.[1]

$$\begin{bmatrix} -3 & 1 & 3 & 0 \\ 2 & -2 & 3 & 0 \\ 1 & 1 & -6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -9/4 & 0 \\ 0 & 1 & -15/4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution of $(P - I)\mathbf{x} = \mathbf{0}$ is $x_1 = \frac{9}{4}x_3$, $x_2 = \frac{15}{4}x_3$, and x_3 is free. Choosing $x_3 = 4$, we obtain a basis for the solution space whose entries are integers, and from this we easily find the steady-state vector whose entries sum to 1:

$$\mathbf{w} = \begin{bmatrix} 9 \\ 15 \\ 4 \end{bmatrix}, \quad \text{and} \quad \mathbf{q} = \begin{bmatrix} 9/28 \\ 15/28 \\ 4/28 \end{bmatrix} \approx \begin{bmatrix} .32 \\ .54 \\ .14 \end{bmatrix}$$

The entries in $\mathbf{q}$ describe the distribution of votes at an election to be held many years from now (assuming the stochastic matrix continues to describe the changes from one election to the next). Thus, eventually, about 54% of the vote will be for the Republican candidate. ■

Numerical Notes

You may have noticed that if $\mathbf{x}_{k+1} = P\mathbf{x}_k$ for $k = 0, 1, \ldots,$ then

$$\mathbf{x}_2 = P\mathbf{x}_1 = P(P\mathbf{x}_0) = P^2\mathbf{x}_0,$$

and, in general,

$$\mathbf{x}_k = P^k\mathbf{x}_0 \quad \text{for } k = 0, 1, \ldots$$

To compute a specific vector such as $\mathbf{x}_3$, fewer arithmetic operations are needed to compute $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, rather than P^3 and $P^3\mathbf{x}_0$. However, if P is small—say, 30×30—the machine computation time is insignificant for both methods, and a command to compute $P^3\mathbf{x}_0$ might be preferred because it requires fewer human keystrokes.

Practice Problems

1. Suppose the residents of a metropolitan region move according to the probabilities in the migration matrix M in Example 1 and a resident is chosen "at random." Then a state vector for a certain year may be interpreted as giving the probabilities that the person is a city resident or a suburban resident at that time.

a. Suppose the person chosen is a city resident now, so that $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. What is the likelihood that the person will live in the suburbs next year?

b. What is the likelihood that the person will be living in the suburbs in two years?

2. Let $P = \begin{bmatrix} .6 & .2 \\ .4 & .8 \end{bmatrix}$ and $\mathbf{q} = \begin{bmatrix} .3 \\ .7 \end{bmatrix}$. Is $\mathbf{q}$ a steady-state vector for P?

3. What percentage of the population in Example 1 will live in the suburbs after many years?

[1] ***Warning:*** Don't multiply only P by 10. Instead multiply the augmented matrix for equation $(P - I)\mathbf{x} = \mathbf{0}$ by 10.

5.9 Exercises

1. A small remote village receives radio broadcasts from two radio stations, a news station and a music station. Of the listeners who are tuned to the news station, 70% will remain listening to the news after the station break that occurs each half hour, while 30% will switch to the music station at the station break. Of the listeners who are tuned to the music station, 60% will switch to the news station at the station break, while 40% will remain listening to the music. Suppose everyone is listening to the news at 8:15 A.M.

a. Give the stochastic matrix that describes how the radio listeners tend to change stations at each station break. Label the rows and columns.

b. Give the initial state vector.

c. What percentage of the listeners will be listening to the music station at 9:25 A.M. (after the station breaks at 8:30 and 9:00 A.M.)?

2. A laboratory animal may eat any one of three foods each day. Laboratory records show that if the animal chooses one food on one trial, it will choose the same food on the next trial with a probability of 50%, and it will choose the other foods on the next trial with equal probabilities of 25%.

a. What is the stochastic matrix for this situation?

b. If the animal chooses food #1 on an initial trial, what is the probability that it will choose food #2 on the second trial after the initial trial?

3. On any given day, a student is either healthy or ill. Of the students who are healthy today, 95% will be healthy tomorrow. Of the students who are ill today, 55% will still be ill tomorrow.

a. What is the stochastic matrix for this situation?

b. Suppose 20% of the students are ill on Monday. What fraction or percentage of the students are likely to be ill on Tuesday? On Wednesday?

c. If a student is well today, what is the probability that he or she will be well two days from now?

4. The weather in Columbus is either good, indifferent, or bad on any given day. If the weather is good today, there is a 60% chance the weather will be good tomorrow, a 30% chance the weather will be indifferent, and a 10% chance the weather will be bad. If the weather is indifferent today, it will be good tomorrow with probability .40 and indifferent with probability .30. Finally, if the weather is bad today, it will be good tomorrow with probability .40 and indifferent with probability .50.

a. What is the stochastic matrix for this situation?

b. Suppose there is a 50% chance of good weather today and a 50% chance of indifferent weather. What are the chances of bad weather tomorrow?

c. Suppose the predicted weather for Monday is 40% indifferent weather and 60% bad weather. What are the chances for good weather on Wednesday?

In Exercises 5–8, find the steady-state vector.

5. $\begin{bmatrix} .1 & .6 \\ .9 & .4 \end{bmatrix}$

6. $\begin{bmatrix} .8 & .5 \\ .2 & .5 \end{bmatrix}$

7. $\begin{bmatrix} .7 & .1 & .1 \\ .2 & .8 & .2 \\ .1 & .1 & .7 \end{bmatrix}$

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

9. Determine if $P = \begin{bmatrix} .2 & 1 \\ .8 & 0 \end{bmatrix}$ is a regular stochastic matrix.

10. Determine if $P = \begin{bmatrix} 1 & .2 \\ 0 & .8 \end{bmatrix}$ is a regular stochastic matrix.

11. a. Find the steady-state vector for the Markov chain in Exercise 1.

b. At some time late in the day, what fraction of the listeners will be listening to the news?

12. Refer to Exercise 2. Which food will the animal prefer after many trials?

13. a. Find the steady-state vector for the Markov chain in Exercise 3.

b. What is the probability that after many days a specific student is ill? Does it matter if that person is ill today?

14. Refer to Exercise 4. In the long run, how likely is it for the weather in Columbus to be good on a given day?

In Exercises 15–20, P is an $n \times n$ stochastic matrix. Mark each statement True or False **(T/F)**. Justify each answer.

15. **(T/F)** The steady state vector is an eigenvector of P.

16. **(T/F)** Every eigenvector of P is a steady state vector.

17. **(T/F)** The all ones vector is an eigenvector of P^T.

18. **(T/F)** The number 2 can be an eigenvalue of a stochastic matrix.

19. **(T/F)** The number 1/2 can be an eigenvalue of a stochastic matrix.

20. **(T/F)** All stochastic matrices are regular.

21. Is $\mathbf{q} = \begin{bmatrix} .6 \\ .8 \end{bmatrix}$ a steady state vector for $A = \begin{bmatrix} .2 & .6 \\ .8 & .4 \end{bmatrix}$? Justify your answer.

22. Is $\mathbf{q} = \begin{bmatrix} .4 \\ .4 \end{bmatrix}$ a steady state vector for $A = \begin{bmatrix} .2 & .8 \\ .8 & .2 \end{bmatrix}$? Justify your answer.

23. Is $\mathbf{q} = \begin{bmatrix} .6 \\ .4 \end{bmatrix}$ a steady state vector for $A = \begin{bmatrix} .4 & .6 \\ .6 & .4 \end{bmatrix}$? Justify your answer.

24. Is $\mathbf{q} = \begin{bmatrix} 3/7 \\ 4/7 \end{bmatrix}$ a steady state vector for $A = \begin{bmatrix} .2 & .6 \\ .8 & .4 \end{bmatrix}$? Justify your answer.

T **25.** Suppose the following matrix describes the likelihood that an individual will switch between an iOS and an Android smartphone:

$$\begin{array}{ccl} \multicolumn{2}{c}{\text{From}} & \text{To} \\ \text{iOS} & \text{Android} & \\ \end{array}$$
$$\begin{bmatrix} .70 & .15 \\ .30 & .85 \end{bmatrix} \begin{array}{l} \text{iOS} \\ \text{Android} \end{array}$$

In the long run, what percentage of smartphone owners would you expect to have an Android operating system?

T **26.** In Detroit, Hertz Rent A Car has a fleet of about 2000 cars. The pattern of rental and return locations is given by the fractions in the table below. On a typical day, about how many cars will be rented or ready to rent from the downtown location?

$$\begin{array}{ccc} \multicolumn{3}{c}{\text{Cars Rented from}} \\ \text{City Airport} & \text{Down-town} & \text{Metro Airport} \end{array} \quad \text{Returned to}$$
$$\begin{bmatrix} .90 & .01 & .09 \\ .01 & .90 & .01 \\ .09 & .09 & .90 \end{bmatrix} \begin{array}{l} \text{City Airport} \\ \text{Downtown} \\ \text{Metro Airport} \end{array}$$

27. Let P be an $n \times n$ stochastic matrix. The following argument shows that the equation $P\mathbf{x} = \mathbf{x}$ has a nontrivial solution. (In fact, a steady-state solution exists with nonnegative entries. A proof is given in some advanced texts.) Justify each assertion below. (Mention a theorem when appropriate.)

a. If all the other rows of $P - I$ are added to the bottom row, the result is a row of zeros.

b. The rows of $P - I$ are linearly dependent.

c. The dimension of the row space of $P - I$ is less than n.

d. $P - I$ has a nontrivial null space.

28. Show that every 2×2 stochastic matrix has at least one steady-state vector. Any such matrix can be written in the form $P = \begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix}$, where α and β are constants between 0 and 1. (There are two linearly independent steady-state vectors if $\alpha = \beta = 0$. Otherwise, there is only one.)

29. Let S be the $1 \times n$ row matrix with a 1 in each column,

$S = [\,1 \quad 1 \quad \cdots \quad 1\,]$

a. Explain why a vector $\mathbf{x}$ in $\mathbb{R}^n$ is a probability vector if and only if its entries are nonnegative and $S\mathbf{x} = 1$. (A 1×1 matrix such as the product $S\mathbf{x}$ is usually written without the matrix bracket symbols.)

b. Let P be an $n \times n$ stochastic matrix. Explain why $SP = S$.

c. Let P be an $n \times n$ stochastic matrix, and let $\mathbf{x}$ be a probability vector. Show that $P\mathbf{x}$ is also a probability vector.

30. Use Exercise 29 to show that if P is an $n \times n$ stochastic matrix, then so is P^2.

T **31.** Examine powers of a regular stochastic matrix.

a. Compute P^k for $k = 2, 3, 4, 5$, when

$$P = \begin{bmatrix} .3355 & .3682 & .3067 & .0389 \\ .2663 & .2723 & .3277 & .5451 \\ .1935 & .1502 & .1589 & .2395 \\ .2047 & .2093 & .2067 & .1765 \end{bmatrix}$$

Display calculations to four decimal places. What happens to the columns of P^k as k increases? Compute the steady-state vector for P.

b. Compute Q^k for $k = 10, 20, \ldots, 80$, when

$$Q = \begin{bmatrix} .97 & .05 & .10 \\ 0 & .90 & .05 \\ .03 & .05 & .85 \end{bmatrix}$$

(Stability for Q^k to four decimal places may require $k = 116$ or more.) Compute the steady-state vector for Q. Conjecture what might be true for any regular stochastic matrix.

c. Use Theorem 11 to explain what you found in parts (a) and (b).

T **32.** Compare two methods for finding the steady-state vector $\mathbf{q}$ of a regular stochastic matrix P: (1) computing $\mathbf{q}$ as in Example 5, or (2) computing P^k for some large value of k and using one of the columns of P^k as an approximation for $\mathbf{q}$. [The *Study Guide* describes a program *nulbasis* that almost automates method (1).]

Experiment with the largest random stochastic matrices your matrix program will allow, and use $k = 100$ or some other large value. For each method, describe the time *you* need to enter the keystrokes and run your program. (Some versions of MATLAB have commands `flops` and `tic ...toc` that record the number of floating point operations and the total elapsed time MATLAB uses.) Contrast the advantages of each method, and state which you prefer.

Solutions to Practice Problems

1. a. Since 5% of the city residents will move to the suburbs within one year, there is a 5% chance of choosing such a person. Without further knowledge about the person, we say that there is a 5% chance the person will move to the suburbs. This fact is contained in the second entry of the state vector $\mathbf{x}_1$, where

$$\mathbf{x}_1 = M\mathbf{x}_0 = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .95 \\ .05 \end{bmatrix}$$

   b. The likelihood that the person will be living in the suburbs after two years is 9.6%, because

$$\mathbf{x}_2 = M\mathbf{x}_1 = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}\begin{bmatrix} .95 \\ .05 \end{bmatrix} = \begin{bmatrix} .904 \\ .096 \end{bmatrix}$$

2. The steady-state vector satisfies $P\mathbf{x} = \mathbf{x}$. Since

$$P\mathbf{q} = \begin{bmatrix} .6 & .2 \\ .4 & .8 \end{bmatrix}\begin{bmatrix} .3 \\ .7 \end{bmatrix} = \begin{bmatrix} .32 \\ .68 \end{bmatrix} \neq \mathbf{q}$$

   we conclude that $\mathbf{q}$ is *not* the steady-state vector for P.

3. M in Example 1 is a regular stochastic matrix because its entries are all strictly positive. So we may use Theorem 11. We already know the steady-state vector from Example 4. Thus the population distribution vectors $\mathbf{x}_k$ converge to

$$\mathbf{q} = \begin{bmatrix} .375 \\ .625 \end{bmatrix}$$

   Eventually 62.5% of the population will live in the suburbs.

CHAPTER 5 PROJECTS

Chapter 5 projects are available online at `bit.ly/30IM8gT`.

A. *Power Method for Finding Eigenvalues*: This project shows how to find the eigenvector associated with the eigenvalue corresponding to the largest eigenvalue.

B. *Integration by Parts*: The purpose of this project is to show how the matrix of a linear transformation relative to a basis $\mathcal{B}$ may be used to find antiderivatives usually found using integration by parts.

C. *Robotics*: In this project, students are asked to find online examples of robots that use 3D rotations to function.

D. *Dynamical Systems and Markov Chains*: This project applies the techniques of discrete dynamical systems to Markov chains.

CHAPTER 5 SUPPLEMENTARY EXERCISES

Throughout these supplementary exercises, A and B represent square matrices of appropriate sizes.

For Exercises 1–23, mark each statement as True or False **(T/F)**. Justify each answer.

1. **(T/F)** If A is invertible and 1 is an eigenvalue for A, then 1 is also an eigenvalue of A^{-1}.

2. **(T/F)** If A is row equivalent to the identity matrix I, then A is diagonalizable.

3. **(T/F)** If A contains a row or column of zeros, then 0 is an eigenvalue of A.

4. **(T/F)** Each eigenvalue of A is also an eigenvalue of A^2.

5. **(T/F)** Each eigenvector of A is also an eigenvector of A^2.

6. **(T/F)** Each eigenvector of an invertible matrix A is also an eigenvector of A^{-1}.

7. **(T/F)** Eigenvalues must be nonzero scalars.

8. **(T/F)** Eigenvectors must be nonzero vectors.

9. **(T/F)** Two eigenvectors corresponding to the same eigenvalue are always linearly dependent.

10. **(T/F)** Similar matrices always have exactly the same eigenvalues.

11. **(T/F)** Similar matrices always have exactly the same eigenvectors.

12. **(T/F)** The sum of two eigenvectors of a matrix A is also an eigenvector of A.

13. **(T/F)** The eigenvalues of an upper triangular matrix A are exactly the nonzero entries on the diagonal of A.

14. **(T/F)** The matrices A and A^T have the same eigenvalues, counting multiplicities.

15. **(T/F)** If a 5×5 matrix A has fewer than 5 distinct eigenvalues, then A is not diagonalizable.

16. **(T/F)** There exists a 2×2 matrix that has no eigenvectors in $\mathbb{R}^2$.

17. **(T/F)** If A is diagonalizable, then the columns of A are linearly independent.

18. **(T/F)** A nonzero vector cannot correspond to two different eigenvalues of A.

19. **(T/F)** A (square) matrix A is invertible if and only if there is a coordinate system in which the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is represented by a diagonal matrix.

20. **(T/F)** If each vector $\mathbf{e}_j$ in the standard basis for $\mathbb{R}^n$ is an eigenvector of A, then A is a diagonal matrix.

21. **(T/F)** If A is similar to a diagonalizable matrix B, then A is also diagonalizable.

22. **(T/F)** If A and B are invertible $n \times n$ matrices, then AB is similar to BA.

23. **(T/F)** An $n \times n$ matrix with n linearly independent eigenvectors is invertible.

24. Show that if $\mathbf{x}$ is an eigenvector of the matrix product AB and $B\mathbf{x} \neq \mathbf{0}$, then $B\mathbf{x}$ is an eigenvector of BA.

25. Suppose $\mathbf{x}$ is an eigenvector of A corresponding to an eigenvalue λ.
    a. Show that $\mathbf{x}$ is an eigenvector of $5I - A$. What is the corresponding eigenvalue?
    b. Show that $\mathbf{x}$ is an eigenvector of $5I - 3A + A^2$. What is the corresponding eigenvalue?

26. Use mathematical induction to show that if λ is an eigenvalue of an $n \times n$ matrix A, with $\mathbf{x}$ a corresponding eigenvector, then, for each positive integer m, λ^m is an eigenvalue of A^m, with $\mathbf{x}$ a corresponding eigenvector.

27. If $p(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^n$, define $p(A)$ to be the matrix formed by replacing each power of t in $p(t)$ by the corresponding power of A (with $A^0 = I$). That is,

$$p(A) = c_0 I + c_1 A + c_2 A^2 + \cdots + c_n A^n$$

Show that if λ is an eigenvalue of A, then one eigenvalue of $p(A)$ is $p(\lambda)$.

28. Suppose $A = PDP^{-1}$, where P is 2×2 and $D = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix}$.
    a. Let $B = 5I - 3A + A^2$. Show that B is diagonalizable by finding a suitable factorization of B.
    b. Given $p(t)$ and $p(A)$ as in Exercise 27, show that $p(A)$ is diagonalizable.

29. Suppose A is diagonalizable and $p(t)$ is the characteristic polynomial of A. Define $p(A)$ as in Exercise 27, and show that $p(A)$ is the zero matrix. This fact, which is also true for *any* square matrix, is called the *Cayley–Hamilton theorem*.

30. a. Let A be a diagonalizable $n \times n$ matrix. Show that if the multiplicity of an eigenvalue λ is n, then $A = \lambda I$.
    b. Use part (a) to show that the matrix $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ is not diagonalizable.

31. Show that $I - A$ is invertible when all the eigenvalues of A are less than 1 in magnitude. [*Hint:* What would be true if $I - A$ were not invertible?]

32. Show that if A is diagonalizable, with all eigenvalues less than 1 in magnitude, then A^k tends to the zero matrix as $k \to \infty$. [*Hint:* Consider $A^k\mathbf{x}$ where $\mathbf{x}$ represents any one of the columns of I.]

33. Let $\mathbf{u}$ be an eigenvector of A corresponding to an eigenvalue λ, and let H be the line in $\mathbb{R}^n$ through $\mathbf{u}$ and the origin.
    a. Explain why H is invariant under A in the sense that $A\mathbf{x}$ is in H whenever $\mathbf{x}$ is in H.
    b. Let K be a one-dimensional subspace of $\mathbb{R}^n$ that is invariant under A. Explain why K contains an eigenvector of A.

34. Let $G = \begin{bmatrix} A & X \\ 0 & B \end{bmatrix}$. Use formula for the determinant in Section 5.2 to explain why $\det G = (\det A)(\det B)$. From this, deduce that the characteristic polynomial of G is the product of the characteristic polynomials of A and B.

Use Exercise 34 to find the eigenvalues of the matrices in Exercises 35 and 36.

35. $A = \begin{bmatrix} 3 & -2 & 8 \\ 0 & 5 & -2 \\ 0 & -4 & 3 \end{bmatrix}$

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

37. Let J be the $n \times n$ matrix of all 1's, and consider $A = (a-b)I + bJ$; that is,

$$A = \begin{bmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{bmatrix}$$

Use the results of Exercise 30 in the Supplementary Exercises for Chapter 3 to show that the eigenvalues of A are $a-b$ and $a+(n-1)b$. What are the multiplicities of these eigenvalues?

38. Apply the result of Exercise 37 to find the eigenvalues of the matrices $\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$ and $\begin{bmatrix} 7 & 3 & 3 & 3 & 3 \\ 3 & 7 & 3 & 3 & 3 \\ 3 & 3 & 7 & 3 & 3 \\ 3 & 3 & 3 & 7 & 3 \\ 3 & 3 & 3 & 3 & 7 \end{bmatrix}$.

39. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. The tr A (the trace of A) is the sum of the diagonal entries in A. Show that the characteristic polynomial of A is

$\lambda^2 - (\text{tr } A)\lambda + \det A$

Then show that the eigenvalues of a 2×2 matrix A are both real if and only if $\det A \le \left(\frac{\text{tr } A}{2}\right)^2$.

40. Let $A = \begin{bmatrix} .4 & -.3 \\ .4 & 1.2 \end{bmatrix}$. Explain why A^k approaches $\begin{bmatrix} -.5 & -.75 \\ 1.0 & 1.50 \end{bmatrix}$ as $k \to \infty$.

Exercises 41–45 concern the polynomial

$p(t) = a_0 + a_1 t + \cdots + a_{n-1} t^{n-1} + t^n$

and an $n \times n$ matrix C_p called the **companion matrix** of p:

$$C_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}$$

41. Write the companion matrix C_p for $p(t) = 6 - 5t + t^2$, and then find the characteristic polynomial of C_p.

42. Let $p(t) = (t-2)(t-3)(t-4) = -24 + 26t - 9t^2 + t^3$. Write the companion matrix for $p(t)$, and use techniques from Chapter 3 to find its characteristic polynomial.

43. Use mathematical induction to prove that for $n \ge 2$,

$$\begin{aligned} \det(C_p - \lambda I) &= (-1)^n (a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n) \\ &= (-1)^n p(\lambda) \end{aligned}$$

[*Hint:* Expanding by cofactors down the first column, show that $\det(C_p - \lambda I)$ has the form $(-\lambda)B + (-1)^n a_0$, where B is a certain polynomial (by the induction assumption).]

44. Let $p(t) = a_0 + a_1 t + a_2 t^2 + t^3$, and let λ be a zero of p.

a. Write the companion matrix for p.

b. Explain why $\lambda^3 = -a_0 - a_1\lambda - a_2\lambda^2$, and show that $(1, \lambda, \lambda^2)$ is an eigenvector of the companion matrix for p.

45. Let p be the polynomial in Exercise 44, and suppose the equation $p(t) = 0$ has distinct roots $\lambda_1, \lambda_2, \lambda_3$. Let V be the Vandermonde matrix

$$V = \begin{bmatrix} 1 & 1 & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 \end{bmatrix}$$

Use Exercise 44 and a theorem from this chapter to deduce that V is invertible (but do not compute V^{-1}). Then explain why $V^{-1}C_pV$ is a diagonal matrix.

T 46. The MATLAB command `roots(p)` computes the roots of the polynomial equation $p(t) = 0$. Read a MATLAB manual, and then describe the basic idea behind the algorithm for the `roots` command.

T 47. Use a matrix program to diagonalize

$$A = \begin{bmatrix} -3 & -2 & 0 \\ 14 & 7 & -1 \\ -6 & -3 & 1 \end{bmatrix}$$

if possible. Use the eigenvalue command to create the diagonal matrix D. If the program has a command that produces eigenvectors, use it to create an invertible matrix P. Then compute $AP - PD$ and PDP^{-1}. Discuss your results.

T 48. Repeat Exercise 47 for $A = \begin{bmatrix} -8 & 5 & -2 & 0 \\ -5 & 2 & 1 & -2 \\ 10 & -8 & 6 & -3 \\ 3 & -2 & 1 & 0 \end{bmatrix}$.

6 Orthogonality and Least Squares

Introductory Example

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Can you tell a puppy from a kitten regardless of breed or coloring? Of course! Both may be furry bundles of joy, but one is clearly a feline and one is clearly a canine to the human eye. Simple. But, as blatantly obvious as it may seem to our trained eyes that are capable of interpreting meaning in the pixels of an image, this turns out to be a significant challenge for a machine. With the advancements in artificial intelligence (AI) and machine learning, computers are rapidly improving their capability to identify the creature on the left in the picture above as a puppy and the one on the right as a cat.

Many industries are now using AI technology to speed up the process of what once took hours of mindless work, such as post office scanners that can read bar codes and handwriting on envelopes to sort the mail with precision and speed. Nordstrom is using machine learning to design, display, organize, and recommend clothing ensembles to customers, exemplifying how even in creative aesthetic fields machine learning can be used to interpret color and shape patterns in pixels and organize visual possibilities into that which our eyes register as pleasing.

When calling a service desk number, one is often greeted by a machine that asks a series of questions and provides suggestions. Only persistence in interacting with this machine gets the caller through to a real person. More and more service calls are answered by machines, making it easier for customers' simple questions to be answered succinctly without waiting time and the obnoxious jingles of hold music. Google has designed an AI assistant that will handle making service calls for you too—booking a restaurant or hair appointment on your behalf.

AI and machine learning comprise developing systems that interpret external data correctly, learn from such data, and use that learning to achieve specific goals and tasks through flexibility and adaptation. Often, the driving engine behind these techniques is linear algebra. In Section 6.2, we see a simple way to design a matrix so that matrix multiplication can identify the correct pattern of blue and white squares. In Sections 6.5, 6.6, and 6.8, we explore techniques used in machine learning.

In order to find an approximate solution to an inconsistent system of equations that has no actual solution, a well-defined notion of nearness is needed. Section 6.1 introduces the concepts of distance and orthogonality in a vector space. Sections 6.2 and 6.3 show how orthogonality can be used to identify the point within a subspace W that is nearest to a point $\mathbf{y}$ lying outside of W. By taking W to be the column space of a matrix, Section 6.5 develops a method for producing approximate ("least-squares") solutions for inconsistent linear systems, an important technique in machine learning, which is discussed in Sections 6.6 and 6.8.

Section 6.4 provides another opportunity to see orthogonal projections at work, creating a matrix factorization widely used in numerical linear algebra. The remaining sections examine some of the many least-squares problems that arise in applications, including those in vector spaces more general than $\mathbb{R}^n$.

6.1 Inner Product, Length, and Orthogonality

Geometric concepts of length, distance, and perpendicularity, which are well known for $\mathbb{R}^2$ and $\mathbb{R}^3$, are defined here for $\mathbb{R}^n$. These concepts provide powerful geometric tools for solving many applied problems, including the least-squares problems mentioned above. All three notions are defined in terms of the inner product of two vectors.

The Inner Product

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then we regard $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices. The transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which we write as a single real number (a scalar) without brackets. The number $\mathbf{u}^T\mathbf{v}$ is called the **inner product** of $\mathbf{u}$ and $\mathbf{v}$, and often it is written as $\mathbf{u}\cdot\mathbf{v}$. This inner product, mentioned in the exercises for Section 2.1, is also referred to as a **dot product**. If

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

then the inner product of $\mathbf{u}$ and $\mathbf{v}$ is

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

EXAMPLE 1 Compute $\mathbf{u}\cdot\mathbf{v}$ and $\mathbf{v}\cdot\mathbf{u}$ for $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}$.

SOLUTION

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = [\,2 \quad -5 \quad -1\,]\begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} = (2)(3) + (-5)(2) + (-1)(-3) = -1$$

$$\mathbf{v}\cdot\mathbf{u} = \mathbf{v}^T\mathbf{u} = [\,3 \quad 2 \quad -3\,]\begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix} = (3)(2) + (2)(-5) + (-3)(-1) = -1$$

It is clear from the calculations in Example 1 why $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$. This commutativity of the inner product holds in general. The following properties of the inner product are easily deduced from properties of the transpose operation in Section 2.1. (See Exercises 29 and 30 at the end of this section.)

THEOREM 1

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

a. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$
b. $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w}$
c. $(c\mathbf{u})\cdot\mathbf{v} = c(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(c\mathbf{v})$
d. $\mathbf{u}\cdot\mathbf{u} \geq 0$, and $\mathbf{u}\cdot\mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Properties (b) and (c) can be combined several times to produce the following useful rule:

$$(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{w} = c_1(\mathbf{u}_1\cdot\mathbf{w}) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{w})$$

The Length of a Vector

If $\mathbf{v}$ is in $\mathbb{R}^n$, with entries $v_1, \ldots, v_n$, then the square root of $\mathbf{v}\cdot\mathbf{v}$ is defined because $\mathbf{v}\cdot\mathbf{v}$ is nonnegative.

DEFINITION

The **length** (or **norm**) of $\mathbf{v}$ is the nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$$

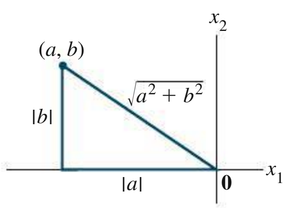


FIGURE 1
Interpretation of $\|\mathbf{v}\|$ as length.

Suppose $\mathbf{v}$ is in $\mathbb{R}^2$, say, $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. If we identify $\mathbf{v}$ with a geometric point in the plane, as usual, then $\|\mathbf{v}\|$ coincides with the standard notion of the length of the line segment from the origin to $\mathbf{v}$. This follows from the Pythagorean Theorem applied to a triangle such as the one in Figure 1.

A similar calculation with the diagonal of a rectangular box shows that the definition of length of a vector $\mathbf{v}$ in $\mathbb{R}^3$ coincides with the usual notion of length.

For any scalar c, the length of $c\mathbf{v}$ is $|c|$ times the length of $\mathbf{v}$. That is,

$$\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$$

(To see this, compute $\|c\mathbf{v}\|^2 = (c\mathbf{v})\cdot(c\mathbf{v}) = c^2\mathbf{v}\cdot\mathbf{v} = c^2\|\mathbf{v}\|^2$ and take square roots.)

A vector whose length is 1 is called a **unit vector**. If we *divide* a nonzero vector $\mathbf{v}$ by its length—that is, multiply by $1/\|\mathbf{v}\|$—we obtain a unit vector $\mathbf{u}$ because the length of $\mathbf{u}$ is $(1/\|\mathbf{v}\|)\|\mathbf{v}\|$. The process of creating $\mathbf{u}$ from $\mathbf{v}$ is sometimes called **normalizing** $\mathbf{v}$, and we say that $\mathbf{u}$ is *in the same direction* as $\mathbf{v}$.

Several examples that follow use the space-saving notation for (column) vectors.

EXAMPLE 2 Let $\mathbf{v} = (1, -2, 2, 0)$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$.

SOLUTION First, compute the length of $\mathbf{v}$:

$$\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v} = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9$$
$$\|\mathbf{v}\| = \sqrt{9} = 3$$

Then, multiply $\mathbf{v}$ by $1/\|\mathbf{v}\|$ to obtain

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{3}\mathbf{v} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}$$

To check that $\|\mathbf{u}\| = 1$, it suffices to show that $\|\mathbf{u}\|^2 = 1$.

$$\|\mathbf{u}\|^2 = \mathbf{u}\cdot\mathbf{u} = \left(\tfrac{1}{3}\right)^2 + \left(-\tfrac{2}{3}\right)^2 + \left(\tfrac{2}{3}\right)^2 + (0)^2$$
$$= \tfrac{1}{9} + \tfrac{4}{9} + \tfrac{4}{9} + 0 = 1$$

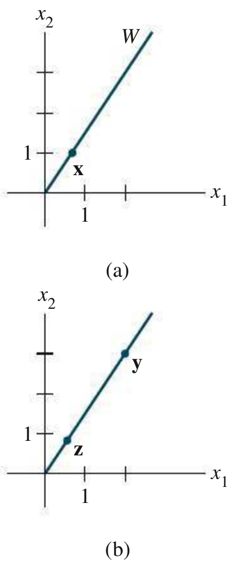


FIGURE 2
Normalizing a vector to produce a unit vector.

EXAMPLE 3 Let W be the subspace of $\mathbb{R}^2$ spanned by $\mathbf{x} = (\frac{2}{3}, 1)$. Find a unit vector $\mathbf{z}$ that is a basis for W.

SOLUTION W consists of all multiples of $\mathbf{x}$, as in Figure 2(a). Any nonzero vector in W is a basis for W. To simplify the calculation, "scale" $\mathbf{x}$ to eliminate fractions. That is, multiply $\mathbf{x}$ by 3 to get

$$\mathbf{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Now compute $\|\mathbf{y}\|^2 = 2^2 + 3^2 = 13$, $\|\mathbf{y}\| = \sqrt{13}$, and normalize $\mathbf{y}$ to get

$$\mathbf{z} = \frac{1}{\sqrt{13}}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{13} \\ 3/\sqrt{13} \end{bmatrix}$$

See Figure 2(b). Another unit vector is $(-2/\sqrt{13}, -3/\sqrt{13})$.

Distance in $\mathbb{R}^n$

We are ready now to describe how close one vector is to another. Recall that if a and b are real numbers, the distance on the number line between a and b is the number $|a - b|$. Two examples are shown in Figure 3. This definition of distance in $\mathbb{R}$ has a direct analogue in $\mathbb{R}^n$.

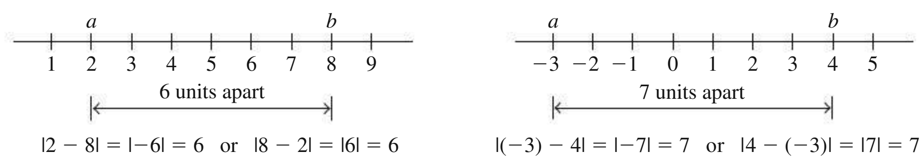


FIGURE 3 Distances in $\mathbb{R}$.

DEFINITION

For **u** and **v** in $\mathbb{R}^n$, the **distance between u and v**, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is,

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$, this definition of distance coincides with the usual formulas for the Euclidean distance between two points, as the next two examples show.

EXAMPLE 4 Compute the distance between the vectors $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$.

SOLUTION Calculate

$$\mathbf{u} - \mathbf{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

The vectors **u**, **v**, and $\mathbf{u} - \mathbf{v}$ are shown in Figure 4. When the vector $\mathbf{u} - \mathbf{v}$ is added to **v**, the result is **u**. Notice that the parallelogram in Figure 4 shows that the distance from **u** to **v** is the same as the distance from $\mathbf{u} - \mathbf{v}$ to **0**. ■

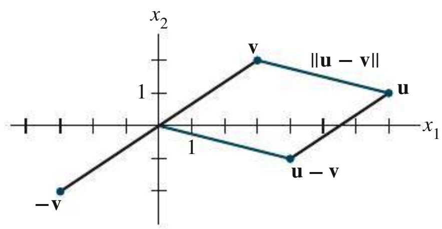


FIGURE 4 The distance between **u** and **v** is the length of $\mathbf{u} - \mathbf{v}$.

EXAMPLE 5 If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$, then

$$\begin{aligned} \text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| &= \sqrt{(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})} \\ &= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2} \end{aligned}$$ ■

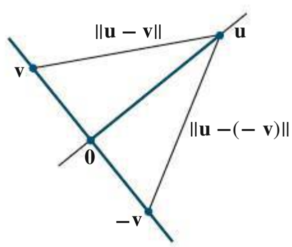


FIGURE 5

Orthogonal Vectors

The rest of this chapter depends on the fact that the concept of perpendicular lines in ordinary Euclidean geometry has an analogue in $\mathbb{R}^n$.

Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors **u** and **v**. The two lines shown in Figure 5 are geometrically perpendicular if and only if the distance from **u** to **v** is the same as the distance from **u** to $-\mathbf{v}$. This is the same as requiring the squares of the distances to be the same. Now

$$\begin{aligned} [\,\text{dist}(\mathbf{u}, -\mathbf{v})\,]^2 &= \|\mathbf{u} - (-\mathbf{v})\|^2 = \|\mathbf{u} + \mathbf{v}\|^2 \\ &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\ &= \mathbf{u} \cdot (\mathbf{u} + \mathbf{v}) + \mathbf{v} \cdot (\mathbf{u} + \mathbf{v}) && \text{Theorem 1(b)} \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} && \text{Theorem 1(a), (b)} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v} && \text{Theorem 1(a)} \qquad (1) \end{aligned}$$

The same calculations with $\mathbf{v}$ and $-\mathbf{v}$ interchanged show that

$$\begin{aligned}[\text{dist}(\mathbf{u},\mathbf{v})]^2 &= \|\mathbf{u}\|^2 + \|-\mathbf{v}\|^2 + 2\mathbf{u}\cdot(-\mathbf{v}) \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u}\cdot\mathbf{v}\end{aligned}$$

The two squared distances are equal if and only if $2\mathbf{u}\cdot\mathbf{v} = -2\mathbf{u}\cdot\mathbf{v}$, which happens if and only if $\mathbf{u}\cdot\mathbf{v} = 0$.

This calculation shows that when vectors $\mathbf{u}$ and $\mathbf{v}$ are identified with geometric points, the corresponding lines through the points and the origin are perpendicular if and only if $\mathbf{u}\cdot\mathbf{v} = 0$. The following definition generalizes to $\mathbb{R}^n$ this notion of perpendicularity (or *orthogonality*, as it is commonly called in linear algebra).

DEFINITION

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u}\cdot\mathbf{v} = 0$.

Observe that the zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $\mathbf{0}^T\mathbf{v} = 0$ for all $\mathbf{v}$.

The next theorem provides a useful fact about orthogonal vectors. The proof follows immediately from the calculation in (1) and the definition of orthogonality. The right triangle shown in Figure 6 provides a visualization of the lengths that appear in the theorem.

THEOREM 2

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

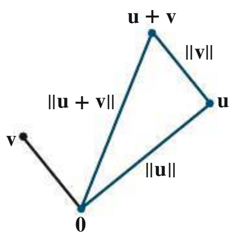


FIGURE 6

Orthogonal Complements

To provide practice using inner products, we introduce a concept here that will be of use in Section 6.3 and elsewhere in the chapter. If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is said to be **orthogonal to** W. The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^\perp$ (and read as "W perpendicular" or simply "W perp").

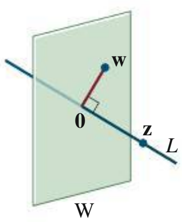


FIGURE 7 A plane and line through $\mathbf{0}$ as orthogonal complements.

EXAMPLE 6 Let W be a plane through the origin in $\mathbb{R}^3$, and let L be the line through the origin and perpendicular to W. If $\mathbf{z}$ and $\mathbf{w}$ are nonzero, $\mathbf{z}$ is on L, and $\mathbf{w}$ is in W, then the line segment from $\mathbf{0}$ to $\mathbf{z}$ is perpendicular to the line segment from $\mathbf{0}$ to $\mathbf{w}$; that is, $\mathbf{z}\cdot\mathbf{w} = 0$. See Figure 7. So each vector on L is orthogonal to every $\mathbf{w}$ in W. In fact, L consists of *all* vectors that are orthogonal to the $\mathbf{w}$'s in W, and W consists of all vectors orthogonal to the $\mathbf{z}$'s in L. That is,

$$L = W^\perp \quad \text{and} \quad W = L^\perp \qquad \blacksquare$$

The following two facts about $W^\perp$, with W a subspace of $\mathbb{R}^n$, are needed later in the chapter. Proofs are suggested in Exercises 37 and 38. Exercises 35–39 provide excellent practice using properties of the inner product.

1. A vector $\mathbf{x}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans W.
2. $W^\perp$ is a subspace of $\mathbb{R}^n$.

The next theorem and Exercise 39 verify the claims made in Section 4.5 concerning the subspaces shown in Figure 8.

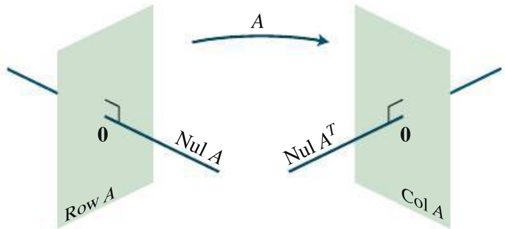


FIGURE 8 The fundamental subspaces determined by an $m \times n$ matrix A.

Remark: A common way to prove that two sets, say S and T, are equal is to show that S is a subset of T and T is a subset of S. The proof of the next theorem that Nul $A =$ (Row $A)^\perp$ is established by showing that Nul A is a subset of (Row $A)^\perp$ and (Row $A)^\perp$ is a subset of Nul A. That is, an arbitrary element $\mathbf{x}$ in Nul A is shown to be in (Row $A)^\perp$, and then an arbitrary element $\mathbf{x}$ in (Row $A)^\perp$ is shown to be in Nul A.

THEOREM 3

Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\text{Row } A)^\perp = \text{Nul } A \quad \text{and} \quad (\text{Col } A)^\perp = \text{Nul } A^T$$

PROOF The row–column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in Nul A, then $\mathbf{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$). Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to Row A. Conversely, if $\mathbf{x}$ is orthogonal to Row A, then $\mathbf{x}$ is certainly orthogonal to each row of A, and hence $A\mathbf{x} = \mathbf{0}$. This proves the first statement of the theorem. Since this statement is true for any matrix, it is true for A^T. That is, the orthogonal complement of the row space of A^T is the null space of A^T. This proves the second statement, because Row $A^T =$ Col A.

Angles in $\mathbb{R}^2$ and $\mathbb{R}^3$ (Optional)

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in either $\mathbb{R}^2$ or $\mathbb{R}^3$, then there is a nice connection between their inner product and the angle ϑ between the two line segments from the origin to the points identified with $\mathbf{u}$ and $\mathbf{v}$. The formula is

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos \vartheta \tag{2}$$

To verify this formula for vectors in $\mathbb{R}^2$, consider the triangle shown in Figure 9, with sides of lengths $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{u} - \mathbf{v}\|$. By the law of cosines,

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\| \, \|\mathbf{v}\| \cos \vartheta$$

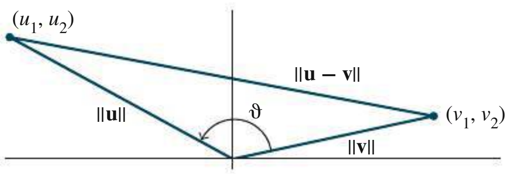


FIGURE 9 The angle between two vectors.

which can be rearranged to produce

$$\begin{aligned}\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\vartheta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2\right] \\ &= \frac{1}{2}\left[u_1^2 + u_2^2 + v_1^2 + v_2^2 - (u_1 - v_1)^2 - (u_2 - v_2)^2\right] \\ &= u_1 v_1 + u_2 v_2 \\ &= \mathbf{u}\cdot\mathbf{v}\end{aligned}$$

The verification for $\mathbb{R}^3$ is similar. When $n > 3$, formula (2) may be used to *define* the angle between two vectors in $\mathbb{R}^n$. In statistics, for instance, the value of $\cos\vartheta$ defined by (2) for suitable vectors $\mathbf{u}$ and $\mathbf{v}$ is what statisticians call a *correlation coefficient*.

Practice Problems

1. Let $\mathbf{a} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$. Compute $\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}$ and $\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\right)\mathbf{a}$.

2. Let $\mathbf{c} = \begin{bmatrix} 4/3 \\ -1 \\ 2/3 \end{bmatrix}$ and $\mathbf{d} = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix}$.

a. Find a unit vector $\mathbf{u}$ in the direction of $\mathbf{c}$.

b. Show that $\mathbf{d}$ is orthogonal to $\mathbf{c}$.

c. Use the results of (a) and (b) to explain why $\mathbf{d}$ must be orthogonal to the unit vector $\mathbf{u}$.

3. Let W be a subspace of R^n. Exercise 38 establishes that $W^\perp$ is also a subspace of R^n. Prove that $\dim W + \dim W^\perp = n$.

6.1 Exercises

Compute the quantities in Exercises 1–8 using the vectors

$$\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix},\quad \mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix},\quad \mathbf{w} = \begin{bmatrix} 3 \\ -1 \\ -5 \end{bmatrix},\quad \mathbf{x} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix}$$

1. $\mathbf{u}\cdot\mathbf{u}$, $\mathbf{v}\cdot\mathbf{u}$, and $\dfrac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$

2. $\mathbf{w}\cdot\mathbf{w}$, $\mathbf{x}\cdot\mathbf{w}$, and $\dfrac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{w}\cdot\mathbf{w}}$

3. $\dfrac{1}{\mathbf{w}\cdot\mathbf{w}}\mathbf{w}$

4. $\dfrac{1}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$

5. $\left(\dfrac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}$

6. $\left(\dfrac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{x}\cdot\mathbf{x}}\right)\mathbf{x}$

7. $\|\mathbf{w}\|$

8. $\|\mathbf{x}\|$

In Exercises 9–12, find a unit vector in the direction of the given vector.

9. $\begin{bmatrix} -30 \\ 40 \end{bmatrix}$

10. $\begin{bmatrix} 3 \\ 6 \\ -3 \end{bmatrix}$

11. $\begin{bmatrix} 7/4 \\ 1/2 \\ 1 \end{bmatrix}$

12. $\begin{bmatrix} 8/3 \\ 1 \end{bmatrix}$

13. Find the distance between $\mathbf{x} = \begin{bmatrix} 10 \\ -3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -1 \\ -5 \end{bmatrix}$.

14. Find the distance between $\mathbf{u} = \begin{bmatrix} 0 \\ -5 \\ 2 \end{bmatrix}$ and $\mathbf{z} = \begin{bmatrix} -4 \\ -1 \\ 4 \end{bmatrix}$.

Determine which pairs of vectors in Exercises 15–18 are orthogonal.

15. $\mathbf{a} = \begin{bmatrix} 8 \\ -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ -3 \end{bmatrix}$

16. $\mathbf{u} = \begin{bmatrix} 12 \\ 3 \\ -5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}$

17. $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ -5 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -4 \\ 1 \\ -2 \\ 6 \end{bmatrix}$

18. $\mathbf{y} = \begin{bmatrix} -3 \\ 7 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{z} = \begin{bmatrix} 1 \\ -8 \\ 15 \\ -7 \end{bmatrix}$

In Exercises 19–28, all vectors are in $\mathbb{R}^n$. Mark each statement True or False (**T/F**). Justify each answer.

19. **(T/F)** $\mathbf{v}\cdot\mathbf{v} = \|\mathbf{v}\|^2$.

20. **(T/F)** $\mathbf{u}\cdot\mathbf{v} - \mathbf{v}\cdot\mathbf{u} = 0$.

21. **(T/F)** If the distance from $\mathbf{u}$ to $\mathbf{v}$ equals the distance from $\mathbf{u}$ to $-\mathbf{v}$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

22. **(T/F)** If $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u}+\mathbf{v}\|^2$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

23. **(T/F)** If vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ span a subspace W and if $\mathbf{x}$ is orthogonal to each $\mathbf{v}_j$ for $j = 1, \dots, p$, then $\mathbf{x}$ is in $W^\perp$.

24. **(T/F)** If $\mathbf{x}$ is orthogonal to every vector in a subspace W then $\mathbf{x}$ is in $W^\perp$.

25. **(T/F)** For any scalar c, $\|c\mathbf{v}\| = c\|\mathbf{v}\|$.

26. **(T/F)** For any scalar c, $\mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$.

27. **(T/F)** For a square matrix A, vectors in Col A are orthogonal to vectors in Nul A.

28. **(T/F)** For an $m \times n$ matrix A, vectors in the null space of A are orthogonal to vectors in the row space of A.

29. Use the transpose definition of the inner product to verify parts (b) and (c) of Theorem 1. Mention the appropriate facts from Chapter 2.

30. Let $\mathbf{u} = (u_1, u_2, u_3)$. Explain why $\mathbf{u}\cdot\mathbf{u} \geq 0$. When is $\mathbf{u}\cdot\mathbf{u} = 0$?

31. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -7 \\ -4 \\ 6 \end{bmatrix}$. Compute and compare $\mathbf{u}\cdot\mathbf{v}$, $\|\mathbf{u}\|^2$, $\|\mathbf{v}\|^2$, and $\|\mathbf{u}+\mathbf{v}\|^2$. Do not use the Pythagorean Theorem.

32. Verify the *parallelogram law* for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

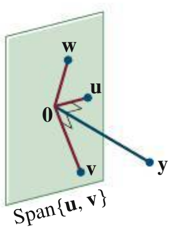
$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$$

33. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. Describe the set H of vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ that are orthogonal to $\mathbf{v}$. [*Hint:* Consider $\mathbf{v} = \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$.]

34. Let $\mathbf{u} = \begin{bmatrix} 5 \\ -6 \\ 7 \end{bmatrix}$, and let W be the set of all $\mathbf{x}$ in $\mathbb{R}^3$ such that $\mathbf{u}\cdot\mathbf{x} = 0$. What theorem in Chapter 4 can be used to show that W is a subspace of $\mathbb{R}^3$? Describe W in geometric language.

35. Suppose a vector $\mathbf{y}$ is orthogonal to vectors $\mathbf{u}$ and $\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to the vector $\mathbf{u}+\mathbf{v}$.

36. Suppose $\mathbf{y}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to every $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$. [*Hint:* An arbitrary $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$ has the form $\mathbf{w} = c_1\mathbf{u} + c_2\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to such a vector $\mathbf{w}$.]



37. Let $W = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$. Show that if $\mathbf{x}$ is orthogonal to each $\mathbf{v}_j$, for $1 \leq j \leq p$, then $\mathbf{x}$ is orthogonal to every vector in W.

38. Let W be a subspace of $\mathbb{R}^n$, and let $W^\perp$ be the set of all vectors orthogonal to W. Show that $W^\perp$ is a subspace of $\mathbb{R}^n$ using the following steps.

a. Take $\mathbf{z}$ in $W^\perp$, and let $\mathbf{u}$ represent any element of W. Then $\mathbf{z}\cdot\mathbf{u} = 0$. Take any scalar c and show that $c\mathbf{z}$ is orthogonal to $\mathbf{u}$. (Since $\mathbf{u}$ was an arbitrary element of W, this will show that $c\mathbf{z}$ is in $W^\perp$.)

b. Take $\mathbf{z}_1$ and $\mathbf{z}_2$ in $W^\perp$, and let $\mathbf{u}$ be any element of W. Show that $\mathbf{z}_1 + \mathbf{z}_2$ is orthogonal to $\mathbf{u}$. What can you conclude about $\mathbf{z}_1 + \mathbf{z}_2$? Why?

c. Finish the proof that $W^\perp$ is a subspace of $\mathbb{R}^n$.

39. Show that if $\mathbf{x}$ is in both W and $W^\perp$, then $\mathbf{x} = \mathbf{0}$.

T **40.** Construct a pair $\mathbf{u}$, $\mathbf{v}$ of random vectors in $\mathbb{R}^4$, and let

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

a. Denote the columns of A by $\mathbf{a}_1, \dots, \mathbf{a}_4$. Compute the length of each column, and compute $\mathbf{a}_1\cdot\mathbf{a}_2$, $\mathbf{a}_1\cdot\mathbf{a}_3$, $\mathbf{a}_1\cdot\mathbf{a}_4$, $\mathbf{a}_2\cdot\mathbf{a}_3$, $\mathbf{a}_2\cdot\mathbf{a}_4$, and $\mathbf{a}_3\cdot\mathbf{a}_4$.

b. Compute and compare the lengths of $\mathbf{u}$, $A\mathbf{u}$, $\mathbf{v}$, and $A\mathbf{v}$.

c. Use equation (2) in this section to compute the cosine of the angle between $\mathbf{u}$ and $\mathbf{v}$. Compare this with the cosine of the angle between $A\mathbf{u}$ and $A\mathbf{v}$.

d. Repeat parts (b) and (c) for two other pairs of random vectors. What do you conjecture about the effect of A on vectors?

T **41.** Generate random vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{v}$ in $\mathbb{R}^4$ with integer entries (and $\mathbf{v} \neq \mathbf{0}$), and compute the quantities

$$\left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v},\ \left(\frac{\mathbf{y}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v},\ \frac{(\mathbf{x}+\mathbf{y})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v},\ \frac{(10\mathbf{x})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$$

Repeat the computations with new random vectors $\mathbf{x}$ and $\mathbf{y}$. What do you conjecture about the mapping $\mathbf{x} \mapsto T(\mathbf{x}) = \left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}$ (for $\mathbf{v} \neq \mathbf{0}$)? Verify your conjecture algebraically.

T **42.** Let $A = \begin{bmatrix} -6 & 3 & -27 & -33 & -13 \\ 6 & -5 & 25 & 28 & 14 \\ 8 & -6 & 34 & 38 & 18 \\ 12 & -10 & 50 & 41 & 23 \\ 14 & -21 & 49 & 29 & 33 \end{bmatrix}$. Construct a matrix N whose columns form a basis for Nul A, and construct a matrix R whose *rows* form a basis for Row A (see Section 4.6 for details). Perform a matrix computation with N and R that illustrates a fact from Theorem 3.

Solutions to Practice Problems

1. $\mathbf{a}\cdot\mathbf{b} = 7$, $\mathbf{a}\cdot\mathbf{a} = 5$. Hence $\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}} = \dfrac{7}{5}$, and $\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\right)\mathbf{a} = \dfrac{7}{5}\mathbf{a} = \begin{bmatrix} -14/5 \\ 7/5 \end{bmatrix}$.

2. a. Scale $\mathbf{c}$, multiplying by 3 to get $\mathbf{y} = \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix}$. Compute $\|\mathbf{y}\|^2 = 29$ and $\|\mathbf{y}\| = \sqrt{29}$. The unit vector in the direction of both $\mathbf{c}$ and $\mathbf{y}$ is $\mathbf{u} = \dfrac{1}{\|\mathbf{y}\|}\mathbf{y} = \begin{bmatrix} 4/\sqrt{29} \\ -3/\sqrt{29} \\ 2/\sqrt{29} \end{bmatrix}$.

   b. $\mathbf{d}$ is orthogonal to $\mathbf{c}$, because

$$\mathbf{d}\cdot\mathbf{c} = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 4/3 \\ -1 \\ 2/3 \end{bmatrix} = \frac{20}{3} - 6 - \frac{2}{3} = 0$$

   c. $\mathbf{d}$ is orthogonal to $\mathbf{u}$, because $\mathbf{u}$ has the form $k\mathbf{c}$ for some k, and

$$\mathbf{d}\cdot\mathbf{u} = \mathbf{d}\cdot(k\mathbf{c}) = k(\mathbf{d}\cdot\mathbf{c}) = k(0) = 0$$

3. If $W \neq \{\mathbf{0}\}$, let $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ be a basis for W, where $1 \le p \le n$. Let A be the $p \times n$ matrix having rows $\mathbf{b}_1^T, \dots, \mathbf{b}_p^T$. It follows that W is the row space of A. Theorem 3 implies that $W^\perp = (\text{Row } A)^\perp = \text{Nul } A$ and hence $\dim W^\perp = \dim \text{Nul } A$. Thus, $\dim W + \dim W^\perp = \dim \text{Row } A + \dim \text{Nul } A = \text{rank } A + \dim \text{Nul } A = n$, by the Rank Theorem. If $W = \{\mathbf{0}\}$, then $W^\perp = \mathbb{R}^n$, and the result follows.

6.2 Orthogonal Sets

A set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

EXAMPLE 1 Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

SOLUTION Consider the three possible pairs of distinct vectors, namely $\{\mathbf{u}_1, \mathbf{u}_2\}$, $\{\mathbf{u}_1, \mathbf{u}_3\}$, and $\{\mathbf{u}_2, \mathbf{u}_3\}$.

$$\mathbf{u}_1\cdot\mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0$$
$$\mathbf{u}_1\cdot\mathbf{u}_3 = 3\left(-\tfrac{1}{2}\right) + 1(-2) + 1\left(\tfrac{7}{2}\right) = 0$$
$$\mathbf{u}_2\cdot\mathbf{u}_3 = -1\left(-\tfrac{1}{2}\right) + 2(-2) + 1\left(\tfrac{7}{2}\right) = 0$$

Each pair of distinct vectors is orthogonal, and so $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set. See Figure 1; the three line segments are mutually perpendicular. ■

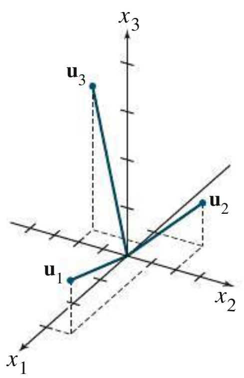


FIGURE 1

THEOREM 4 If $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

PROOF If $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \dots, c_p$, then

$$\begin{aligned} 0 = \mathbf{0}\cdot\mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= (c_1\mathbf{u}_1)\cdot\mathbf{u}_1 + (c_2\mathbf{u}_2)\cdot\mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) + c_2(\mathbf{u}_2\cdot\mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{u}_1) \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) \end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_p$. Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \dots, c_p$ must be zero. Thus S is linearly independent. ■

DEFINITION

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

The next theorem suggests why an orthogonal basis is much nicer than other bases. The weights in a linear combination can be computed easily.

THEOREM 5

Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y}\cdot\mathbf{u}_j}{\mathbf{u}_j\cdot\mathbf{u}_j} \qquad (j = 1, \dots, p)$$

PROOF As in the preceding proof, the orthogonality of $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ shows that

$$\mathbf{y}\cdot\mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 = c_1(\mathbf{u}_1\cdot\mathbf{u}_1)$$

Since $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero, the equation above can be solved for c_1. To find c_j for $j = 2, \dots, p$, compute $\mathbf{y}\cdot\mathbf{u}_j$ and solve for c_j. ■

EXAMPLE 2 The set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ in Example 1 is an orthogonal basis for $\mathbb{R}^3$. Express the vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S.

SOLUTION Compute

$$\begin{aligned} \mathbf{y}\cdot\mathbf{u}_1 &= 11, & \mathbf{y}\cdot\mathbf{u}_2 &= -12, & \mathbf{y}\cdot\mathbf{u}_3 &= -33 \\ \mathbf{u}_1\cdot\mathbf{u}_1 &= 11, & \mathbf{u}_2\cdot\mathbf{u}_2 &= 6, & \mathbf{u}_3\cdot\mathbf{u}_3 &= 33/2 \end{aligned}$$

By Theorem 5,

$$\begin{aligned} \mathbf{y} &= \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 \\ &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{33/2}\mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 \end{aligned}$$ ■

Notice how easy it is to compute the weights needed to build $\mathbf{y}$ from an orthogonal basis. If the basis were not orthogonal, it would be necessary to solve a system of linear equations in order to find the weights, as in Chapter 1.

We turn next to a construction that will become a key step in many calculations involving orthogonality, and it will lead to a geometric interpretation of Theorem 5.

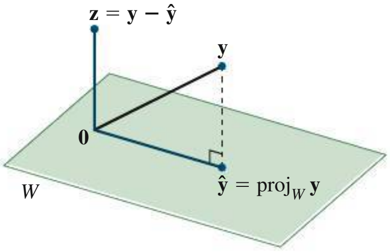


FIGURE 2
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

An Orthogonal Projection

Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$. We wish to write

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$. See Figure 2. Given any scalar α, let $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$, so that (1) is satisfied. Then $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\mathbf{u}$ if and only if

$$0 = (\mathbf{y} - \alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - (\alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - \alpha(\mathbf{u}\cdot\mathbf{u})$$

That is, (1) is satisfied with $\mathbf{z}$ orthogonal to $\mathbf{u}$ if and only if $\alpha = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$ and $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of y onto u**, and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**.

If c is any nonzero scalar and if $\mathbf{u}$ is replaced by $c\mathbf{u}$ in the definition of $\hat{\mathbf{y}}$, then the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{u}$ is exactly the same as the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ (Exercise 39). Hence this projection is determined by the *subspace* L spanned by $\mathbf{u}$ (the line through $\mathbf{u}$ and $\mathbf{0}$). Sometimes $\hat{\mathbf{y}}$ is denoted by $\text{proj}_L\,\mathbf{y}$ and is called the **orthogonal projection of y onto** L. That is,

$$\boxed{\hat{\mathbf{y}} = \text{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}} \tag{2}$$

EXAMPLE 3 Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

SOLUTION Compute

$$\mathbf{y}\cdot\mathbf{u} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40$$

$$\mathbf{u}\cdot\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20$$

The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{40}{20}\mathbf{u} = 2\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

and the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

The sum of these two vectors is $\mathbf{y}$. That is,

$$\underset{\substack{\uparrow \\ \mathbf{y}}}{\begin{bmatrix} 7 \\ 6 \end{bmatrix}} = \underset{\substack{\uparrow \\ \hat{\mathbf{y}}}}{\begin{bmatrix} 8 \\ 4 \end{bmatrix}} + \underset{\substack{\uparrow \\ (\mathbf{y} - \hat{\mathbf{y}})}}{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}$$

This decomposition of $\mathbf{y}$ is illustrated in Figure 3. *Note:* If the calculations above are correct, then $\{\hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}}\}$ will be an orthogonal set. As a check, compute

$$\hat{\mathbf{y}}\cdot(\mathbf{y} - \hat{\mathbf{y}}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix}\cdot\begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0 \qquad \blacksquare$$

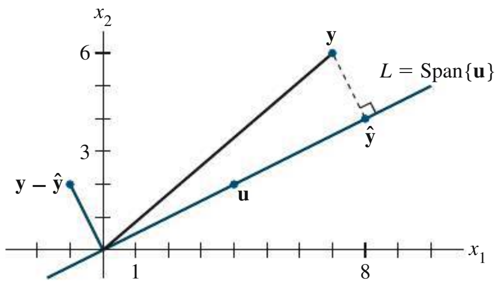


FIGURE 3 The orthogonal projection of $\mathbf{y}$ onto a line L through the origin.

Since the line segment in Figure 3 between $\mathbf{y}$ and $\hat{\mathbf{y}}$ is perpendicular to L, by construction of $\hat{\mathbf{y}}$, the point identified with $\hat{\mathbf{y}}$ is the closest point of L to $\mathbf{y}$. (This can be proved from geometry. We will assume this for $\mathbb{R}^2$ now and prove it for $\mathbb{R}^n$ in Section 6.3.)

EXAMPLE 4 Find the distance in Figure 3 from $\mathbf{y}$ to L.

SOLUTION The distance from $\mathbf{y}$ to L is the length of the perpendicular line segment from $\mathbf{y}$ to the orthogonal projection $\hat{\mathbf{y}}$. This length equals the length of $\mathbf{y} - \hat{\mathbf{y}}$. Thus the distance is

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}$$ ■

A Geometric Interpretation of Theorem 5

The formula for the orthogonal projection $\hat{\mathbf{y}}$ in (2) has the same appearance as each of the terms in Theorem 5. Thus Theorem 5 decomposes a vector $\mathbf{y}$ into a sum of orthogonal projections onto one-dimensional subspaces.

It is easy to visualize the case in which $W = \mathbb{R}^2 = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, with $\mathbf{u}_1$ and $\mathbf{u}_2$ orthogonal. Any $\mathbf{y}$ in $\mathbb{R}^2$ can be written in the form

$$\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 \tag{3}$$

The first term in (3) is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_1$ (the line through $\mathbf{u}_1$ and the origin), and the second term is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_2$. Thus (3) expresses $\mathbf{y}$ as the sum of its projections onto the (orthogonal) axes determined by $\mathbf{u}_1$ and $\mathbf{u}_2$. See Figure 4.

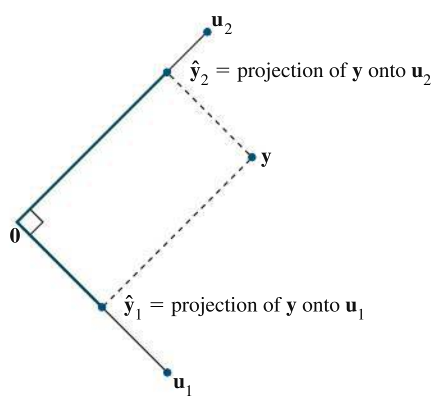


FIGURE 4 A vector decomposed into the sum of two projections.

Theorem 5 decomposes each $\mathbf{y}$ in Span $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ into the sum of p projections onto one-dimensional subspaces that are mutually orthogonal.

Decomposing a Force into Component Forces

The decomposition in Figure 4 can occur in physics when some sort of force is applied to an object. Choosing an appropriate coordinate system allows the force to be represented by a vector $\mathbf{y}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$. Often the problem involves some particular direction of interest, which is represented by another vector $\mathbf{u}$. For instance, if the object is moving in a straight line when the force is applied, the vector $\mathbf{u}$ might point in the direction of movement, as in Figure 5. A key step in the problem is to decompose the force into a component in the direction of $\mathbf{u}$ and a component orthogonal to $\mathbf{u}$. The calculations would be analogous to those previously made in Example 3.

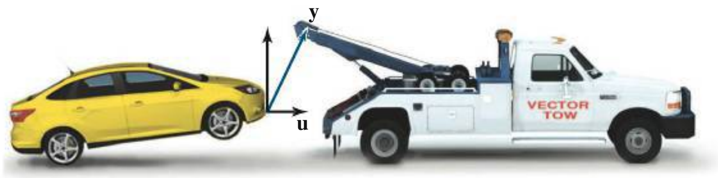


FIGURE 5

Orthonormal Sets

A set $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an **orthonormal basis** for W, since the set is automatically linearly independent, by Theorem 4.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Any nonempty subset of $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is orthonormal, too. Here is a more complicated example.

EXAMPLE 5 Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

SOLUTION Compute

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = -3/\sqrt{66} + 2/\sqrt{66} + 1/\sqrt{66} = 0$$
$$\mathbf{v}_1 \cdot \mathbf{v}_3 = -3/\sqrt{726} - 4/\sqrt{726} + 7/\sqrt{726} = 0$$
$$\mathbf{v}_2 \cdot \mathbf{v}_3 = 1/\sqrt{396} - 8/\sqrt{396} + 7/\sqrt{396} = 0$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set. Also,

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 9/11 + 1/11 + 1/11 = 1$$
$$\mathbf{v}_2 \cdot \mathbf{v}_2 = 1/6 + 4/6 + 1/6 = 1$$
$$\mathbf{v}_3 \cdot \mathbf{v}_3 = 1/66 + 16/66 + 49/66 = 1$$

which shows that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are unit vectors. Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal set. Since the set is linearly independent, its three vectors form a basis for $\mathbb{R}^3$. See Figure 6.

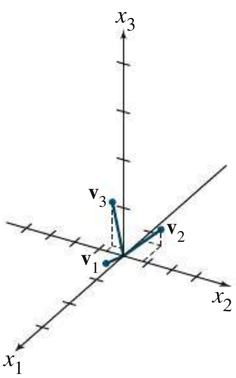


FIGURE 6

■

When the vectors in an orthogonal set of nonzero vectors are *normalized* to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set. See Exercise 40. It is easy to check that the vectors in Figure 6 (Example 5) are simply the unit vectors in the directions of the vectors in Figure 1 (Example 1).

Matrices whose columns form an orthonormal set are important in applications and in computer algorithms for matrix computations. Their main properties are given in Theorems 6 and 7.

THEOREM 6

An $m \times n$ matrix U has orthonormal columns if and only if $U^TU = I$.

PROOF To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$. The proof of the general case is essentially the same. Let $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,]$ and compute

$$U^TU = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix} \tag{4}$$

The entries in the matrix at the right are inner products, using transpose notation. The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0 \tag{5}$$

The columns of U all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1, \quad \mathbf{u}_2^T\mathbf{u}_2 = 1, \quad \mathbf{u}_3^T\mathbf{u}_3 = 1 \tag{6}$$

The theorem follows immediately from (4)–(6). ■

THEOREM 7

Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then

a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
b. $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$
c. $(U\mathbf{x})\cdot(U\mathbf{y}) = 0$ if and only if $\mathbf{x}\cdot\mathbf{y} = 0$

Properties (a) and (c) say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality. These properties are crucial for many computer algorithms. See Exercise 33 for the proof of Theorem 7.

EXAMPLE 6 Let $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. Notice that U has orthonormal columns and

$$U^TU = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Verify that $\|U\mathbf{x}\| = \|\mathbf{x}\|$.

SOLUTION

$$U\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$

$$\|U\mathbf{x}\| = \sqrt{9+1+1} = \sqrt{11}$$

$$\|\mathbf{x}\| = \sqrt{2+9} = \sqrt{11}$$ ■

Theorems 6 and 7 are particularly useful when applied to *square* matrices. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. By Theorem 6, such a matrix has orthonormal columns.[1] It is easy to see that any *square* matrix with orthonormal columns is an orthogonal matrix. Surprisingly, such a matrix must have orthonormal *rows*, too. See Exercises 35 and 36. Orthogonal matrices will appear frequently in Chapter 7.

EXAMPLE 7 The matrix

$$U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}$$

is an orthogonal matrix because it is square and because its columns are orthonormal, by Example 5. Verify that the rows are orthonormal, too! ■

Practice Problems

1. Let $\mathbf{u}_1 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$. Show that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for $\mathbb{R}^2$.
2. Let $\mathbf{y}$ and L be as in Example 3 and Figure 3. Compute the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto L using $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ instead of the $\mathbf{u}$ in Example 3.
3. Let U and $\mathbf{x}$ be as in Example 6, and let $\mathbf{y} = \begin{bmatrix} -3\sqrt{2} \\ 6 \end{bmatrix}$. Verify that $U\mathbf{x}\cdot U\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$.
4. Let U be an $n \times n$ matrix with orthonormal columns. Show that $\det U = \pm 1$.

6.2 Exercises

In Exercises 1–6, determine which sets of vectors are orthogonal.

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

2. $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}$

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

4. $\begin{bmatrix} 2 \\ -5 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$

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

6. $\begin{bmatrix} 5 \\ -4 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -4 \\ 1 \\ -3 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 5 \\ -1 \end{bmatrix}$

In Exercises 7–10, show that $\{\mathbf{u}_1, \mathbf{u}_2\}$ or $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^2$ or $\mathbb{R}^3$, respectively. Then express $\mathbf{x}$ as a linear combination of the $\mathbf{u}$'s.

7. $\mathbf{u}_1 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 9 \\ -7 \end{bmatrix}$

[1] A better name might be *orthonormal matrix*, and this term is found in some statistics texts. However, *orthogonal matrix* is the standard term in linear algebra.

8. $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$

9. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 8 \\ -4 \\ -3 \end{bmatrix}$

10. $\mathbf{u}_1 = \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 5 \\ -3 \\ 1 \end{bmatrix}$

11. Compute the orthogonal projection of $\begin{bmatrix} 1 \\ 7 \end{bmatrix}$ onto the line through $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ and the origin.

12. Compute the orthogonal projection of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ onto the line through $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and the origin.

13. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ -7 \end{bmatrix}$. Write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

14. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$. Write $\mathbf{y}$ as the sum of a vector in Span $\{\mathbf{u}\}$ and a vector orthogonal to $\mathbf{u}$.

15. Let $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$. Compute the distance from $\mathbf{y}$ to the line through $\mathbf{u}$ and the origin.

16. Let $\mathbf{y} = \begin{bmatrix} -3 \\ 9 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Compute the distance from $\mathbf{y}$ to the line through $\mathbf{u}$ and the origin.

In Exercises 17–22, determine which sets of vectors are orthonormal. If a set is only orthogonal, normalize the vectors to produce an orthonormal set.

17. $\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$

18. $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$

19. $\begin{bmatrix} -.6 \\ .8 \end{bmatrix}, \begin{bmatrix} .8 \\ .6 \end{bmatrix}$

20. $\begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}, \begin{bmatrix} 1/3 \\ 2/3 \\ 0 \end{bmatrix}$

21. $\begin{bmatrix} 1/\sqrt{10} \\ 3/\sqrt{20} \\ 3/\sqrt{20} \end{bmatrix}, \begin{bmatrix} 3/\sqrt{10} \\ -1/\sqrt{20} \\ -1/\sqrt{20} \end{bmatrix}, \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

22. $\begin{bmatrix} 1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -2/3 \\ 1/3 \\ -2/3 \end{bmatrix}$

In Exercises 23–32, all vectors are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** Not every linearly independent set in $\mathbb{R}^n$ is an orthogonal set.

24. **(T/F)** Not every orthogonal set in $\mathbb{R}^n$ is linearly independent.

25. **(T/F)** If $\mathbf{y}$ is a linear combination of nonzero vectors from an orthogonal set, then the weights in the linear combination can be computed without row operations on a matrix.

26. **(T/F)** If a set $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ has the property that $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$, then S is an orthonormal set.

27. **(T/F)** If the vectors in an orthogonal set of nonzero vectors are normalized, then some of the new vectors may not be orthogonal.

28. **(T/F)** If the columns of an $m \times n$ matrix A are orthonormal, then the linear mapping $\mathbf{x} \mapsto A\mathbf{x}$ preserves lengths.

29. **(T/F)** A matrix with orthonormal columns is an orthogonal matrix.

30. **(T/F)** The orthogonal projection of $\mathbf{y}$ onto $\mathbf{v}$ is the same as the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{v}$ whenever $c \neq 0$.

31. **(T/F)** If L is a line through $\mathbf{0}$ and if $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto L, then $\|\hat{\mathbf{y}}\|$ gives the distance from $\mathbf{y}$ to L.

32. **(T/F)** An orthogonal matrix is invertible.

33. Prove Theorem 7. [*Hint:* For (a), compute $\|U\mathbf{x}\|^2$, or prove (b) first.]

34. Suppose W is a subspace of $\mathbb{R}^n$ spanned by n nonzero orthogonal vectors. Explain why $W = \mathbb{R}^n$.

35. Let U be a square matrix with orthonormal columns. Explain why U is invertible. (Mention the theorems you use.)

36. Let U be an $n \times n$ orthogonal matrix. Show that the rows of U form an orthonormal basis of $\mathbb{R}^n$.

37. Let U and V be $n \times n$ orthogonal matrices. Explain why UV is an orthogonal matrix. [That is, explain why UV is invertible and its inverse is $(UV)^T$.]

38. Let U be an orthogonal matrix, and construct V by interchanging some of the columns of U. Explain why V is an orthogonal matrix.

39. Show that the orthogonal projection of a vector $\mathbf{y}$ onto a line L through the origin in $\mathbb{R}^2$ does not depend on the choice of the nonzero $\mathbf{u}$ in L used in the formula for $\hat{\mathbf{y}}$. To do this, suppose $\mathbf{y}$ and $\mathbf{u}$ are given and $\hat{\mathbf{y}}$ has been computed by formula (2) in this section. Replace $\mathbf{u}$ in that formula by $c\mathbf{u}$, where c is an unspecified nonzero scalar. Show that the new formula gives the same $\hat{\mathbf{y}}$.

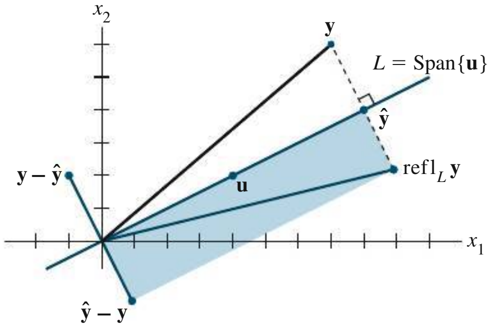
40. Let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be an orthogonal set of nonzero vectors, and let c_1, c_2 be any nonzero scalars. Show that $\{c_1\mathbf{v}_1, c_2\mathbf{v}_2\}$ is also an orthogonal set. Since orthogonality of a set is defined in terms of pairs of vectors, this shows that if the vectors in an orthogonal set are normalized, the new set will still be orthogonal.

41. Given $\mathbf{u} \neq \mathbf{0}$ in $\mathbb{R}^n$, let $L = \text{Span}\{\mathbf{u}\}$. Show that the mapping $\mathbf{x} \mapsto \text{proj}_L\,\mathbf{x}$ is a linear transformation.

42. Given $\mathbf{u} \neq \mathbf{0}$ in $\mathbb{R}^n$, let $L = \text{Span}\{\mathbf{u}\}$. For $\mathbf{y}$ in $\mathbb{R}^n$, the **reflection of y in** L is the point $\text{refl}_L\,\mathbf{y}$ defined by

$$\text{refl}_L\,\mathbf{y} = 2\ \text{proj}_L\,\mathbf{y} - \mathbf{y}$$

See the figure, which shows that $\text{refl}_L\,\mathbf{y}$ is the sum of $\hat{\mathbf{y}} = \text{proj}_L\,\mathbf{y}$ and $\hat{\mathbf{y}} - \mathbf{y}$. Show that the mapping $\mathbf{y} \mapsto \text{refl}_L\,\mathbf{y}$ is a linear transformation.



The reflection of $\mathbf{y}$ in a line through the origin.

T 43. Show that the columns of the matrix A are orthogonal by making an appropriate matrix calculation. State the calculation you use.

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

T 44. In parts (a)–(d), let U be the matrix formed by normalizing each column of the matrix A in Exercise 43.

a. Compute U^TU and UU^T. How do they differ?

b. Generate a random vector $\mathbf{y}$ in $\mathbb{R}^8$, and compute $\mathbf{p} = UU^T\mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \mathbf{p}$. Explain why $\mathbf{p}$ is in Col A. Verify that $\mathbf{z}$ is orthogonal to $\mathbf{p}$.

c. Verify that $\mathbf{z}$ is orthogonal to each column of U.

d. Notice that $\mathbf{y} = \mathbf{p} + \mathbf{z}$, with $\mathbf{p}$ in Col A. Explain why $\mathbf{z}$ is in $(\text{Col } A)^\perp$. (The significance of this decomposition of $\mathbf{y}$ will be explained in the next section.)

Solutions to Practice Problems

1. The vectors are orthogonal because

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = -2/5 + 2/5 = 0$$

They are unit vectors because

$$\|\mathbf{u}_1\|^2 = (-1/\sqrt{5})^2 + (2/\sqrt{5})^2 = 1/5 + 4/5 = 1$$
$$\|\mathbf{u}_2\|^2 = (2/\sqrt{5})^2 + (1/\sqrt{5})^2 = 4/5 + 1/5 = 1$$

In particular, the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly independent, and hence is a basis for $\mathbb{R}^2$ since there are two vectors in the set.

2. When $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$,

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{20}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 4\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

This is the same $\hat{\mathbf{y}}$ found in Example 3. The orthogonal projection does not depend on the $\mathbf{u}$ chosen on the line. See Exercise 39.

3. $U\mathbf{y} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}\begin{bmatrix} -3\sqrt{2} \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ 2 \end{bmatrix}$

Also, from Example 6, $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$ and $U\mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$. Hence

$$U\mathbf{x}\cdot U\mathbf{y} = 3 + 7 + 2 = 12, \quad \text{and} \quad \mathbf{x}\cdot\mathbf{y} = -6 + 18 = 12$$

STUDY GUIDE offers additional resources for mastering the concepts around an orthogonal basis.

4. Since U is an $n \times n$ matrix with orthonormal columns, by Theorem 6, $U^T U = I$. Taking the determinant of the left side of this equation, and applying Theorems 5 and 6 from Section 3.2 results in $\det U^T U = (\det U^T)(\det U) = (\det U)(\det U) = (\det U)^2$. Recall $\det I = 1$. Putting the two sides of the equation back together results in $(\det U)^2 = 1$ and hence $\det U = \pm 1$.

6.3 Orthogonal Projections

The orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$. Given a vector $\mathbf{y}$ and a subspace W in $\mathbb{R}^n$, there is a vector $\hat{\mathbf{y}}$ in W such that (1) $\hat{\mathbf{y}}$ is the unique vector in W for which $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W, and (2) $\hat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$. See Figure 1. These two properties of $\hat{\mathbf{y}}$ provide the key to finding least-squares solutions of linear systems.

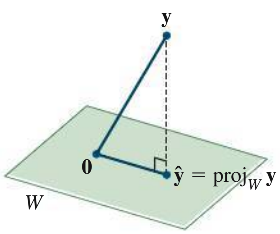


FIGURE 1

To prepare for the first theorem, observe that whenever a vector $\mathbf{y}$ is written as a linear combination of vectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ in $\mathbb{R}^n$, the terms in the sum for $\mathbf{y}$ can be grouped into two parts so that $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$$

where $\mathbf{z}_1$ is a linear combination of some of the $\mathbf{u}_i$ and $\mathbf{z}_2$ is a linear combination of the rest of the $\mathbf{u}_i$. This idea is particularly useful when $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is an orthogonal basis. Recall from Section 6.1 that $W^\perp$ denotes the set of all vectors orthogonal to a subspace W.

EXAMPLE 1 Let $\{\mathbf{u}_1, \dots, \mathbf{u}_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and let

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_5\mathbf{u}_5$$

Consider the subspace $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, and write $\mathbf{y}$ as the sum of a vector $\mathbf{z}_1$ in W and a vector $\mathbf{z}_2$ in $W^\perp$.

SOLUTION Write

$$\mathbf{y} = \underbrace{c_1\mathbf{u}_1 + c_2\mathbf{u}_2}_{\mathbf{z}_1} + \underbrace{c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5}_{\mathbf{z}_2}$$

where $\quad \mathbf{z}_1 = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \quad$ is in Span $\{\mathbf{u}_1, \mathbf{u}_2\}$

and $\quad \mathbf{z}_2 = c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5 \quad$ is in Span $\{\mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$.

To show that $\mathbf{z}_2$ is in $W^\perp$, it suffices to show that $\mathbf{z}_2$ is orthogonal to the vectors in the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for W. (See Section 6.1.) Using properties of the inner product, compute

$$\begin{aligned} \mathbf{z}_2 \cdot \mathbf{u}_1 &= (c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5) \cdot \mathbf{u}_1 \\ &= c_3\mathbf{u}_3 \cdot \mathbf{u}_1 + c_4\mathbf{u}_4 \cdot \mathbf{u}_1 + c_5\mathbf{u}_5 \cdot \mathbf{u}_1 \\ &= 0 \end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_3$, $\mathbf{u}_4$, and $\mathbf{u}_5$. A similar calculation shows that $\mathbf{z}_2 \cdot \mathbf{u}_2 = 0$. Thus $\mathbf{z}_2$ is in $W^\perp$. ■

The next theorem shows that the decomposition $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$ in Example 1 can be computed without having an orthogonal basis for $\mathbb{R}^n$. It is enough to have an orthogonal basis only for W.

THEOREM 8

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

The vector $\hat{\mathbf{y}}$ in (2) is called the **orthogonal projection of y onto** W and often is written as $\text{proj}_W\, \mathbf{y}$. See Figure 2. When W is a one-dimensional subspace, the formula for $\hat{\mathbf{y}}$ matches the formula given in Section 6.2.

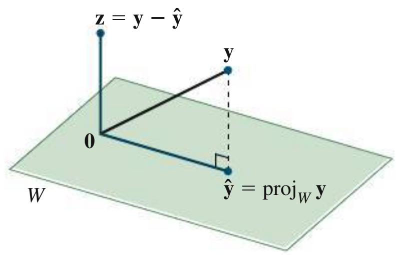


FIGURE 2 The orthogonal projection of $\mathbf{y}$ onto W.

PROOF Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be any orthogonal basis for W, and define $\hat{\mathbf{y}}$ by (2).[1] Then $\hat{\mathbf{y}}$ is in W because $\hat{\mathbf{y}}$ is a linear combination of the basis $\mathbf{u}_1, \dots, \mathbf{u}_p$. Let $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. Since $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_p$, it follows from (2) that

$$\begin{aligned}\mathbf{z}\cdot\mathbf{u}_1 = (\mathbf{y} - \hat{\mathbf{y}})\cdot\mathbf{u}_1 &= \mathbf{y}\cdot\mathbf{u}_1 - \left(\frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1\cdot\mathbf{u}_1 - 0 - \cdots - 0 \\ &= \mathbf{y}\cdot\mathbf{u}_1 - \mathbf{y}\cdot\mathbf{u}_1 = 0\end{aligned}$$

Thus $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$. Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u}_j$ in the basis for W. Hence $\mathbf{z}$ is orthogonal to every vector in W. That is, $\mathbf{z}$ is in $W^{\perp}$.

To show that the decomposition in (1) is unique, suppose $\mathbf{y}$ can also be written as $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$, with $\hat{\mathbf{y}}_1$ in W and $\mathbf{z}_1$ in $W^{\perp}$. Then $\hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ (since both sides equal $\mathbf{y}$), and so

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$$

This equality shows that the vector $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1$ is in W and in $W^{\perp}$ (because $\mathbf{z}_1$ and $\mathbf{z}$ are both in $W^{\perp}$, and $W^{\perp}$ is a subspace). Hence $\mathbf{v}\cdot\mathbf{v} = 0$, which shows that $\mathbf{v} = \mathbf{0}$. This proves that $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$ and also $\mathbf{z}_1 = \mathbf{z}$. ■

The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W and not on the particular basis used in (2).

[1] We may assume that W is not the zero subspace, for otherwise $W^{\perp} = \mathbb{R}^n$ and (1) is simply $\mathbf{y} = \mathbf{0} + \mathbf{y}$. The next section will show that any nonzero subspace of $\mathbb{R}^n$ has an orthogonal basis.

EXAMPLE 2 Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

SOLUTION The orthogonal projection of $\mathbf{y}$ onto W is

$$\begin{aligned} \hat{\mathbf{y}} &= \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 \\ &= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} \end{aligned}$$

Also

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

Theorem 8 ensures that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^{\perp}$. To check the calculations, however, it is a good idea to verify that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and hence to all of W. The desired decomposition of $\mathbf{y}$ is

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

■

A Geometric Interpretation of the Orthogonal Projection

When W is a one-dimensional subspace, the formula (2) for $\text{proj}_W\,\mathbf{y}$ contains just one term. Thus, when $\dim W > 1$, each term in (2) is itself an orthogonal projection of $\mathbf{y}$ onto a one-dimensional subspace spanned by one of the $\mathbf{u}$'s in the basis for W. Figure 3 illustrates this when W is a subspace of $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Here $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ denote the projections of $\mathbf{y}$ onto the lines spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively. The orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto W is the sum of the projections of $\mathbf{y}$ onto one-dimensional subspaces that are orthogonal to each other. The vector $\hat{\mathbf{y}}$ in Figure 3 corresponds to the vector $\mathbf{y}$ in Figure 4 of Section 6.2, because now it is $\hat{\mathbf{y}}$ that is in W.

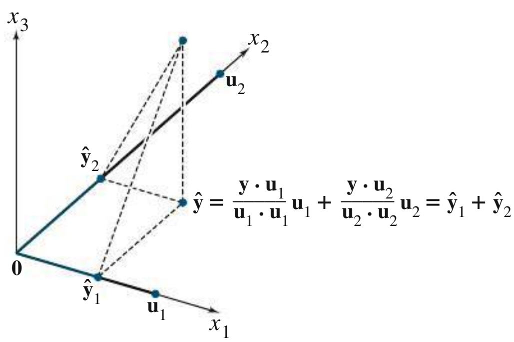


FIGURE 3 The orthogonal projection of $\mathbf{y}$ is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal.

Properties of Orthogonal Projections

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal basis for W and if $\mathbf{y}$ happens to be in W, then the formula for $\text{proj}_W\, \mathbf{y}$ is exactly the same as the representation of $\mathbf{y}$ given in Theorem 5 in Section 6.2. In this case, $\text{proj}_W\, \mathbf{y} = \mathbf{y}$.

If $\mathbf{y}$ is in $W = \text{Span}\,\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then $\text{proj}_W\, \mathbf{y} = \mathbf{y}$.

This fact also follows from the next theorem.

THEOREM 9

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{3}$$

for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

The vector $\hat{\mathbf{y}}$ in Theorem 9 is called **the best approximation to y by elements of** W. Later sections in the text will examine problems where a given $\mathbf{y}$ must be replaced, or *approximated*, by a vector $\mathbf{v}$ in some fixed subspace W. The distance from $\mathbf{y}$ to $\mathbf{v}$, given by $\|\mathbf{y} - \mathbf{v}\|$, can be regarded as the "error" of using $\mathbf{v}$ in place of $\mathbf{y}$. Theorem 9 says that this error is minimized when $\mathbf{v} = \hat{\mathbf{y}}$.

Inequality (3) leads to a new proof that $\hat{\mathbf{y}}$ does not depend on the particular orthogonal basis used to compute it. If a different orthogonal basis for W was used to construct an orthogonal projection of $\mathbf{y}$, then this projection would also be the closest point in W to $\mathbf{y}$, namely $\hat{\mathbf{y}}$.

PROOF Take $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$. See Figure 4. Then $\hat{\mathbf{y}} - \mathbf{v}$ is in W. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W. In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ (which is in W). Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2$$

(See the right triangle outlined in teal in Figure 4. The length of each side is labeled.) Now $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$ because $\hat{\mathbf{y}} - \mathbf{v} \neq \mathbf{0}$, and so inequality (3) follows immediately. ■

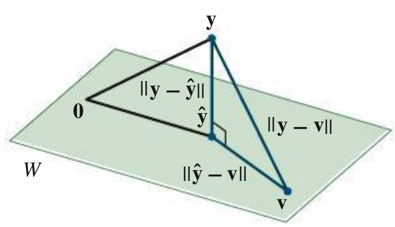


FIGURE 4 The orthogonal projection of $\mathbf{y}$ onto W is the closest point in W to $\mathbf{y}$.

EXAMPLE 3 If $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, and $W = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$, as in Example 2, then the closest point in W to $\mathbf{y}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$ ■

EXAMPLE 4 The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. Find the distance from $\mathbf{y}$ to $W = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

SOLUTION By the Best Approximation Theorem, the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \text{proj}_W\,\mathbf{y}$. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$$\hat{\mathbf{y}} = \frac{15}{30}\mathbf{u}_1 + \frac{-21}{6}\mathbf{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = 3^2 + 6^2 = 45$$

The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$. ■

The final theorem in this section shows how formula (2) for $\text{proj}_W\,\mathbf{y}$ is simplified when the basis for W is an orthonormal set.

THEOREM 10

If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W\,\mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p \tag{4}$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\,]$, then

$$\text{proj}_W\,\mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \tag{5}$$

PROOF Formula (4) follows immediately from (2) in Theorem 8. Also, (4) shows that $\text{proj}_W\,\mathbf{y}$ is a linear combination of the columns of U using the weights $\mathbf{y}\cdot\mathbf{u}_1$, $\mathbf{y}\cdot\mathbf{u}_2, \dots, \mathbf{y}\cdot\mathbf{u}_p$. The weights can be written as $\mathbf{u}_1^T\mathbf{y}, \mathbf{u}_2^T\mathbf{y}, \dots, \mathbf{u}_p^T\mathbf{y}$, showing that they are the entries in $U^T\mathbf{y}$ and justifying (5). ■

Suppose U is an $n \times p$ matrix with orthonormal columns, and let W be the column space of U. Then

$$U^TU\mathbf{x} = I_p\mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^p \qquad \text{Theorem 6}$$

$$UU^T\mathbf{y} = \text{proj}_W\,\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \qquad \text{Theorem 10}$$

If U is an $n \times n$ (square) matrix with orthonormal columns, then U is an *orthogonal* matrix, the column space W is all of $\mathbb{R}^n$, and $UU^T\mathbf{y} = I\mathbf{y} = \mathbf{y}$ for all $\mathbf{y}$ in $\mathbb{R}^n$.

Although formula (4) is important for theoretical purposes, in practice it usually involves calculations with square roots of numbers (in the entries of the $\mathbf{u}_i$). Formula (2) is recommended for hand calculations.

Example 9 of Section 2.1 illustrates how matrix multiplication and transposition are used to detect a specified pattern illustrated using blue and white squares. Now that we have more experience working with bases for W and $W^{\perp}$, we are ready to discuss how to set up the matrix M in Figure 6. Let $\mathbf{w}$ be the vector generated from a pattern of blue and white squares by turning each blue square into a 1 and each white square into a 0, and then lining up each column below the column before it. See Figure 5.

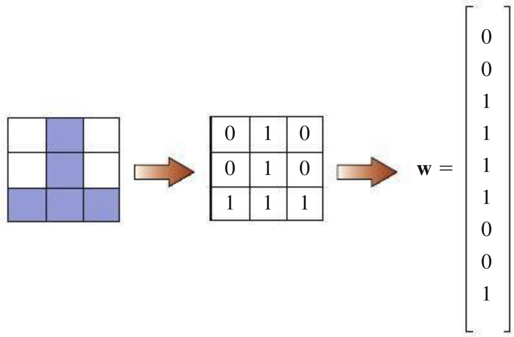


FIGURE 5 Creating a vector from colored squares.

Let $W = \text{span}\{\mathbf{w}\}$. Choose a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}\}$ for $W^{\perp}$. Create the matrix

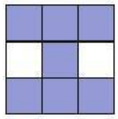
$B = \begin{bmatrix} \mathbf{v}_1{}^T \\ \mathbf{v}_2{}^T \\ \vdots \\ \mathbf{v}_{n-1}{}^T \end{bmatrix}$. Notice $B\mathbf{u} = \mathbf{0}$ if and only if $\mathbf{u}$ is orthogonal to a set of basis vectors for $W^{\perp}$, which happens if and only if $\mathbf{u}$ is in W. Set $M = B^T B$. Then $\mathbf{u}^T M\mathbf{u} = \mathbf{u}^T B^T B\mathbf{u} = (B\mathbf{u})^T B\mathbf{u}$. By Theorem 1, $(B\mathbf{u})^T B\mathbf{u} = 0$ if and only if $B\mathbf{u} = 0$, and hence $\mathbf{u}^T M\mathbf{u} = 0$ if and only if $\mathbf{u} \in W$. But there are only two vectors in W consisting of zeros and ones: $1\mathbf{w} = \mathbf{w}$ and $0\mathbf{w} = \mathbf{0}$. Thus we can conclude that if $\mathbf{u}^T M\mathbf{u} = 0$, but $\mathbf{u}^T\mathbf{u} \neq 0$, then $\mathbf{u} = \mathbf{w}$. See Figure 6.

EXAMPLE 5 Find a matrix M that can be used in Figure 6 to identify the perp symbol.

SOLUTION First change the symbol into a vector. Set $\mathbf{w} = [0\ \ 0\ 1\ \ 1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 1]^T$. Next set $W = \text{span}\{\mathbf{w}\}$ and find a basis for $W^{\perp}$: solving $\mathbf{x}^T\mathbf{w} = 0$ creates the homogeneous system of equations:

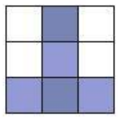
$$x_3 + x_4 + x_5 + x_6 + x_9 = 0$$

Treating x_3 as the basic variable and the remaining variables as free variables we get a basis for $W^{\perp}$. Transposing each vector in the basis and inserting it as a row of B we get



$$\mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{w}^T M \mathbf{w} = 2 \text{ and } \mathbf{w}^T \mathbf{w} = 7$$

This pattern is not the perpendicular symbol since $\mathbf{w}^T M \mathbf{w} \neq 0$.



$$\mathbf{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{w}^T M \mathbf{w} = 0 \text{ and } \mathbf{w}^T \mathbf{w} = 5$$

This pattern is the perpendicular symbol since $\mathbf{w}^T M \mathbf{w} = 0$, but $\mathbf{w}^T \mathbf{w} \neq 0$.

FIGURE 6 How AI detects the perp symbol.

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } M = B^T B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Notice $\mathbf{w}^T M \mathbf{w} = 0$, but $\mathbf{w}^T \mathbf{w} \neq 0$. ∎

Practice Problems

1. Let $\mathbf{u}_1 = \begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Use the fact that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal to compute $\text{proj}_W \mathbf{y}$.
2. Let W be a subspace of $\mathbb{R}^n$. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^n$ and let $\mathbf{z} = \mathbf{x} + \mathbf{y}$. If $\mathbf{u}$ is the projection of $\mathbf{x}$ onto W and $\mathbf{v}$ is the projection of $\mathbf{y}$ onto W, show that $\mathbf{u} + \mathbf{v}$ is the projection of $\mathbf{z}$ onto W.

6.3 Exercises

In Exercises 1 and 2, you may assume that $\{\mathbf{u}_1, \dots, \mathbf{u}_4\}$ is an orthogonal basis for $\mathbb{R}^4$.

1. $\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \\ -4 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 3 \\ 5 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ -4 \end{bmatrix}$, $\mathbf{u}_4 = \begin{bmatrix} 5 \\ -3 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 10 \\ -8 \\ 2 \\ 0 \end{bmatrix}$. Write $\mathbf{x}$ as the sum of two vectors, one in Span $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and the other in Span $\{\mathbf{u}_4\}$.

2. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ -1 \end{bmatrix}$, $\mathbf{u}_4 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ -2 \\ 2 \end{bmatrix}$. Write $\mathbf{v}$ as the sum of two vectors, one in Span $\{\mathbf{u}_1\}$ and the other in Span $\{\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

In Exercises 3–6, verify that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal set, and then find the orthogonal projection of $\mathbf{y}$ onto Span $\{\mathbf{u}_1, \mathbf{u}_2\}$.

3. $\mathbf{y} = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

4. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -4 \\ 3 \\ 0 \end{bmatrix}$

5. $\mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 6 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$

6. $\mathbf{y} = \begin{bmatrix} 4 \\ 4 \\ 1 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} -4 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

In Exercises 7–10, let W be the subspace spanned by the $\mathbf{u}$'s, and write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

7. $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}$

8. $\mathbf{y} = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}$

9. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ 3 \\ -1 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 3 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

10. $\mathbf{y} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 4 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

In Exercises 11 and 12, find the closest point to $\mathbf{y}$ in the subspace W spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.

11. $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 5 \\ 1 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

12. $\mathbf{y} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 13 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix}$

In Exercises 13 and 14, find the best approximation to $\mathbf{z}$ by vectors of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

13. $\mathbf{z} = \begin{bmatrix} 3 \\ -7 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ -3 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$

14. $\mathbf{z} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \\ -3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 5 \\ -2 \\ 4 \\ 2 \end{bmatrix}$

15. Let $\mathbf{y} = \begin{bmatrix} 5 \\ -9 \\ 5 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} -3 \\ -5 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$. Find the distance from $\mathbf{y}$ to the plane in $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$.

16. Let $\mathbf{y}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ be as in Exercise 12. Find the distance from $\mathbf{y}$ to the subspace of $\mathbb{R}^4$ spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.

17. Let $\mathbf{y} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

a. Let $U = [\,\mathbf{u}_1 \;\; \mathbf{u}_2\,]$. Compute U^TU and UU^T.

b. Compute $\text{proj}_W\,\mathbf{y}$ and $(UU^T)\mathbf{y}$.

18. Let $\mathbf{y} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1\}$.

a. Let U be the 2×1 matrix whose only column is $\mathbf{u}_1$. Compute U^TU and UU^T.

b. Compute $\text{proj}_W\,\mathbf{y}$ and $(UU^T)\mathbf{y}$.

19. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal but that $\mathbf{u}_3$ is not orthogonal to $\mathbf{u}_1$ or $\mathbf{u}_2$. It can be shown that $\mathbf{u}_3$ is not in the subspace W spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Use this fact to construct a nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$.

20. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be as in Exercise 19, and let $\mathbf{u}_4 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. It can be shown that $\mathbf{u}_4$ is not in the subspace W spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Use this fact to construct a nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$.

In Exercises 21–30, all vectors and subspaces are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$ and to $\mathbf{u}_2$ and if $W = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$, then $\mathbf{z}$ must be in $W^{\perp}$.

22. **(T/F)** For each $\mathbf{y}$ and each subspace W, the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$ is orthogonal to W.

23. **(T/F)** The orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto a subspace W can sometimes depend on the orthogonal basis for W used to compute $\hat{\mathbf{y}}$.

24. **(T/F)** If $\mathbf{y}$ is in a subspace W, then the orthogonal projection of $\mathbf{y}$ onto W is $\mathbf{y}$ itself.

25. **(T/F)** The best approximation to $\mathbf{y}$ by elements of a subspace W is given by the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$.

26. **(T/F)** If W is a subspace of $\mathbb{R}^n$ and if $\mathbf{v}$ is in both W and $W^{\perp}$, then $\mathbf{v}$ must be the zero vector.

27. **(T/F)** In the Orthogonal Decomposition Theorem, each term in formula (2) for $\hat{\mathbf{y}}$ is itself an orthogonal projection of $\mathbf{y}$ onto a subspace of W.

28. **(T/F)** If $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$, where $\mathbf{z}_1$ is in a subspace W and $\mathbf{z}_2$ is in $W^{\perp}$, then $\mathbf{z}_1$ must be the orthogonal projection of $\mathbf{y}$ onto W.

29. **(T/F)** If the columns of an $n \times p$ matrix U are orthonormal, then $UU^T\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto the column space of U.

30. **(T/F)** If an $n \times p$ matrix U has orthonormal columns, then $UU^T\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

31. Let A be an $m \times n$ matrix. Prove that every vector $\mathbf{x}$ in $\mathbb{R}^n$ can be written in the form $\mathbf{x} = \mathbf{p} + \mathbf{u}$, where $\mathbf{p}$ is in Row A and $\mathbf{u}$ is in Nul A. Also, show that if the equation $A\mathbf{x} = \mathbf{b}$ is consistent, then there is a unique $\mathbf{p}$ in Row A such that $A\mathbf{p} = \mathbf{b}$.

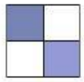
32. Let W be a subspace of $\mathbb{R}^n$ with an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_q\}$ be an orthogonal basis for $W^{\perp}$.

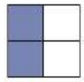
a. Explain why $\{\mathbf{w}_1, \ldots, \mathbf{w}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q\}$ is an orthogonal set.

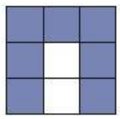
b. Explain why the set in part (a) spans $\mathbb{R}^n$.

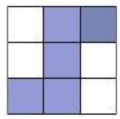
c. Show that $\dim W + \dim W^{\perp} = n$.

In Exercises 33–36, first change the given pattern into a vector $\mathbf{w}$ of zeros and ones and then use the method illustrated in Example 5 to find a matrix M so that $\mathbf{w}^T M \mathbf{w} = 0$, but $\mathbf{u}^T M \mathbf{u} \neq 0$ for all other nonzero vectors $\mathbf{u}$ of zeros and ones.

33. 

34. 

35. 

36. 

T 37. Let U be the 8×4 matrix in Exercise 43 in Section 6.2. Find the closest point to $\mathbf{y} = (1, 1, 1, 1, 1, 1, 1, 1)$ in Col U. Write the keystrokes or commands you use to solve this problem.

T 38. Let U be the matrix in Exercise 37. Find the distance from $\mathbf{b} = (1, 1, 1, 1, -1, -1, -1, -1)$ to Col U.

Solution to Practice Problems

1. Compute

$$\text{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \frac{88}{66}\mathbf{u}_1 + \frac{-2}{6}\mathbf{u}_2$$

$$= \frac{4}{3}\begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix} = \mathbf{y}$$

In this case, $\mathbf{y}$ happens to be a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$, so $\mathbf{y}$ is in W. The closest point in W to $\mathbf{y}$ is $\mathbf{y}$ itself.

2. Using Theorem 10, let U be a matrix whose columns consist of an orthonormal basis for W. Then $\text{proj}_W\, \mathbf{z} = UU^T\, \mathbf{z} = UU^T\,(\mathbf{x} + \mathbf{y}) = UU^T\mathbf{x} + UU^T\mathbf{y} = \text{proj}_W \mathbf{x} + \text{proj}_W \mathbf{y} = \mathbf{u} + \mathbf{v}$.

6.4 The Gram–Schmidt Process

The Gram–Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of $\mathbb{R}^n$. The first two examples of the process are aimed at hand calculation.

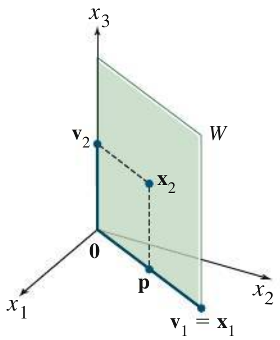


FIGURE 1
Construction of an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

EXAMPLE 1 Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

SOLUTION The subspace W is shown in Figure 1, along with $\mathbf{x}_1, \mathbf{x}_2$, and the projection $\mathbf{p}$ of $\mathbf{x}_2$ onto $\mathbf{x}_1$. The component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$ is $\mathbf{x}_2 - \mathbf{p}$, which is in W because it is formed from $\mathbf{x}_2$ and a multiple of $\mathbf{x}_1$. Let $\mathbf{v}_1 = \mathbf{x}_1$ and

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of nonzero vectors in W. Since $\dim W = 2$, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. ■

The next example fully illustrates the Gram–Schmidt process. Study it carefully.

EXAMPLE 2 Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

SOLUTION

Step 1. Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{Span}\{\mathbf{x}_1\} = \text{Span}\{\mathbf{v}_1\}$.

Step 2. Let $\mathbf{v}_2$ be the vector produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is, let

$$\begin{aligned} \mathbf{v}_2 &= \mathbf{x}_2 - \text{proj}_{W_1} \mathbf{x}_2 \\ &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \qquad \text{Since } \mathbf{v}_1 = \mathbf{x}_1 \\ &= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \end{aligned}$$

As in Example 1, $\mathbf{v}_2$ is the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

Step 2′ (optional). If appropriate, scale $\mathbf{v}_2$ to simplify later computations. Since $\mathbf{v}_2$ has fractional entries, it is convenient to scale it by a factor of 4 and replace $\{\mathbf{v}_1, \mathbf{v}_2\}$ by the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2' = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Step 3. Let $\mathbf{v}_3$ be the vector produced by subtracting from $\mathbf{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$ to compute this projection onto W_2:

$$\text{proj}_{W_2}\,\mathbf{x}_3 = \underbrace{\boxed{\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1}}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_1} + \underbrace{\boxed{\frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2'}}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_2'} = \frac{2}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

Then $\mathbf{v}_3$ is the component of $\mathbf{x}_3$ orthogonal to W_2, namely

$$\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{W_2}\,\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

See Figure 2 for a diagram of this construction. Observe that $\mathbf{v}_3$ is in W, because $\mathbf{x}_3$ and $\text{proj}_{W_2}\mathbf{x}_3$ are both in W. Thus $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal set of nonzero vectors and hence a linearly independent set in W. Note that W is three-dimensional since it was defined by a basis of three vectors. Hence, by the Basis Theorem in Section 4.5, $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal basis for W. ■

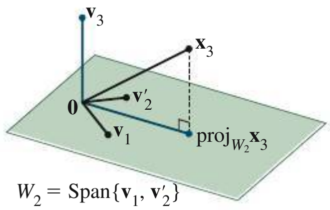


FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

The proof of the next theorem shows that this strategy really works. Scaling of vectors is not mentioned because that is used only to simplify hand calculations.

THEOREM 11

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\,\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \tag{1}$$

PROOF For $1 \le k \le p$, let $W_k = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$. Set $\mathbf{v}_1 = \mathbf{x}_1$, so that $\text{Span}\{\mathbf{v}_1\} = \text{Span}\{\mathbf{x}_1\}$. Suppose, for some $k < p$, we have constructed $\mathbf{v}_1, \dots, \mathbf{v}_k$ so that $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an orthogonal basis for W_k. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1} \tag{2}$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Note that $\text{proj}_{W_k} \mathbf{x}_{k+1}$ is in W_k and hence also in W_{k+1}. Since $\mathbf{x}_{k+1}$ is in W_{k+1}, so is $\mathbf{v}_{k+1}$ (because W_{k+1} is a subspace and is closed under subtraction). Furthermore, $\mathbf{v}_{k+1} \neq \mathbf{0}$ because $\mathbf{x}_{k+1}$ is not in $W_k = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$. Hence $\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$ is an orthogonal set of nonzero vectors in the $(k+1)$-dimensional space W_{k+1}. By the Basis Theorem in Section 4.5, this set is an orthogonal basis for W_{k+1}. Hence $W_{k+1} = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$. When $k + 1 = p$, the process stops. ■

Theorem 11 shows that any nonzero subspace W of $\mathbb{R}^n$ has an orthogonal basis, because an ordinary basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ is always available (by Theorem 12 in Section 4.5), and the Gram–Schmidt process depends only on the existence of orthogonal projections onto subspaces of W that already have orthogonal bases.

Orthonormal Bases

An orthonormal basis is constructed easily from an orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$: simply normalize (i.e., "scale") all the $\mathbf{v}_k$. When working problems by hand, this is easier than normalizing each $\mathbf{v}_k$ as soon as it is found (because it avoids unnecessary writing of square roots).

EXAMPLE 3 Example 1 constructed the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

An orthonormal basis is

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{45}}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$ ■

QR Factorization of Matrices

If an $m \times n$ matrix A has linearly independent columns $\mathbf{x}_1, \dots, \mathbf{x}_n$, then applying the Gram–Schmidt process (with normalizations) to $\mathbf{x}_1, \dots, \mathbf{x}_n$ amounts to *factoring* A, as described in the next theorem. This factorization is widely used in computer algorithms for various computations, such as solving equations (discussed in Section 6.5) and finding eigenvalues (mentioned in the exercises for Section 5.2).

THEOREM 12

The QR Factorization

If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

PROOF The columns of A form a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ for Col A. Construct an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $W = \text{Col } A$ with property (1) in Theorem 11. This basis may be constructed by the Gram–Schmidt process or some other means. Let

$$Q = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n\,]$$

For $k = 1, \dots, n, \mathbf{x}_k$ is in Span $\{\mathbf{x}_1, \dots, \mathbf{x}_k\} = \text{Span}\,\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$. So there are constants, $r_{1k}, \dots, r_{kk}$, such that

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\,\mathbf{u}_{k+1} + \cdots + 0\,\mathbf{u}_n$$

We may assume that $r_{kk} \geq 0$. (If $r_{kk} < 0$, multiply both r_{kk} and $\mathbf{u}_k$ by -1.) This shows that $\mathbf{x}_k$ is a linear combination of the columns of Q using as weights the entries in the vector

$$\mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

That is, $\mathbf{x}_k = Q\mathbf{r}_k$ for $k = 1, \dots, n$. Let $R = [\,\mathbf{r}_1 \quad \cdots \quad \mathbf{r}_n\,]$. Then

$$A = [\,\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n\,] = [\,Q\mathbf{r}_1 \quad \cdots \quad Q\mathbf{r}_n\,] = QR$$

The fact that R is invertible follows easily from the fact that the columns of A are linearly independent (Exercise 23). Since R is clearly upper triangular, its nonnegative diagonal entries must be positive. ■

EXAMPLE 4 Find a QR factorization of $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

SOLUTION The columns of A are the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ in Example 2. An orthogonal basis for Col $A = \text{Span}\,\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ was found in that example:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2' = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

To simplify the arithmetic that follows, scale $\mathbf{v}_3$ by letting $\mathbf{v}_3' = 3\mathbf{v}_3$. Then normalize the three vectors to obtain $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$, and use these vectors as the columns of Q:

$$Q = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}$$

By construction, the first k columns of Q are an orthonormal basis of Span $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$. From the proof of Theorem 12, $A = QR$ for some R. To find R, observe that $Q^TQ = I$, because the columns of Q are orthonormal. Hence

$$Q^TA = Q^T(QR) = IR = R$$

and

$$R = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{bmatrix}$$ ■

Numerical Notes

1. When the Gram–Schmidt process is run on a computer, roundoff error can build up as the vectors $\mathbf{u}_k$ are calculated, one by one. For j and k large but unequal, the inner products $\mathbf{u}_j^T\mathbf{u}_k$ may not be sufficiently close to zero. This loss of orthogonality can be reduced substantially by rearranging the order of the calculations.[1] However, a different computer-based QR factorization is usually preferred to this modified Gram–Schmidt method because it yields a more accurate orthonormal basis, even though the factorization requires about twice as much arithmetic.
2. To produce a QR factorization of a matrix A, a computer program usually left-multiplies A by a sequence of orthogonal matrices until A is transformed into an upper triangular matrix. This construction is analogous to the left-multiplication by elementary matrices that produces an LU factorization of A.

Practice Problems

1. Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1/3 \\ 1/3 \\ -2/3 \end{bmatrix}$. Construct an orthonormal basis for W.
2. Suppose $A = QR$, where Q is an $m \times n$ matrix with orthogonal columns and R is an $n \times n$ matrix. Show that if the columns of A are linearly dependent, then R cannot be invertible.

6.4 Exercises

In Exercises 1–6, the given set is a basis for a subspace W. Use the Gram–Schmidt process to produce an orthogonal basis for W.

1. $\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$

2. $\begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \\ -7 \end{bmatrix}$

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

4. $\begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 14 \\ -7 \end{bmatrix}$

5. $\begin{bmatrix} 1 \\ -4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 \\ -7 \\ -4 \\ 1 \end{bmatrix}$

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

[1] See *Fundamentals of Matrix Computations*, by David S. Watkins (New York: John Wiley & Sons, 1991), pp. 167–180.

7. Find an orthonormal basis of the subspace spanned by the vectors in Exercise 3.

8. Find an orthonormal basis of the subspace spanned by the vectors in Exercise 4.

Find an orthogonal basis for the column space of each matrix in Exercises 9–12.

9. $\begin{bmatrix} 3 & -5 & 1 \\ 1 & 1 & 1 \\ -1 & 5 & -2 \\ 3 & -7 & 8 \end{bmatrix}$ 10. $\begin{bmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{bmatrix}$

11. $\begin{bmatrix} 1 & 2 & 5 \\ -1 & 1 & -4 \\ -1 & 4 & -3 \\ 1 & -4 & 7 \\ 1 & 2 & 1 \end{bmatrix}$ 12. $\begin{bmatrix} 1 & 3 & 5 \\ -1 & -3 & 1 \\ 0 & 2 & 3 \\ 1 & 5 & 2 \\ 1 & 5 & 8 \end{bmatrix}$

In Exercises 13 and 14, the columns of Q were obtained by applying the Gram–Schmidt process to the columns of A. Find an upper triangular matrix R such that $A = QR$. Check your work.

13. $A = \begin{bmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{bmatrix}, Q = \begin{bmatrix} 5/6 & -1/6 \\ 1/6 & 5/6 \\ -3/6 & 1/6 \\ 1/6 & 3/6 \end{bmatrix}$

14. $A = \begin{bmatrix} -2 & 3 \\ 5 & 7 \\ 2 & -2 \\ 4 & 6 \end{bmatrix}, Q = \begin{bmatrix} -2/7 & 5/7 \\ 5/7 & 2/7 \\ 2/7 & -4/7 \\ 4/7 & 2/7 \end{bmatrix}$

15. Find a QR factorization of the matrix in Exercise 11.

16. Find a QR factorization of the matrix in Exercise 12.

In Exercises 17–22, all vectors and subspaces are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for W, then multiplying $\mathbf{v}_3$ by a scalar c gives a new orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, c\mathbf{v}_3\}$.

18. **(T/F)** If $W = \text{Span}\,\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ linearly independent, and if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set in W, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for W.

19. **(T/F)** The Gram–Schmidt process produces from a linearly independent set $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ an orthogonal set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ with the property that for each k, the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ span the same subspace as that spanned by $\mathbf{x}_1, \dots, \mathbf{x}_k$.

20. **(T/F)** If $\mathbf{x}$ is not in a subspace W, then $\mathbf{x} - \text{proj}_W \mathbf{x}$ is not zero.

21. **(T/F)** If $A = QR$, where Q has orthonormal columns, then $R = Q^T A$.

22. **(T/F)** In a QR factorization, say $A = QR$ (when A has linearly independent columns), the columns of Q form an orthonormal basis for the column space of A.

23. Suppose $A = QR$, where Q is $m \times n$ and R is $n \times n$. Show that if the columns of A are linearly independent, then R must be invertible. [*Hint:* Study the equation $R\mathbf{x} = \mathbf{0}$ and use the fact that $A = QR$.]

24. Suppose $A = QR$, where R is an invertible matrix. Show that A and Q have the same column space. [*Hint:* Given $\mathbf{y}$ in Col A, show that $\mathbf{y} = Q\mathbf{x}$ for some $\mathbf{x}$. Also, given $\mathbf{y}$ in Col Q, show that $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x}$.]

25. Given $A = QR$ as in Theorem 12, describe how to find an orthogonal $m \times m$ (square) matrix Q_1 and an invertible $n \times n$ upper triangular matrix R such that

$$A = Q_1 \begin{bmatrix} R \\ 0 \end{bmatrix}$$

The MATLAB `qr` command supplies this "full" QR factorization when rank $A = n$.

26. Let $\mathbf{u}_1, \dots, \mathbf{u}_p$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $T(\mathbf{x}) = \text{proj}_W \mathbf{x}$. Show that T is a linear transformation.

27. Suppose $A = QR$ is a QR factorization of an $m \times n$ matrix A (with linearly independent columns). Partition A as $[A_1 \quad A_2]$, where A_1 has p columns. Show how to obtain a QR factorization of A_1, and explain why your factorization has the appropriate properties.

T 28. Use the Gram–Schmidt process as in Example 2 to produce an orthogonal basis for the column space of

$$A = \begin{bmatrix} -10 & 13 & 7 & -11 \\ 2 & 1 & -5 & 3 \\ -6 & 3 & 13 & -3 \\ 16 & -16 & -2 & 5 \\ 2 & 1 & -5 & -7 \end{bmatrix}$$

T 29. Use the method in this section to produce a QR factorization of the matrix in Exercise 28.

T 30. For a matrix program, the Gram–Schmidt process works better with orthonormal vectors. Starting with $\mathbf{x}_1, \dots, \mathbf{x}_p$ as in Theorem 11, let $A = [\,\mathbf{x}_1 \ \cdots \ \mathbf{x}_p\,]$. Suppose Q is an $n \times k$ matrix whose columns form an orthonormal basis for the subspace W_k spanned by the first k columns of A. Then for $\mathbf{x}$ in $\mathbb{R}^n$, $QQ^T\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto W_k (Theorem 10 in Section 6.3). If $\mathbf{x}_{k+1}$ is the next column of A, then equation (2) in the proof of Theorem 11 becomes

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - Q(Q^T\mathbf{x}_{k+1})$$

(The parentheses above reduce the number of arithmetic operations.) Let $\mathbf{u}_{k+1} = \mathbf{v}_{k+1}/\|\mathbf{v}_{k+1}\|$. The new Q for the next step is $[\,Q \quad \mathbf{u}_{k+1}\,]$. Use this procedure to compute the QR factorization of the matrix in Exercise 28. Write the keystrokes or commands you use.

Solution to Practice Problems

1. Let $\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - 0\mathbf{v}_1 = \mathbf{x}_2$. So $\{\mathbf{x}_1, \mathbf{x}_2\}$ is already orthogonal. All that is needed is to normalize the vectors. Let

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Instead of normalizing $\mathbf{v}_2$ directly, normalize $\mathbf{v}_2' = 3\mathbf{v}_2$ instead:

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2'\|}\mathbf{v}_2' = \frac{1}{\sqrt{1^2 + 1^2 + (-2)^2}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$$

Then $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for W.

2. Since the columns of A are linearly dependent, there is a nontrivial vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. But then $QR\mathbf{x} = \mathbf{0}$. Applying Theorem 7 from Section 6.2 results in $\|R\mathbf{x}\| = \|QR\mathbf{x}\| = \|\mathbf{0}\| = 0$. But $\|R\mathbf{x}\| = 0$ implies $R\mathbf{x} = \mathbf{0}$, by Theorem 1 from Section 6.1. Thus there is a nontrivial vector $\mathbf{x}$ such that $R\mathbf{x} = \mathbf{0}$ and hence, by the Invertible Matrix Theorem, R cannot be invertible.

6.5 Least-Squares Problems

Inconsistent systems arise often in applications. When a solution is demanded and none exists, the best one can do is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

Think of $A\mathbf{x}$ as an *approximation* to $\mathbf{b}$. The smaller the distance between $\mathbf{b}$ and $A\mathbf{x}$, given by $\|\mathbf{b} - A\mathbf{x}\|$, the better the approximation. The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible. The adjective "least-squares" arises from the fact that $\|\mathbf{b} - A\mathbf{x}\|$ is the square root of a sum of squares.

DEFINITION

If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

The most important aspect of the least-squares problem is that no matter what $\mathbf{x}$ we select, the vector $A\mathbf{x}$ will necessarily be in the column space, Col A. So we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point in Col A to $\mathbf{b}$. See Figure 1. (Of course, if $\mathbf{b}$ happens to be in Col A, then $\mathbf{b}$ *is* $A\mathbf{x}$ for some $\mathbf{x}$, and such an $\mathbf{x}$ is a "least-squares solution.")

Solution of the General Least-Squares Problem

Given A and $\mathbf{b}$ as above, apply the Best Approximation Theorem in Section 6.3 to the subspace Col A. Let

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\,\mathbf{b}$$

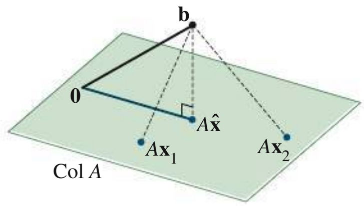


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Because $\hat{\mathbf{b}}$ is in the column space of A, the equation $A\mathbf{x} = \hat{\mathbf{b}}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{1}$$

Since $\hat{\mathbf{b}}$ is the closest point in Col A to $\mathbf{b}$, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\hat{\mathbf{x}}$ satisfies (1). Such an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathbf{b}}$ out of the columns of A. See Figure 2. [There are many solutions of (1) if the equation has free variables.]

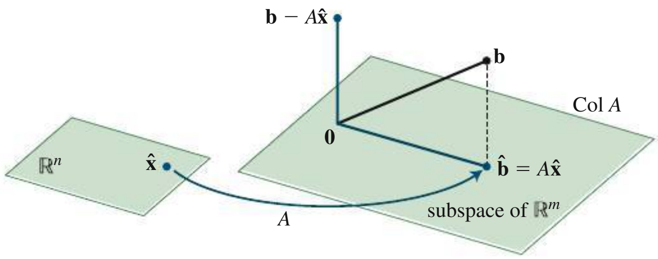


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Suppose $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. By the Orthogonal Decomposition Theorem in Section 6.3, the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. If $\mathbf{a}_j$ is any column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$, and $\mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \tag{2}$$

(This equation also follows from Theorem 3 in Section 6.1.) Thus

$$\begin{aligned} A^T\mathbf{b} - A^TA\hat{\mathbf{x}} &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} &= A^T\mathbf{b} \end{aligned}$$

These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$\boxed{A^TA\mathbf{x} = A^T\mathbf{b}} \tag{3}$$

The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$. A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

THEOREM 13

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

PROOF As shown, the set of least-squares solutions is nonempty and each least-squares solution $\hat{\mathbf{x}}$ satisfies the normal equations. Conversely, suppose $\hat{\mathbf{x}}$ satisfies $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $\hat{\mathbf{x}}$ satisfies (2), which shows that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T and hence is orthogonal to the columns of A. Since the columns of A span Col A, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to all of Col A. Hence the equation

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

is a decomposition of $\mathbf{b}$ into the sum of a vector in Col A and a vector orthogonal to Col A. By the uniqueness of the orthogonal decomposition, $A\hat{\mathbf{x}}$ must be the orthogonal projection of $\mathbf{b}$ onto Col A. That is, $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, and $\hat{\mathbf{x}}$ is a least-squares solution. ■

EXAMPLE 1 Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

SOLUTION To use normal equations (3), compute:

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Row operations can be used to solve this system, but since A^TA is invertible and 2×2, it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then to solve $A^TA\mathbf{x} = A^T\mathbf{b}$ as

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{aligned}$$ ■

In many calculations, A^TA is invertible, but this is not always the case. The next example involves a matrix of the sort that appears in what are called *analysis of variance* problems in statistics.

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

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

SOLUTION Compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

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

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$ ■

The next theorem gives useful criteria for determining when there is only one least-squares solution of $A\mathbf{x} = \mathbf{b}$. (Of course, the orthogonal projection $\hat{\mathbf{b}}$ is always unique.)

THEOREM 14

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

The main elements of a proof of Theorem 14 are outlined in Exercises 27–29, which also review concepts from Chapter 4. Formula (4) for $\hat{\mathbf{x}}$ is useful mainly for theoretical purposes and for hand calculations when A^TA is a 2×2 invertible matrix.

When a least-squares solution $\hat{\mathbf{x}}$ is used to produce $A\hat{\mathbf{x}}$ as an approximation to $\mathbf{b}$, the distance from $\mathbf{b}$ to $A\hat{\mathbf{x}}$ is called the **least-squares error** of this approximation.

EXAMPLE 3 Given A and $\mathbf{b}$ as in Example 1, determine the least-squares error in the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

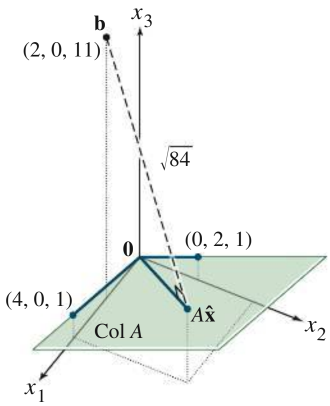


FIGURE 3

SOLUTION From Example 1,

$$\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \quad \text{and} \quad A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

Hence

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

and

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$$

The least-squares error is $\sqrt{84}$. For any $\mathbf{x}$ in $\mathbb{R}^2$, the distance between $\mathbf{b}$ and the vector $A\mathbf{x}$ is at least $\sqrt{84}$. See Figure 3. Note that the least-squares solution $\hat{\mathbf{x}}$ itself does not appear in the figure. ■

Alternative Calculations of Least-Squares Solutions

The next example shows how to find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ when the columns of A are orthogonal. Such matrices often appear in linear regression problems, discussed in the next section.

EXAMPLE 4 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

SOLUTION Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 \tag{5}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

Now that $\hat{\mathbf{b}}$ is known, we can solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. But this is trivial, since we already know what weights to place on the columns of A to produce $\hat{\mathbf{b}}$. It is clear from (5) that

$$\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$ ■

In some cases, the normal equations for a least-squares problem can be *ill-conditioned*; that is, small errors in the calculations of the entries of A^TA can sometimes cause relatively large errors in the solution $\hat{\mathbf{x}}$. If the columns of A are linearly independent, the least-squares solution can often be computed more reliably through a QR factorization of A (described in Section 6.4).[1]

[1] The QR method is compared with the standard normal equation method in G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. (Baltimore: Johns Hopkins Press, 1996), pp. 230–231.

THEOREM 15

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A as in Theorem 12. Then, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b} \tag{6}$$

PROOF Let $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$. Then

$$A\hat{\mathbf{x}} = QR\hat{\mathbf{x}} = QRR^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b}$$

By Theorem 12, the columns of Q form an orthonormal basis for Col A. Hence, by Theorem 10, $QQ^T\mathbf{b}$ is the orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto Col A. Then $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, which shows that $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$. The uniqueness of $\hat{\mathbf{x}}$ follows from Theorem 14. ■

Numerical Notes

Since R in Theorem 15 is upper triangular, $\hat{\mathbf{x}}$ should be calculated as the exact solution of the equation

$$R\mathbf{x} = Q^T\mathbf{b} \tag{7}$$

It is much faster to solve (7) by back-substitution or row operations than to compute R^{-1} and use (6).

EXAMPLE 5 Find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

SOLUTION The QR factorization of A can be obtained as in Section 6.4.

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

Then

$$Q^T\mathbf{b} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

The least-squares solution $\hat{\mathbf{x}}$ satisfies $R\mathbf{x} = Q^T\mathbf{b}$; that is,

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

This equation is solved easily and yields $\hat{\mathbf{x}} = \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}$. ■

Practice Problems

1. Let $A = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$. Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$, and compute the associated least-squares error.
2. What can you say about the least-squares solution of $A\mathbf{x} = \mathbf{b}$ when $\mathbf{b}$ is orthogonal to the columns of A?

6.5 Exercises

In Exercises 1–4, find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ by (a) constructing the normal equations for $\hat{\mathbf{x}}$ and (b) solving for $\hat{\mathbf{x}}$.

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

2. $A = \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix}$

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

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

In Exercises 5 and 6, describe all least-squares solutions of the equation $A\mathbf{x} = \mathbf{b}$.

5. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix}$

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

7. Compute the least-squares error associated with the least-squares solution found in Exercise 3.

8. Compute the least-squares error associated with the least-squares solution found in Exercise 4.

In Exercises 9–12, find (a) the orthogonal projection of $\mathbf{b}$ onto Col A and (b) a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

9. $A = \begin{bmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4 \\ -2 \\ -3 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$

11. $A = \begin{bmatrix} 4 & 0 & 1 \\ 1 & -5 & 1 \\ 6 & 1 & 0 \\ 1 & -1 & -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 9 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 5 \\ 6 \\ 6 \end{bmatrix}$

13. Let $A = \begin{bmatrix} 3 & 4 \\ -2 & 1 \\ 3 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 11 \\ -9 \\ 5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Could $\mathbf{u}$ possibly be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

14. Let $A = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Is it possible that at least one of $\mathbf{u}$ or $\mathbf{v}$ could be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

In Exercises 15 and 16, use the factorization $A = QR$ to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

15. $A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ 2/3 & 2/3 \\ 1/3 & -2/3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix}$

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

In Exercises 17–26, A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$. Mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

18. **(T/F)** If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution.

19. **(T/F)** A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto Col A.

20. **(T/F)** A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that $\|\mathbf{b} - A\mathbf{x}\| \leq \|\mathbf{b} - A\hat{\mathbf{x}}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

21. **(T/F)** Any solution of $A^T A\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

22. **(T/F)** If the columns of A are linearly independent, then the equation $A\mathbf{x} = \mathbf{b}$ has exactly one least-squares solution.

23. **(T/F)** The least-squares solution of $A\mathbf{x} = \mathbf{b}$ is the point in the column space of A closest to $\mathbf{b}$.

24. **(T/F)** A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a list of weights that, when applied to the columns of A, produces the orthogonal projection of $\mathbf{b}$ onto Col A.

25. **(T/F)** The normal equations always provide a reliable method for computing least-squares solutions.

26. **(T/F)** If A has a QR factorization, say $A = QR$, then the best way to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ is to compute $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

27. Let A be an $m \times n$ matrix. Use the steps below to show that a vector $\mathbf{x}$ in $\mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{0}$ if and only if $A^TA\mathbf{x} = \mathbf{0}$. This will show that Nul A = Nul A^TA.

a. Show that if $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = \mathbf{0}$.

b. Suppose $A^TA\mathbf{x} = \mathbf{0}$. Explain why $\mathbf{x}^TA^TA\mathbf{x} = \mathbf{0}$, and use this to show that $A\mathbf{x} = \mathbf{0}$.

28. Let A be an $m \times n$ matrix such that A^TA is invertible. Show that the columns of A are linearly independent. [*Careful:* You may not assume that A is invertible; it may not even be square.]

29. Let A be an $m \times n$ matrix whose columns are linearly independent. [*Careful:* A need not be square.]

a. Use Exercise 27 to show that A^TA is an invertible matrix.

b. Explain why A must have at least as many rows as columns.

c. Determine the rank of A.

30. Use Exercise 27 to show that rank A^TA = rank A. [*Hint:* How many columns does A^TA have? How is this connected with the rank of A^TA?]

31. Suppose A is $m \times n$ with linearly independent columns and $\mathbf{b}$ is in $\mathbb{R}^m$. Use the normal equations to produce a formula for $\hat{\mathbf{b}}$, the projection of $\mathbf{b}$ onto Col A. [*Hint:* Find $\hat{\mathbf{x}}$ first. The formula does not require an orthogonal basis for Col A.]

32. Find a formula for the least-squares solution of $A\mathbf{x} = \mathbf{b}$ when the columns of A are orthonormal.

33. Describe all least-squares solutions of the system

$$x + y = 2$$
$$x + y = 4$$

T **34.** Example 2 in Section 4.8 displayed a low-pass linear filter that changed a signal $\{y_k\}$ into $\{y_{k+1}\}$ and changed a higher-frequency signal $\{w_k\}$ into the zero signal, where $y_k = \cos(\pi k/4)$ and $w_k = \cos(3\pi k/4)$. The following calculations will design a filter with approximately those properties. The filter equation is

$$a_0 y_{k+2} + a_1 y_{k+1} + a_2 y_k = z_k \qquad \text{for all } k \tag{8}$$

Because the signals are periodic, with period 8, it suffices to study equation (8) for $k = 0, \ldots, 7$. The action on the two signals described above translates into two sets of eight equations, shown below:

$$\begin{array}{cc} & \begin{array}{ccc} y_{k+2} & y_{k+1} & y_k \end{array} \\ \begin{array}{c} k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array} & \begin{bmatrix} 0 & .7 & 1 \\ -.7 & 0 & .7 \\ -1 & -.7 & 0 \\ -.7 & -1 & -.7 \\ 0 & -.7 & -1 \\ .7 & 0 & -.7 \\ 1 & .7 & 0 \\ .7 & 1 & .7 \end{bmatrix} \end{array} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{array}{c} y_{k+1} \\ \begin{bmatrix} .7 \\ 0 \\ -.7 \\ -1 \\ -.7 \\ 0 \\ .7 \\ 1 \end{bmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{array}{ccc} w_{k+2} & w_{k+1} & w_k \end{array} \\ \begin{array}{c} k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array} & \begin{bmatrix} 0 & -.7 & 1 \\ .7 & 0 & -.7 \\ -1 & .7 & 0 \\ .7 & -1 & .7 \\ 0 & .7 & -1 \\ -.7 & 0 & .7 \\ 1 & -.7 & 0 \\ -.7 & 1 & -.7 \end{bmatrix} \end{array} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Write an equation $A\mathbf{x} = \mathbf{b}$, where A is a 16×3 matrix formed from the two coefficient matrices above and where $\mathbf{b}$ in $\mathbb{R}^{16}$ is formed from the two right sides of the equations. Find a_0, a_1, and a_2 given by the least-squares solution of $A\mathbf{x} = \mathbf{b}$. (The .7 in the data above was used as an approximation for $\sqrt{2}/2$, to illustrate how a typical computation in an applied problem might proceed. If .707 were used instead, the resulting filter coefficients would agree to at least seven decimal places with $\sqrt{2}/4$, $1/2$, and $\sqrt{2}/4$, the values produced by exact arithmetic calculations.)

Solutions to Practice Problems

1. First, compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ -3 & 5 & 7 \\ -3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 9 & 0 \\ 9 & 83 & 28 \\ 0 & 28 & 14 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ -3 & 5 & 7 \\ -3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix} = \begin{bmatrix} -3 \\ -65 \\ -28 \end{bmatrix}$$

Next, row reduce the augmented matrix for the normal equations, $A^TA\mathbf{x} = A^T\mathbf{b}$:

$$\begin{bmatrix} 3 & 9 & 0 & -3 \\ 9 & 83 & 28 & -65 \\ 0 & 28 & 14 & -28 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 56 & 28 & -56 \\ 0 & 28 & 14 & -28 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & -3/2 & 2 \\ 0 & 1 & 1/2 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The general least-squares solution is $x_1 = 2 + \frac{3}{2}x_3$, $x_2 = -1 - \frac{1}{2}x_3$, with x_3 free. For one specific solution, take $x_3 = 0$ (for example), and get

$$\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

To find the least-squares error, compute

$$\hat{\mathbf{b}} = A\hat{\mathbf{x}} = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$$

It turns out that $\hat{\mathbf{b}} = \mathbf{b}$, so $\|\mathbf{b} - \hat{\mathbf{b}}\| = 0$. The least-squares error is zero because $\mathbf{b}$ happens to be in Col A.

2. If $\mathbf{b}$ is orthogonal to the columns of A, then the projection of $\mathbf{b}$ onto the column space of A is $\mathbf{0}$. In this case, a least-squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ satisfies $A\hat{\mathbf{x}} = \mathbf{0}$.

6.6 Machine Learning and Linear Models

Machine Learning

Machine learning uses linear models in situations where the machine is being *trained* to predict the outcome (dependent variables) based on the values of the inputs (independent variables). The machine is given a set of training data where the values of the independent and dependent variables are known. The machine then *learns* the relationship between the independent variables and the dependent variables. One type of learning is to fit a curve, such as a least-squares line or parabola, to the data. Once the machine has learned the pattern from the training data, it can then estimate the value of the output based on a given value for the input.

Least-Squares Lines

A common task in science and engineering is to analyze and understand relationships among several quantities that vary. This section describes a variety of situations in which data are used to build or verify a formula that predicts the value of one variable as a function of other variables. In each case, the problem will amount to solving a least-squares problem.

For easy application of the discussion to real problems that you may encounter later in your career, we choose notation that is commonly used in the statistical analysis of scientific and engineering data. Instead of $A\mathbf{x} = \mathbf{b}$, we write $X\boldsymbol{\beta} = \mathbf{y}$ and refer to X as the **design matrix**, $\boldsymbol{\beta}$ as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

The simplest relation between two variables x and y is the linear equation $y = \beta_0 + \beta_1 x$.[1] Experimental data often produce points $(x_1, y_1), \dots, (x_n, y_n)$ that, when graphed, seem to lie close to a line. We want to determine the parameters β_0 and β_1 that make the line as "close" to the points as possible.

Suppose β_0 and β_1 are fixed, and consider the line $y = \beta_0 + \beta_1 x$ in Figure 1. Corresponding to each data point (x_j, y_j) there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinate. We call y_j the *observed* value of y and $\beta_0 + \beta_1 x_j$ the *predicted* y-value (determined by the line). The difference between an observed y-value and a predicted y-value is called a *residual*.

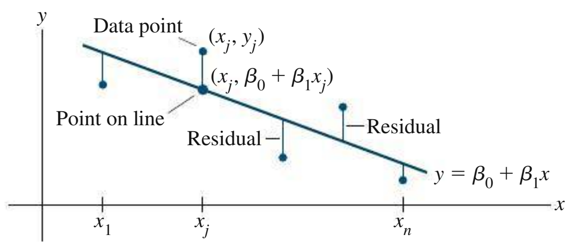


FIGURE 1 Fitting a line to experimental data.

There are several ways to measure how "close" the line is to the data. The usual choice (primarily because the mathematical calculations are simple) is to add the squares of the residuals. The **least-squares line** is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals. This line is also called a **line of regression of y on x**, because any errors in the data are assumed to be only in the y-coordinates. The coefficients β_0, β_1 of the line are called (linear) **regression coefficients**.[2]

If the data points were on the line, the parameters β_0 and β_1 would satisfy the equations

Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

[1] This notation is commonly used for least-squares lines instead of $y = mx + b$.

[2] If the measurement errors are in x instead of y, simply interchange the coordinates of the data (x_j, y_j) before plotting the points and computing the regression line. If both coordinates are subject to possible error, then you might choose the line that minimizes the sum of the squares of the *orthogonal* (perpendicular) distances from the points to the line.

We can write this system as

$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{1}$$

Of course, if the data points don't lie on a line, then there are no parameters β_0, β_1 for which the predicted y-values in $X\boldsymbol{\beta}$ equal the observed y-values in $\mathbf{y}$, and $X\boldsymbol{\beta} = \mathbf{y}$ has no solution. This is a least-squares problem, $A\mathbf{x} = \mathbf{b}$, with different notation!

The square of the distance between the vectors $X\boldsymbol{\beta}$ and $\mathbf{y}$ is precisely the sum of the squares of the residuals. The $\boldsymbol{\beta}$ that minimizes this sum also minimizes the distance between $X\boldsymbol{\beta}$ and $\mathbf{y}$. *Computing the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$ is equivalent to finding the $\boldsymbol{\beta}$ that determines the least-squares line in Figure 1.*

EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points (2, 1), (5, 2), (7, 3), and (8, 3).

SOLUTION Use the x-coordinates of the data to build the design matrix X in (1) and the y-coordinates to build the observation vector $\mathbf{y}$:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, obtain the normal equations (with the new notation):

$$X^T X\boldsymbol{\beta} = X^T\mathbf{y}$$

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14}x$$

See Figure 2. ■

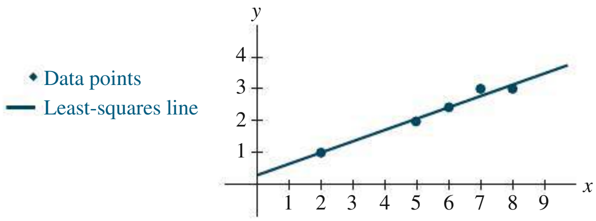


FIGURE 2 The least-squares line $y = \frac{2}{7} + \frac{5}{14}x$.

EXAMPLE 2 If a machine learns the data from Example 1 by creating a least-squares line, what outcome will it predict for the inputs 4 and 6?

SOLUTION The machine would perform the same calculations as in Example 1 to arrive at the least-squares line

$$y = \frac{2}{7} + \frac{5}{14}x$$

as a reasonable pattern to use to predict the outcomes.

For the value $x = 4$, the machine will predict an output of $y = \dfrac{2}{7} + \dfrac{5}{14}(4) = \dfrac{12}{7}$.

For the value $x = 6$, the machine will predict an output of $y = \dfrac{2}{7} + \dfrac{5}{14}(6) = \dfrac{17}{7}$.

See Figure 3.

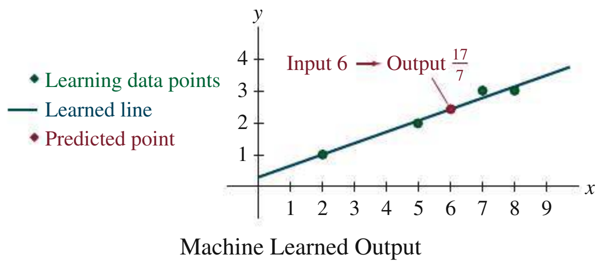


FIGURE 3 Machine-learned output.

A common practice before computing a least-squares line is to compute the average $\overline{x}$ of the original x-values and form a new variable $x^* = x - \overline{x}$. The new x-data are said to be in **mean-deviation form**. In this case, the two columns of the design matrix will be orthogonal. Solution of the normal equations is simplified, just as in Example 4 in Section 6.5. See Exercises 23 and 24.

The General Linear Model

In some applications, it is necessary to fit data points with something other than a straight line. In the examples that follow, the matrix equation is still $X\boldsymbol{\beta} = \mathbf{y}$, but the specific form of X changes from one problem to the next. Statisticians usually introduce a **residual vector** $\boldsymbol{\epsilon}$, defined by $\boldsymbol{\epsilon} = \mathbf{y} - X\boldsymbol{\beta}$, and write

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Any equation of this form is referred to as a **linear model**. Once X and $\mathbf{y}$ are determined, the goal is to minimize the length of $\boldsymbol{\epsilon}$, which amounts to finding a least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$. In each case, the least-squares solution $\hat{\boldsymbol{\beta}}$ is a solution of the normal equations

$$X^T X \boldsymbol{\beta} = X^T \mathbf{y}$$

Least-Squares Fitting of Other Curves

When data points $(x_1, y_1), \dots, (x_n, y_n)$ on a scatter plot do not lie close to any line, it may be appropriate to postulate some other functional relationship between x and y.

The next two examples show how to fit data by curves that have the general form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x) \tag{2}$$

where $f_0, \dots, f_k$ are known functions and $\beta_0, \dots, \beta_k$ are parameters that must be determined. As we will see, equation (2) describes a linear model because it is linear in the unknown parameters.

For a particular value of x, (2) gives a predicted, or "fitted," value of y. The difference between the observed value and the predicted value is the residual. The parameters $\beta_0, \dots, \beta_k$ must be determined so as to minimize the sum of the squares of the residuals.

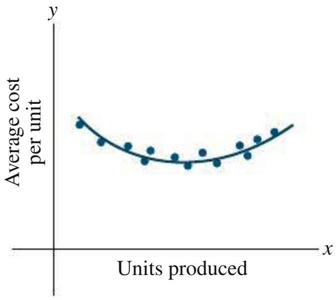


FIGURE 4
Average cost curve.

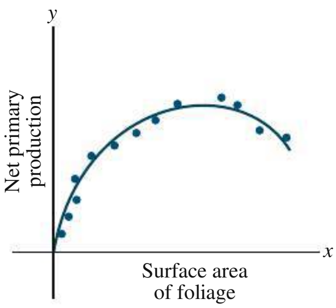


FIGURE 5
Production of nutrients.

EXAMPLE 3 Suppose data points $(x_1, y_1), \dots, (x_n, y_n)$ appear to lie along some sort of parabola instead of a straight line. For instance, if the x-coordinate denotes the production level for a company, and y denotes the average cost per unit of operating at a level of x units per day, then a typical average cost curve looks like a parabola that opens upward (Figure 4). In ecology, a parabolic curve that opens downward is used to model the net primary production of nutrients in a plant, as a function of the surface area of the foliage (Figure 5). Suppose we wish to approximate the data by an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{3}$$

Describe the linear model that produces a "least-squares fit" of the data by equation (3).

SOLUTION Equation (3) describes the ideal relationship. Suppose the actual values of the parameters are $\beta_0, \beta_1, \beta_2$. Then the coordinates of the first data point (x_1, y_1) satisfy an equation of the form

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1$$

where ϵ_1 is the residual error between the observed value y_1 and the predicted y-value $\beta_0 + \beta_1 x_1 + \beta_2 x_1^2$. Each data point determines a similar equation:

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_2 \\
&\vdots \qquad\qquad\quad \vdots \\
y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \epsilon_n
\end{aligned}$$

It is a simple matter to write this system of equations in the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. To find X, inspect the first few rows of the system and look for the pattern.

$$\begin{array}{ccccc} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} & = & \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} & + & \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \\ \mathbf{y} & = & X \qquad\quad \boldsymbol{\beta} & + & \boldsymbol{\epsilon} \end{array}$$

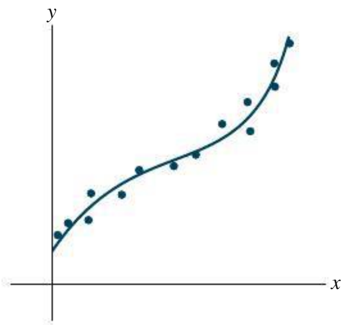


FIGURE 6
Data points along a cubic curve.

EXAMPLE 4 If data points tend to follow a pattern such as in Figure 6, then an appropriate model might be an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Such data, for instance, could come from a company's total costs, as a function of the level of production. Describe the linear model that gives a least-squares fit of this type to data $(x_1, y_1), \dots, (x_n, y_n)$.

SOLUTION By an analysis similar to that in Example 2, we obtain

$$\mathbf{y} = \overset{\substack{\text{Observation} \\ \text{vector}}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}, \quad X = \overset{\substack{\text{Design} \\ \text{matrix}}}{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}}, \quad \boldsymbol{\beta} = \overset{\substack{\text{Parameter} \\ \text{vector}}}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}, \quad \boldsymbol{\epsilon} = \overset{\substack{\text{Residual} \\ \text{vector}}}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

Multiple Regression

Suppose an experiment involves two independent variables—say, u and v—and one dependent variable, y. A simple equation for predicting y from u and v has the form

$$y = \beta_0 + \beta_1 u + \beta_2 v \tag{4}$$

A more general prediction equation might have the form

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 u^2 + \beta_4 uv + \beta_5 v^2 \tag{5}$$

This equation is used in geology, for instance, to model erosion surfaces, glacial cirques, soil pH, and other quantities. In such cases, the least-squares fit is called a *trend surface*.

Equations (4) and (5) both lead to a linear model because they are linear in the unknown parameters (even though u and v are multiplied). In general, a linear model will arise whenever y is to be predicted by an equation of the form

$$y = \beta_0 f_0(u, v) + \beta_1 f_1(u, v) + \cdots + \beta_k f_k(u, v)$$

with $f_0, \dots, f_k$ any sort of known functions and $\beta_0, \dots, \beta_k$ unknown weights.

EXAMPLE 5 In geography, local models of terrain are constructed from data $(u_1, v_1, y_1), \dots, (u_n, v_n, y_n)$, where u_j, v_j, and y_j are latitude, longitude, and altitude, respectively. Describe the linear model based on (4) that gives a least-squares fit to such data. The solution is called the *least-squares plane*. See Figure 7.

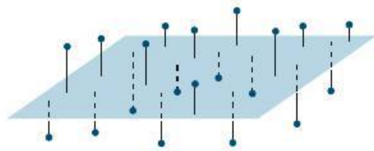


FIGURE 7 A least-squares plane.

SOLUTION We expect the data to satisfy the following equations:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 u_1 + \beta_2 v_1 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 u_2 + \beta_2 v_2 + \epsilon_2 \\ &\vdots \qquad\qquad \vdots \\ y_n &= \beta_0 + \beta_1 u_n + \beta_2 v_n + \epsilon_n \end{aligned}$$

This system has the matrix form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

$$\underset{\text{Observation vector}}{\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}, \quad \underset{\text{Design matrix}}{X = \begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}}, \quad \underset{\text{Parameter vector}}{\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}, \quad \underset{\text{Residual vector}}{\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

Example 5 shows that the linear model for multiple regression has the same abstract form as the model for the simple regression in the earlier examples. Linear algebra gives us the power to understand the general principle behind all the linear models. Once X is defined properly, the normal equations for $\boldsymbol{\beta}$ have the same matrix form, no matter how many variables are involved. Thus, for any linear model where X^TX is invertible, the least-squares $\hat{\boldsymbol{\beta}}$ is given by $(X^TX)^{-1}X^T\mathbf{y}$.

STUDY GUIDE offers additional resources for understanding the geometry of a linear model.

Practice Problem

When the monthly sales of a product are subject to seasonal fluctuations, a curve that approximates the sales data might have the form

$$y = \beta_0 + \beta_1 x + \beta_2 \sin(2\pi x/12)$$

where x is the time in months. The term $\beta_0 + \beta_1 x$ gives the basic sales trend, and the sine term reflects the seasonal changes in sales. Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above. Assume the data are $(x_1, y_1), \dots, (x_n, y_n)$.

6.6 Exercises

In Exercises 1–4, find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the given data points.

1. $(0, 1), (1, 1), (2, 2), (3, 2)$
2. $(1, 0), (2, 1), (4, 2), (5, 3)$
3. $(-1, 0), (0, 1), (1, 2), (2, 4)$
4. $(2, 3), (3, 2), (5, 1), (6, 0)$
5. If a machine learns the least-squares line that best fits the data in Exercise 1, what will the machine pick for the value of y when $x = 4$?
6. If a machine learns the least-squares line that best fits the data in Exercise 2, what will the machine pick for the value of y when $x = 3$?
7. If a machine learns the least-squares line that best fits the data in Exercise 1, what will the machine pick for the value of y when $x = 3$? How closely does this match the data point at $x = 3$ fed into the machine?
8. If a machine learns the least-squares line that best fits the data in Exercise 2, what will the machine pick for the value of y when $x = 2$? How closely does this match the data point at $x = 2$ fed into the machine?
9. If you enter the data from Exercise 1 into a machine and it returns a y value of 20 when $x = 2.5$, should you trust the machine? Justify your answer.
10. If you enter the data from Exercise 2 into a machine and it returns a y value of -4 when $x = 2.5$, should you trust the machine? Justify your answer.
11. Let X be the design matrix used to find the least-squares line to fit data $(x_1, y_1), \dots, (x_n, y_n)$. Use a theorem in Section 6.5 to show that the normal equations have a unique solution if and only if the data include at least two data points with different x-coordinates.
12. Let X be the design matrix in Example 2 corresponding to a least-squares fit of a parabola to data $(x_1, y_1), \dots, (x_n, y_n)$. Suppose x_1, x_2, and x_3 are distinct. Explain why there is only one parabola that fits the data best, in a least-squares sense. (See Exercise 11.)
13. A certain experiment produces the data $(1, 1.8)$, $(2, 2.7)$, $(3, 3.4)$, $(4, 3.8)$, $(5, 3.9)$. Describe the model that produces a least-squares fit of these points by a function of the form

    $$y = \beta_1 x + \beta_2 x^2$$

    Such a function might arise, for example, as the revenue from the sale of x units of a product, when the amount offered for sale affects the price to be set for the product.

    a. Give the design matrix, the observation vector, and the unknown parameter vector.

    T b. Find the associated least-squares curve for the data.

    c. If a machine learned the curve you found in (b), what output would it provide for an input of $x = 6$?

14. A simple curve that often makes a good model for the variable costs of a company, as a function of the sales level x, has the form $y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$. There is no constant term because fixed costs are not included.

    a. Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above, with data $(x_1, y_1), \dots, (x_n, y_n)$.

    T b. Find the least-squares curve of the form above to fit the data $(4, 1.58)$, $(6, 2.08)$, $(8, 2.5)$, $(10, 2.8)$, $(12, 3.1)$, $(14, 3.4)$, $(16, 3.8)$, and $(18, 4.32)$, with values in thousands. If possible, produce a graph that shows the data points and the graph of the cubic approximation.

    c. If a machine learned the curve you found in (b), what output would it provide for an input of $x = 9$?

15. A certain experiment produces the data $(1, 7.9)$, $(2, 5.4)$, and $(3, -.9)$. Describe the model that produces a least-squares fit of these points by a function of the form

    $$y = A\cos x + B\sin x$$

16. Suppose radioactive substances A and B have decay constants of .02 and .07, respectively. If a mixture of these two substances at time $t = 0$ contains M_A grams of A and M_B grams of B, then a model for the total amount y of the mixture present at time t is

    $$y = M_A e^{-.02t} + M_B e^{-.07t} \tag{6}$$

    Suppose the initial amounts M_A and M_B are unknown, but a scientist is able to measure the total amounts present at several times and records the following points (t_i, y_i): $(10, 21.34)$, $(11, 20.68)$, $(12, 20.05)$, $(14, 18.87)$, and $(15, 18.30)$.

    a. Describe a linear model that can be used to estimate M_A and M_B.

    T b. Find the least-squares curve based on (6).

Halley's Comet last appeared in 1986 and will reappear in 2061.

T 17. According to Kepler's first law, a comet should have an elliptic, parabolic, or hyperbolic orbit (with gravitational attractions from the planets ignored). In suitable polar coordinates, the position (r, ϑ) of a comet satisfies an equation of the form

$$r = \beta + e(r \cdot \cos \vartheta)$$

where β is a constant and e is the *eccentricity* of the orbit, with $0 \le e < 1$ for an ellipse, $e = 1$ for a parabola, and $e > 1$ for a hyperbola. Suppose observations of a newly discovered comet provide the data below. Determine the type of orbit, and predict where the comet will be when $\vartheta = 4.6$ (radians).[3]

ϑ	.88	1.10	1.42	1.77	2.14
r	3.00	2.30	1.65	1.25	1.01

T 18. A healthy child's systolic blood pressure p (in millimeters of mercury) and weight w (in pounds) are approximately related by the equation

$$\beta_0 + \beta_1 \ln w = p$$

Use the following experimental data to estimate the systolic blood pressure of a healthy child weighing 100 pounds.

w	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88
p	91	98	103	110	112

T 19. To measure the takeoff performance of an airplane, the horizontal position of the plane was measured every second, from $t = 0$ to $t = 12$. The positions (in feet) were: 0, 8.8, 29.9, 62.0, 104.7, 159.1, 222.0, 294.5, 380.4, 471.1, 571.7, 686.8, and 809.2.

a. Find the least-squares cubic curve $y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ for these data.

b. If a machine learned the curve given in part (a), what would it estimate the velocity of the plane to be when $t = 4.5$ seconds?

20. Let $\overline{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ and $\overline{y} = \frac{1}{n}(y_1 + \cdots + y_n)$. Show that the least-squares line for the data $(x_1, y_1), \ldots, (x_n, y_n)$ must pass through $(\overline{x}, \overline{y})$. That is, show that $\overline{x}$ and $\overline{y}$ satisfy the linear equation $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$. [*Hint:* Derive this equation from the vector equation $\mathbf{y} = X\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$. Denote the first column of X by $\mathbf{1}$. Use the fact that the residual vector $\boldsymbol{\epsilon}$ is orthogonal to the column space of X and hence is orthogonal to $\mathbf{1}$.]

[3] The basic idea of least-squares fitting of data is due to K. F. Gauss (and, independently, to A. Legendre), whose initial rise to fame occurred in 1801 when he used the method to determine the path of the asteroid *Ceres*. Forty days after the asteroid was discovered, it disappeared behind the sun. Gauss predicted it would appear ten months later and gave its location. The accuracy of the prediction astonished the European scientific community.

Given data for a least-squares problem, $(x_1, y_1), \ldots, (x_n, y_n)$, the following abbreviations are helpful:

$$\textstyle\sum x = \sum_{i=1}^n x_i, \quad \sum x^2 = \sum_{i=1}^n x_i^2,$$
$$\textstyle\sum y = \sum_{i=1}^n y_i, \quad \sum xy = \sum_{i=1}^n x_i y_i$$

The normal equations for a least-squares line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ may be written in the form

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \textstyle\sum x &= \textstyle\sum y \\ \hat{\beta}_0 \textstyle\sum x + \hat{\beta}_1 \textstyle\sum x^2 &= \textstyle\sum xy \end{aligned} \tag{7}$$

21. Derive the normal equations (7) from the matrix form given in this section.

22. Use a matrix inverse to solve the system of equations in (7) and thereby obtain formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$ that appear in many statistics texts.

23. a. Rewrite the data in Example 1 with new x-coordinates in mean deviation form. Let X be the associated design matrix. Why are the columns of X orthogonal?

b. Write the normal equations for the data in part (a), and solve them to find the least-squares line, $y = \beta_0 + \beta_1 x^*$, where $x^* = x - 5.5$.

24. Suppose the x-coordinates of the data $(x_1, y_1), \ldots, (x_n, y_n)$ are in mean deviation form, so that $\sum x_i = 0$. Show that if X is the design matrix for the least-squares line in this case, then $X^T X$ is a diagonal matrix.

Exercises 25 and 26 involve a design matrix X with two or more columns and a least-squares solution $\hat{\boldsymbol{\beta}}$ of $\mathbf{y} = X\boldsymbol{\beta}$. Consider the following numbers.

(i) $\|X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares of the "regression term." Denote this number by SS(R).

(ii) $\|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares for the error term. Denote this number by SS(E).

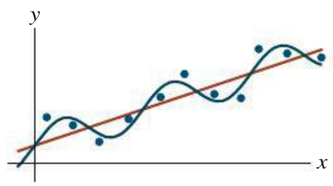
(iii) $\|\mathbf{y}\|^2$—the "total" sum of the squares of the y-values. Denote this number by SS(T).

Every statistics text that discusses regression and the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ introduces these numbers, though terminology and notation vary somewhat. To simplify matters, assume that the mean of the y-values is zero. In this case, SS(T) is proportional to what is called the *variance* of the set of y-values.

25. Justify the equation SS(T) = SS(R) + SS(E). [*Hint:* Use a theorem, and explain why the hypotheses of the theorem are satisfied.] This equation is extremely important in statistics, both in regression theory and in the analysis of variance.

26. Show that $\|X\hat{\boldsymbol{\beta}}\|^2 = \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$. [*Hint:* Rewrite the left side and use the fact that $\hat{\boldsymbol{\beta}}$ satisfies the normal equations.] This formula for SS(R) is used in statistics. From this and from Exercise 25, obtain the standard formula for SS(E):

$$\text{SS(E)} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$$



Sales trend with seasonal fluctuations.

Solution to Practice Problem

Construct X and $\boldsymbol{\beta}$ so that the kth row of $X\boldsymbol{\beta}$ is the predicted y-value that corresponds to the data point (x_k, y_k), namely

$$\beta_0 + \beta_1 x_k + \beta_2 \sin(2\pi x_k/12)$$

It should be clear that

$$X = \begin{bmatrix} 1 & x_1 & \sin(2\pi x_1/12) \\ \vdots & \vdots & \vdots \\ 1 & x_n & \sin(2\pi x_n/12) \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

6.7 Inner Product Spaces

Notions of length, distance, and orthogonality are often important in applications involving a vector space. For $\mathbb{R}^n$, these concepts were based on the properties of the inner product listed in Theorem 1 of Section 6.1. For other spaces, we need analogues of the inner product with the same properties. The conclusions of Theorem 1 now become *axioms* in the following definition.

DEFINITION

An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms, for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

A vector space with an inner product is called an **inner product space**.

The vector space $\mathbb{R}^n$ with the standard inner product is an inner product space, and nearly everything discussed in this chapter for $\mathbb{R}^n$ carries over to inner product spaces. The examples in this section and the next lay the foundation for a variety of applications treated in courses in engineering, physics, mathematics, and statistics.

EXAMPLE 1 Fix any two positive numbers—say, 4 and 5—and for vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in $\mathbb{R}^2$, set

$$\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2 \tag{1}$$

Show that equation (1) defines an inner product.

SOLUTION Certainly Axiom 1 is satisfied, because $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2 = 4v_1u_1 + 5v_2u_2 = \langle \mathbf{v}, \mathbf{u} \rangle$. If $\mathbf{w} = (w_1, w_2)$, then

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 4(u_1 + v_1)w_1 + 5(u_2 + v_2)w_2 \\ &= 4u_1w_1 + 5u_2w_2 + 4v_1w_1 + 5v_2w_2 \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

This verifies Axiom 2. For Axiom 3, compute

$$\langle c\mathbf{u}, \mathbf{v}\rangle = 4(cu_1)v_1 + 5(cu_2)v_2 = c(4u_1v_1 + 5u_2v_2) = c\langle \mathbf{u}, \mathbf{v}\rangle$$

For Axiom 4, note that $\langle \mathbf{u}, \mathbf{u}\rangle = 4u_1^2 + 5u_2^2 \geq 0$, and $4u_1^2 + 5u_2^2 = 0$ only if $u_1 = u_2 = 0$, that is, if $\mathbf{u} = \mathbf{0}$. Also, $\langle \mathbf{0}, \mathbf{0}\rangle = 0$. So (1) defines an inner product on $\mathbb{R}^2$. ■

Inner products similar to (1) can be defined on $\mathbb{R}^n$. They arise naturally in connection with "weighted least-squares" problems, in which weights are assigned to the various entries in the sum for the inner product in such a way that more importance is given to the more reliable measurements.

From now on, when an inner product space involves polynomials or other functions, we will write the functions in the familiar way, rather than use the boldface type for vectors. Nevertheless, it is important to remember that each function *is* a vector when it is treated as an element of a vector space.

EXAMPLE 2 Let $t_0, \ldots, t_n$ be distinct real numbers. For p and q in $\mathbb{P}_n$, define

$$\langle p, q\rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \cdots + p(t_n)q(t_n) \tag{2}$$

Inner product Axioms 1–3 are readily checked. For Axiom 4, note that

$$\langle p, p\rangle = [p(t_0)]^2 + [p(t_1)]^2 + \cdots + [p(t_n)]^2 \geq 0$$

Also, $\langle \mathbf{0}, \mathbf{0}\rangle = 0$. (The boldface zero here denotes the zero polynomial, the zero vector in $\mathbb{P}_n$.) If $\langle p, p\rangle = 0$, then p must vanish at $n + 1$ points: $t_0, \ldots, t_n$. This is possible only if p is the zero polynomial, because the degree of p is less than $n + 1$. Thus (2) defines an inner product on $\mathbb{P}_n$. ■

EXAMPLE 3 Let V be $\mathbb{P}_2$, with the inner product from Example 2, where $t_0 = 0$, $t_1 = \frac{1}{2}$, and $t_2 = 1$. Let $p(t) = 12t^2$ and $q(t) = 2t - 1$. Compute $\langle p, q\rangle$ and $\langle q, q\rangle$.

SOLUTION

$$\begin{aligned}
\langle p, q\rangle &= p(0)q(0) + p\left(\tfrac{1}{2}\right)q\left(\tfrac{1}{2}\right) + p(1)q(1) \\
&= (0)(-1) + (3)(0) + (12)(1) = 12 \\
\langle q, q\rangle &= [q(0)]^2 + [q\left(\tfrac{1}{2}\right)]^2 + [q(1)]^2 \\
&= (-1)^2 + (0)^2 + (1)^2 = 2
\end{aligned}$$ ■

Lengths, Distances, and Orthogonality

Let V be an inner product space, with the inner product denoted by $\langle \mathbf{u}, \mathbf{v}\rangle$. Just as in $\mathbb{R}^n$, we define the **length**, or **norm**, of a vector $\mathbf{v}$ to be the scalar

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$$

Equivalently, $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v}\rangle$. (This definition makes sense because $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$, but the definition *does not* say that $\langle \mathbf{v}, \mathbf{v}\rangle$ is a "sum of squares," because $\mathbf{v}$ need not be an element of $\mathbb{R}^n$.)

A **unit vector** is one whose length is 1. The **distance between u and v** is $\|\mathbf{u} - \mathbf{v}\|$. Vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.

EXAMPLE 4 Let $\mathbb{P}_2$ have the inner product (2) of Example 3. Compute the lengths of the vectors $p(t) = 12t^2$ and $q(t) = 2t - 1$.

SOLUTION

$$\begin{aligned}\|p\|^2 &= \langle p, p \rangle = [p(0)]^2 + \left[p\left(\tfrac{1}{2}\right)\right]^2 + [p(1)]^2 \\ &= 0 + [3]^2 + [12]^2 = 153 \\ \|p\| &= \sqrt{153}\end{aligned}$$

From Example 3, $\langle q, q \rangle = 2$. Hence $\|q\| = \sqrt{2}$. ■

The Gram–Schmidt Process

The existence of orthogonal bases for finite-dimensional subspaces of an inner product space can be established by the Gram–Schmidt process, just as in $\mathbb{R}^n$. Certain orthogonal bases that arise frequently in applications can be constructed by this process.

The orthogonal projection of a vector onto a subspace W with an orthogonal basis can be constructed as usual. The projection does not depend on the choice of orthogonal basis, and it has the properties described in the Orthogonal Decomposition Theorem and the Best Approximation Theorem.

EXAMPLE 5 Let V be $\mathbb{P}_4$ with the inner product in Example 2, involving evaluation of polynomials at -2, -1, 0, 1, and 2, and view $\mathbb{P}_2$ as a subspace of V. Produce an orthogonal basis for $\mathbb{P}_2$ by applying the Gram–Schmidt process to the polynomials 1, t, and t^2.

SOLUTION The inner product depends only on the values of a polynomial at $-2, \dots, 2$, so we list the values of each polynomial as a vector in $\mathbb{R}^5$, underneath the name of the polynomial:[1]

$$\begin{array}{lccc} \text{Polynomial:} & 1 & t & t^2 \\ \text{Vector of values:} & \begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}, & \begin{bmatrix}-2\\-1\\0\\1\\2\end{bmatrix}, & \begin{bmatrix}4\\1\\0\\1\\4\end{bmatrix}\end{array}$$

The inner product of two polynomials in V equals the (standard) inner product of their corresponding vectors in $\mathbb{R}^5$. Observe that t is orthogonal to the constant function 1. So take $p_0(t) = 1$ and $p_1(t) = t$. For p_2, use the vectors in $\mathbb{R}^5$ to compute the projection of t^2 onto Span$\{p_0, p_1\}$:

$$\begin{aligned}\langle t^2, p_0 \rangle &= \langle t^2, 1 \rangle = 4 + 1 + 0 + 1 + 4 = 10 \\ \langle p_0, p_0 \rangle &= 5 \\ \langle t^2, p_1 \rangle &= \langle t^2, t \rangle = -8 + (-1) + 0 + 1 + 8 = 0\end{aligned}$$

The orthogonal projection of t^2 onto Span$\{1, t\}$ is $\frac{10}{5} p_0 + 0 p_1$. Thus

$$p_2(t) = t^2 - 2p_0(t) = t^2 - 2$$

[1] Each polynomial in $\mathbb{P}_4$ is uniquely determined by its value at the five numbers $-2, \dots, 2$. In fact, the correspondence between p and its vector of values is an isomorphism, that is, a one-to-one mapping onto $\mathbb{R}^5$ that preserves linear combinations.

An orthogonal basis for the subspace $\mathbb{P}_2$ of V is

$$\begin{array}{lccc} \text{Polynomial} & p_0 & p_1 & p_2 \\ \text{Vector of values} & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, & \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, & \begin{bmatrix} 2 \\ -1 \\ -2 \\ -1 \\ 2 \end{bmatrix} \end{array} \tag{3}$$

■

Best Approximation in Inner Product Spaces

A common problem in applied mathematics involves a vector space V whose elements are functions. The problem is to approximate a function f in V by a function g from a specified subspace W of V. The "closeness" of the approximation of f depends on the way $\|f - g\|$ is defined. We will consider only the case in which the distance between f and g is determined by an inner product. In this case, the *best approximation to f by functions in W* is the orthogonal projection of f onto the subspace W.

EXAMPLE 6 Let V be $\mathbb{P}_4$ with the inner product in Example 5, and let p_0, p_1, and p_2 be the orthogonal basis found in Example 5 for the subspace $\mathbb{P}_2$. Find the best approximation to $p(t) = 5 - \frac{1}{2}t^4$ by polynomials in $\mathbb{P}_2$.

SOLUTION The values of p_0, p_1, and p_2 at the numbers $-2, -1, 0, 1,$ and 2 are listed in $\mathbb{R}^5$ vectors in (3) above. The corresponding values for p are -3, 9/2, 5, 9/2, and -3. Compute

$$\begin{array}{lll} \langle p, p_0\rangle = 8, & \langle p, p_1\rangle = 0, & \langle p, p_2\rangle = -31 \\ \langle p_0, p_0\rangle = 5, & & \langle p_2, p_2\rangle = 14 \end{array}$$

Then the best approximation in V to p by polynomials in $\mathbb{P}_2$ is

$$\begin{aligned} \hat{p} = \text{proj}_{\mathbb{P}_2} p &= \frac{\langle p, p_0\rangle}{\langle p_0, p_0\rangle} p_0 + \frac{\langle p, p_1\rangle}{\langle p_1, p_1\rangle} p_1 + \frac{\langle p, p_2\rangle}{\langle p_2, p_2\rangle} p_2 \\ &= \tfrac{8}{5} p_0 + \tfrac{-31}{14} p_2 = \tfrac{8}{5} - \tfrac{31}{14}(t^2 - 2). \end{aligned}$$

This polynomial is the closest to p of all polynomials in $\mathbb{P}_2$, when the distance between polynomials is measured only at $-2, -1, 0, 1,$ and 2. See Figure 1. ■

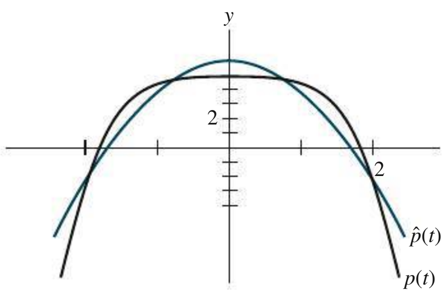


FIGURE 1

The polynomials p_0, p_1, and p_2 in Examples 5 and 6 belong to a class of polynomials that are referred to in statistics as *orthogonal polynomials*.[2] The orthogonality refers to the type of inner product described in Example 2.

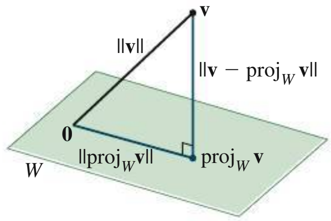


FIGURE 2
The hypotenuse is the longest side.

Two Inequalities

Given a vector $\mathbf{v}$ in an inner product space V and given a finite-dimensional subspace W, we may apply the Pythagorean Theorem to the orthogonal decomposition of $\mathbf{v}$ with respect to W and obtain

$$\|\mathbf{v}\|^2 = \|\operatorname{proj}_W \mathbf{v}\|^2 + \|\mathbf{v} - \operatorname{proj}_W \mathbf{v}\|^2$$

See Figure 2. In particular, this shows that the norm of the projection of $\mathbf{v}$ onto W does not exceed the norm of $\mathbf{v}$ itself. This simple observation leads to the following important inequality.

THEOREM 16

The Cauchy–Schwarz Inequality

For all $\mathbf{u}, \mathbf{v}$ in V,

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\| \tag{4}$$

PROOF If $\mathbf{u} = \mathbf{0}$, then both sides of (4) are zero, and hence the inequality is true in this case. (See Practice Problem 1.) If $\mathbf{u} \ne \mathbf{0}$, let W be the subspace spanned by $\mathbf{u}$. Recall that $\|c\mathbf{u}\| = |c|\,\|\mathbf{u}\|$ for any scalar c. Thus

$$\|\operatorname{proj}_W \mathbf{v}\| = \left\| \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\langle \mathbf{u}, \mathbf{u}\rangle}\mathbf{u} \right\| = \frac{|\langle \mathbf{v}, \mathbf{u}\rangle|}{|\langle \mathbf{u}, \mathbf{u}\rangle|}\|\mathbf{u}\| = \frac{|\langle \mathbf{v}, \mathbf{u}\rangle|}{\|\mathbf{u}\|^2}\|\mathbf{u}\| = \frac{|\langle \mathbf{u}, \mathbf{v}\rangle|}{\|\mathbf{u}\|}$$

Since $\|\operatorname{proj}_W \mathbf{v}\| \le \|\mathbf{v}\|$, we have $\dfrac{|\langle \mathbf{u}, \mathbf{v}\rangle|}{\|\mathbf{u}\|} \le \|\mathbf{v}\|$, which gives (4). ■

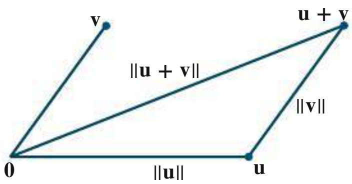


FIGURE 3
The lengths of the sides of a triangle.

The Cauchy–Schwarz inequality is useful in many branches of mathematics. A few simple applications are presented in the exercises. Our main need for this inequality here is to prove another fundamental inequality involving norms of vectors. See Figure 3.

THEOREM 17

The Triangle Inequality

For all $\mathbf{u}, \mathbf{v}$ in V,

$$\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$$

PROOF

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle + 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle \\
&\le \|\mathbf{u}\|^2 + 2|\langle \mathbf{u}, \mathbf{v}\rangle| + \|\mathbf{v}\|^2 \\
&\le \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \qquad \text{Cauchy–Schwarz} \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned}$$

The triangle inequality follows immediately by taking square roots of both sides. ■

[2] See *Statistics and Experimental Design in Engineering and the Physical Sciences*, 2nd ed., by Norman L. Johnson and Fred C. Leone (New York: John Wiley & Sons, 1977). Tables there list "Orthogonal Polynomials," which are simply the values of the polynomial at numbers such as −2, −1, 0, 1, and 2.

An Inner Product for $C[a, b]$ (Calculus required)

Probably the most widely used inner product space for applications is the vector space $C[a, b]$ of all continuous functions on an interval $a \le t \le b$, with an inner product that we will describe.

We begin by considering a polynomial p and any integer n larger than or equal to the degree of p. Then p is in $\mathbb{P}_n$, and we may compute a "length" for p using the inner product of Example 2 involving evaluation at $n + 1$ points in $[a, b]$. However, this length of p captures the behavior at only those $n + 1$ points. Since p is in $\mathbb{P}_n$ for all large n, we could use a much larger n, with many more points for the "evaluation" inner product. See Figure 4.

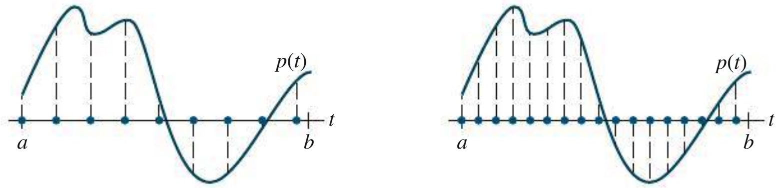
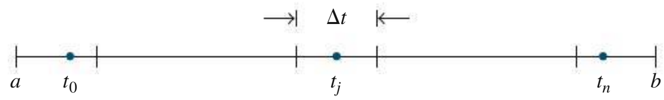


FIGURE 4 Using different numbers of evaluation points in $[a, b]$ to compute $\|p\|^2$.

Let us partition $[a, b]$ into $n + 1$ subintervals of length $\Delta t = (b - a)/(n + 1)$, and let $t_0, \dots, t_n$ be arbitrary points in these subintervals.



If n is large, the inner product on $\mathbb{P}_n$ determined by $t_0, \dots, t_n$ will tend to give a large value to $\langle p, p\rangle$, so we scale it down and divide by $n + 1$. Observe that $1/(n + 1) = \Delta t/(b - a)$, and define

$$\langle p, q\rangle = \frac{1}{n+1}\sum_{j=0}^{n} p(t_j)q(t_j) = \frac{1}{b-a}\left[\sum_{j=0}^{n} p(t_j)q(t_j)\Delta t\right]$$

Now, let n increase without bound. Since polynomials p and q are continuous functions, the expression in brackets is a Riemann sum that approaches a definite integral, and we are led to consider the *average value of* $p(t)q(t)$ on the interval $[a, b]$:

$$\frac{1}{b-a}\int_a^b p(t)q(t)\,dt$$

This quantity is defined for polynomials of any degree (in fact, for all continuous functions), and it has all the properties of an inner product, as the next example shows. The scale factor $1/(b - a)$ is inessential and is often omitted for simplicity.

EXAMPLE 7 For f, g in $C[a, b]$, set

$$\langle f, g\rangle = \int_a^b f(t)g(t)\,dt \tag{5}$$

Show that (5) defines an inner product on $C[a, b]$.

SOLUTION Inner product Axioms 1–3 follow from elementary properties of definite integrals. For Axiom 4, observe that

$$\langle f, f\rangle = \int_a^b [f(t)]^2\, dt \geq 0$$

The function $[f(t)]^2$ is continuous and nonnegative on $[a, b]$. If the definite integral of $[f(t)]^2$ is zero, then $[f(t)]^2$ must be identically zero on $[a, b]$, by a theorem in advanced calculus, in which case f is the zero function. Thus $\langle f, f\rangle = 0$ implies that f is the zero function on $[a, b]$. So (5) defines an inner product on $C[a, b]$. ■

EXAMPLE 8 Let V be the space $C[0, 1]$ with the inner product of Example 7, and let W be the subspace spanned by the polynomials $p_1(t) = 1$, $p_2(t) = 2t - 1$, and $p_3(t) = 12t^2$. Use the Gram–Schmidt process to find an orthogonal basis for W.

SOLUTION Let $q_1 = p_1$, and compute

$$\langle p_2, q_1\rangle = \int_0^1 (2t-1)(1)\, dt = (t^2 - t)\Big|_0^1 = 0$$

So p_2 is already orthogonal to q_1, and we can take $q_2 = p_2$. For the projection of p_3 onto $W_2 = \text{Span}\,\{q_1, q_2\}$, compute

$$\begin{aligned}
\langle p_3, q_1\rangle &= \int_0^1 12t^2 \cdot 1\, dt = 4t^3\Big|_0^1 = 4\\
\langle q_1, q_1\rangle &= \int_0^1 1 \cdot 1\, dt = t\Big|_0^1 = 1\\
\langle p_3, q_2\rangle &= \int_0^1 12t^2(2t-1)\, dt = \int_0^1 (24t^3 - 12t^2)\, dt = 2\\
\langle q_2, q_2\rangle &= \int_0^1 (2t-1)^2\, dt = \frac{1}{6}(2t-1)^3\Big|_0^1 = \frac{1}{3}
\end{aligned}$$

Then

$$\text{proj}_{W_2}\, p_3 = \frac{\langle p_3, q_1\rangle}{\langle q_1, q_1\rangle} q_1 + \frac{\langle p_3, q_2\rangle}{\langle q_2, q_2\rangle} q_2 = \frac{4}{1} q_1 + \frac{2}{1/3} q_2 = 4q_1 + 6q_2$$

and

$$q_3 = p_3 - \text{proj}_{W_2}\, p_3 = p_3 - 4q_1 - 6q_2$$

As a function, $q_3(t) = 12t^2 - 4 - 6(2t-1) = 12t^2 - 12t + 2$. The orthogonal basis for the subspace W is $\{q_1, q_2, q_3\}$. ■

Practice Problems

Use the inner product axioms to verify the following statements.

1. $\langle \mathbf{v}, \mathbf{0}\rangle = \langle \mathbf{0}, \mathbf{v}\rangle = 0$.
2. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{w}\rangle$.

6.7 Exercises

1. Let $\mathbb{R}^2$ have the inner product of Example 1, and let $\mathbf{x} = (1, 1)$ and $\mathbf{y} = (5, -1)$.
   a. Find $\|\mathbf{x}\|$, $\|\mathbf{y}\|$, and $|\langle \mathbf{x}, \mathbf{y} \rangle|^2$.
   b. Describe all vectors (z_1, z_2) that are orthogonal to $\mathbf{y}$.

2. Let $\mathbb{R}^2$ have the inner product of Example 1. Show that the Cauchy–Schwarz inequality holds for $\mathbf{x} = (3, -2)$ and $\mathbf{y} = (-2, 1)$. [*Suggestion:* Study $|\langle \mathbf{x}, \mathbf{y} \rangle|^2$.]

Exercises 3–8 refer to $\mathbb{P}_2$ with the inner product given by evaluation at -1, 0, and 1. (See Example 2.)

3. Compute $\langle p, q \rangle$, where $p(t) = 4 + t$, $q(t) = 5 - 4t^2$.

4. Compute $\langle p, q \rangle$, where $p(t) = 3t - t^2$, $q(t) = 3 + 2t^2$.

5. Compute $\|p\|$ and $\|q\|$, for p and q in Exercise 3.

6. Compute $\|p\|$ and $\|q\|$, for p and q in Exercise 4.

7. Compute the orthogonal projection of q onto the subspace spanned by p, for p and q in Exercise 3.

8. Compute the orthogonal projection of q onto the subspace spanned by p, for p and q in Exercise 4.

9. Let $\mathbb{P}_3$ have the inner product given by evaluation at -3, -1, 1, and 3. Let $p_0(t) = 1$, $p_1(t) = t$, and $p_2(t) = t^2$.
   a. Compute the orthogonal projection of p_2 onto the subspace spanned by p_0 and p_1.
   b. Find a polynomial q that is orthogonal to p_0 and p_1, such that $\{p_0, p_1, q\}$ is an orthogonal basis for Span$\{p_0, p_1, p_2\}$. Scale the polynomial q so that its vector of values at $(-3, -1, 1, 3)$ is $(1, -1, -1, 1)$.

10. Let $\mathbb{P}_3$ have the inner product as in Exercise 9, with p_0, p_1, and q the polynomials described there. Find the best approximation to $p(t) = t^3$ by polynomials in Span$\{p_0, p_1, q\}$.

11. Let p_0, p_1, and p_2 be the orthogonal polynomials described in Example 5, where the inner product on $\mathbb{P}_4$ is given by evaluation at -2, -1, 0, 1, and 2. Find the orthogonal projection of t^3 onto Span$\{p_0, p_1, p_2\}$.

12. Find a polynomial p_3 such that $\{p_0, p_1, p_2, p_3\}$ (see Exercise 11) is an orthogonal basis for the subspace $\mathbb{P}_3$ of $\mathbb{P}_4$. Scale the polynomial p_3 so that its vector of values is $(-1, 2, 0, -2, 1)$.

13. Let A be any invertible $n \times n$ matrix. Show that for $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$, the formula $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v}) = (A\mathbf{u})^T(A\mathbf{v})$ defines an inner product on $\mathbb{R}^n$.

14. Let T be a one-to-one linear transformation from a vector space V into $\mathbb{R}^n$. Show that for $\mathbf{u}$, $\mathbf{v}$ in V, the formula $\langle \mathbf{u}, \mathbf{v} \rangle = T(\mathbf{u}) \cdot T(\mathbf{v})$ defines an inner product on V.

Use the inner product axioms and other results of this section to verify the statements in Exercises 15–18.

15. $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$ for all scalars c.

16. If $\{\mathbf{u}, \mathbf{v}\}$ is an orthonormal set in V, then $\|\mathbf{u} - \mathbf{v}\| = \sqrt{2}$.

17. $\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4}\|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u} - \mathbf{v}\|^2$.

18. $\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$.

In Exercises 19–24, $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors. Mark each statement True or False **(T/F)**. Justify each answer.

19. **(T/F)** If $\langle \mathbf{u}, \mathbf{u} \rangle = 0$, then $\mathbf{u} = \mathbf{0}$.

20. **(T/F)** If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, then either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$.

21. **(T/F)** $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle$.

22. **(T/F)** $\langle c\mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$.

23. **(T/F)** $|\langle \mathbf{u}, \mathbf{u} \rangle| = \langle \mathbf{u}, \mathbf{u} \rangle$.

24. **(T/F)** $|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$.

25. Given $a \ge 0$ and $b \ge 0$, let $\mathbf{u} = \begin{bmatrix} \sqrt{a} \\ \sqrt{b} \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} \sqrt{b} \\ \sqrt{a} \end{bmatrix}$. Use the Cauchy–Schwarz inequality to compare the geometric mean $\sqrt{ab}$ with the arithmetic mean $(a + b)/2$.

26. Let $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Use the Cauchy–Schwarz inequality to show that
$$\left(\frac{a+b}{2}\right)^2 \le \frac{a^2 + b^2}{2}$$

Exercises 27–30 refer to $V = C[0, 1]$, with the inner product given by an integral, as in Example 7.

27. Compute $\langle f, g \rangle$, where $f(t) = 1 - 3t^2$ and $g(t) = t - t^3$.

28. Compute $\langle f, g \rangle$, where $f(t) = 5t - 3$ and $g(t) = t^3 - t^2$.

29. Compute $\|f\|$ for f in Exercise 27.

30. Compute $\|g\|$ for g in Exercise 28.

31. Let V be the space $C[-1, 1]$ with the inner product of Example 7. Find an orthogonal basis for the subspace spanned by the polynomials 1, t, and t^2. The polynomials in this basis are called *Legendre polynomials*.

32. Let V be the space $C[-2, 2]$ with the inner product of Example 7. Find an orthogonal basis for the subspace spanned by the polynomials 1, t, and t^2.

T 33. Let $\mathbb{P}_4$ have the inner product as in Example 5, and let p_0, p_1, p_2 be the orthogonal polynomials from that example. Using your matrix program, apply the Gram–Schmidt process to the set $\{p_0, p_1, p_2, t^3, t^4\}$ to create an orthogonal basis for $\mathbb{P}_4$.

T 34. Let V be the space $C[0, 2\pi]$ with the inner product of Example 7. Use the Gram–Schmidt process to create an orthogonal basis for the subspace spanned by $\{1, \cos t, \cos^2 t, \cos^3 t\}$. Use a matrix program or computational program to compute the appropriate definite integrals.

Solutions to Practice Problems

1. By Axiom 1, $\langle \mathbf{v}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle$. Then $\langle \mathbf{0}, \mathbf{v} \rangle = \langle 0\mathbf{v}, \mathbf{v} \rangle = 0\langle \mathbf{v}, \mathbf{v} \rangle$, by Axiom 3, so $\langle \mathbf{0}, \mathbf{v} \rangle = 0$.
2. By Axioms 1, 2, and then 1 again, $\langle \mathbf{u}, \mathbf{v}+\mathbf{w} \rangle = \langle \mathbf{v}+\mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.

6.8 Applications of Inner Product Spaces

The examples in this section suggest how the inner product spaces defined in Section 6.7 arise in practical problems. Like in Section 6.6, important components of machine learning are analyzed.

Weighted Least-Squares

Let $\mathbf{y}$ be a vector of n observations, $y_1, \dots, y_n$, and suppose we wish to approximate $\mathbf{y}$ by a vector $\hat{\mathbf{y}}$ that belongs to some specified subspace of $\mathbb{R}^n$. (In Section 6.5, $\hat{\mathbf{y}}$ was written as $A\mathbf{x}$ so that $\hat{\mathbf{y}}$ was in the column space of A.) Denote the entries in $\hat{\mathbf{y}}$ by $\hat{y}_1, \dots, \hat{y}_n$. Then the *sum of the squares for error*, or SS(E), in approximating $\mathbf{y}$ by $\hat{\mathbf{y}}$ is

$$\text{SS(E)} = (y_1 - \hat{y}_1)^2 + \cdots + (y_n - \hat{y}_n)^2 \tag{1}$$

This is simply $\|\mathbf{y} - \hat{\mathbf{y}}\|^2$, using the standard length in $\mathbb{R}^n$.

Now suppose the measurements that produced the entries in $\mathbf{y}$ are not equally reliable. The entries in $\mathbf{y}$ might be computed from various samples of measurements, with unequal sample sizes. Then it becomes appropriate to weight the squared errors in (1) in such a way that more importance is assigned to the more reliable measurements.[1] If the weights are denoted by $w_1^2, \dots, w_n^2$, then the weighted sum of the squares for error is

$$\text{Weighted SS(E)} = w_1^2(y_1 - \hat{y}_1)^2 + \cdots + w_n^2(y_n - \hat{y}_n)^2 \tag{2}$$

This is the square of the length of $\mathbf{y} - \hat{\mathbf{y}}$, where the length is derived from an inner product analogous to that in Example 1 in Section 6.7, namely

$$\langle \mathbf{x}, \mathbf{y} \rangle = w_1^2 x_1 y_1 + \cdots + w_n^2 x_n y_n$$

It is sometimes convenient to transform a weighted least-squares problem into an equivalent ordinary least-squares problem. Let W be the diagonal matrix with (positive) $w_1, \dots, w_n$ on its diagonal, so that

$$W\mathbf{y} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & w_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_1 y_1 \\ w_2 y_2 \\ \vdots \\ w_n y_n \end{bmatrix}$$

with a similar expression for $W\hat{\mathbf{y}}$. Observe that the jth term in (2) can be written as

$$w_j^2(y_j - \hat{y}_j)^2 = (w_j y_j - w_j \hat{y}_j)^2$$

[1] Note for readers with a background in statistics: Suppose the errors in measuring the y_i are independent random variables with means equal to zero and variances of $\sigma_1^2, \dots, \sigma_n^2$. Then the appropriate weights in (2) are $w_i^2 = 1/\sigma_i^2$. The larger the variance of the error, the smaller the weight.

It follows that the weighted SS(E) in (2) is the square of the ordinary length in $\mathbb{R}^n$ of $W\mathbf{y} - W\hat{\mathbf{y}}$, which we write as $\|W\mathbf{y} - W\hat{\mathbf{y}}\|^2$.

Now suppose the approximating vector $\hat{\mathbf{y}}$ is to be constructed from the columns of a matrix A. Then we seek an $\hat{\mathbf{x}}$ that makes $A\hat{\mathbf{x}} = \hat{\mathbf{y}}$ as close to $\mathbf{y}$ as possible. However, the measure of closeness is the weighted error,

$$\|W\mathbf{y} - W\hat{\mathbf{y}}\|^2 = \|W\mathbf{y} - WA\hat{\mathbf{x}}\|^2$$

Thus $\hat{\mathbf{x}}$ is the (ordinary) least-squares solution of the equation

$$WA\mathbf{x} = W\mathbf{y}$$

The normal equation for the least-squares solution is

$$(WA)^T WA\mathbf{x} = (WA)^T W\mathbf{y}$$

EXAMPLE 1 Find the least-squares line $y = \beta_0 + \beta_1 x$ that best fits the data $(-2, 3)$, $(-1, 5)$, $(0, 5)$, $(1, 4)$, and $(2, 3)$. Suppose the errors in measuring the y-values of the last two data points are greater than for the other points. Weight these data half as much as the rest of the data.

SOLUTION As in Section 6.6, write X for the matrix A and $\boldsymbol{\beta}$ for the vector $\mathbf{x}$, and obtain

$$X = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 4 \\ 3 \end{bmatrix}$$

For a weighting matrix, choose W with diagonal entries 2, 2, 2, 1, and 1. Left-multiplication by W scales the rows of X and $\mathbf{y}$:

$$WX = \begin{bmatrix} 2 & -4 \\ 2 & -2 \\ 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad W\mathbf{y} = \begin{bmatrix} 6 \\ 10 \\ 10 \\ 4 \\ 3 \end{bmatrix}$$

For the normal equation, compute

$$(WX)^T WX = \begin{bmatrix} 14 & -9 \\ -9 & 25 \end{bmatrix} \quad \text{and} \quad (WX)^T W\mathbf{y} = \begin{bmatrix} 59 \\ -34 \end{bmatrix}$$

and solve

$$\begin{bmatrix} 14 & -9 \\ -9 & 25 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 59 \\ -34 \end{bmatrix}$$

The solution of the normal equation is (to two significant digits) $\beta_0 = 4.3$ and $\beta_1 = .20$. The desired line is

$$y = 4.3 + .20x$$

In contrast, the ordinary least-squares line for these data is

$$y = 4.0 - .10x$$

Both lines are displayed in Figure 1. ■

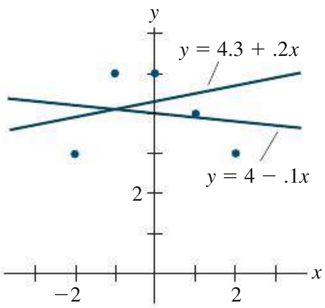


FIGURE 1
Weighted and ordinary least-squares lines.

Trend Analysis of Data

Let f represent an unknown function whose values are known (perhaps only approximately) at $t_0, \dots, t_n$. If there is a "linear trend" in the data $f(t_0), \dots, f(t_n)$, then we might expect to approximate the values of f by a function of the form $\beta_0 + \beta_1 t$. If there is a "quadratic trend" to the data, then we would try a function of the form $\beta_0 + \beta_1 t + \beta_2 t^2$. This was discussed in Section 6.6, from a different point of view.

In some statistical problems, it is important to be able to separate the linear trend from the quadratic trend (and possibly cubic or higher-order trends). For instance, suppose engineers are analyzing the performance of a new car, and $f(t)$ represents the distance between the car at time t and some reference point. If the car is traveling at constant velocity, then the graph of $f(t)$ should be a straight line whose slope is the car's velocity. If the gas pedal is suddenly pressed to the floor, the graph of $f(t)$ will change to include a quadratic term and possibly a cubic term (due to the acceleration). To analyze the ability of the car to pass another car, for example, engineers may want to separate the quadratic and cubic components from the linear term.

If the function is approximated by a curve of the form $y = \beta_0 + \beta_1 t + \beta_2 t^2$, the coefficient β_2 may not give the desired information about the quadratic trend in the data, because it may not be "independent" in a statistical sense from the other β_i. To make what is known as a **trend analysis** of the data, we introduce an inner product on the space $\mathbb{P}_n$ analogous to that given in Example 2 in Section 6.7. For p, q in $\mathbb{P}_n$, define

$$\langle p, q \rangle = p(t_0)q(t_0) + \cdots + p(t_n)q(t_n)$$

In practice, statisticians seldom need to consider trends in data of degree higher than cubic or quartic. So let p_0, p_1, p_2, p_3 denote an orthogonal basis of the subspace $\mathbb{P}_3$ of $\mathbb{P}_n$, obtained by applying the Gram–Schmidt process to the polynomials 1, t, t^2, and t^3. There is a polynomial g in $\mathbb{P}_n$ whose values at $t_0, \dots, t_n$ coincide with those of the unknown function f. Let $\hat{g}$ be the orthogonal projection (with respect to the given inner product) of g onto $\mathbb{P}_3$, say,

$$\hat{g} = c_0 p_0 + c_1 p_1 + c_2 p_2 + c_3 p_3$$

Then $\hat{g}$ is called a cubic **trend function**, and $c_0, \dots, c_3$ are the **trend coefficients** of the data. The coefficient c_1 measures the linear trend, c_2 the quadratic trend, and c_3 the cubic trend. It turns out that if the data have certain properties, these coefficients are statistically independent.

Since $p_0, \dots, p_3$ are orthogonal, the trend coefficients may be computed one at a time, independently of one another. (Recall that $c_i = \langle g, p_i \rangle / \langle p_i, p_i \rangle$.) We can ignore p_3 and c_3 if we want only the quadratic trend. And if, for example, we needed to determine the quartic trend, we would have to find (via Gram–Schmidt) only a polynomial p_4 in $\mathbb{P}_4$ that is orthogonal to $\mathbb{P}_3$ and compute $\langle g, p_4 \rangle / \langle p_4, p_4 \rangle$.

EXAMPLE 2 The simplest and most common use of trend analysis occurs when the points $t_0, \dots, t_n$ can be adjusted so that they are evenly spaced and sum to zero. Fit a quadratic trend function to the data $(-2, 3)$, $(-1, 5)$, $(0, 5)$, $(1, 4)$, and $(2, 3)$.

SOLUTION The t-coordinates are suitably scaled to use the orthogonal polynomials found in Example 5 of Section 6.7:

$$\begin{array}{lcccc} \text{Polynomial} & p_0 & p_1 & p_2 & \text{Data: } g \\ \text{Vector of values} & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, & \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, & \begin{bmatrix} 2 \\ -1 \\ -2 \\ -1 \\ 2 \end{bmatrix}, & \begin{bmatrix} 3 \\ 5 \\ 5 \\ 4 \\ 3 \end{bmatrix} \end{array}$$

The calculations involve only these vectors, not the specific formulas for the orthogonal polynomials. The best approximation to the data by polynomials in $\mathbb{P}_2$ is the orthogonal projection given by

$$\begin{aligned} \hat{p} &= \frac{\langle g, p_0 \rangle}{\langle p_0, p_0 \rangle} p_0 + \frac{\langle g, p_1 \rangle}{\langle p_1, p_1 \rangle} p_1 + \frac{\langle g, p_2 \rangle}{\langle p_2, p_2 \rangle} p_2 \\ &= \tfrac{20}{5} p_0 - \tfrac{1}{10} p_1 - \tfrac{7}{14} p_2 \end{aligned}$$

and

$$\hat{p}(t) = 4 - .1t - .5(t^2 - 2) \tag{3}$$

Since the coefficient of p_2 is not extremely small, it would be reasonable to conclude that the trend is at least quadratic. This is confirmed by the graph in Figure 2. ■

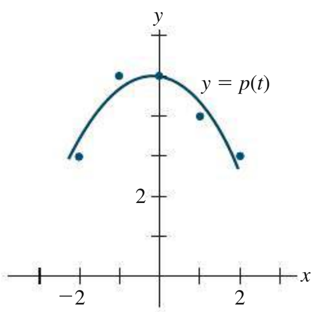


FIGURE 2
Approximation by a quadratic trend function.

Fourier Series (Calculus required)

Continuous functions are often approximated by linear combinations of sine and cosine functions. For instance, a continuous function might represent a sound wave, an electric signal of some type, or the movement of a vibrating mechanical system.

For simplicity, we consider functions on $0 \le t \le 2\pi$. It turns out that any function in $C[0, 2\pi]$ can be approximated as closely as desired by a function of the form

$$\frac{a_0}{2} + a_1 \cos t + \cdots + a_n \cos nt + b_1 \sin t + \cdots + b_n \sin nt \tag{4}$$

for a sufficiently large value of n. The function (4) is called a **trigonometric polynomial**. If a_n and b_n are not both zero, the polynomial is said to be of **order n**. The connection between trigonometric polynomials and other functions in $C[0, 2\pi]$ depends on the fact that for any $n \ge 1$, the set

$$\{1, \cos t, \cos 2t, \ldots, \cos nt, \sin t, \sin 2t, \ldots, \sin nt\} \tag{5}$$

is orthogonal with respect to the inner product

$$\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)\,dt \tag{6}$$

This orthogonality is verified as in the following example and in Exercises 5 and 6.

EXAMPLE 3 Let $C[0, 2\pi]$ have the inner product (6), and let m and n be unequal positive integers. Show that $\cos mt$ and $\cos nt$ are orthogonal.

SOLUTION Use a trigonometric identity. When $m \neq n$,

$$\begin{aligned}\langle \cos mt, \cos nt\rangle &= \int_0^{2\pi} \cos mt \cos nt\, dt \\ &= \frac{1}{2}\int_0^{2\pi} [\cos(mt+nt) + \cos(mt-nt)]\, dt \\ &= \frac{1}{2}\left[\frac{\sin(mt+nt)}{m+n} + \frac{\sin(mt-nt)}{m-n}\right]\Bigg|_0^{2\pi} = 0\end{aligned}$$

■

Let W be the subspace of $C[0, 2\pi]$ spanned by the functions in (5). Given f in $C[0, 2\pi]$, the best approximation to f by functions in W is called the ***n*th-order Fourier approximation** to f on $[0, 2\pi]$. Since the functions in (5) are orthogonal, the best approximation is given by the orthogonal projection onto W. In this case, the coefficients a_k and b_k in (4) are called the **Fourier coefficients** of f. The standard formula for an orthogonal projection shows that

$$a_k = \frac{\langle f, \cos kt\rangle}{\langle \cos kt, \cos kt\rangle}, \quad b_k = \frac{\langle f, \sin kt\rangle}{\langle \sin kt, \sin kt\rangle}, \quad k \geq 1$$

Exercise 7 asks you to show that $\langle \cos kt, \cos kt\rangle = \pi$ and $\langle \sin kt, \sin kt\rangle = \pi$. Thus

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt\, dt, \quad b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt\, dt \tag{7}$$

The coefficient of the (constant) function 1 in the orthogonal projection is

$$\frac{\langle f, 1\rangle}{\langle 1, 1\rangle} = \frac{1}{2\pi}\int_0^{2\pi} f(t)\cdot 1\, dt = \frac{1}{2}\left[\frac{1}{\pi}\int_0^{2\pi} f(t)\cos(0\cdot t)\, dt\right] = \frac{a_0}{2}$$

where a_0 is defined by (7) for $k = 0$. This explains why the constant term in (4) is written as $a_0/2$.

EXAMPLE 4 Find the nth-order Fourier approximation to the function $f(t) = t$ on the interval $[0, 2\pi]$.

SOLUTION Compute

$$\frac{a_0}{2} = \frac{1}{2}\cdot\frac{1}{\pi}\int_0^{2\pi} t\, dt = \frac{1}{2\pi}\left[\frac{1}{2}t^2\Big|_0^{2\pi}\right] = \pi$$

and for $k > 0$, using integration by parts,

$$\begin{aligned}a_k &= \frac{1}{\pi}\int_0^{2\pi} t\cos kt\, dt = \frac{1}{\pi}\left[\frac{1}{k^2}\cos kt + \frac{t}{k}\sin kt\right]_0^{2\pi} = 0 \\ b_k &= \frac{1}{\pi}\int_0^{2\pi} t\sin kt\, dt = \frac{1}{\pi}\left[\frac{1}{k^2}\sin kt - \frac{t}{k}\cos kt\right]_0^{2\pi} = -\frac{2}{k}\end{aligned}$$

Thus the nth-order Fourier approximation of $f(t) = t$ is

$$\pi - 2\sin t - \sin 2t - \frac{2}{3}\sin 3t - \cdots - \frac{2}{n}\sin nt$$

Figure 3 shows the third- and fourth-order Fourier approximations of f. ■

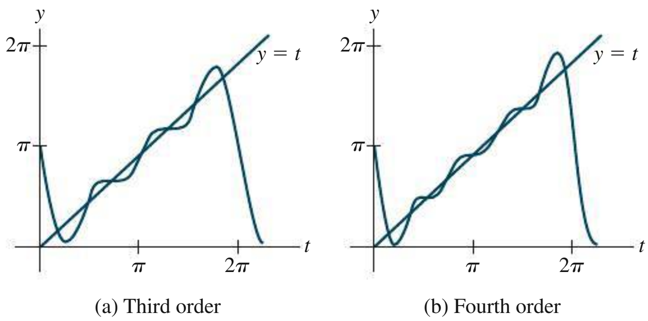


FIGURE 3 Fourier approximations of the function $f(t) = t$.

The norm of the difference between f and a Fourier approximation is called the **mean square error** in the approximation. (The term *mean* refers to the fact that the norm is determined by an integral.) It can be shown that the mean square error approaches zero as the order of the Fourier approximation increases. For this reason, it is common to write

$$f(t) = \frac{a_0}{2} + \sum_{m=1}^{\infty}(a_m \cos mt + b_m \sin mt)$$

This expression for $f(t)$ is called the **Fourier series** for f on $[0, 2\pi]$. The term $a_m \cos mt$, for example, is the projection of f onto the one-dimensional subspace spanned by $\cos mt$.

Practice Problems

1. Let $q_1(t) = 1$, $q_2(t) = t$, and $q_3(t) = 3t^2 - 4$. Verify that $\{q_1, q_2, q_3\}$ is an orthogonal set in $C[-2, 2]$ with the inner product of Example 7 in Section 6.7 (integration from -2 to 2).
2. Find the first-order and third-order Fourier approximations to

$$f(t) = 3 - 2\sin t + 5\sin 2t - 6\cos 2t$$

6.8 Exercises

1. Find the least-squares line $y = \beta_0 + \beta_1 x$ that best fits the data $(-2, 0)$, $(-1, 0)$, $(0, 2)$, $(1, 4)$, and $(2, 4)$, assuming that the first and last data points are less reliable. Weight them half as much as the three interior points.

2. Suppose 5 out of 25 data points in a weighted least-squares problem have a y-measurement that is less reliable than the others, and they are to be weighted half as much as the other 20 points. One method is to weight the 20 points by a factor of 1 and the other 5 by a factor of $\frac{1}{2}$. A second method is to weight the 20 points by a factor of 2 and the other 5 by a factor of 1. Do the two methods produce different results? Explain.

3. Fit a cubic trend function to the data in Example 2. The orthogonal cubic polynomial is $p_3(t) = \frac{5}{6}t^3 - \frac{17}{6}t$.

4. To make a trend analysis of six evenly spaced data points, one can use orthogonal polynomials with respect to evaluation at the points $t = -5, -3, -1, 1, 3$, and 5.

   a. Show that the first three orthogonal polynomials are

   $$p_0(t) = 1, \quad p_1(t) = t, \quad \text{and} \quad p_2(t) = \tfrac{3}{8}t^2 - \tfrac{35}{8}$$

   (The polynomial p_2 has been scaled so that its values at the evaluation points are small integers.)

   b. Fit a quadratic trend function to the data

   $$(-5, 1), (-3, 1), (-1, 4), (1, 4), (3, 6), (5, 8)$$

In Exercises 5–14, the space is $C[0, 2\pi]$ with the inner product (6).

5. Show that $\sin mt$ and $\sin nt$ are orthogonal when $m \neq n$.

6. Show that $\sin mt$ and $\cos nt$ are orthogonal for all positive integers m and n.

7. Show that $\|\cos kt\|^2 = \pi$ and $\|\sin kt\|^2 = \pi$ for $k > 0$.

8. Find the third-order Fourier approximation to $f(t) = t - 1$.

9. Find the third-order Fourier approximation to $f(t) = 2\pi - t$.

10. Find the third-order Fourier approximation to the *square wave function* $f(t) = 1$ for $0 \le t < \pi$ and $f(t) = -1$ for $\pi \le t < 2\pi$.

11. Find the third-order Fourier approximation to $\sin^2 t$, without performing any integration calculations.

12. Find the third-order Fourier approximation to $\cos^3 t$, without performing any integration calculations.

13. Explain why a Fourier coefficient of the sum of two functions is the sum of the corresponding Fourier coefficients of the two functions.

14. Suppose the first few Fourier coefficients of some function f in $C[0, 2\pi]$ are a_0, a_1, a_2, and b_1, b_2, b_3. Which of the following trigonometric polynomials is closer to f? Defend your answer.

$$g(t) = \frac{a_0}{2} + a_1 \cos t + a_2 \cos 2t + b_1 \sin t$$

$$h(t) = \frac{a_0}{2} + a_1 \cos t + a_2 \cos 2t + b_1 \sin t + b_2 \sin 2t$$

T 15. Refer to the data in Exercise 19 in Section 6.6, concerning the takeoff performance of an airplane. Suppose the possible measurement errors become greater as the speed of the airplane increases, and let W be the diagonal weighting matrix whose diagonal entries are 1, 1, 1, .9, .9, .8, .7, .6, .5, .4, .3, .2, and .1. Find the cubic curve that fits the data with minimum weighted least-squares error, and use it to estimate the velocity of the plane when $t = 4.5$ seconds.

T 16. Let f_4 and f_5 be the fourth-order and fifth-order Fourier approximations in $C[0, 2\pi]$ to the square wave function in Exercise 10. Produce separate graphs of f_4 and f_5 on the interval $[0, 2\pi]$, and produce a graph of f_5 on $[-2\pi, 2\pi]$.

STUDY GUIDE offers additional resources for mastering orthogonal projections.

Solutions to Practice Problems

1. Compute

$$\langle q_1, q_2 \rangle = \int_{-2}^{2} 1 \cdot t\, dt = \frac{1}{2}t^2 \Big|_{-2}^{2} = 0$$

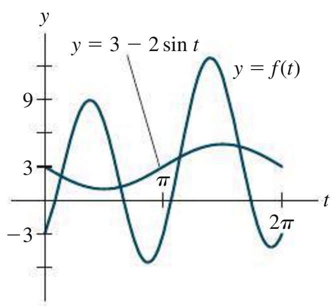
$$\langle q_1, q_3 \rangle = \int_{-2}^{2} 1 \cdot (3t^2 - 4)\, dt = (t^3 - 4t)\Big|_{-2}^{2} = 0$$

$$\langle q_2, q_3 \rangle = \int_{-2}^{2} t \cdot (3t^2 - 4)\, dt = \left(\frac{3}{4}t^4 - 2t^2\right)\Big|_{-2}^{2} = 0$$

2. The third-order Fourier approximation to f is the best approximation in $C[0, 2\pi]$ to f by functions (vectors) in the subspace spanned by 1, $\cos t$, $\cos 2t$, $\cos 3t$, $\sin t$, $\sin 2t$, and $\sin 3t$. But f is obviously *in* this subspace, so f is its own best approximation:

$$f(t) = 3 - 2\sin t + 5\sin 2t - 6\cos 2t$$

For the first-order approximation, the closest function to f in the subspace $W = \text{Span}\{1, \cos t, \sin t\}$ is $3 - 2\sin t$. The other two terms in the formula for $f(t)$ are orthogonal to the functions in W, so they contribute nothing to the integrals that give the Fourier coefficients for a first-order approximation.



First- and third-order approximations to $f(t)$.

CHAPTER 6 PROJECTS

Chapter 6 projects are available online at bit.ly/30IM8gT.

A. *The QR Method for Finding Eigenvalues*: This project shows how the QR factorization of a matrix may be used to calculate the eigenvalues of the matrix.

B. *Finding the Roots of a Polynomial with Eigenvalues*: This project shows how the real roots of a polynomial can be calculated by finding the eigenvalues of a particular matrix. These eigenvalues will be found by the QR method.

CHAPTER 6 SUPPLEMENTARY EXERCISES

The statements in Exercises 1–19 refer to vectors in $\mathbb{R}^n$ (or $\mathbb{R}^m$) with the standard inner product. Mark each statement True or False **(T/F)**. Justify each answer.

1. **(T/F)** The length of every vector is a positive number.

2. **(T/F)** A vector $\mathbf{v}$ and its negative, $-\mathbf{v}$, have equal lengths.

3. **(T/F)** The distance between $\mathbf{u}$ and $\mathbf{v}$ is $\|\mathbf{u}-\mathbf{v}\|$.

4. **(T/F)** If r is any scalar, then $\|r\mathbf{v}\| = r\|\mathbf{v}\|$.

5. **(T/F)** If two vectors are orthogonal, they are linearly independent.

6. **(T/F)** If $\mathbf{x}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$, then $\mathbf{x}$ must be orthogonal to $\mathbf{u}-\mathbf{v}$.

7. **(T/F)** If $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

8. **(T/F)** If $\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

9. **(T/F)** The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is a scalar multiple of $\mathbf{y}$.

10. **(T/F)** If a vector $\mathbf{y}$ coincides with its orthogonal projection onto a subspace W, then $\mathbf{y}$ is in W.

11. **(T/F)** The set of all vectors in $\mathbb{R}^n$ orthogonal to one fixed vector is a subspace of $\mathbb{R}^n$.

12. **(T/F)** If W is a subspace of $\mathbb{R}^n$, then W and $W^{\perp}$ have no vectors in common.

13. **(T/F)** If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set and if c_1, c_2, and c_3 are scalars, then $\{c_1\mathbf{v}_1, c_2\mathbf{v}_2, c_3\mathbf{v}_3\}$ is an orthogonal set.

14. **(T/F)** If a matrix U has orthonormal columns, then $UU^T = I$.

15. **(T/F)** A square matrix with orthogonal columns is an orthogonal matrix.

16. **(T/F)** If a square matrix has orthonormal columns, then it also has orthonormal rows.

17. **(T/F)** If W is a subspace, then $\|\operatorname{proj}_W \mathbf{v}\|^2 + \|\mathbf{v} - \operatorname{proj}_W \mathbf{v}\|^2 = \|\mathbf{v}\|^2$.

18. **(T/F)** A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is the vector $A\hat{\mathbf{x}}$ in Col A closest to $\mathbf{b}$, so that $\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$ for all $\mathbf{x}$.

19. **(T/F)** The normal equations for a least-squares solution of $A\mathbf{x} = \mathbf{b}$ are given by $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.

20. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be an orthonormal set. Verify the following equality by induction, beginning with $p = 2$. If $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$, then $\|\mathbf{x}\|^2 = |c_1|^2 + \cdots + |c_p|^2$

21. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be an orthonormal set in $\mathbb{R}^n$. Verify the following inequality, called *Bessel's inequality*, which is true for each $\mathbf{x}$ in $\mathbb{R}^n$:

$$\|\mathbf{x}\|^2 \ge |\mathbf{x}\cdot\mathbf{v}_1|^2 + |\mathbf{x}\cdot\mathbf{v}_2|^2 + \cdots + |\mathbf{x}\cdot\mathbf{v}_p|^2$$

22. Let U be an $n \times n$ orthogonal matrix. Show that if $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$, then so is $\{U\mathbf{v}_1, \dots, U\mathbf{v}_n\}$.

23. Show that if an $n \times n$ matrix U satisfies $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, then U is an orthogonal matrix.

24. Show that if U is an orthogonal matrix, then any real eigenvalue of U must be ± 1.

25. A *Householder matrix*, or an *elementary reflector*, has the form $Q = I - 2\mathbf{u}\mathbf{u}^T$ where $\mathbf{u}$ is a unit vector. Show that Q is an orthogonal matrix. (Elementary reflectors are often used in computer programs to produce a QR factorization of a matrix A. If A has linearly independent columns, then left-multiplication by a sequence of elementary reflectors can produce an upper triangular matrix.)

26. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation that preserves lengths; that is, $\|T(\mathbf{x})\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.
    a. Show that T also preserves orthogonality; that is, $T(\mathbf{x})\cdot T(\mathbf{y}) = 0$ whenever $\mathbf{x}\cdot\mathbf{y} = 0$.
    b. Show that the standard matrix of T is an orthogonal matrix.

27. Let $\mathbf{u}$ and $\mathbf{v}$ be linearly independent vectors in $\mathbb{R}^n$ that are *not* orthogonal. Describe how to find the best approximation to $\mathbf{z}$ in $\mathbb{R}^n$ by vectors of the form $x_1\mathbf{u} + x_2\mathbf{v}$ without first constructing an orthogonal basis for Span $\{\mathbf{u}, \mathbf{v}\}$.

28. Suppose the columns of A are linearly independent. Determine what happens to the least-squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ when $\mathbf{b}$ is replaced by $c\mathbf{b}$ for some nonzero scalar c.

29. If a, b, and c are distinct numbers, then the following system is inconsistent because the graphs of the equations are parallel planes. Show that the set of all least-squares solutions of the system is precisely the plane whose equation is $x - 2y + 5z = (a + b + c)/3$.

$$\begin{aligned} x - 2y + 5z &= a \\ x - 2y + 5z &= b \\ x - 2y + 5z &= c \end{aligned}$$

30. Consider the problem of finding an eigenvalue of an $n \times n$ matrix A when an approximate eigenvector $\mathbf{v}$ is known. Since $\mathbf{v}$ is not exactly correct, the equation

$$A\mathbf{v} = \lambda\mathbf{v} \tag{1}$$

will probably not have a solution. However, λ can be estimated by a least-squares solution when (1) is viewed properly. Think of $\mathbf{v}$ as an $n \times 1$ matrix V, think of λ as a vector in $\mathbb{R}^1$, and denote the vector $A\mathbf{v}$ by the symbol $\mathbf{b}$. Then (1) becomes $\mathbf{b} = \lambda V$, which may also be written as $V\lambda = \mathbf{b}$. Find the least-squares solution of this system of n equations in the one unknown λ, and write this solution using the original symbols. The resulting estimate for λ is called a *Rayleigh quotient*.

31. Use the steps below to prove the following relations among the four fundamental subspaces determined by an $m \times n$ matrix A.

$\text{Row } A = (\text{Nul } A)^{\perp}, \quad \text{Col } A = (\text{Nul } A^T)^{\perp}$

a. Show that Row A is contained in $(\text{Nul } A)^{\perp}$. (Show that if $\mathbf{x}$ is in Row A, then $\mathbf{x}$ is orthogonal to every $\mathbf{u}$ in Nul A.)

b. Suppose rank $A = r$. Find dim Nul A and dim $(\text{Nul } A)^{\perp}$, and then deduce from part (a) that Row $A = (\text{Nul } A)^{\perp}$. [*Hint:* Study the exercises for Section 6.3.]

c. Explain why Col $A = (\text{Nul } A^T)^{\perp}$.

32. Explain why an equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is orthogonal to all solutions of the equation $A^T\mathbf{x} = \mathbf{0}$.

Exercises 33 and 34 concern the (real) *Schur factorization* of an $n \times n$ matrix A in the form $A = URU^T$, where U is an orthogonal matrix and R is an $n \times n$ upper triangular matrix.[1]

33. Show that if A admits a (real) Schur factorization, $A = URU^T$, then A has n real eigenvalues, counting multiplicities.

34. Let A be an $n \times n$ matrix with n real eigenvalues, counting multiplicities, denoted by $\lambda_1, \dots, \lambda_n$. It can be shown that A admits a (real) Schur factorization. Parts (a) and (b) show the key ideas in the proof. The rest of the proof amounts to repeating (a) and (b) for successively smaller matrices, and then piecing together the results.

a. Let $\mathbf{u}_1$ be a unit eigenvector corresponding to λ_1, let $\mathbf{u}_2, \dots, \mathbf{u}_n$ be any other vectors such that $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$, and then let $U = [\,\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \cdots \;\; \mathbf{u}_n\,]$. Show that the first column of U^TAU is $\lambda_1\mathbf{e}_1$, where $\mathbf{e}_1$ is the first column of the $n \times n$ identity matrix.

b. Part (a) implies that U^TAU has the form shown below. Explain why the eigenvalues of A_1 are $\lambda_2, \dots, \lambda_n$. [*Hint:* See the Supplementary Exercises for Chapter 5.]

$$U^TAU = \begin{bmatrix} \lambda_1 & * & * & * & * \\ 0 & & & & \\ \vdots & & A_1 & & \\ 0 & & & & \end{bmatrix}$$

[1] If complex numbers are allowed, *every* $n \times n$ matrix A admits a (complex) Schur factorization, $A = URU^{-1}$, where R is upper triangular and U^{-1} is the *conjugate* transpose of U. This very useful fact is discussed in *Matrix Analysis*, by Roger A. Horn and Charles R. Johnson (Cambridge: Cambridge University Press, 1985), pp. 79–100.

T When the right side of an equation $A\mathbf{x} = \mathbf{b}$ is changed slightly—say, to $A\mathbf{x} = \mathbf{b} + \Delta\mathbf{b}$ for some vector $\Delta\mathbf{b}$—the solution changes from $\mathbf{x}$ to $\mathbf{x} + \Delta\mathbf{x}$, where $\Delta\mathbf{x}$ satisfies $A(\Delta\mathbf{x}) = \Delta\mathbf{b}$. The quotient $\|\Delta\mathbf{b}\|/\|\mathbf{b}\|$ is called the **relative change** in $\mathbf{b}$ (or the **relative error** in $\mathbf{b}$ when $\Delta\mathbf{b}$ represents possible error in the entries of $\mathbf{b}$). The relative change in the solution is $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$. When A is invertible, the **condition number** of A, written as cond(A), produces a bound on how large the relative change in $\mathbf{x}$ can be:

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \text{cond}(A)\cdot\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{2}$$

In Exercises 35–38, solve $A\mathbf{x} = \mathbf{b}$ and $A(\Delta\mathbf{x}) = \Delta\mathbf{b}$, and show that the inequality (2) holds in each case. (See the discussion of *ill-conditioned* matrices in Exercises 49–51 in Section 2.3.)

T 35. $A = \begin{bmatrix} 4.5 & 3.1 \\ 1.6 & 1.1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 19.249 \\ 6.843 \end{bmatrix}$, $\Delta\mathbf{b} = \begin{bmatrix} .001 \\ -.003 \end{bmatrix}$

T 36. $A = \begin{bmatrix} 4.5 & 3.1 \\ 1.6 & 1.1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} .500 \\ -1.407 \end{bmatrix}$, $\Delta\mathbf{b} = \begin{bmatrix} .001 \\ -.003 \end{bmatrix}$

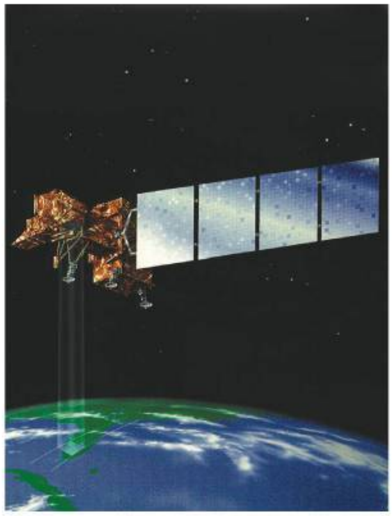
T 37. $A = \begin{bmatrix} 7 & -6 & -4 & 1 \\ -5 & 1 & 0 & -2 \\ 10 & 11 & 7 & -3 \\ 19 & 9 & 7 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} .100 \\ 2.888 \\ -1.404 \\ 1.462 \end{bmatrix}$,

$$\Delta\mathbf{b} = 10^{-4}\begin{bmatrix} .49 \\ -1.28 \\ 5.78 \\ 8.04 \end{bmatrix}$$

T 38. $A = \begin{bmatrix} 7 & -6 & -4 & 1 \\ -5 & 1 & 0 & -2 \\ 10 & 11 & 7 & -3 \\ 19 & 9 & 7 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4.230 \\ -11.043 \\ 49.991 \\ 69.536 \end{bmatrix}$,

$$\Delta\mathbf{b} = 10^{-4}\begin{bmatrix} .27 \\ 7.76 \\ -3.77 \\ 3.93 \end{bmatrix}$$

7 Symmetric Matrices and Quadratic Forms



Introductory Example

MULTICHANNEL IMAGE PROCESSING

Around the world in little more than 80 *minutes*, the two Landsat satellites streak silently across the sky in near polar orbits, recording images of terrain and coastline, in swaths 185 kilometers wide. Every 16 days, each satellite passes over almost every square kilometer of the earth's surface, so any location can be monitored every 8 days.

The Landsat images are useful for many purposes. Developers and urban planners use them to study the rate and direction of urban growth, industrial development, and other changes in land usage. Rural countries can analyze soil moisture, classify the vegetation in remote regions, and locate inland lakes and streams. Governments can detect and assess damage from natural disasters, such as forest fires, lava flows, floods, and hurricanes. Environmental agencies can identify pollution from smokestacks and measure water temperatures in lakes and rivers near power plants.

Sensors aboard the satellite acquire seven simultaneous images of any region on earth to be studied. The sensors record energy from separate wavelength bands—three in the visible light spectrum and four in infrared and thermal bands. Each image is digitized and stored as a rectangular array of numbers, each number indicating the signal intensity at a corresponding small point (or *pixel*) on the image. Each of the seven images is one channel of a *multichannel* or *multispectral image*.

The seven Landsat images of one fixed region typically contain much redundant information, since some features will appear in several images. Yet other features, because of their color or temperature, may reflect light that is recorded by only one or two sensors. One goal of multichannel image processing is to view the data in a way that extracts information better than studying each image separately.

Principal component analysis is an effective way to suppress redundant information and provide in only one or two composite images most of the information from the initial data. Roughly speaking, the goal is to find a special linear combination of the images, that is, a list of weights that at each pixel combine all seven corresponding image values into one new value. The weights are chosen in a way that makes the range of light intensities—the *scene variance*—in the composite image (called the *first principal component*) greater than that in any of the original images. Additional *component* images can also be constructed by criteria that will be explained in Section 7.5.

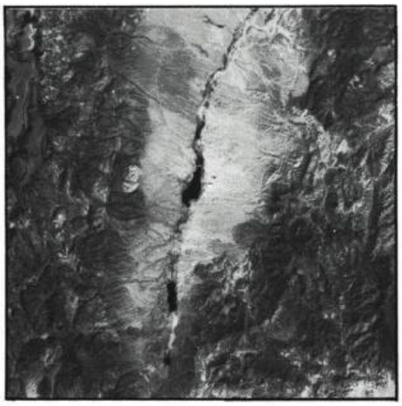
Principal component analysis is illustrated in the photos on the next page, taken over Railroad Valley, Nevada. Images from three Landsat spectral bands are shown in (a)–(c). The total information in the three bands is rearranged in the three principal component images in (d)–(f). The first component (d) displays (or "explains") 93.5% of the scene variance present in the initial data.

In this way, the three-channel initial data have been reduced to one-channel data, with a loss in some sense of only 6.5% of the scene variance.

Earth Satellite Corporation of Rockville, Maryland, which kindly supplied the photos shown here, is experimenting with images from 224 separate spectral bands. Principal component analysis, essential for such massive data sets, typically reduces the data to about 15 usable principal components.

(a) Spectral band 1: Visible blue.

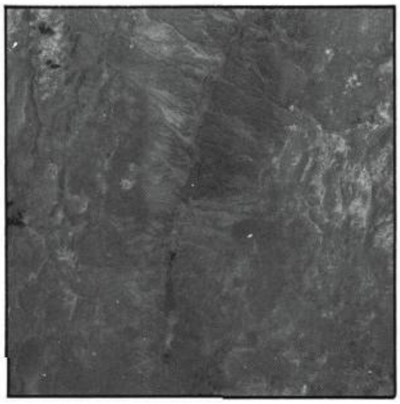
(b) Spectral band 4: Near infrared.



(c) Spectral band 7: Mid-infrared.

(d) Principal component 1: 93.5%.

(e) Principal component 2: 5.3%.



(f) Principal component 3: 1.2%.

Symmetric matrices arise more often in applications, in one way or another, than any other major class of matrices. The theory is rich and beautiful, depending in an essential way on both diagonalization from Chapter 5 and orthogonality from Chapter 6. The diagonalization of a symmetric matrix, described in Section 7.1, is the foundation for the discussion in Sections 7.2 and 7.3 concerning quadratic forms. Section 7.3, in turn, is needed for the final two sections on the singular value decomposition and on the image processing described in the introductory example. Throughout the chapter, all vectors and matrices have real entries.

7.1 Diagonalization of Symmetric Matrices

A **symmetric** matrix is a matrix A such that $A^T = A$. Such a matrix is necessarily square. Its main diagonal entries are arbitrary, but its other entries occur in pairs—on opposite sides of the main diagonal.

EXAMPLE 1 Of the following matrices, only the first three are symmetric:

$$\text{Symmetric:}\quad \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix},\quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix},\quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

$$\text{Nonsymmetric:}\quad \begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix},\quad \begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix},\quad \begin{bmatrix} 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$$ ■

To begin the study of symmetric matrices, it is helpful to review the diagonalization process of Section 5.3.

EXAMPLE 2 If possible, diagonalize the matrix $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$.

SOLUTION The characteristic equation of A is

$$0 = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$$

Standard calculations produce a basis for each eigenspace:

$$\lambda = 8:\ \mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix};\qquad \lambda = 6:\ \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix};\qquad \lambda = 3:\ \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

These three vectors form a basis for $\mathbb{R}^3$. In fact, it is easy to check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an *orthogonal* basis for $\mathbb{R}^3$. Experience from Chapter 6 suggests that an *orthonormal* basis might be useful for calculations, so here are the normalized (unit) eigenvectors.

$$\mathbf{u}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix},\quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix},\quad \mathbf{u}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Let

$$P = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix},\quad D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Then $A = PDP^{-1}$, as usual. But this time, since P is square and has orthonormal columns, P is an *orthogonal* matrix, and P^{-1} is simply P^T. (See Section 6.2.) ■

Theorem 1 explains why the eigenvectors in Example 2 are orthogonal—they correspond to distinct eigenvalues.

THEOREM 1

If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

PROOF Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2. To show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, compute

$$\begin{aligned}
\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 && \text{Since } \mathbf{v}_1 \text{ is an eigenvector} \\
&= (\mathbf{v}_1^T A^T)\mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) && \text{Since } A^T = A \\
&= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) && \text{Since } \mathbf{v}_2 \text{ is an eigenvector} \\
&= \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2
\end{aligned}$$

Hence $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. But $\lambda_1 - \lambda_2 \neq 0$, so $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. ■

The special type of diagonalization in Example 2 is crucial for the theory of symmetric matrices. An $n \times n$ matrix A is said to be **orthogonally diagonalizable** if there are an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1} \tag{1}$$

Such a diagonalization requires n linearly independent and orthonormal eigenvectors. When is this possible? If A is orthogonally diagonalizable as in (1), then

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A$$

Thus A is symmetric! Theorem 2 below shows that, conversely, every symmetric matrix is orthogonally diagonalizable. The proof is much harder and is omitted; the main idea for a proof will be given after Theorem 3.

THEOREM 2

An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.

This theorem is rather amazing, because the work in Chapter 5 would suggest that it is usually impossible to tell when a matrix is diagonalizable. But this is not the case for symmetric matrices.

The next example treats a matrix whose eigenvalues are not all distinct.

EXAMPLE 3 Orthogonally diagonalize the matrix $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$, whose characteristic equation is

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

SOLUTION The usual calculations produce bases for the eigenspaces:

$$\lambda = 7: \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}; \qquad \lambda = -2: \mathbf{v}_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$$

Although $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, they are not orthogonal. Recall from Section 6.2 that the projection of $\mathbf{v}_2$ onto $\mathbf{v}_1$ is $\dfrac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$, and the component of $\mathbf{v}_2$ orthogonal to $\mathbf{v}_1$ is

$$\mathbf{z}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} - \frac{-1/2}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{z}_2\}$ is an orthogonal set in the eigenspace for $\lambda = 7$. (Note that $\mathbf{z}_2$ is a linear combination of the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, so $\mathbf{z}_2$ is in the eigenspace. This construction of $\mathbf{z}_2$ is just the Gram–Schmidt process of Section 6.4.) Since the eigenspace is two-dimensional (with basis $\mathbf{v}_1, \mathbf{v}_2$), the orthogonal set $\{\mathbf{v}_1, \mathbf{z}_2\}$ is an *orthogonal basis* for the eigenspace, by the Basis Theorem. (See Section 2.9 or 4.5.)

Normalize $\mathbf{v}_1$ and $\mathbf{z}_2$ to obtain the following orthonormal basis for the eigenspace for $\lambda = 7$:

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}$$

An orthonormal basis for the eigenspace for $\lambda = -2$ is

$$\mathbf{u}_3 = \frac{1}{\|2\mathbf{v}_3\|}2\mathbf{v}_3 = \frac{1}{3}\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$$

By Theorem 1, $\mathbf{u}_3$ is orthogonal to the other eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$. Hence $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set. Let

$$P = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \\ 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \end{bmatrix}, \quad D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Then P orthogonally diagonalizes A, and $A = PDP^{-1}$. ■

In Example 3, the eigenvalue 7 has multiplicity two and the eigenspace is two-dimensional. This fact is not accidental, as the next theorem shows.

The Spectral Theorem

The set of eigenvalues of a matrix A is sometimes called the *spectrum* of A, and the following description of the eigenvalues is called a *spectral theorem.*

THEOREM 3

The Spectral Theorem for Symmetric Matrices

An $n \times n$ symmetric matrix A has the following properties:

a. A has n real eigenvalues, counting multiplicities.
b. The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.
c. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.
d. A is orthogonally diagonalizable.

Part (a) follows from Exercise 28 in Section 5.5. Part (b) follows easily from part (d). (See Exercise 37.) Part (c) is Theorem 1. Because of (a), a proof of (d) can be given using Exercise 38 and the Schur factorization discussed in Supplementary Exercise 34 in Chapter 6. The details are omitted.

Spectral Decomposition

Suppose $A = PDP^{-1}$, where the columns of P are orthonormal eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ of A and the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$\begin{aligned} A = PDP^T &= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \end{aligned}$$

Using the column–row expansion of a product (Theorem 10 in Section 2.4), we can write

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T \tag{2}$$

This representation of A is called a **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A. Each term in (2) is an $n \times n$ matrix of rank 1. For example, every column of $\lambda_1\mathbf{u}_1\mathbf{u}_1^T$ is a multiple of $\mathbf{u}_1$. Furthermore, each matrix $\mathbf{u}_j\mathbf{u}_j^T$ is a **projection matrix** in the sense that for each $\mathbf{x}$ in $\mathbb{R}^n$, the vector $(\mathbf{u}_j\mathbf{u}_j^T)\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto the subspace spanned by $\mathbf{u}_j$. (See Exercise 41.)

EXAMPLE 4 Construct a spectral decomposition of the matrix A that has the orthogonal diagonalization

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

SOLUTION Denote the columns of P by $\mathbf{u}_1$ and $\mathbf{u}_2$. Then

$$A = 8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T$$

To verify this decomposition of A, compute

$$\mathbf{u}_1\mathbf{u}_1^T = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}$$

$$\mathbf{u}_2\mathbf{u}_2^T = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{bmatrix}$$

and

$$8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T = \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = A \quad \blacksquare$$

Numerical Notes

When A is symmetric and not too large, modern high-performance computer algorithms calculate eigenvalues and eigenvectors with great precision. They apply a sequence of similarity transformations to A involving orthogonal matrices. The diagonal entries of the transformed matrices converge rapidly to the eigenvalues of A. (See the Numerical Notes in Section 5.2.) Using orthogonal matrices generally prevents numerical errors from accumulating during the process. When A is symmetric, the sequence of orthogonal matrices combines to form an orthogonal matrix whose columns are eigenvectors of A.

A nonsymmetric matrix cannot have a full set of orthogonal eigenvectors, but the algorithm still produces fairly accurate eigenvalues. After that, nonorthogonal techniques are needed to calculate eigenvectors.

Practice Problems

1. Show that if A is a symmetric matrix, then A^2 is symmetric.
2. Show that if A is orthogonally diagonalizable, then so is A^2.

7.1 Exercises

Determine which of the matrices in Exercises 1–6 are symmetric.

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

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

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

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

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

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

Determine which of the matrices in Exercises 7–12 are orthogonal. If orthogonal, find the inverse.

7. $\begin{bmatrix} .6 & .8 \\ .8 & -.6 \end{bmatrix}$

8. $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

9. $\begin{bmatrix} -4/5 & 3/5 \\ 3/5 & 4/5 \end{bmatrix}$

10. $\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ 2/3 & 1/3 & -2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}$

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

12. $\begin{bmatrix} .5 & .5 & -.5 & -.5 \\ .5 & .5 & .5 & .5 \\ .5 & -.5 & -.5 & .5 \\ .5 & -.5 & .5 & -.5 \end{bmatrix}$

Orthogonally diagonalize the matrices in Exercises 13–22, giving an orthogonal matrix P and a diagonal matrix D. To save you time, the eigenvalues in Exercises 17–22 are the following: (17) $-4, 4, 7$; (18) $-3, -6, 9$; (19) $-2, 7$; (20) $-3, 15$; (21) $1, 5, 9$; (22) $3, 5$.

13. $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

14. $\begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$

15. $\begin{bmatrix} 3 & 4 \\ 4 & 9 \end{bmatrix}$

16. $\begin{bmatrix} 6 & -2 \\ -2 & 9 \end{bmatrix}$

17. $\begin{bmatrix} 1 & 1 & 5 \\ 1 & 5 & 1 \\ 5 & 1 & 1 \end{bmatrix}$

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

19. $\begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$

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

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

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

23. Let $A = \begin{bmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Verify that 5 is an eigenvalue of A and $\mathbf{v}$ is an eigenvector. Then orthogonally diagonalize A.

24. Let $A = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. Verify that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. Then orthogonally diagonalize A.

In Exercises 25–32, mark each statement True or False **(T/F)**. Justify each answer.

25. **(T/F)** An $n \times n$ matrix that is orthogonally diagonalizable must be symmetric.

26. **(T/F)** There are symmetric matrices that are not orthogonally diagonalizable.

27. **(T/F)** An orthogonal matrix is orthogonally diagonalizable.

28. **(T/F)** If $B = PDP^T$, where $P^T = P^{-1}$ and D is a diagonal matrix, then B is a symmetric matrix.

29. **(T/F)** For a nonzero $\mathbf{v}$ in $\mathbb{R}^n$, the matrix $\mathbf{v}\mathbf{v}^T$ is called a projection matrix.

30. **(T/F)** If $A^T = A$ and if vectors $\mathbf{u}$ and $\mathbf{v}$ satisfy $A\mathbf{u} = 3\mathbf{u}$ and $A\mathbf{v} = 4\mathbf{v}$, then $\mathbf{u} \cdot \mathbf{v} = 0$.

31. **(T/F)** An $n \times n$ symmetric matrix has n distinct real eigenvalues.

32. **(T/F)** The dimension of an eigenspace of a symmetric matrix is sometimes less than the multiplicity of the corresponding eigenvalue.

33. Show that if A is an $n \times n$ symmetric matrix, then $(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$ for all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$.

34. Suppose A is a symmetric $n \times n$ matrix and B is any $n \times m$ matrix. Show that B^TAB, B^TB, and BB^T are symmetric matrices.

35. Suppose A is invertible and orthogonally diagonalizable. Explain why A^{-1} is also orthogonally diagonalizable.

36. Suppose A and B are both orthogonally diagonalizable and $AB = BA$. Explain why AB is also orthogonally diagonalizable.

37. Let $A = PDP^{-1}$, where P is orthogonal and D is diagonal, and let λ be an eigenvalue of A of multiplicity k. Then λ appears k times on the diagonal of D. Explain why the dimension of the eigenspace for λ is k.

38. Suppose $A = PRP^{-1}$, where P is orthogonal and R is upper triangular. Show that if A is symmetric, then R is symmetric and hence is actually a diagonal matrix.

39. Construct a spectral decomposition of A from Example 2.

40. Construct a spectral decomposition of A from Example 3.

41. Let $\mathbf{u}$ be a unit vector in $\mathbb{R}^n$, and let $B = \mathbf{u}\mathbf{u}^T$.

a. Given any $\mathbf{x}$ in $\mathbb{R}^n$, compute $B\mathbf{x}$ and show that $B\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathbf{u}$, as described in Section 6.2.

b. Show that B is a symmetric matrix and $B^2 = B$.

c. Show that $\mathbf{u}$ is an eigenvector of B. What is the corresponding eigenvalue?

42. Let B be an $n \times n$ symmetric matrix such that $B^2 = B$. Any such matrix is called a **projection matrix** (or an **orthogonal projection matrix**). Given any $\mathbf{y}$ in $\mathbb{R}^n$, let $\hat{\mathbf{y}} = B\mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

a. Show that $\mathbf{z}$ is orthogonal to $\hat{\mathbf{y}}$.

b. Let W be the column space of B. Show that $\mathbf{y}$ is the sum of a vector in W and a vector in $W^\perp$. Why does this prove that $B\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto the column space of B?

Orthogonally diagonalize the matrices in Exercises 43–46. To practice the methods of this section, do not use an eigenvector routine from your matrix program. Instead, use the program to find the eigenvalues, and, for each eigenvalue λ, find an orthonormal basis for $\text{Nul}(A - \lambda I)$, as in Examples 2 and 3.

T 43. $\begin{bmatrix} 6 & 2 & 9 & -6 \\ 2 & 6 & -6 & 9 \\ 9 & -6 & 6 & 2 \\ -6 & 9 & 2 & 6 \end{bmatrix}$

T 44. $\begin{bmatrix} .63 & -.18 & -.06 & -.04 \\ -.18 & .84 & -.04 & .12 \\ -.06 & -.04 & .72 & -.12 \\ -.04 & .12 & -.12 & .66 \end{bmatrix}$

T 45. $\begin{bmatrix} .31 & .58 & .08 & .44 \\ .58 & -.56 & .44 & -.58 \\ .08 & .44 & .19 & -.08 \\ .44 & -.58 & -.08 & .31 \end{bmatrix}$

T 46. $\begin{bmatrix} 8 & 2 & 2 & -6 & 9 \\ 2 & 8 & 2 & -6 & 9 \\ 2 & 2 & 8 & -6 & 9 \\ -6 & -6 & -6 & 24 & 9 \\ 9 & 9 & 9 & 9 & -21 \end{bmatrix}$

Solutions to Practice Problems

1. $(A^2)^T = (AA)^T = A^TA^T$, by a property of transposes. By hypothesis, $A^T = A$. So $(A^2)^T = AA = A^2$, which shows that A^2 is symmetric.

2. If A is orthogonally diagonalizable, then A is symmetric, by Theorem 2. By Practice Problem 1, A^2 is symmetric and hence is orthogonally diagonalizable (Theorem 2).

7.2 Quadratic Forms

Until now, our attention in this text has focused on linear equations, except for the sums of squares encountered in Chapter 6 when computing $\mathbf{x}^T\mathbf{x}$. Such sums and more general expressions, called *quadratic forms*, occur frequently in applications of linear algebra to engineering (in design criteria and optimization) and signal processing (as output noise power). They also arise, for example, in physics (as potential and kinetic energy), differential geometry (as normal curvature of surfaces), economics (as utility functions), and statistics (in confidence ellipsoids). Some of the mathematical background for such applications flows easily from our work on symmetric matrices.

A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

The simplest example of a nonzero quadratic form is $Q(\mathbf{x}) = \mathbf{x}^TI\mathbf{x} = \|\mathbf{x}\|^2$. Examples 1 and 2 show the connection between any symmetric matrix A and the quadratic form $\mathbf{x}^TA\mathbf{x}$.

EXAMPLE 1 Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Compute $\mathbf{x}^TA\mathbf{x}$ for the following matrices:

a. $A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$ b. $A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$

SOLUTION

a. $\mathbf{x}^TA\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 4x_1 \\ 3x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2.$

b. There are two -2 entries in A. Watch how they enter the calculations. The $(1, 2)$-entry in A is in boldface type.

$$\begin{aligned}\mathbf{x}^TA\mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 3 & \mathbf{-2} \\ -2 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 3x_1 - \mathbf{2}x_2 \\ -2x_1 + 7x_2 \end{bmatrix} \\ &= x_1(3x_1 - \mathbf{2}x_2) + x_2(-2x_1 + 7x_2) \\ &= 3x_1^2 - \mathbf{2}x_1x_2 - 2x_2x_1 + 7x_2^2 \\ &= 3x_1^2 - 4x_1x_2 + 7x_2^2\end{aligned}$$ ■

The presence of $-4x_1x_2$ in the quadratic form in Example 1(b) is due to the -2 entries off the diagonal in the matrix A. In contrast, the quadratic form associated with the diagonal matrix A in Example 1(a) has no x_1x_2 *cross-product* term.

EXAMPLE 2 For $\mathbf{x}$ in $\mathbb{R}^3$, let $Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$. Write this quadratic form as $\mathbf{x}^TA\mathbf{x}$.

SOLUTION The coefficients of x_1^2, x_2^2, x_3^2 go on the diagonal of A. To make A symmetric, the coefficient of x_ix_j for $i \neq j$ must be split evenly between the (i, j)- and (j, i)-entries in A. The coefficient of x_1x_3 is 0. It is readily checked that

$$Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$ ■

EXAMPLE 3 Let $Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2$. Compute the value of $Q(\mathbf{x})$ for $\mathbf{x} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

SOLUTION

$$Q(-3, 1) = (-3)^2 - 8(-3)(1) - 5(1)^2 = 28$$
$$Q(2, -2) = (2)^2 - 8(2)(-2) - 5(-2)^2 = 16$$
$$Q(1, -3) = (1)^2 - 8(1)(-3) - 5(-3)^2 = -20$$

In some cases, quadratic forms are easier to use when they have no cross-product terms—that is, when the matrix of the quadratic form is a diagonal matrix. Fortunately, the cross-product term can be eliminated by making a suitable change of variable.

Change of Variable in a Quadratic Form

If $\mathbf{x}$ represents a variable vector in $\mathbb{R}^n$, then a **change of variable** is an equation of the form

$$\mathbf{x} = P\mathbf{y}, \qquad \text{or equivalently,} \qquad \mathbf{y} = P^{-1}\mathbf{x} \tag{1}$$

where P is an invertible matrix and $\mathbf{y}$ is a new variable vector in $\mathbb{R}^n$. Here $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to the basis of $\mathbb{R}^n$ determined by the columns of P. (See Section 4.4.)

If the change of variable (1) is made in a quadratic form $\mathbf{x}^T A\mathbf{x}$, then

$$\mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T P^T AP\mathbf{y} = \mathbf{y}^T (P^T AP)\mathbf{y} \tag{2}$$

and the new matrix of the quadratic form is $P^T AP$. Since A is symmetric, Theorem 2 guarantees that there is an *orthogonal* matrix P such that $P^T AP$ is a diagonal matrix D, and the quadratic form in (2) becomes $\mathbf{y}^T D\mathbf{y}$. This is the strategy of the next example.

EXAMPLE 4 Make a change of variable that transforms the quadratic form in Example 3 into a quadratic form with no cross-product term.

SOLUTION The matrix of the quadratic form in Example 3 is

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

The first step is to orthogonally diagonalize A. Its eigenvalues turn out to be $\lambda = 3$ and $\lambda = -7$. Associated unit eigenvectors are

$$\lambda = 3: \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}; \qquad \lambda = -7: \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

These vectors are automatically orthogonal (because they correspond to distinct eigenvalues) and so provide an orthonormal basis for $\mathbb{R}^2$. Let

$$P = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}, \qquad D = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}$$

Then $A = PDP^{-1}$ and $D = P^{-1}AP = P^T AP$, as pointed out earlier. A suitable change of variable is

$$\mathbf{x} = P\mathbf{y}, \qquad \text{where } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Then

$$\begin{aligned} x_1^2 - 8x_1x_2 - 5x_2^2 &= \mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) \\ &= \mathbf{y}^T P^T AP\mathbf{y} = \mathbf{y}^T D\mathbf{y} \\ &= 3y_1^2 - 7y_2^2 \end{aligned}$$

■

To illustrate the meaning of the equality of quadratic forms in Example 4, we can compute $Q(\mathbf{x})$ for $\mathbf{x} = (2, -2)$ using the new quadratic form. First, since $\mathbf{x} = P\mathbf{y}$,

$$\mathbf{y} = P^{-1}\mathbf{x} = P^T\mathbf{x}$$

so

$$\mathbf{y} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 6/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$$

Hence

$$\begin{aligned} 3y_1^2 - 7y_2^2 &= 3(6/\sqrt{5})^2 - 7(-2/\sqrt{5})^2 = 3(36/5) - 7(4/5) \\ &= 80/5 = 16 \end{aligned}$$

This is the value of $Q(\mathbf{x})$ in Example 3 when $\mathbf{x} = (2, -2)$. See Figure 1.

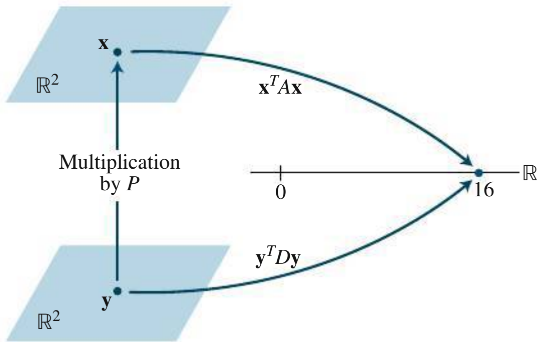


FIGURE 1 Change of variable in $\mathbf{x}^T A\mathbf{x}$.

Example 4 illustrates the following theorem. The proof of the theorem was essentially given before Example 4.

THEOREM 4

The Principal Axes Theorem

Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $\mathbf{x} = P\mathbf{y}$, 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 term.

The columns of P in the theorem are called the **principal axes** of the quadratic form $\mathbf{x}^T A\mathbf{x}$. The vector $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to the orthonormal basis of $\mathbb{R}^n$ given by these principal axes.

A Geometric View of Principal Axes

Suppose $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is an invertible 2×2 symmetric matrix, and let c be a constant. It can be shown that the set of all $\mathbf{x}$ in $\mathbb{R}^2$ that satisfy

$$\mathbf{x}^T A\mathbf{x} = c \tag{3}$$

either corresponds to an ellipse (or circle), a hyperbola, two intersecting lines, or a single point, or contains no points at all. If A is a diagonal matrix, the graph is in *standard position*, such as in Figure 2. If A is not a diagonal matrix, the graph of equation (3) is

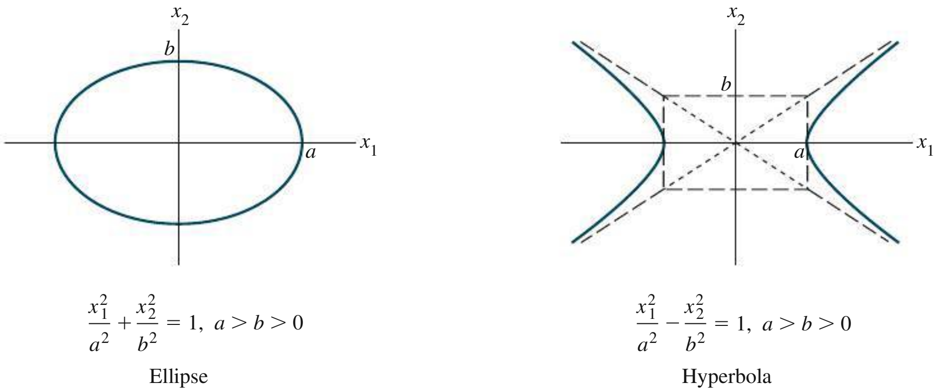


FIGURE 2 An ellipse and a hyperbola in standard position.

rotated out of standard position, as in Figure 3. Finding the *principal axes* (determined by the eigenvectors of A) amounts to finding a new coordinate system with respect to which the graph is in standard position.

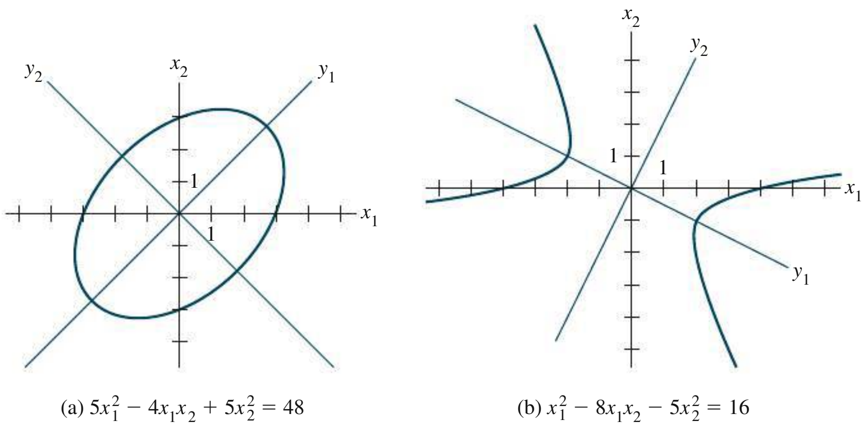


FIGURE 3 An ellipse and a hyperbola *not* in standard position.

The hyperbola in Figure 3(b) is the graph of the equation $\mathbf{x}^T A\mathbf{x} = 16$, where A is the matrix in Example 4. The positive y_1-axis in Figure 3(b) is in the direction of the first column of the matrix P in Example 4, and the positive y_2-axis is in the direction of the second column of P.

EXAMPLE 5 The ellipse in Figure 3(a) is the graph of the equation $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$. Find a change of variable that removes the cross-product term from the equation.

SOLUTION The matrix of the quadratic form is $A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$. The eigenvalues of A turn out to be 3 and 7, with corresponding unit eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Let $P = [\,\mathbf{u}_1 \quad \mathbf{u}_2\,] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$. Then P orthogonally diagonalizes A, so the change of variable $\mathbf{x} = P\mathbf{y}$ produces the quadratic form $\mathbf{y}^T D\mathbf{y} = 3y_1^2 + 7y_2^2$. The new axes for this change of variable are shown in Figure 3(a). ■

Classifying Quadratic Forms

When A is an $n \times n$ matrix, the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a real-valued function with domain $\mathbb{R}^n$. Figure 4 displays the graphs of four quadratic forms with domain $\mathbb{R}^2$. For each point $\mathbf{x} = (x_1, x_2)$ in the domain of a quadratic form Q, the graph displays the point (x_1, x_2, z) where $z = Q(\mathbf{x})$. Notice that except at $\mathbf{x} = \mathbf{0}$, the values of $Q(\mathbf{x})$ are all positive in Figure 4(a) and all negative in Figure 4(d). The horizontal cross-sections of the graphs are ellipses in Figures 4(a) and 4(d) and hyperbolas in Figure 4(c).

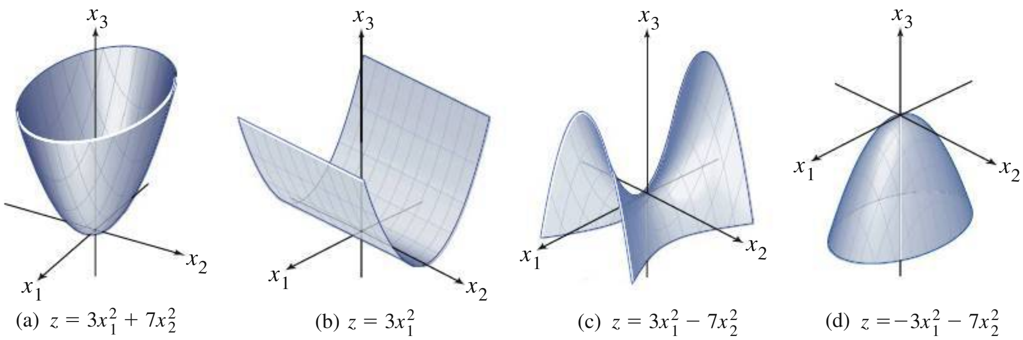


FIGURE 4 Graphs of quadratic forms.

The simple 2×2 examples in Figure 4 illustrate the following definitions.

DEFINITION

A quadratic form Q is

a. **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,

b. **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,

c. **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values.

Also, Q is said to be **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, and to be **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$. The quadratic forms in parts (a) and (b) of Figure 4 are both positive semidefinite, but the form in (a) is better described as positive definite.

Theorem 5 characterizes some quadratic forms in terms of eigenvalues.

THEOREM 5

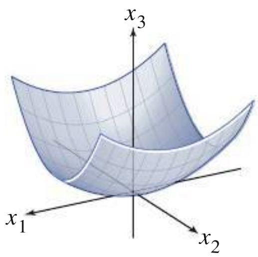
Quadratic Forms and Eigenvalues

Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\mathbf{x}^T A\mathbf{x}$ is

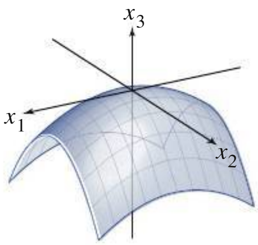
a. positive definite if and only if the eigenvalues of A are all positive,

b. negative definite if and only if the eigenvalues of A are all negative, or

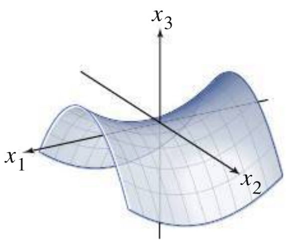
c. indefinite if and only if A has both positive and negative eigenvalues.



Positive definite



Negative definite



Indefinite

PROOF By the Principal Axes Theorem, there exists an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ such that

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{4}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A. Since P is invertible, there is a one-to-one correspondence between all nonzero $\mathbf{x}$ and all nonzero $\mathbf{y}$. Thus the values of $Q(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{0}$ coincide with the values of the expression on the right side of (4), which is obviously controlled by the signs of the eigenvalues $\lambda_1, \dots, \lambda_n$, in the three ways described in the theorem. ■

EXAMPLE 6 Is $Q(\mathbf{x}) = 3x_1^2 + 2x_2^2 + x_3^2 + 4x_1x_2 + 4x_2x_3$ positive definite?

SOLUTION Because of all the plus signs, this form "looks" positive definite. But the matrix of the form is

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

and the eigenvalues of A turn out to be 5, 2, and -1. So Q is an indefinite quadratic form, not positive definite. ■

The classification of a quadratic form is often carried over to the matrix of the form. Thus a **positive definite matrix** A is a *symmetric* matrix for which the quadratic form $\mathbf{x}^T A\mathbf{x}$ is positive definite. Other terms, such as **positive semidefinite matrix**, are defined analogously.

Numerical Notes

A fast way to determine whether a symmetric matrix A is positive definite is to attempt to factor A in the form $A = R^T R$, where R is upper triangular with positive diagonal entries. (A slightly modified algorithm for an LU factorization is one approach.) Such a *Cholesky factorization* is possible if and only if A is positive definite. See Supplementary Exercise 23 at the end of Chapter 7.

Practice Problem

Describe a positive semidefinite matrix A in terms of its eigenvalues.

7.2 Exercises

1. Compute the quadratic form $\mathbf{x}^T A\mathbf{x}$, when $A = \begin{bmatrix} 5 & 1/3 \\ 1/3 & 1 \end{bmatrix}$ and

a. $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ b. $\mathbf{x} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$ c. $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

2. Compute the quadratic form $\mathbf{x}^T A\mathbf{x}$, for $A = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 1 & 2 \\ 0 & 2 & 0 \end{bmatrix}$ and

a. $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ b. $\mathbf{x} = \begin{bmatrix} -2 \\ -1 \\ 5 \end{bmatrix}$ c. $\mathbf{x} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$

3. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^2$.

a. $3x_1^2 - 4x_1x_2 + 5x_2^2$ b. $3x_1^2 + 2x_1x_2$

4. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^2$.

a. $5x_1^2 + 16x_1x_2 - 5x_2^2$ b. $2x_1x_2$

5. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^3$.
   a. $3x_1^2 + 2x_2^2 - 5x_3^2 - 6x_1x_2 + 8x_1x_3 - 4x_2x_3$
   b. $6x_1x_2 + 4x_1x_3 - 10x_2x_3$

6. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^3$.
   a. $3x_1^2 - 2x_2^2 + 5x_3^2 + 4x_1x_2 - 6x_1x_3$
   b. $4x_3^2 - 2x_1x_2 + 4x_2x_3$

7. Make a change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $x_1^2 + 10x_1x_2 + x_2^2$ into a quadratic form with no cross-product term. Give P and the new quadratic form.

8. Let A be the matrix of the quadratic form

$$9x_1^2 + 7x_2^2 + 11x_3^2 - 8x_1x_2 + 8x_1x_3$$

It can be shown that the eigenvalues of A are 3, 9, and 15. Find an orthogonal matrix P such that the change of variable $\mathbf{x} = P\mathbf{y}$ transforms $\mathbf{x}^TA\mathbf{x}$ into a quadratic form with no cross-product term. Give P and the new quadratic form.

Classify the quadratic forms in Exercises 9–18. Then make a change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form into one with no cross-product term. Write the new quadratic form. Construct P using the methods of Section 7.1.

9. $4x_1^2 - 4x_1x_2 + 4x_2^2$

10. $2x_1^2 + 6x_1x_2 - 6x_2^2$

11. $2x_1^2 - 4x_1x_2 - x_2^2$

12. $-x_1^2 - 2x_1x_2 - x_2^2$

13. $x_1^2 - 6x_1x_2 + 9x_2^2$

14. $3x_1^2 + 4x_1x_2$

T 15. $-3x_1^2 - 7x_2^2 - 10x_3^2 - 10x_4^2 + 4x_1x_2 + 4x_1x_3 + 4x_1x_4 + 6x_3x_4$

T 16. $4x_1^2 + 4x_2^2 + 4x_3^2 + 4x_4^2 + 8x_1x_2 + 8x_3x_4 - 6x_1x_4 + 6x_2x_3$

T 17. $11x_1^2 + 11x_2^2 + 11x_3^2 + 11x_4^2 + 16x_1x_2 - 12x_1x_4 + 12x_2x_3 + 16x_3x_4$

T 18. $2x_1^2 + 2x_2^2 - 6x_1x_2 - 6x_1x_3 - 6x_1x_4 - 6x_2x_3 - 6x_2x_4 - 2x_3x_4$

19. What is the largest possible value of the quadratic form $5x_1^2 + 8x_2^2$ if $\mathbf{x} = (x_1, x_2)$ and $\mathbf{x}^T\mathbf{x} = 1$, that is, if $x_1^2 + x_2^2 = 1$? (Try some examples of $\mathbf{x}$.)

20. What is the largest value of the quadratic form $5x_1^2 - 3x_2^2$ if $\mathbf{x}^T\mathbf{x} = 1$?

In Exercises 21–30, matrices are $n \times n$ and vectors are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** The matrix of a quadratic form is a symmetric matrix.

22. **(T/F)** The expression $\|\mathbf{x}\|^2$ is not a quadratic form.

23. **(T/F)** A quadratic form has no cross-product terms if and only if the matrix of the quadratic form is a diagonal matrix.

24. **(T/F)** If A is symmetric and P is an orthogonal matrix, then the change of variable $\mathbf{x} = P\mathbf{y}$ transforms $\mathbf{x}^TA\mathbf{x}$ into a quadratic form with no cross-product term.

25. **(T/F)** The principal axes of a quadratic form $\mathbf{x}^TA\mathbf{x}$ are eigenvectors of A.

26. **(T/F)** If the eigenvalues of a symmetric matrix A are all positive, then the quadratic form $\mathbf{x}^TA\mathbf{x}$ is positive definite.

27. **(T/F)** A positive definite quadratic form Q satisfies $Q(\mathbf{x}) > 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

28. **(T/F)** An indefinite quadratic form is neither positive semidefinite nor negative semidefinite.

29. **(T/F)** A Cholesky factorization of a symmetric matrix A has the form $A = R^TR$, for an upper triangular matrix R with positive diagonal entries.

30. **(T/F)** If A is symmetric and the quadratic form $\mathbf{x}^TA\mathbf{x}$ has only negative values for $\mathbf{x} \neq \mathbf{0}$, then the eigenvalues of A are all positive.

Exercises 31 and 32 show how to classify a quadratic form $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, when $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ and $\det A \neq 0$, without finding the eigenvalues of A.

31. If λ_1 and λ_2 are the eigenvalues of A, then the characteristic polynomial of A can be written in two ways: $\det(A - \lambda I)$ and $(\lambda - \lambda_1)(\lambda - \lambda_2)$. Use this fact to show that $\lambda_1 + \lambda_2 = a + d$ (the diagonal entries of A) and $\lambda_1\lambda_2 = \det A$.

32. Verify the following statements:
    a. Q is positive definite if $\det A > 0$ and $a > 0$.
    b. Q is negative definite if $\det A > 0$ and $a < 0$.
    c. Q is indefinite if $\det A < 0$.

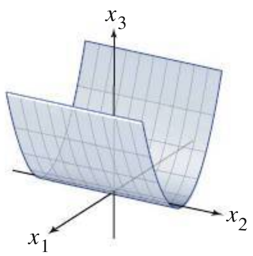
33. Show that if B is $m \times n$, then B^TB is positive semidefinite; and if B is $n \times n$ and invertible, then B^TB is positive definite.

34. Show that if an $n \times n$ matrix A is positive definite, then there exists a positive definite matrix B such that $A = B^TB$. [*Hint:* Write $A = PDP^T$, with $P^T = P^{-1}$. Produce a diagonal matrix C such that $D = C^TC$, and let $B = PCP^T$. Show that B works.]

35. Let A and B be symmetric $n \times n$ matrices whose eigenvalues are all positive. Show that the eigenvalues of $A + B$ are all positive. [*Hint:* Consider quadratic forms.]

36. Let A be an $n \times n$ invertible symmetric matrix. Show that if the quadratic form $\mathbf{x}^TA\mathbf{x}$ is positive definite, then so is the quadratic form $\mathbf{x}^TA^{-1}\mathbf{x}$. [*Hint:* Consider eigenvalues.]

STUDY GUIDE offers additional resources on diagonalization and quadratic forms.



Positive semidefinite

Solution to Practice Problem

Make an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$, and write

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

as in equation (4). If an eigenvalue—say, λ_i—were negative, then $\mathbf{x}^T A\mathbf{x}$ would be negative for the $\mathbf{x}$ corresponding to $\mathbf{y} = \mathbf{e}_i$ (the ith column of I_n). So the eigenvalues of a positive semidefinite quadratic form must all be nonnegative. Conversely, if the eigenvalues are nonnegative, the expansion above shows that $\mathbf{x}^T A\mathbf{x}$ must be positive semidefinite.

7.3 Constrained Optimization

Engineers, economists, scientists, and mathematicians often need to find the maximum or minimum value of a quadratic form $Q(\mathbf{x})$ for $\mathbf{x}$ in some specified set. Typically, the problem can be arranged so that $\mathbf{x}$ varies over the set of unit vectors. This *constrained optimization problem* has an interesting and elegant solution. Example 6 and the discussion in Section 7.5 will illustrate how such problems arise in practice.

The requirement that a vector $\mathbf{x}$ in $\mathbb{R}^n$ be a unit vector can be stated in several equivalent ways:

$$\|\mathbf{x}\| = 1, \qquad \|\mathbf{x}\|^2 = 1, \qquad \mathbf{x}^T\mathbf{x} = 1$$

and

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \tag{1}$$

The expanded version (1) of $\mathbf{x}^T\mathbf{x} = 1$ is commonly used in applications.

When a quadratic form Q has no cross-product terms, it is easy to find the maximum and minimum of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

EXAMPLE 1 Find the maximum and minimum values of $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$.

SOLUTION Since x_2^2 and x_3^2 are nonnegative, note that

$$4x_2^2 \le 9x_2^2 \qquad \text{and} \qquad 3x_3^2 \le 9x_3^2$$

and hence

$$\begin{aligned} Q(\mathbf{x}) &= 9x_1^2 + 4x_2^2 + 3x_3^2 \\ &\le 9x_1^2 + 9x_2^2 + 9x_3^2 \\ &= 9(x_1^2 + x_2^2 + x_3^2) \\ &= 9 \end{aligned}$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. So the maximum value of $Q(\mathbf{x})$ cannot exceed 9 when $\mathbf{x}$ is a unit vector. Furthermore, $Q(\mathbf{x}) = 9$ when $\mathbf{x} = (1, 0, 0)$. Thus 9 is the maximum value of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

To find the minimum value of $Q(\mathbf{x})$, observe that

$$9x_1^2 \ge 3x_1^2, \qquad 4x_2^2 \ge 3x_2^2$$

and hence

$$Q(\mathbf{x}) \ge 3x_1^2 + 3x_2^2 + 3x_3^2 = 3(x_1^2 + x_2^2 + x_3^2) = 3$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. Also, $Q(\mathbf{x}) = 3$ when $x_1 = 0$, $x_2 = 0$, and $x_3 = 1$. So 3 is the minimum value of $Q(\mathbf{x})$ when $\mathbf{x}^T\mathbf{x} = 1$. ■

It is easy to see in Example 1 that the matrix of the quadratic form Q has eigenvalues 9, 4, and 3 and that the greatest and least eigenvalues equal, respectively, the (constrained) maximum and minimum of $Q(\mathbf{x})$. The same holds true for any quadratic form, as we shall see.

EXAMPLE 2 Let $A = \begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}$, and let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ for $\mathbf{x}$ in $\mathbb{R}^2$. Figure 1 displays the graph of Q. Figure 2 shows only the portion of the graph inside a cylinder; the intersection of the cylinder with the surface is the set of points (x_1, x_2, z) such that $z = Q(x_1, x_2)$ and $x_1^2 + x_2^2 = 1$. The "heights" of these points are the constrained values of $Q(\mathbf{x})$. Geometrically, the constrained optimization problem is to locate the highest and lowest points on the intersection curve.

The two highest points on the curve are 7 units above the x_1x_2-plane, occurring where $x_1 = 0$ and $x_2 = \pm 1$. These points correspond to the eigenvalue 7 of A and the eigenvectors $\mathbf{x} = (0, 1)$ and $-\mathbf{x} = (0, -1)$. Similarly, the two lowest points on the curve are 3 units above the x_1x_2-plane. They correspond to the eigenvalue 3 and the eigenvectors $(1, 0)$ and $(-1, 0)$. ■

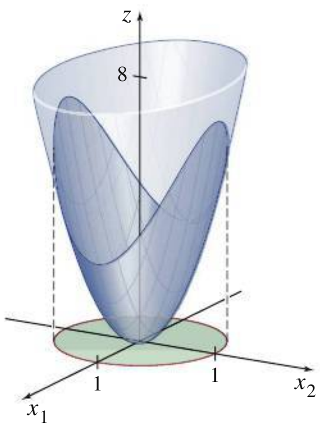


FIGURE 1 $z = 3x_1^2 + 7x_2^2$.

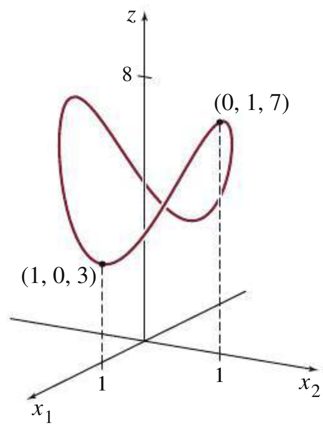


FIGURE 2 The intersection of $z = 3x_1^2 + 7x_2^2$ and the cylinder $x_1^2 + x_2^2 = 1$.

Every point on the intersection curve in Figure 2 has a z-coordinate between 3 and 7, and for any number t between 3 and 7, there is a unit vector $\mathbf{x}$ such that $Q(\mathbf{x}) = t$. In other words, the set of all possible values of $\mathbf{x}^T A\mathbf{x}$, for $\|\mathbf{x}\| = 1$, is the closed interval $3 \le t \le 7$.

It can be shown that for any symmetric matrix A, the set of all possible values of $\mathbf{x}^T A\mathbf{x}$, for $\|\mathbf{x}\| = 1$, is a closed interval on the real axis. (See Exercise 13.) Denote the left and right endpoints of this interval by m and M, respectively. That is, let

$$m = \min \{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\}, \quad M = \max \{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\} \tag{2}$$

Exercise 12 asks you to prove that if λ is an eigenvalue of A, then $m \le \lambda \le M$. The next theorem says that m and M are themselves eigenvalues of A, just as in Example 2.[1]

[1] The use of *minimum* and *maximum* in (2), and *least* and *greatest* in the theorem, refers to the natural ordering of the real numbers, not to magnitudes.

THEOREM 6

Let A be a symmetric matrix, and define m and M as in (2). Then M is the greatest eigenvalue λ_1 of A and m is the least eigenvalue of A. The value of $\mathbf{x}^T A\mathbf{x}$ is M when $\mathbf{x}$ is a unit eigenvector $\mathbf{u}_1$ corresponding to M. The value of $\mathbf{x}^T A\mathbf{x}$ is m when $\mathbf{x}$ is a unit eigenvector corresponding to m.

PROOF Orthogonally diagonalize A as PDP^{-1}. We know that

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} \quad \text{when } \mathbf{x} = P\mathbf{y} \tag{3}$$

Also,

$$\|\mathbf{x}\| = \|P\mathbf{y}\| = \|\mathbf{y}\| \quad \text{for all } \mathbf{y}$$

because $P^T P = I$ and $\|P\mathbf{y}\|^2 = (P\mathbf{y})^T(P\mathbf{y}) = \mathbf{y}^T P^T P\mathbf{y} = \mathbf{y}^T\mathbf{y} = \|\mathbf{y}\|^2$. In particular, $\|\mathbf{y}\| = 1$ if and only if $\|\mathbf{x}\| = 1$. Thus $\mathbf{x}^T A\mathbf{x}$ and $\mathbf{y}^T D\mathbf{y}$ assume the same set of values as $\mathbf{x}$ and $\mathbf{y}$ range over the set of all unit vectors.

To simplify notation, suppose that A is a 3×3 matrix with eigenvalues $a \ge b \ge c$. Arrange the (eigenvector) columns of P so that $P = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,]$ and

$$D = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

Given any unit vector $\mathbf{y}$ in $\mathbb{R}^3$ with coordinates y_1, y_2, y_3, observe that

$$\begin{aligned} ay_1^2 &= ay_1^2 \\ by_2^2 &\le ay_2^2 \\ cy_3^2 &\le ay_3^2 \end{aligned}$$

and obtain these inequalities:

$$\begin{aligned} \mathbf{y}^T D\mathbf{y} &= ay_1^2 + by_2^2 + cy_3^2 \\ &\le ay_1^2 + ay_2^2 + ay_3^2 \\ &= a(y_1^2 + y_2^2 + y_3^2) \\ &= a\|\mathbf{y}\|^2 = a \end{aligned}$$

Thus $M \le a$, by definition of M. However, $\mathbf{y}^T D\mathbf{y} = a$ when $\mathbf{y} = \mathbf{e}_1 = (1, 0, 0)$, so in fact $M = a$. By (3), the $\mathbf{x}$ that corresponds to $\mathbf{y} = \mathbf{e}_1$ is the eigenvector $\mathbf{u}_1$ of A, because

$$\mathbf{x} = P\mathbf{e}_1 = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,]\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{u}_1$$

Thus $M = a = \mathbf{e}_1^T D\mathbf{e}_1 = \mathbf{u}_1^T A\mathbf{u}_1$, which proves the statement about M. A similar argument shows that m is the least eigenvalue, c, and this value of $\mathbf{x}^T A\mathbf{x}$ is attained when $\mathbf{x} = P\mathbf{e}_3 = \mathbf{u}_3$. ■

EXAMPLE 3 Let $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$. Find the maximum value of the quadratic form $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$, and find a unit vector at which this maximum value is attained.

SOLUTION By Theorem 6, the desired maximum value is the greatest eigenvalue of A. The characteristic equation turns out to be

$$0 = -\lambda^3 + 10\lambda^2 - 27\lambda + 18 = -(\lambda - 6)(\lambda - 3)(\lambda - 1)$$

The greatest eigenvalue is 6.

The constrained maximum of $\mathbf{x}^T A\mathbf{x}$ is attained when $\mathbf{x}$ is a unit eigenvector for $\lambda = 6$. Solve $(A - 6I)\mathbf{x} = 0$ and find an eigenvector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Set $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$. ■

In Theorem 7 and in later applications, the values of $\mathbf{x}^T A\mathbf{x}$ are computed with additional constraints on the unit vector $\mathbf{x}$.

THEOREM 7

Let A, λ_1, and $\mathbf{u}_1$ be as in Theorem 6. Then the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to the constraints

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0$$

is the second greatest eigenvalue, λ_2, and this maximum is attained when $\mathbf{x}$ is an eigenvector $\mathbf{u}_2$ corresponding to λ_2.

Theorem 7 can be proved by an argument similar to the one above in which the theorem is reduced to the case where the matrix of the quadratic form is diagonal. The next example gives an idea of the proof for the case of a diagonal matrix.

EXAMPLE 4 Find the maximum value of $9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraints $\mathbf{x}^T\mathbf{x} = 1$ and $\mathbf{x}^T\mathbf{u}_1 = 0$, where $\mathbf{u}_1 = (1, 0, 0)$. Note that $\mathbf{u}_1$ is a unit eigenvector corresponding to the greatest eigenvalue $\lambda = 9$ of the matrix of the quadratic form.

SOLUTION If the coordinates of $\mathbf{x}$ are x_1, x_2, x_3, then the constraint $\mathbf{x}^T\mathbf{u}_1 = 0$ means simply that $x_1 = 0$. For such a unit vector, $x_2^2 + x_3^2 = 1$, and

$$\begin{aligned} 9x_1^2 + 4x_2^2 + 3x_3^2 &= 4x_2^2 + 3x_3^2 \\ &\le 4x_2^2 + 4x_3^2 \\ &= 4(x_2^2 + x_3^2) \\ &= 4 \end{aligned}$$

Thus the constrained maximum of the quadratic form does not exceed 4. And this value is attained for $\mathbf{x} = (0, 1, 0)$, which is an eigenvector for the second greatest eigenvalue of the matrix of the quadratic form. ■

EXAMPLE 5 Let A be the matrix in Example 3 and let $\mathbf{u}_1$ be a unit eigenvector corresponding to the greatest eigenvalue of A. Find the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to the conditions

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0 \tag{4}$$

SOLUTION From Example 3, the second greatest eigenvalue of A is $\lambda = 3$. Solve $(A - 3I)\mathbf{x} = \mathbf{0}$ to find an eigenvector, and normalize it to obtain

$$\mathbf{u}_2 = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$$

The vector $\mathbf{u}_2$ is automatically orthogonal to $\mathbf{u}_1$ because the vectors correspond to different eigenvalues. Thus the maximum of $\mathbf{x}^T A\mathbf{x}$ subject to the constraints in (4) is 3, attained when $\mathbf{x} = \mathbf{u}_2$. ■

The next theorem generalizes Theorem 7 and, together with Theorem 6, gives a useful characterization of *all* the eigenvalues of A. The proof is omitted.

THEOREM 8

Let A be a symmetric $n \times n$ matrix with an orthogonal diagonalization $A = PDP^{-1}$, where the entries on the diagonal of D are arranged so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and where the columns of P are corresponding unit eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Then for $k = 2, \ldots, n$, the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to the constraints

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0, \quad \ldots, \quad \mathbf{x}^T\mathbf{u}_{k-1} = 0$$

is the eigenvalue λ_k, and this maximum is attained at $\mathbf{x} = \mathbf{u}_k$.

Theorem 8 will be helpful in Sections 7.4 and 7.5. The following application requires only Theorem 6.

EXAMPLE 6 During the next year, a county government is planning to repair x hundred miles of public roads and bridges and to improve y hundred acres of parks and recreation areas. The county must decide how to allocate its resources (funds, equipment, labor, etc.) between these two projects. If it is more cost effective to work simultaneously on both projects rather than on only one, then x and y might satisfy a *constraint* such as

$$4x^2 + 9y^2 \leq 36$$

See Figure 3. Each point (x, y) in the shaded *feasible set* represents a possible public works schedule for the year. The points on the constraint curve, $4x^2 + 9y^2 = 36$, use the maximum amounts of resources available.

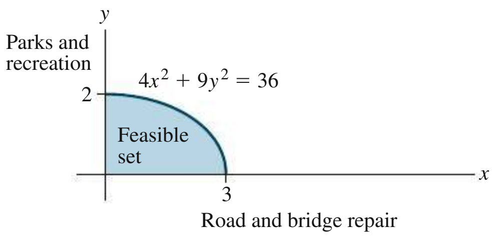


FIGURE 3 Public works schedules.

In choosing its public works schedule, the county wants to consider the opinions of the county residents. To measure the value, or *utility*, that the residents would assign to the various work schedules (x, y), economists sometimes use a function such as

$$q(x, y) = xy$$

The set of points (x, y) at which $q(x, y)$ is a constant is called an *indifference curve*. Three such curves are shown in Figure 4. Points along an indifference curve correspond to alternatives that county residents as a group would find equally valuable.[2] Find the public works schedule that maximizes the utility function q.

[2] Indifference curves are discussed in Michael D. Intriligator, Ronald G. Bodkin, and Cheng Hsiao, *Econometric Models, Techniques, and Applications* (Upper Saddle River, NJ: Prentice Hall, 1996).

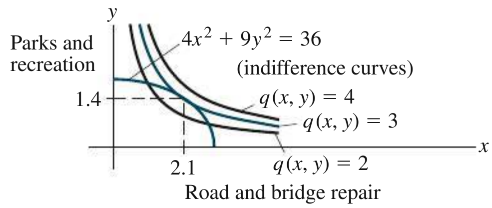


FIGURE 4 The optimum public works schedule is (2.1, 1.4).

SOLUTION The constraint equation $4x^2 + 9y^2 = 36$ does not describe a set of unit vectors, but a change of variable can fix that problem. Rewrite the constraint in the form

$$\left(\frac{x}{3}\right)^2 + \left(\frac{y}{2}\right)^2 = 1$$

and define

$$x_1 = \frac{x}{3}, \quad x_2 = \frac{y}{2}, \qquad \text{that is,} \quad x = 3x_1 \quad \text{and} \quad y = 2x_2$$

Then the constraint equation becomes

$$x_1^2 + x_2^2 = 1$$

and the utility function becomes $q(3x_1, 2x_2) = (3x_1)(2x_2) = 6x_1x_2$. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Then the problem is to maximize $Q(\mathbf{x}) = 6x_1x_2$ subject to $\mathbf{x}^T\mathbf{x} = 1$. Note that $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, where

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

The eigenvalues of A are ± 3, with eigenvectors $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ for $\lambda = 3$ and $\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ for $\lambda = -3$. Thus the maximum value of $Q(\mathbf{x}) = q(x_1, x_2)$ is 3, attained when $x_1 = 1/\sqrt{2}$ and $x_2 = 1/\sqrt{2}$.

In terms of the original variables, the optimum public works schedule is $x = 3x_1 = 3/\sqrt{2} \approx 2.1$ hundred miles of roads and bridges and $y = 2x_2 = \sqrt{2} \approx 1.4$ hundred acres of parks and recreational areas. The optimum public works schedule is the point where the constraint curve and the indifference curve $q(x, y) = 3$ just meet. Points (x, y) with a higher utility lie on indifference curves that do not touch the constraint curve. See Figure 4. ■

Practice Problems

1. Let $Q(\mathbf{x}) = 3x_1^2 + 3x_2^2 + 2x_1x_2$. Find a change of variable that transforms Q into a quadratic form with no cross-product term, and give the new quadratic form.
2. With Q as in Problem 1, find the maximum value of $Q(\mathbf{x})$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$, and find a unit vector at which the maximum is attained.

7.3 Exercises

In Exercises 1 and 2, find the change of variable $\mathbf{x} = P\mathbf{y}$ that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into $\mathbf{y}^T D\mathbf{y}$ as shown.

1. $5x_1^2+6x_2^2+7x_3^2+4x_1x_2-4x_2x_3=9y_1^2+6y_2^2+3y_3^2$

2. $3x_1^2+3x_2^2+5x_3^2+6x_1x_2+2x_1x_3+2x_2x_3=7y_1^2+4y_2^2$

Hint: $\mathbf{x}$ and $\mathbf{y}$ must have the same number of coordinates, so the quadratic form shown here must have a coefficient of zero for y_3^2.

In Exercises 3–6, find (a) the maximum value of $Q(\mathbf{x})$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$, (b) a unit vector $\mathbf{u}$ where this maximum is attained, and (c) the maximum of $Q(\mathbf{x})$ subject to the constraints $\mathbf{x}^T\mathbf{x} = 1$ and $\mathbf{x}^T\mathbf{u} = 0$.

3. $Q(\mathbf{x}) = 5x_1^2 + 6x_2^2 + 7x_3^2 + 4x_1x_2 - 4x_2x_3$ (See Exercise 1.)

4. $Q(\mathbf{x}) = 3x_1^2+3x_2^2+5x_3^2+6x_1x_2+2x_1x_3+2x_2x_3$ (See Exercise 2.)

5. $Q(\mathbf{x}) = x_1^2 + x_2^2 - 10x_1x_2$

6. $Q(\mathbf{x}) = 3x_1^2 + 9x_2^2 + 8x_1x_2$

7. Let $Q(\mathbf{x}) = -2x_1^2 - x_2^2 + 4x_1x_2 + 4x_2x_3$. Find a unit vector $\mathbf{x}$ in $\mathbb{R}^3$ at which $Q(\mathbf{x})$ is maximized, subject to $\mathbf{x}^T\mathbf{x} = 1$. [*Hint:* The eigenvalues of the matrix of the quadratic form Q are 2, -1, and -4.]

8. Let $Q(\mathbf{x}) = 7x_1^2 + x_2^2 + 7x_3^2 - 8x_1x_2 - 4x_1x_3 - 8x_2x_3$. Find a unit vector $\mathbf{x}$ in $\mathbb{R}^3$ at which $Q(\mathbf{x})$ is maximized, subject to $\mathbf{x}^T\mathbf{x} = 1$. [*Hint:* The eigenvalues of the matrix of the quadratic form Q are 9 and -3.]

9. Find the maximum value of $Q(\mathbf{x}) = 7x_1^2 + 3x_2^2 - 2x_1x_2$, subject to the constraint $x_1^2 + x_2^2 = 1$. (Do not go on to find a vector where the maximum is attained.)

10. Find the maximum value of $Q(\mathbf{x}) = -3x_1^2 + 5x_2^2 - 2x_1x_2$, subject to the constraint $x_1^2 + x_2^2 = 1$. (Do not go on to find a vector where the maximum is attained.)

11. Suppose $\mathbf{x}$ is a unit eigenvector of a matrix A corresponding to an eigenvalue 3. What is the value of $\mathbf{x}^T A\mathbf{x}$?

12. Let λ be any eigenvalue of a symmetric matrix A. Justify the statement made in this section that $m \le \lambda \le M$, where m and M are defined as in (2). [*Hint:* Find an $\mathbf{x}$ such that $\lambda = \mathbf{x}^T A\mathbf{x}$.]

13. Let A be an $n \times n$ symmetric matrix, let M and m denote the maximum and minimum values of the quadratic form $\mathbf{x}^T A\mathbf{x}$, where $\mathbf{x}^T\mathbf{x} = 1$, and denote corresponding unit eigenvectors by $\mathbf{u}_1$ and $\mathbf{u}_n$. The following calculations show that given any number t between M and m, there is a unit vector $\mathbf{x}$ such that $t = \mathbf{x}^T A\mathbf{x}$. Verify that $t = (1-\alpha)m + \alpha M$ for some number α between 0 and 1. Then let $\mathbf{x} = \sqrt{1-\alpha}\,\mathbf{u}_n + \sqrt{\alpha}\,\mathbf{u}_1$, and show that $\mathbf{x}^T\mathbf{x} = 1$ and $\mathbf{x}^T A\mathbf{x} = t$.

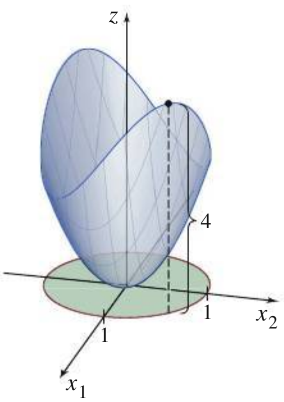
In Exercises 14–17, follow the instructions given for Exercises 3–6.

T **14.** $3x_1x_2 + 5x_1x_3 + 7x_1x_4 + 7x_2x_3 + 5x_2x_4 + 3x_3x_4$

T **15.** $4x_1^2-6x_1x_2-10x_1x_3-10x_1x_4-6x_2x_3-6x_2x_4-2x_3x_4$

T **16.** $-6x_1^2-10x_2^2-13x_3^2-13x_4^2-4x_1x_2-4x_1x_3-4x_1x_4+6x_3x_4$

T **17.** $x_1x_2 + 3x_1x_3 + 30x_1x_4 + 30x_2x_3 + 3x_2x_4 + x_3x_4$



The maximum value of $Q(\mathbf{x})$ subject to $\mathbf{x}^T\mathbf{x} = 1$ is 4.

Solutions to Practice Problems

1. The matrix of the quadratic form is $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. It is easy to find the eigenvalues, 4 and 2, and corresponding unit eigenvectors, $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. So the desired change of variable is $\mathbf{x} = P\mathbf{y}$, where $P = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$. (A common error here is to forget to normalize the eigenvectors.) The new quadratic form is $\mathbf{y}^T D\mathbf{y} = 4y_1^2 + 2y_2^2$.

2. The maximum of $Q(\mathbf{x})$, for a unit vector $\mathbf{x}$, is 4 and the maximum is attained at the unit eigenvector $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. [A common incorrect answer is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. This vector maximizes the quadratic form $\mathbf{y}^T D\mathbf{y}$ instead of $Q(\mathbf{x})$.]

7.4 The Singular Value Decomposition

The diagonalization theorems in Sections 5.3 and 7.1 play a part in many interesting applications. Unfortunately, as we know, not all matrices can be factored as $A = PDP^{-1}$ with D diagonal. However, a factorization $A = QDP^{-1}$ *is* possible for any $m \times n$ matrix A! A special factorization of this type, called the *singular value decomposition*, is one of the most useful matrix factorizations in applied linear algebra.

The singular value decomposition is based on the following property of the ordinary diagonalization that can be imitated for rectangular matrices: The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretches or shrinks certain vectors (the eigenvectors). If $A\mathbf{x} = \lambda\mathbf{x}$ and $\|\mathbf{x}\| = 1$, then

$$\|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\| = |\lambda| \tag{1}$$

If λ_1 is the eigenvalue with the greatest magnitude, then a corresponding unit eigenvector $\mathbf{v}_1$ identifies a direction in which the stretching effect of A is greatest. That is, the length of $A\mathbf{x}$ is maximized when $\mathbf{x} = \mathbf{v}_1$, and $\|A\mathbf{v}_1\| = |\lambda_1|$, by (1). This description of $\mathbf{v}_1$ and $|\lambda_1|$ has an analogue for rectangular matrices that will lead to the singular value decomposition.

EXAMPLE 1 If $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$, then the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps the unit sphere $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$ in $\mathbb{R}^3$ onto an ellipse in $\mathbb{R}^2$, shown in Figure 1. Find a unit vector $\mathbf{x}$ at which the length $\|A\mathbf{x}\|$ is maximized, and compute this maximum length.

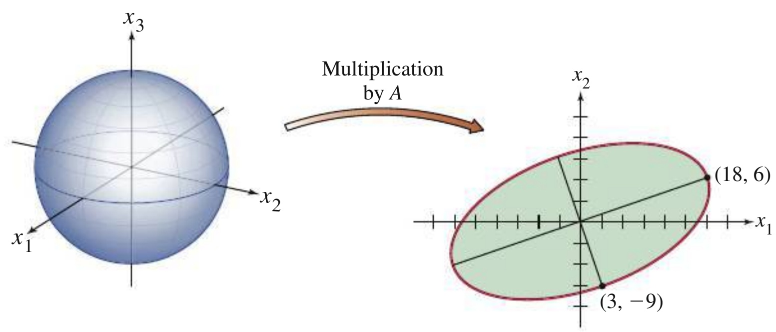


FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

SOLUTION The quantity $\|A\mathbf{x}\|^2$ is maximized at the same $\mathbf{x}$ that maximizes $\|A\mathbf{x}\|$, and $\|A\mathbf{x}\|^2$ is easier to study. Observe that

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T(A^T A)\mathbf{x}$$

Also, $A^T A$ is a symmetric matrix, since $(A^T A)^T = A^T A^{TT} = A^T A$. So the problem now is to maximize the quadratic form $\mathbf{x}^T(A^T A)\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$. By Theorem 6 in Section 7.3, the maximum value is the greatest eigenvalue λ_1 of $A^T A$. Also, the maximum value is attained at a unit eigenvector of $A^T A$ corresponding to λ_1.

For the matrix A in this example,

$$A^T A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

The eigenvalues of A^TA are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$. Corresponding unit eigenvectors are, respectively,

$$\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

The maximum value of $\|A\mathbf{x}\|^2$ is 360, attained when $\mathbf{x}$ is the unit vector $\mathbf{v}_1$. The vector $A\mathbf{v}_1$ is a point on the ellipse in Figure 1 farthest from the origin, namely

$$A\mathbf{v}_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}$$

For $\|\mathbf{x}\| = 1$, the maximum value of $\|A\mathbf{x}\|$ is $\|A\mathbf{v}_1\| = \sqrt{360} = 6\sqrt{10}$. ■

Example 1 suggests that the effect of A on the unit sphere in $\mathbb{R}^3$ is related to the quadratic form $\mathbf{x}^T(A^TA)\mathbf{x}$. In fact, the entire geometric behavior of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is captured by this quadratic form, as we shall see.

The Singular Values of an $m \times n$ Matrix

Let A be an $m \times n$ matrix. Then A^TA is symmetric and can be orthogonally diagonalized. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA, and let $\lambda_1, \dots, \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \le i \le n$,

$$\begin{aligned} \|A\mathbf{v}_i\|^2 &= (A\mathbf{v}_i)^TA\mathbf{v}_i = \mathbf{v}_i^T A^TA\mathbf{v}_i \\ &= \mathbf{v}_i^T(\lambda_i\mathbf{v}_i) \qquad \text{Since } \mathbf{v}_i \text{ is an eigenvector of } A^TA \\ &= \lambda_i \qquad \text{Since } \mathbf{v}_i \text{ is a unit vector} \end{aligned} \tag{2}$$

So the eigenvalues of A^TA are all nonnegative. By renumbering, if necessary, we may assume that the eigenvalues are arranged so that

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$$

The **singular values** of A are the square roots of the eigenvalues of A^TA, denoted by $\sigma_1, \dots, \sigma_n$, and they are arranged in decreasing order. That is, $\sigma_i = \sqrt{\lambda_i}$ for $1 \le i \le n$. By equation (2), *the singular values of A are the lengths of the vectors $A\mathbf{v}_1, \dots, A\mathbf{v}_n$.*

EXAMPLE 2 Let A be the matrix in Example 1. Since the eigenvalues of A^TA are 360, 90, and 0, the singular values of A are

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \quad \sigma_2 = \sqrt{90} = 3\sqrt{10}, \quad \sigma_3 = 0$$

From Example 1, the first singular value of A is the maximum of $\|A\mathbf{x}\|$ over all unit vectors, and the maximum is attained at the unit eigenvector $\mathbf{v}_1$. Theorem 7 in Section 7.3 shows that the second singular value of A is the maximum of $\|A\mathbf{x}\|$ over all unit vectors that are *orthogonal to* $\mathbf{v}_1$, and this maximum is attained at the second unit eigenvector, $\mathbf{v}_2$ (Exercise 22). For the $\mathbf{v}_2$ in Example 1,

$$A\mathbf{v}_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}$$

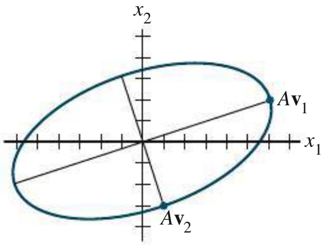


FIGURE 2

This point is on the minor axis of the ellipse in Figure 1, just as $A\mathbf{v}_1$ is on the major axis. (See Figure 2.) The first two singular values of A are the lengths of the major and minor semiaxes of the ellipse. ■

The fact that $A\mathbf{v}_1$ and $A\mathbf{v}_2$ are orthogonal in Figure 2 is no accident, as the next theorem shows.

THEOREM 9

Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA, arranged so that the corresponding eigenvalues of A^TA satisfy $\lambda_1 \geq \cdots \geq \lambda_n$, and suppose A has r nonzero singular values. Then $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthogonal basis for Col A, and rank $A = r$.

PROOF Because $\mathbf{v}_i$ and $\lambda_j \mathbf{v}_j$ are orthogonal for $i \neq j$,

$$(A\mathbf{v}_i)^T(A\mathbf{v}_j) = \mathbf{v}_i^T A^T A\mathbf{v}_j = \mathbf{v}_i^T(\lambda_j \mathbf{v}_j) = 0$$

Thus $\{A\mathbf{v}_1, \dots, A\mathbf{v}_n\}$ is an orthogonal set. Furthermore, since the lengths of the vectors $A\mathbf{v}_1, \dots, A\mathbf{v}_n$ are the singular values of A, and since there are r nonzero singular values, $A\mathbf{v}_i \neq \mathbf{0}$ if and only if $1 \leq i \leq r$. So $A\mathbf{v}_1, \dots, A\mathbf{v}_r$ are linearly independent vectors, and they are in Col A. Finally, for any $\mathbf{y}$ in Col A—say, $\mathbf{y} = A\mathbf{x}$—we can write $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, and

$$\begin{aligned} \mathbf{y} = A\mathbf{x} &= c_1 A\mathbf{v}_1 + \cdots + c_r A\mathbf{v}_r + c_{r+1}A\mathbf{v}_{r+1} + \cdots + c_n A\mathbf{v}_n \\ &= c_1 A\mathbf{v}_1 + \cdots + c_r A\mathbf{v}_r + 0 + \cdots + 0 \end{aligned}$$

Thus $\mathbf{y}$ is in Span$\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$, which shows that $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthogonal basis for Col A. Hence rank $A = \dim \text{Col}\, A = r$. ■

Numerical Notes

In some cases, the rank of A may be very sensitive to small changes in the entries of A. The obvious method of counting the number of pivot columns in A does not work well if A is row reduced by a computer. Roundoff error often creates an echelon form with full rank.

In practice, the most reliable way to estimate the rank of a large matrix A is to count the number of nonzero singular values. In this case, extremely small nonzero singular values are assumed to be zero for all practical purposes, and the *effective rank* of the matrix is the number obtained by counting the remaining nonzero singular values.[1]

[1] In general, rank estimation is not a simple problem. For a discussion of the subtle issues involved, see Philip E. Gill, Walter Murray, and Margaret H. Wright, *Numerical Linear Algebra and Optimization*, vol. 1 (Redwood City, CA: Addison-Wesley, 1991), Sec. 5.8.

The Singular Value Decomposition

The decomposition of A involves an $m \times n$ "diagonal" matrix Σ of the form

$$\Sigma = \begin{array}{ll} \left[\begin{array}{cc} D & 0 \\ 0 & 0 \end{array}\right] & \begin{array}{l} \\ \leftarrow m - r \text{ rows} \end{array} \\ \begin{array}{ll} & \uparrow\, n - r \text{ columns} \end{array} & \end{array} \tag{3}$$

where D is an $r \times r$ diagonal matrix for some r not exceeding the smaller of m and n. (If r equals m or n or both, some or all of the zero matrices do not appear.)

THEOREM 10

The Singular Value Decomposition

Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix Σ as in (3) for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

$$A = U \Sigma V^T$$

Any factorization $A = U\Sigma V^T$, with U and V orthogonal, Σ as in (3), and positive diagonal entries in D, is called a **singular value decomposition** (or **SVD**) of A. The matrices U and V are not uniquely determined by A, but the diagonal entries of Σ are necessarily the singular values of A. See Exercise 19. The columns of U in such a decomposition are called **left singular vectors** of A, and the columns of V are called **right singular vectors** of A.

PROOF Let λ_i and $\mathbf{v}_i$ be as in Theorem 9, so that $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthogonal basis for Col A. Normalize each $A\mathbf{v}_i$ to obtain an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, where

$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

and

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \qquad (1 \leq i \leq r) \tag{4}$$

Now extend $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ of $\mathbb{R}^m$, and let

$$U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m\,] \quad \text{and} \quad V = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,]$$

By construction, U and V are orthogonal matrices. Also, from (4),

$$AV = [\,A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\,] = [\,\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\,]$$

Let D be the diagonal matrix with diagonal entries $\sigma_1, \dots, \sigma_r$, and let Σ be as in (3) above. Then

$$U\Sigma = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m\,]\left[\begin{array}{cccc|c} \sigma_1 & & & 0 & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ 0 & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right]$$

$$\begin{aligned} &= [\,\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\,] \\ &= AV \end{aligned}$$

Since V is an orthogonal matrix, $U\Sigma V^T = AVV^T = A$. ■

The next two examples focus attention on the internal structure of a singular value decomposition. An efficient and numerically stable algorithm for this decomposition would use a different approach. See the Numerical Note at the end of the section.

EXAMPLE 3 Use the results of Examples 1 and 2 to construct a singular value decomposition of $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$.

SOLUTION A construction can be divided into three steps.

STUDY GUIDE offers additional resources for learning to compute an SVD.

Step 1. Find an orthogonal diagonalization of A^TA. That is, find the eigenvalues of A^TA and a corresponding orthonormal set of eigenvectors. If A had only two columns, the calculations could be done by hand. Larger matrices usually require a matrix program. However, for the matrix A here, the eigendata for A^TA are provided in Example 1.

Step 2. Set up V and Σ. Arrange the eigenvalues of A^TA in decreasing order. In Example 1, the eigenvalues are already listed in decreasing order: 360, 90, and 0. The corresponding unit eigenvectors, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, are the right singular vectors of A. Using Example 1, construct

$$V = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,] = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$$

The square roots of the eigenvalues are the singular values:

$$\sigma_1 = 6\sqrt{10}, \quad \sigma_2 = 3\sqrt{10}, \quad \sigma_3 = 0$$

The nonzero singular values are the diagonal entries of D. The matrix Σ is the same size as A, with D in its upper left corner and with 0's elsewhere.

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}, \qquad \Sigma = [\,D \quad 0\,] = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

Step 3. Construct U. When A has rank r, the first r columns of U are the normalized vectors obtained from $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$. In this example, A has two nonzero singular values, so rank $A = 2$. Recall from equation (2) and the paragraph before Example 2 that $\|A\mathbf{v}_1\| = \sigma_1$ and $\|A\mathbf{v}_2\| = \sigma_2$. Thus

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{6\sqrt{10}}\begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{3\sqrt{10}}\begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$$

Note that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is already a basis for $\mathbb{R}^2$. Thus no additional vectors are needed for U, and $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2\,]$. The singular value decomposition of A is

$$A = \underset{\substack{\uparrow \\ U}}{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}} \underset{\substack{\uparrow \\ \Sigma}}{\begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}} \underset{\substack{\uparrow \\ V^T}}{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}} \qquad \blacksquare$$

EXAMPLE 4 Find a singular value decomposition of $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$.

SOLUTION First, compute $A^TA = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$. The eigenvalues of A^TA are 18 and 0, with corresponding unit eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

These unit vectors form the columns of V:

$$V = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

The singular values are $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$. Since there is only one nonzero singular value, the "matrix" D may be written as a single number. That is, $D = 3\sqrt{2}$. The matrix Σ is the same size as A, with D in its upper left corner:

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

To construct U, first construct $A\mathbf{v}_1$ and $A\mathbf{v}_2$:

$$A\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad A\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

As a check on the calculations, verify that $\|A\mathbf{v}_1\| = \sigma_1 = 3\sqrt{2}$. Of course, $A\mathbf{v}_2 = \mathbf{0}$ because $\|A\mathbf{v}_2\| = \sigma_2 = 0$. The only column found for U so far is

$$\mathbf{u}_1 = \frac{1}{3\sqrt{2}} A\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

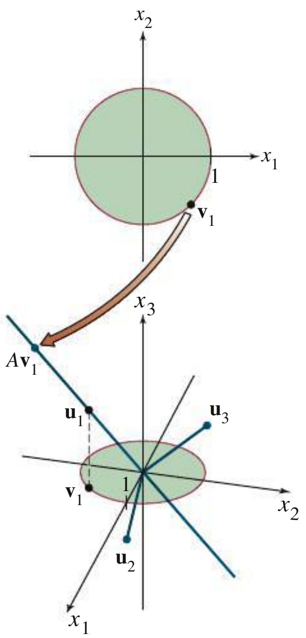


FIGURE 3

The other columns of U are found by extending the set $\{\mathbf{u}_1\}$ to an orthonormal basis for $\mathbb{R}^3$. In this case, we need two orthogonal unit vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ that are orthogonal to $\mathbf{u}_1$. (See Figure 3.) Each vector must satisfy $\mathbf{u}_1^T\mathbf{x} = 0$, which is equivalent to the equation $x_1 - 2x_2 + 2x_3 = 0$. A basis for the solution set of this equation is

$$\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

(Check that $\mathbf{w}_1$ and $\mathbf{w}_2$ are each orthogonal to $\mathbf{u}_1$.) Apply the Gram–Schmidt process (with normalizations) to $\{\mathbf{w}_1, \mathbf{w}_2\}$, and obtain

$$\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

Finally, set $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,]$, take Σ and V^T from above, and write

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

■

Applications of the Singular Value Decomposition

The SVD is often used to estimate the rank of a matrix, as noted above. Several other numerical applications are described briefly below, and an application to image processing is presented in Section 7.5.

EXAMPLE 5 (The Condition Number) Most numerical calculations involving an equation $A\mathbf{x} = \mathbf{b}$ are as reliable as possible when the SVD of A is used. The two orthogonal matrices U and V do not affect lengths of vectors or angles between vectors (Theorem 7 in Section 6.2). Any possible instabilities in numerical calculations are identified in Σ. If the singular values of A are extremely large or small, roundoff errors are almost inevitable, but an error analysis is aided by knowing the entries in Σ and V.

If A is an invertible $n \times n$ matrix, then the ratio σ_1/σ_n of the largest and smallest singular values gives the **condition number** of A. Exercises 50–52 in Section 2.3 showed how the condition number affects the sensitivity of a solution of $A\mathbf{x} = \mathbf{b}$ to changes (or errors) in the entries of A. (Actually, a "condition number" of A can be computed in several ways, but the definition given here is widely used for studying $A\mathbf{x} = \mathbf{b}$.) ■

EXAMPLE 6 (Bases for Fundamental Subspaces) Given an SVD for an $m \times n$ matrix A, let $\mathbf{u}_1, \dots, \mathbf{u}_m$ be the left singular vectors, $\mathbf{v}_1, \dots, \mathbf{v}_n$ the right singular vectors, and $\sigma_1, \dots, \sigma_n$ the singular values, and let r be the rank of A. By Theorem 9,

$$\{\mathbf{u}_1, \dots, \mathbf{u}_r\} \tag{5}$$

is an orthonormal basis for Col A.

Recall from Theorem 3 in Section 6.1 that $(\text{Col } A)^{\perp} = \text{Nul } A^T$. Hence

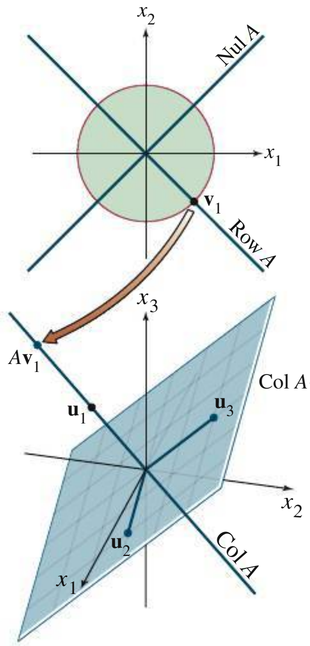
$$\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\} \tag{6}$$

is an orthonormal basis for Nul A^T.

Since $\|A\mathbf{v}_i\| = \sigma_i$ for $1 \le i \le n$, and σ_i is 0 if and only if $i > r$, the vectors $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ span a subspace of Nul A of dimension $n - r$. By the Rank Theorem, dim Nul $A = n - \text{rank } A$. It follows that

$$\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\} \tag{7}$$

is an orthonormal basis for Nul A, by the Basis Theorem (in Section 4.5).



The fundamental subspaces in Example 4.

From (5) and (6), the orthogonal complement of Nul A^T is Col A. Interchanging A and A^T, note that $(\text{Nul } A)^{\perp} = \text{Col } A^T = \text{Row } A$. Hence, from (7),

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \tag{8}$$

is an orthonormal basis for Row A.

Figure 4 summarizes (5)–(8), but shows the orthogonal basis $\{\sigma_1\mathbf{u}_1, \ldots, \sigma_r\mathbf{u}_r\}$ for Col A instead of the normalized basis, to remind you that $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$ for $1 \le i \le r$. Explicit orthonormal bases for the four fundamental subspaces determined by A are useful in some calculations, particularly in constrained optimization problems. ■

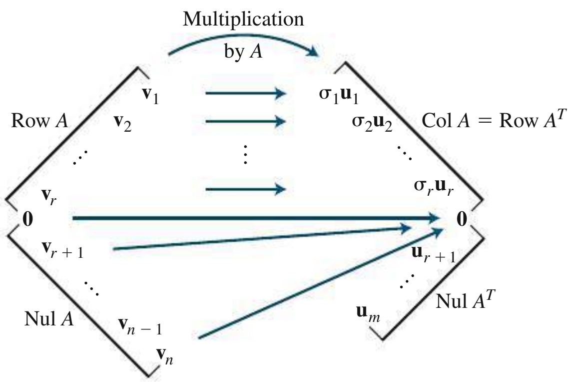


FIGURE 4 The four fundamental subspaces and the action of A.

The four fundamental subspaces and the concept of singular values provide the final statements of the Invertible Matrix Theorem. (Recall that statements about A^T have been omitted from the theorem to avoid nearly doubling the number of statements.) The other statements were given in Sections 2.3, 2.9, 3.2, 4.5, and 5.2.

THEOREM

The Invertible Matrix Theorem (concluded)

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix:

s. $(\text{Col } A)^{\perp} = \{\mathbf{0}\}$.

t. $(\text{Nul } A)^{\perp} = \mathbb{R}^n$.

u. Row $A = \mathbb{R}^n$.

v. A has n nonzero singular values.

EXAMPLE 7 (Reduced SVD and the Pseudoinverse of A) When Σ contains rows or columns of zeros, a more compact decomposition of A is possible. Using the notation established above, let $r = \text{rank } A$, and partition U and V into submatrices whose first blocks contain r columns:

$$U = [\, U_r \quad U_{m-r} \,], \quad \text{where } U_r = [\, \mathbf{u}_1 \quad \cdots \quad \mathbf{u}_r \,]$$
$$V = [\, V_r \quad V_{n-r} \,], \quad \text{where } V_r = [\, \mathbf{v}_1 \quad \cdots \quad \mathbf{v}_r \,]$$

Then U_r is $m \times r$ and V_r is $n \times r$. (To simplify notation, we consider U_{m-r} or V_{n-r} even though one of them may have no columns.) Then partitioned matrix multiplication shows that

$$A = [\,U_r \quad U_{m-r}\,]\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T \tag{9}$$

This factorization of A is called a **reduced singular value decomposition** of A. Since the diagonal entries in D are nonzero, D is invertible. The following matrix is called the **pseudoinverse** (also, the **Moore–Penrose inverse**) of A:

$$A^+ = V_r D^{-1} U_r^T \tag{10}$$

Supplementary Exercises 28–30 at the end of the chapter explore some of the properties of the reduced singular value decomposition and the pseudoinverse. ■

EXAMPLE 8 (Least-Squares Solution) Given the equation $A\mathbf{x} = \mathbf{b}$, use the pseudoinverse of A in (10) to define

$$\hat{\mathbf{x}} = A^+\mathbf{b} = V_r D^{-1} U_r^T \mathbf{b}$$

Then, from the SVD in (9),

$$\begin{aligned} A\hat{\mathbf{x}} &= (U_r D V_r^T)(V_r D^{-1} U_r^T \mathbf{b}) \\ &= U_r D D^{-1} U_r^T \mathbf{b} \qquad \text{Because } V_r^T V_r = I_r \\ &= U_r U_r^T \mathbf{b} \end{aligned}$$

It follows from (5) that $U_r U_r^T \mathbf{b}$ is the orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto Col A. (See Theorem 10 in Section 6.3.) Thus $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$. In fact, this $\hat{\mathbf{x}}$ has the smallest length among all least-squares solutions of $A\mathbf{x} = \mathbf{b}$. See Supplementary Exercise 30. ■

Numerical Notes

Examples 1–4 and the exercises illustrate the concept of singular values and suggest how to perform calculations by hand. In practice, the computation of A^TA should be avoided, since any errors in the entries of A are squared in the entries of A^TA. There exist fast iterative methods that produce the singular values and singular vectors of A accurately to many decimal places.

Practice Problems

1. Given a singular value decomposition, $A = U\Sigma V^T$, find an SVD of A^T. How are the singular values of A and A^T related?
2. For any $n \times n$ matrix A, use the SVD to show that there is an $n \times n$ orthogonal matrix Q such that $A^TA = Q^T(A^TA)Q$.

Remark: Practice Problem 2 establishes that for any $n \times n$ matrix A, the matrices AA^T and A^TA are *orthogonally similar*.

7.4 Exercises

Find the singular values of the matrices in Exercises 1–4.

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

3. $\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$ **4.** $\begin{bmatrix} 3 & 0 \\ 8 & 3 \end{bmatrix}$

Find an SVD of each matrix in Exercises 5–12. [*Hint:* In Exercise 11, one choice for U is $\begin{bmatrix} -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \end{bmatrix}$. In Exercise 12, one column of U can be $\begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$.]

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

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

9. $\begin{bmatrix} 3 & -3 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$ **10.** $\begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix}$

11. $\begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}$ **12.** $\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}$

13. Find the SVD of $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$ [*Hint:* Work with A^T.]

14. In Exercise 7, find a unit vector $\mathbf{x}$ at which $A\mathbf{x}$ has maximum length.

15. Suppose the factorization below is an SVD of a matrix A, with the entries in U and V rounded to two decimal places.

$$A = \begin{bmatrix} .40 & -.78 & .47 \\ .37 & -.33 & -.87 \\ -.84 & -.52 & -.16 \end{bmatrix} \begin{bmatrix} 7.10 & 0 & 0 \\ 0 & 3.10 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} .30 & -.51 & -.81 \\ .76 & .64 & -.12 \\ .58 & -.58 & .58 \end{bmatrix}$$

a. What is the rank of A?

b. Use this decomposition of A, with no calculations, to write a basis for Col A and a basis for Nul A. [*Hint:* First write the columns of V.]

16. Repeat Exercise 15 for the following SVD of a 3×4 matrix A:

$$A = \begin{bmatrix} -.86 & -.11 & -.50 \\ .31 & .68 & -.67 \\ .41 & -.73 & -.55 \end{bmatrix} \begin{bmatrix} 12.48 & 0 & 0 & 0 \\ 0 & 6.34 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} .66 & -.03 & -.35 & .66 \\ -.13 & -.90 & -.39 & -.13 \\ .65 & .08 & -.16 & -.73 \\ -.34 & .42 & -.84 & -.08 \end{bmatrix}$$

In Exercises 17–24, A is an $m \times n$ matrix with a singular value decomposition $A = U\Sigma V^T$, where U is an $m \times m$ orthogonal matrix, Σ is an $m \times n$ "diagonal" matrix with r positive entries and no negative entries, and V is an $n \times n$ orthogonal matrix. Justify each answer.

17. Show that if A is square, then $|\det A|$ is the product of the singular values of A.

18. Suppose A is square and invertible. Find a singular value decomposition of A^{-1}.

19. Show that the columns of V are eigenvectors of A^TA, the columns of U are eigenvectors of AA^T, and the diagonal entries of Σ are the singular values of A. [*Hint:* Use the SVD to compute A^TA and AA^T.]

20. Show that if P is an orthogonal $m \times m$ matrix, then PA has the same singular values as A.

21. Justify the statement in Example 2 that the second singular value of a matrix A is the maximum of $\|A\mathbf{x}\|$ as $\mathbf{x}$ varies over all unit vectors orthogonal to $\mathbf{v}_1$, with $\mathbf{v}_1$ a right singular vector corresponding to the first singular value of A. [*Hint:* Use Theorem 7 in Section 7.3.]

22. Show that if A is an $n \times n$ positive definite matrix, then an orthogonal diagonalization $A = PDP^T$ is a singular value decomposition of A.

23. Let $U = [\,\mathbf{u}_1 \ \cdots \ \mathbf{u}_m\,]$ and $V = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_n\,]$, where the $\mathbf{u}_i$ and $\mathbf{v}_i$ are as in Theorem 10. Show that

$$A = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T.$$

24. Using the notation of Exercise 23, show that $A^T\mathbf{u}_j = \sigma_j\mathbf{v}_j$ for $1 \le j \le r = \text{rank } A$.

25. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Describe how to find a basis $\mathcal{B}$ for $\mathbb{R}^n$ and a basis $\mathcal{C}$ for $\mathbb{R}^m$ such that the matrix for T relative to $\mathcal{B}$ and $\mathcal{C}$ is an $m \times n$ "diagonal" matrix.

Compute an SVD of each matrix in Exercises 26 and 27. Report the final matrix entries accurate to two decimal places. Use the method of Examples 3 and 4.

26. $A = \begin{bmatrix} -18 & 13 & -4 & 4 \\ 2 & 19 & -4 & 12 \\ -14 & 11 & -12 & 8 \\ -2 & 21 & 4 & 8 \end{bmatrix}$

T 27. $A = \begin{bmatrix} 6 & -8 & -4 & 5 & -4 \\ 2 & 7 & -5 & -6 & 4 \\ 0 & -1 & -8 & 2 & 2 \\ -1 & -2 & 4 & 4 & -8 \end{bmatrix}$

T 28. Compute the singular values of the 4×4 matrix in Exercise 9 in Section 2.3, and compute the condition number σ_1/σ_4.

T 29. Compute the singular values of the 5×5 matrix in Exercise 10 in Section 2.3, and compute the condition number σ_1/σ_5.

Solutions to Practice Problems

1. If $A = U\Sigma V^T$, where Σ is $m \times n$, then $A^T = (V^T)^T \Sigma^T U^T = V\Sigma^T U^T$. This is an SVD of A^T because V and U are orthogonal matrices and Σ^T is an $n \times m$ "diagonal" matrix. Since Σ and Σ^T have the same nonzero diagonal entries, A and A^T have the same nonzero singular values. [*Note:* If A is $2 \times n$, then AA^T is only 2×2 and its eigenvalues may be easier to compute (by hand) than the eigenvalues of A^TA.]
2. Use the SVD to write $A = U\Sigma V^T$, where U and V are $n \times n$ orthogonal matrices and Σ is an $n \times n$ diagonal matrix. Notice that $U^TU = I = V^TV$ and $\Sigma^T = \Sigma$, since U and V are orthogonal matrices and Σ is a diagonal matrix. Substituting the SVD for A into AA^T and A^TA results in

$$AA^T = U\Sigma V^T(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T = U\Sigma^2 U^T,$$

and

$$A^TA = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T = V\Sigma^2 V^T.$$

Let $Q = VU^T$. Then

$$\begin{aligned} Q^T(A^TA)Q &= (VU^T)^T(V\Sigma^2V^T)(VU^T) = UV^TV\Sigma^2V^TVU^T \\ &= U\Sigma^2U^T = AA^T. \end{aligned}$$

7.5 Applications to Image Processing and Statistics

The satellite photographs in this chapter's introduction provide an example of multidimensional, or *multivariate*, data—information organized so that each datum in the data set is identified with a point (vector) in $\mathbb{R}^n$. The main goal of this section is to explain a technique, called *principal component analysis*, used to analyze such multivariate data. The calculations will illustrate the use of orthogonal diagonalization and the singular value decomposition.

Principal component analysis can be applied to any data that consist of lists of measurements made on a collection of objects or individuals. For instance, consider a chemical process that produces a plastic material. To monitor the process, 300 samples are taken of the material produced, and each sample is subjected to a battery of eight tests, such as melting point, density, ductility, tensile strength, and so on. The laboratory report for each sample is a vector in $\mathbb{R}^8$, and the set of such vectors forms an 8×300 matrix, called the **matrix of observations**.

Loosely speaking, we can say that the process control data are eight-dimensional. The next two examples describe data that can be visualized graphically.

EXAMPLE 1 An example of two-dimensional data is given by a set of weights and heights of N college students. Let $\mathbf{X}_j$ denote the **observation vector** in $\mathbb{R}^2$ that lists the weight and height of the jth student. If w denotes weight and h height, then the matrix of observations has the form

$$\begin{array}{cccc} \begin{bmatrix} w_1 & w_2 & \cdots & w_N \\ h_1 & h_2 & \cdots & h_N \end{bmatrix} \\ \uparrow \quad \uparrow \qquad\quad \uparrow \\ \mathbf{X}_1 \quad \mathbf{X}_2 \qquad \mathbf{X}_N \end{array}$$

The set of observation vectors can be visualized as a two-dimensional *scatter plot*. See Figure 1. ■

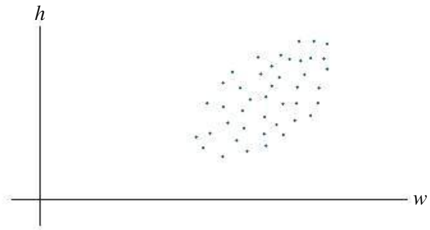


FIGURE 1 A scatter plot of observation vectors $\mathbf{X}_1, \dots, \mathbf{X}_N$.

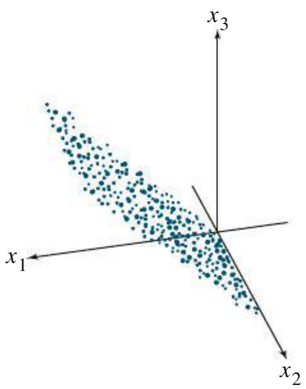


FIGURE 2
A scatter plot of spectral data for a satellite image.

EXAMPLE 2 The first three photographs of Railroad Valley, Nevada, shown in the chapter introduction can be viewed as *one* image of the region, with *three spectral components*, because simultaneous measurements of the region were made at three separate wavelengths. Each photograph gives different information about the same physical region. For instance, the first pixel in the upper-left corner of each photograph corresponds to the same place on the ground (about 30 meters by 30 meters). To each pixel there corresponds an observation vector in $\mathbb{R}^3$ that lists the signal intensities for that pixel in the three spectral bands.

Typically, the image is 2000×2000 pixels, so there are 4 million pixels in the image. The data for the image form a matrix with 3 rows and 4 million columns (with columns arranged in any convenient order). In this case, the "multidimensional" character of the data refers to the three *spectral* dimensions rather than the two *spatial* dimensions that naturally belong to any photograph. The data can be visualized as a cluster of 4 million points in $\mathbb{R}^3$, perhaps as in Figure 2. ■

Mean and Covariance

To prepare for principal component analysis, let $[\,\mathbf{X}_1 \ \cdots \ \mathbf{X}_N\,]$ be a $p \times N$ matrix of observations, such as described above. The **sample mean**, $\mathbf{M}$, of the observation vectors $\mathbf{X}_1, \dots, \mathbf{X}_N$ is given by

$$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$$

For the data in Figure 1, the sample mean is the point in the "center" of the scatter plot. For $k = 1, \dots, N$, let

$$\hat{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{M}$$

The columns of the $p \times N$ matrix

$$B = [\,\hat{\mathbf{X}}_1 \quad \hat{\mathbf{X}}_2 \quad \cdots \quad \hat{\mathbf{X}}_N\,]$$

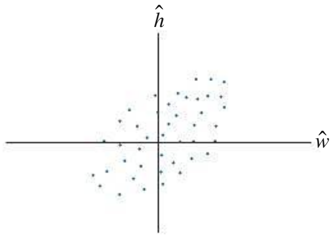


FIGURE 3
Weight–height data in mean-deviation form.

have a zero sample mean, and B is said to be in **mean-deviation form**. When the sample mean is subtracted from the data in Figure 1, the resulting scatter plot has the form in Figure 3.

The (**sample**) **covariance matrix** is the $p \times p$ matrix S defined by

$$S = \frac{1}{N-1}BB^T$$

Since any matrix of the form BB^T is positive semidefinite, so is S. (See Exercise 33 in Section 7.2 with B and B^T interchanged.)

EXAMPLE 3 Three measurements are made on each of four individuals in a random sample from a population. The observation vectors are

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix}, \quad \mathbf{X}_3 = \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix}, \quad \mathbf{X}_4 = \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix}$$

Compute the sample mean and the covariance matrix.

SOLUTION The sample mean is

$$\mathbf{M} = \frac{1}{4}\left(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix} + \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix} + \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix} \right) = \frac{1}{4}\begin{bmatrix} 20 \\ 16 \\ 20 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 5 \end{bmatrix}$$

Subtract the sample mean from $\mathbf{X}_1, \dots, \mathbf{X}_4$ to obtain

$$\hat{\mathbf{X}}_1 = \begin{bmatrix} -4 \\ -2 \\ -4 \end{bmatrix}, \quad \hat{\mathbf{X}}_2 = \begin{bmatrix} -1 \\ -2 \\ 8 \end{bmatrix}, \quad \hat{\mathbf{X}}_3 = \begin{bmatrix} 2 \\ 4 \\ -4 \end{bmatrix}, \quad \hat{\mathbf{X}}_4 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

and

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

The sample covariance matrix is

$$\begin{aligned} S &= \frac{1}{3}\begin{bmatrix} -4 & -1 & 2 & 3 \\ -2 & -2 & 4 & 0 \\ -4 & 8 & -4 & 0 \end{bmatrix}\begin{bmatrix} -4 & -2 & -4 \\ -1 & -2 & 8 \\ 2 & 4 & -4 \\ 3 & 0 & 0 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 30 & 18 & 0 \\ 18 & 24 & -24 \\ 0 & -24 & 96 \end{bmatrix} = \begin{bmatrix} 10 & 6 & 0 \\ 6 & 8 & -8 \\ 0 & -8 & 32 \end{bmatrix} \end{aligned}$$

■

To discuss the entries in $S = [s_{ij}]$, let $\mathbf{X}$ represent a vector that varies over the set of observation vectors and denote the coordinates of $\mathbf{X}$ by $x_1, \dots, x_p$. Then x_1, for example, is a scalar that varies over the set of first coordinates of $\mathbf{X}_1, \dots, \mathbf{X}_N$. For $j = 1, \dots, p$, the diagonal entry s_{jj} in S is called the **variance** of x_j.

The variance of x_j measures the spread of the values of x_j. (See Exercise 13.) In Example 3, the variance of x_1 is 10 and the variance of x_3 is 32. The fact that 32 is more than 10 indicates that the set of third entries in the response vectors contains a wider spread of values than the set of first entries.

The **total variance** of the data is the sum of the variances on the diagonal of S. In general, the sum of the diagonal entries of a square matrix S is called the **trace** of the matrix, written $\text{tr}(S)$. Thus

$$\{\text{total variance}\} = \text{tr}(S)$$

The entry s_{ij} in S for $i \neq j$ is called the **covariance** of x_i and x_j. Observe that in Example 3, the covariance between x_1 and x_3 is 0 because the $(1, 3)$-entry in S is 0. Statisticians say that x_1 and x_3 are **uncorrelated**. Analysis of the multivariate data in $\mathbf{X}_1, \ldots, \mathbf{X}_N$ is greatly simplified when most or all of the variables $x_1, \ldots, x_p$ are uncorrelated, that is, when the covariance matrix of $\mathbf{X}_1, \ldots, \mathbf{X}_N$ is diagonal or nearly diagonal.

Principal Component Analysis

For simplicity, assume that the matrix $[\,\mathbf{X}_1 \ \cdots \ \mathbf{X}_N\,]$ is already in mean-deviation form. The goal of principal component analysis is to find an orthogonal $p \times p$ matrix $P = [\,\mathbf{u}_1 \ \cdots \ \mathbf{u}_p\,]$ that determines a change of variable, $\mathbf{X} = P\mathbf{Y}$, or

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

with the property that the new variables $y_1, \ldots, y_p$ are uncorrelated and are arranged in order of decreasing variance.

The orthogonal change of variable $\mathbf{X} = P\mathbf{Y}$ means that each observation vector $\mathbf{X}_k$ receives a "new name," $\mathbf{Y}_k$, such that $\mathbf{X}_k = P\mathbf{Y}_k$. Notice that $\mathbf{Y}_k$ is the coordinate vector of $\mathbf{X}_k$ with respect to the columns of P, and $\mathbf{Y}_k = P^{-1}\mathbf{X}_k = P^T\mathbf{X}_k$ for $k = 1, \ldots, N$.

It is not difficult to verify that for any orthogonal P, the covariance matrix of $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ is P^TSP (Exercise 11). So the desired orthogonal matrix P is one that makes P^TSP diagonal. Let D be a diagonal matrix with the eigenvalues $\lambda_1, \ldots, \lambda_p$ of S on the diagonal, arranged so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and let P be an orthogonal matrix whose columns are the corresponding unit eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$. Then $S = PDP^T$ and $P^TSP = D$.

The unit eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$ of the covariance matrix S are called the **principal components** of the data (in the matrix of observations). The **first principal component** is the eigenvector corresponding to the largest eigenvalue of S, the **second principal component** is the eigenvector corresponding to the second largest eigenvalue, and so on.

The first principal component $\mathbf{u}_1$ determines the new variable y_1 in the following way. Let $c_1, \ldots, c_p$ be the entries in $\mathbf{u}_1$. Since $\mathbf{u}_1^T$ is the first row of P^T, the equation $\mathbf{Y} = P^T\mathbf{X}$ shows that

$$y_1 = \mathbf{u}_1^T\mathbf{X} = c_1x_1 + c_2x_2 + \cdots + c_px_p$$

Thus y_1 is a linear combination of the original variables $x_1, \ldots, x_p$, using the entries in the eigenvector $\mathbf{u}_1$ as weights. In a similar fashion, $\mathbf{u}_2$ determines the variable y_2, and so on.

EXAMPLE 4 The initial data for the multispectral image of Railroad Valley (Example 2) consisted of 4 million vectors in $\mathbb{R}^3$. The associated covariance matrix is[1]

$$S = \begin{bmatrix} 2382.78 & 2611.84 & 2136.20 \\ 2611.84 & 3106.47 & 2553.90 \\ 2136.20 & 2553.90 & 2650.71 \end{bmatrix}$$

[1] Data for Example 4 and Exercises 5 and 6 were provided by Earth Satellite Corporation, Rockville, Maryland.

Find the principal components of the data, and list the new variable determined by the first principal component.

SOLUTION The eigenvalues of S and the associated principal components (the unit eigenvectors) are

$$\lambda_1 = 7614.23 \qquad \lambda_2 = 427.63 \qquad \lambda_3 = 98.10$$
$$\mathbf{u}_1 = \begin{bmatrix} .5417 \\ .6295 \\ .5570 \end{bmatrix} \quad \mathbf{u}_2 = \begin{bmatrix} -.4894 \\ -.3026 \\ .8179 \end{bmatrix} \quad \mathbf{u}_3 = \begin{bmatrix} .6834 \\ -.7157 \\ .1441 \end{bmatrix}$$

Using two decimal places for simplicity, the variable for the first principal component is

$$y_1 = .54x_1 + .63x_2 + .56x_3$$

This equation was used to create photograph (d) in the chapter introduction. The variables x_1, x_2, and x_3 are the signal intensities in the three spectral bands. The values of x_1, converted to a gray scale between black and white, produced photograph (a). Similarly, the values of x_2 and x_3 produced photographs (b) and (c), respectively. At each pixel in photograph (d), the gray scale value is computed from y_1, a weighted linear combination of x_1, x_2, and x_3. In this sense, photograph (d) "displays" the first principal component of the data. ■

In Example 4, the covariance matrix for the transformed data, using variables y_1, y_2, and y_3, is

$$D = \begin{bmatrix} 7614.23 & 0 & 0 \\ 0 & 427.63 & 0 \\ 0 & 0 & 98.10 \end{bmatrix}$$

Although D is obviously simpler than the original covariance matrix S, the merit of constructing the new variables is not yet apparent. However, the variances of the variables y_1, y_2, and y_3 appear on the diagonal of D, and obviously the first variance in D is much larger than the other two. As we shall see, this fact will permit us to view the data as essentially one-dimensional rather than three-dimensional.

Reducing the Dimension of Multivariate Data

Principal component analysis is potentially valuable for applications in which most of the variation, or dynamic range, in the data is due to variations in *only a few* of the new variables, $y_1, \dots, y_p$.

It can be shown that an orthogonal change of variables, $\mathbf{X} = P\mathbf{Y}$, does not change the total variance of the data. (Roughly speaking, this is true because left-multiplication by P does not change the lengths of vectors or the angles between them. See Exercise 12.) This means that if $S = PDP^T$, then

$$\begin{Bmatrix} \text{total variance} \\ \text{of } x_1, \dots, x_p \end{Bmatrix} = \begin{Bmatrix} \text{total variance} \\ \text{of } y_1, \dots, y_p \end{Bmatrix} = \text{tr}(D) = \lambda_1 + \cdots + \lambda_p$$

The variance of y_j is λ_j, and the quotient $\lambda_j / \text{tr}(S)$ measures the fraction of the total variance that is "explained" or "captured" by y_j.

EXAMPLE 5 Compute the various percentages of variance of the Railroad Valley multispectral data that are displayed in the principal component photographs, (d)–(f), shown in the chapter introduction.

SOLUTION The total variance of the data is

$$\text{tr}(D) = 7614.23 + 427.63 + 98.10 = 8139.96$$

[Verify that this number also equals $\text{tr}(S)$.] The percentages of the total variance explained by the principal components are

First component	Second component	Third component
$\dfrac{7614.23}{8139.96} = 93.5\%$	$\dfrac{427.63}{8139.96} = 5.3\%$	$\dfrac{98.10}{8139.96} = 1.2\%$

In a sense, 93.5% of the information collected by Landsat for the Railroad Valley region is displayed in photograph (d), with 5.3% in (e) and only 1.2% remaining for (f). ■

The calculations in Example 5 show that the data have practically no variance in the third (new) coordinate. The values of y_3 are all close to zero. Geometrically, the data points lie nearly in the plane $y_3 = 0$, and their locations can be determined fairly accurately by knowing only the values of y_1 and y_2. In fact, y_2 also has relatively small variance, which means that the points lie approximately along a line, and the data are essentially one-dimensional. See Figure 2, in which the data resemble a popsicle stick.

Characterizations of Principal Component Variables

If $y_1, \dots, y_p$ arise from a principal component analysis of a $p \times N$ matrix of observations, then the variance of y_1 is as large as possible in the following sense: If $\mathbf{u}$ is any unit vector and if $y = \mathbf{u}^T\mathbf{X}$, then the variance of the values of y as $\mathbf{X}$ varies over the original data $\mathbf{X}_1, \dots, \mathbf{X}_N$ turns out to be $\mathbf{u}^T S\mathbf{u}$. By Theorem 8 in Section 7.3, the maximum value of $\mathbf{u}^T S\mathbf{u}$, over all unit vectors $\mathbf{u}$, is the largest eigenvalue λ_1 of S, and this variance is attained when $\mathbf{u}$ is the corresponding eigenvector $\mathbf{u}_1$. In the same way, Theorem 8 shows that y_2 has maximum possible variance among all variables $y = \mathbf{u}^T\mathbf{X}$ that are *uncorrelated* with y_1. Likewise, y_3 has maximum possible variance among all variables uncorrelated with both y_1 and y_2, and so on.

Numerical Notes

The singular value decomposition is the main tool for performing principal component analysis in practical applications. If B is a $p \times N$ matrix of observations in mean-deviation form, and if $A = \left(1/\sqrt{N-1}\right)B^T$, then A^TA is the covariance matrix, S. The squares of the singular values of A are the p eigenvalues of S, and the right singular vectors of A are the principal components of the data.

As mentioned in Section 7.4, iterative calculation of the SVD of A is faster and more accurate than an eigenvalue decomposition of S. This is particularly true, for instance, in the hyperspectral image processing (with $p = 224$) mentioned in the chapter introduction. Principal component analysis is completed in seconds on specialized workstations.

Practice Problems

The following table lists the weights and heights of five boys:

Boy	#1	#2	#3	#4	#5
Weight (lb)	120	125	125	135	145
Height (in.)	61	60	64	68	72

1. Find the covariance matrix for the data.
2. Make a principal component analysis of the data to find a single *size index* that explains most of the variation in the data.

7.5 Exercises

In Exercises 1 and 2, convert the matrix of observations to mean-deviation form, and construct the sample covariance matrix.

1. $\begin{bmatrix} 19 & 22 & 6 & 3 & 2 & 20 \\ 12 & 6 & 9 & 15 & 13 & 5 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 5 & 2 & 6 & 7 & 3 \\ 3 & 11 & 6 & 8 & 15 & 11 \end{bmatrix}$

3. Find the principal components of the data for Exercise 1.

4. Find the principal components of the data for Exercise 2.

T 5. A Landsat image with three spectral components was made of Homestead Air Force Base in Florida (after the base was hit by Hurricane Andrew in 1992). The covariance matrix of the data is shown below. Find the first principal component of the data, and compute the percentage of the total variance that is contained in this component.

$$S = \begin{bmatrix} 164.12 & 32.73 & 81.04 \\ 32.73 & 539.44 & 249.13 \\ 81.04 & 249.13 & 189.11 \end{bmatrix}$$

T 6. The covariance matrix below was obtained from a Landsat image of the Columbia River in Washington, using data from three spectral bands. Let x_1, x_2, x_3 denote the spectral components of each pixel in the image. Find a new variable of the form $y_1 = c_1x_1 + c_2x_2 + c_3x_3$ that has maximum possible variance, subject to the constraint that $c_1^2 + c_2^2 + c_3^2 = 1$. What percentage of the total variance in the data is explained by y_1?

$$S = \begin{bmatrix} 29.64 & 18.38 & 5.00 \\ 18.38 & 20.82 & 14.06 \\ 5.00 & 14.06 & 29.21 \end{bmatrix}$$

7. Let x_1, x_2 denote the variables for the two-dimensional data in Exercise 1. Find a new variable y_1 of the form $y_1 = c_1x_1 + c_2x_2$, with $c_1^2 + c_2^2 = 1$, such that y_1 has maximum possible variance over the given data. How much of the variance in the data is explained by y_1?

8. Repeat Exercise 7 for the data in Exercise 2.

9. Suppose three tests are administered to a random sample of college students. Let $\mathbf{X}_1, \dots, \mathbf{X}_N$ be observation vectors in $\mathbb{R}^3$ that list the three scores of each student, and for $j = 1, 2, 3$, let x_j denote a student's score on the jth exam. Suppose the covariance matrix of the data is

$$S = \begin{bmatrix} 5 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 7 \end{bmatrix}$$

Let y be an "index" of student performance, with $y = c_1x_1 + c_2x_2 + c_3x_3$ and $c_1^2 + c_2^2 + c_3^2 = 1$. Choose c_1, c_2, c_3 so that the variance of y over the data set is as large as possible. [*Hint:* The eigenvalues of the sample covariance matrix are $\lambda = 3, 6,$ and 9.]

T 10. Repeat Exercise 9 with $S = \begin{bmatrix} 5 & 4 & 2 \\ 4 & 11 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

11. Given multivariate data $\mathbf{X}_1, \dots, \mathbf{X}_N$ (in $\mathbb{R}^p$) in mean-deviation form, let P be a $p \times p$ matrix, and define $\mathbf{Y}_k = P^T\mathbf{X}_k$ for $k = 1, \dots, N$.
   a. Show that $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ are in mean-deviation form. [*Hint:* Let $\mathbf{w}$ be the vector in $\mathbb{R}^N$ with a 1 in each entry. Then $[\,\mathbf{X}_1 \;\; \cdots \;\; \mathbf{X}_N\,]\,\mathbf{w} = \mathbf{0}$ (the zero vector in $\mathbb{R}^p$).]
   b. Show that if the covariance matrix of $\mathbf{X}_1, \dots, \mathbf{X}_N$ is S, then the covariance matrix of $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ is P^TSP.

12. Let $\mathbf{X}$ denote a vector that varies over the columns of a $p \times N$ matrix of observations, and let P be a $p \times p$ orthogonal matrix. Show that the change of variable $\mathbf{X} = P\mathbf{Y}$ does not change the total variance of the data. [*Hint:* By Exercise 11, it suffices to show that $\operatorname{tr}(P^TSP) = \operatorname{tr}(S)$. Use a property of the trace mentioned in Exercise 27 in Section 5.4.]

13. The sample covariance matrix is a generalization of a formula for the variance of a sample of N scalar measurements, say, $t_1, \dots, t_N$. If m is the average of $t_1, \dots, t_N$, then the *sample variance* is given by

$$\frac{1}{N-1}\sum_{k=1}^{n}(t_k - m)^2 \qquad (1)$$

Show how the sample covariance matrix, S, defined prior to Example 3, may be written in a form similar to (1). [*Hint:* Use partitioned matrix multiplication to write S as $1/(N-1)$ times the sum of N matrices of size $p \times p$. For $1 \le k \le N$, write $\mathbf{X}_k - \mathbf{M}$ in place of $\hat{\mathbf{X}}_k$.]

Solutions to Practice Problems

1. First arrange the data in mean-deviation form. The sample mean vector is easily seen to be $\mathbf{M} = \begin{bmatrix} 130 \\ 65 \end{bmatrix}$. Subtract $\mathbf{M}$ from the observation vectors (the columns in the table) and obtain

$$B = \begin{bmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{bmatrix}$$

Then the sample covariance matrix is

$$S = \frac{1}{5-1}\begin{bmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{bmatrix}\begin{bmatrix} -10 & -4 \\ -5 & -5 \\ -5 & -1 \\ 5 & 3 \\ 15 & 7 \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 400 & 190 \\ 190 & 100 \end{bmatrix} = \begin{bmatrix} 100.0 & 47.5 \\ 47.5 & 25.0 \end{bmatrix}$$

2. The eigenvalues of S are (to two decimal places)

$$\lambda_1 = 123.02 \quad \text{and} \quad \lambda_2 = 1.98$$

The unit eigenvector corresponding to λ_1 is $\mathbf{u} = \begin{bmatrix} .900 \\ .436 \end{bmatrix}$. (Since S is 2×2, the computations can be done by hand if a matrix program is not available.) For the *size index*, set

$$y = .900\hat{w} + .436\hat{h}$$

where $\hat{w}$ and $\hat{h}$ are weight and height, respectively, in mean-deviation form. The variance of this index over the data set is 123.02. Because the total variance is $\text{tr}(S) = 100 + 25 = 125$, the size index accounts for practically all (98.4%) of the variance of the data.

The original data for Practice Problem 1 and the line determined by the first principal component $\mathbf{u}$ are shown in Figure 4. (In parametric vector form, the line

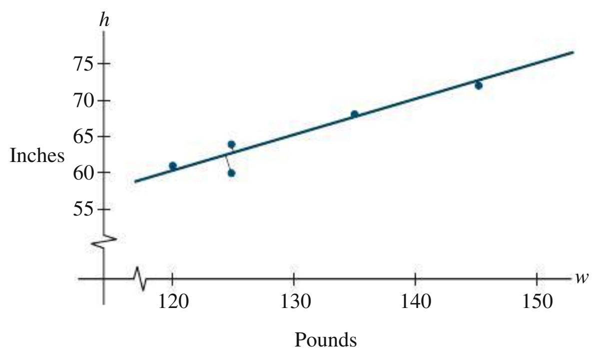


FIGURE 4 An orthogonal regression line determined by the first principal component of the data.

is $\mathbf{x} = \mathbf{M} + t\mathbf{u}$.) It can be shown that the line is the best approximation to the data, in the sense that the sum of the squares of the *orthogonal* distances to the line is minimized. In fact, principal component analysis is equivalent to what is termed *orthogonal regression*, but that is a story for another day.

CHAPTER 7 PROJECTS

Chapter 7 projects are available online at bit.ly/30IM8gT.

A. *Conic Sections and Quadric Surfaces*: This project shows how quadratic forms and the Principal Axes Theorem may be used to classify conic sections and quadric surfaces.

B. *Extrema for Functions of Several Variables*: This project shows how quadratic forms may be used to investigate maximum and minimum values of functions of several variables.

CHAPTER 7 SUPPLEMENTARY EXERCISES

Mark each statement True or False. Justify each answer. In each part, A represents an $n \times n$ matrix.

1. **(T/F)** If A is orthogonally diagonalizable, then A is symmetric.

2. **(T/F)** If A is an orthogonal matrix, then A is symmetric.

3. **(T/F)** If A is an orthogonal matrix, then $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

4. **(T/F)** The principal axes of a quadratic form $\mathbf{x}^T A\mathbf{x}$ can be the columns of any matrix P that diagonalizes A.

5. **(T/F)** If P is an $n \times n$ matrix with orthogonal columns, then $P^T = P^{-1}$.

6. **(T/F)** If every coefficient in a quadratic form is positive, then the quadratic form is positive definite.

7. **(T/F)** If $\mathbf{x}^T A\mathbf{x} > 0$ for some $\mathbf{x}$, then the quadratic form $\mathbf{x}^T A\mathbf{x}$ is positive definite.

8. **(T/F)** By a suitable change of variable, any quadratic form can be changed into one with no cross-product term.

9. **(T/F)** The largest value of a quadratic form $\mathbf{x}^T A\mathbf{x}$, for $\|\mathbf{x}\| = 1$, is the largest entry on the diagonal of A.

10. **(T/F)** The maximum value of a positive definite quadratic form $\mathbf{x}^T A\mathbf{x}$ is the greatest eigenvalue of A.

11. **(T/F)** A positive definite quadratic form can be changed into a negative definite form by a suitable change of variable $\mathbf{x} = P\mathbf{u}$, for some orthogonal matrix P.

12. **(T/F)** An indefinite quadratic form is one whose eigenvalues are not definite.

13. **(T/F)** If P is an $n \times n$ orthogonal matrix, then the change of variable $\mathbf{x} = P\mathbf{u}$ transforms $\mathbf{x}^T A\mathbf{x}$ into a quadratic form whose matrix is $P^{-1}AP$.

14. **(T/F)** If U is $m \times n$ with orthogonal columns, then $UU^T\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto Col U.

15. **(T/F)** If B is $m \times n$ and $\mathbf{x}$ is a unit vector in $\mathbb{R}^n$, then $\|B\mathbf{x}\| \le \sigma_1$, where σ_1 is the first singular value of B.

16. **(T/F)** A singular value decomposition of an $m \times n$ matrix B can be written as $B = P\Sigma Q$, where P is an $m \times m$ orthogonal matrix, Q is an $n \times n$ orthogonal matrix, and Σ is an $m \times n$ "diagonal" matrix.

17. **(T/F)** If A is $n \times n$, then A and $A^T A$ have the same singular values.

18. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be an orthonormal basis for $\mathbb{R}^n$, and let $\lambda_1, \dots, \lambda_n$ be any real scalars. Define

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

   a. Show that A is symmetric.

   b. Show that $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A.

19. Let A be an $n \times n$ symmetric matrix of rank r. Explain why the spectral decomposition of A represents A as the sum of r rank 1 matrices.

20. Let A be an $n \times n$ symmetric matrix.

   a. Show that $(\text{Col } A)^\perp = \text{Nul } A$. [*Hint:* See Section 6.1.]

   b. Show that each $\mathbf{y}$ in $\mathbb{R}^n$ can be written in the form $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, with $\hat{\mathbf{y}}$ in Col A and $\mathbf{z}$ in Nul A.

21. Show that if $\mathbf{v}$ is an eigenvector of an $n \times n$ matrix A and $\mathbf{v}$ corresponds to a nonzero eigenvalue of A, then $\mathbf{v}$ is in Col A. [*Hint:* Use the definition of an eigenvector.]

22. Let A be an $n \times n$ symmetric matrix. Use Exercise 21 and an eigenvector basis for $\mathbb{R}^n$ to give a second proof of the decomposition in Exercise 20(b).

23. Prove that an $n \times n$ matrix A is positive definite if and only if A admits a *Cholesky factorization*, namely $A = R^T R$ for

some invertible upper triangular matrix R whose diagonal entries are all positive. [*Hint:* Use a QR factorization and Exercise 34 in Section 7.2.]

24. Use Exercise 23 to show that if A is positive definite, then A has an LU factorization, $A = LU$, where U has positive pivots on its diagonal. (The converse is true, too.)

If A is $m \times n$, then the matrix $G = A^TA$ is called the *Gram matrix of* A. In this case, the entries of G are the inner products of the columns of A. (See Exercises 25 and 26.)

25. Show that the Gram matrix of any matrix A is positive semidefinite, with the same rank as A. (See the Exercises in Section 6.5.)

26. Show that if an $n \times n$ matrix G is positive semidefinite and has rank r, then G is the Gram matrix of some $r \times n$ matrix A. This is called a *rank-revealing factorization* of G. [*Hint:* Consider the spectral decomposition of G, and first write G as BB^T for an $n \times r$ matrix B.]

27. Prove that any $n \times n$ matrix A admits a *polar decomposition* of the form $A = PQ$, where P is an $n \times n$ positive semidefinite matrix with the same rank as A and where Q is an $n \times n$ orthogonal matrix. [*Hint:* Use a singular value decomposition, $A = U\Sigma V^T$, and observe that $A = (U\Sigma U^T)(UV^T)$.] This decomposition is used, for instance, in mechanical engineering to model the deformation of a material. The matrix P describes the stretching or compression of the material (in the directions of the eigenvectors of P), and Q describes the rotation of the material in space.

Exercises 28–30 concern an $m \times n$ matrix A with a reduced singular value decomposition, $A = U_r D V_r^T$, and the pseudoinverse $A^+ = V_r D^{-1} U_r^T$.

28. Verify the properties of A^+:

a. For each $\mathbf{y}$ in $\mathbb{R}^m$, $AA^+\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto Col A.

b. For each $\mathbf{x}$ in $\mathbb{R}^n$, $A^+A\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto Row A.

c. $AA^+A = A$ and $A^+AA^+ = A^+$.

29. Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent, and let $\mathbf{x}^+ = A^+\mathbf{b}$. By Exercise 31 in Section 6.3, there is exactly one vector $\mathbf{p}$ in Row A such that $A\mathbf{p} = \mathbf{b}$. The following steps prove that $\mathbf{x}^+ = \mathbf{p}$ and $\mathbf{x}^+$ is the *minimum length solution* of $A\mathbf{x} = \mathbf{b}$.

a. Show that $\mathbf{x}^+$ is in Row A. [*Hint:* Write $\mathbf{b}$ as $A\mathbf{x}$ for some $\mathbf{x}$, and use Exercise 28.]

b. Show that $\mathbf{x}^+$ is a solution of $A\mathbf{x} = \mathbf{b}$.

c. Show that if $\mathbf{u}$ is any solution of $A\mathbf{x} = \mathbf{b}$, then $\|\mathbf{x}^+\| \leq \|\mathbf{u}\|$, with equality only if $\mathbf{u} = \mathbf{x}^+$.

30. Given any $\mathbf{b}$ in $\mathbb{R}^m$, adapt Exercise 28 to show that $A^+\mathbf{b}$ is the *least-squares solution of minimum length*. [*Hint:* Consider the equation $A\mathbf{x} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto Col A.]

T In Exercises 31 and 32, construct the pseudoinverse of A. Begin by using a matrix program to produce the SVD of A, or, if that is not available, begin with an orthogonal diagonalization of A^TA. Use the pseudoinverse to solve $A\mathbf{x} = \mathbf{b}$, for $\mathbf{b} = (6, -1, -4, 6)$, and let $\hat{\mathbf{x}}$ be the solution. Make a calculation to verify that $\hat{\mathbf{x}}$ is in Row A. Find a nonzero vector $\mathbf{u}$ in Nul A, and verify that $\|\hat{\mathbf{x}}\| < \|\hat{\mathbf{x}} + \mathbf{u}\|$, which must be true by Exercise 29(c).

T 31. $A = \begin{bmatrix} -3 & -3 & -6 & 6 & 1 \\ -1 & -1 & -1 & 1 & -2 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & -1 & 1 & -1 \end{bmatrix}$

T 32. $A = \begin{bmatrix} 4 & 0 & -1 & -2 & 0 \\ -5 & 0 & 3 & 5 & 0 \\ 2 & 0 & -1 & -2 & 0 \\ 6 & 0 & -3 & -6 & 0 \end{bmatrix}$

Appendix A
Uniqueness of the Reduced Echelon Form

THEOREM

Uniqueness of the Reduced Echelon Form

Each $m \times n$ matrix A is row equivalent to a unique reduced echelon matrix U.

PROOF The proof uses the idea from Section 4.3 that the columns of row-equivalent matrices have exactly the same linear dependence relations.

The row reduction algorithm shows that there exists at least one such matrix U. Suppose that A is row equivalent to matrices U and V in reduced echelon form. The leftmost nonzero entry in a row of U is a "leading 1." Call the location of such a leading 1 a pivot position, and call the column that contains it a pivot column. (This definition uses only the echelon nature of U and V and does not assume the uniqueness of the reduced echelon form.)

The pivot columns of U and V are precisely the nonzero columns that are *not* linearly dependent on the columns to their left. (This condition is satisfied automatically by a *first* column if it is nonzero.) Since U and V are row equivalent (both being row equivalent to A), their columns have the same linear dependence relations. Hence, the pivot columns of U and V appear in the same locations. If there are r such columns, then since U and V are in reduced echelon form, their pivot columns are the first r columns of the $m \times m$ identity matrix. Thus, *corresponding pivot columns of U and V are equal*.

Finally, consider any nonpivot column of U, say column j. This column is either zero or a linear combination of the pivot columns to its left (because those pivot columns are a basis for the space spanned by the columns to the left of column j). Either case can be expressed by writing $U\mathbf{x} = \mathbf{0}$ for some $\mathbf{x}$ whose jth entry is 1. Then $V\mathbf{x} = \mathbf{0}$, too, which says that column j of V is either zero or the *same* linear combination of the pivot columns of V to *its* left. Since corresponding pivot columns of U and V are equal, columns j of U and V are also equal. This holds for all nonpivot columns, so $V = U$, which proves that U is unique.

Appendix B
Complex Numbers

A **complex number** is a number written in the form

$$z = a + bi$$

where a and b are real numbers and i is a formal symbol satisfying the relation $i^2 = -1$. The number a is the **real part** of z, denoted by $\operatorname{Re} z$, and b is the **imaginary part** of z, denoted by $\operatorname{Im} z$. Two complex numbers are considered equal if and only if their real and imaginary parts are equal. For example, if $z = 5 + (-2)i$, then $\operatorname{Re} z = 5$ and $\operatorname{Im} z = -2$. For simplicity, we write $z = 5 - 2i$.

A real number a is considered as a special type of complex number, by identifying a with $a + 0i$. Furthermore, arithmetic operations on real numbers can be extended to the set of complex numbers.

The **complex number system**, denoted by $\mathbb{C}$, is the set of all complex numbers, together with the following operations of addition and multiplication:

$$(a + bi) + (c + di) = (a + c) + (b + d)i \tag{1}$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i \tag{2}$$

These rules reduce to ordinary addition and multiplication of real numbers when b and d are zero in (1) and (2). It is readily checked that the usual laws of arithmetic for $\mathbb{R}$ also hold for $\mathbb{C}$. For this reason, multiplication is usually computed by algebraic expansion, as in the following example.

EXAMPLE 1

$$\begin{aligned}(5 - 2i)(3 + 4i) &= 15 + 20i - 6i - 8i^2 \\ &= 15 + 14i - 8(-1) \\ &= 23 + 14i\end{aligned}$$

That is, multiply each term of $5 - 2i$ by each term of $3 + 4i$, use $i^2 = -1$, and write the result in the form $a + bi$. ■

Subtraction of complex numbers z_1 and z_2 is defined by

$$z_1 - z_2 = z_1 + (-1)z_2$$

In particular, we write $-z$ in place of $(-1)z$.

The **conjugate** of $z = a + bi$ is the complex number $\overline{z}$ (read as "z bar"), defined by

$$\overline{z} = a - bi$$

Obtain $\overline{z}$ from z by reversing the sign of the imaginary part.

EXAMPLE 2 The conjugate of $-3 + 4i$ is $-3 - 4i$; write $\overline{-3 + 4i} = -3 - 4i$. ■

Observe that if $z = a + bi$, then

$$z\overline{z} = (a + bi)(a - bi) = a^2 - abi + bai - b^2i^2 = a^2 + b^2 \tag{3}$$

Since $z\overline{z}$ is real and nonnegative, it has a square root. The **absolute value** (or **modulus**) of z is the real number $|z|$ defined by

$$|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}$$

If z is a real number, then $z = a + 0i$, and $|z| = \sqrt{a^2}$, which equals the ordinary absolute value of a.

Some useful properties of conjugates and absolute value are listed below; w and z denote complex numbers.

1. $\overline{z} = z$ if and only if z is a real number.
2. $\overline{w + z} = \overline{w} + \overline{z}$.
3. $\overline{wz} = \overline{w}\,\overline{z}$; in particular, $\overline{rz} = r\overline{z}$ if r is a real number.
4. $z\overline{z} = |z|^2 \geq 0$.
5. $|wz| = |w||z|$.
6. $|w + z| \leq |w| + |z|$.

If $z \neq 0$, then $|z| > 0$ and z has a multiplicative inverse, denoted by $1/z$ or z^{-1} and given by

$$\frac{1}{z} = z^{-1} = \frac{\overline{z}}{|z|^2}$$

Of course, a quotient w/z simply means $w \cdot (1/z)$.

EXAMPLE 3 Let $w = 3 + 4i$ and $z = 5 - 2i$. Compute $z\overline{z}$, $|z|$, and w/z.

SOLUTION From equation (3),

$$z\overline{z} = 5^2 + (-2)^2 = 25 + 4 = 29$$

For the absolute value, $|z| = \sqrt{z\overline{z}} = \sqrt{29}$. To compute w/z, first multiply both the numerator and the denominator by $\overline{z}$, the conjugate of the denominator. Because of (3), this eliminates the i in the denominator:

$$\begin{aligned}
\frac{w}{z} &= \frac{3 + 4i}{5 - 2i} \\
&= \frac{3 + 4i}{5 - 2i} \cdot \frac{5 + 2i}{5 + 2i} \\
&= \frac{15 + 6i + 20i - 8}{5^2 + (-2)^2} \\
&= \frac{7 + 26i}{29} \\
&= \frac{7}{29} + \frac{26}{29}i
\end{aligned}$$
■

Geometric Interpretation

Each complex number $z = a + bi$ corresponds to a point (a, b) in the plane $\mathbb{R}^2$, as in Figure 1. The horizontal axis is called the **real axis** because the points $(a, 0)$ on it correspond to the real numbers. The vertical axis is the **imaginary axis** because the points $(0, b)$ on it correspond to the **pure imaginary numbers** of the form $0 + bi$, or simply bi. The conjugate of z is the mirror image of z in the real axis. The absolute value of z is the distance from (a, b) to the origin.

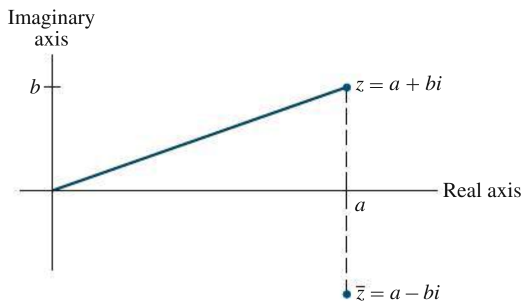


FIGURE 1 The complex conjugate is a mirror image.

Addition of complex numbers $z = a + bi$ and $w = c + di$ corresponds to vector addition of (a, b) and (c, d) in $\mathbb{R}^2$, as in Figure 2.

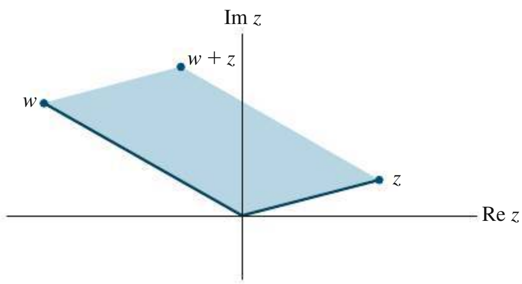


FIGURE 2 Addition of complex numbers.

To give a graphical representation of complex multiplication, we use **polar coordinates** in $\mathbb{R}^2$. Given a nonzero complex number $z = a + bi$, let φ be the angle between the positive real axis and the point (a, b), as in Figure 3 where $-\pi < \varphi \leq \pi$. The angle φ is called the **argument** of z; we write $\varphi = \arg z$. From trigonometry,

$$a = |z| \cos \varphi, \qquad b = |z| \sin \varphi$$

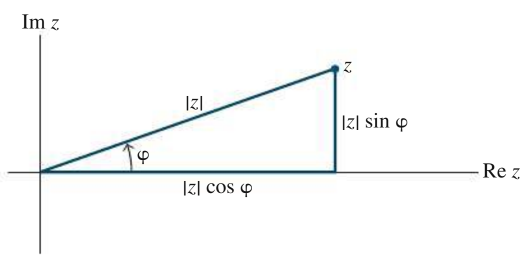


FIGURE 3 Polar coordinates of z.

and so

$$z = a + bi = |z|(\cos\varphi + i\sin\varphi)$$

If w is another nonzero complex number, say,

$$w = |w|\,(\cos\vartheta + i\sin\vartheta)$$

then, using standard trigonometric identities for the sine and cosine of the sum of two angles, one can verify that

$$wz = |w|\,|z|\,[\cos(\vartheta+\varphi) + i\sin(\vartheta+\varphi)] \tag{4}$$

See Figure 4. A similar formula may be written for quotients in polar form. The formulas for products and quotients can be stated in words as follows.

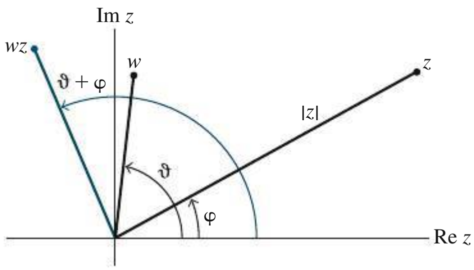
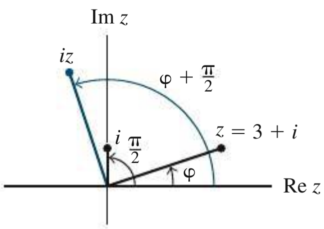


FIGURE 4 Multiplication with polar coordinates.

> The product of two nonzero complex numbers is given in polar form by the product of their absolute values and the sum of their arguments. The quotient of two nonzero complex numbers is given by the quotient of their absolute values and the difference of their arguments.



Multiplication by i.

EXAMPLE 4

a. If w has absolute value 1, then $w = \cos\vartheta + i\sin\vartheta$, where ϑ is the argument of w. Multiplication of any nonzero number z by w simply rotates z through the angle ϑ.

b. The argument of i itself is $\pi/2$ radians, so multiplication of z by i rotates z through an angle of $\pi/2$ radians. For example, $3+i$ is rotated into $(3+i)i = -1+3i$. ■

Powers of a Complex Number

Formula (4) applies when $z = w = r(\cos\varphi + i\sin\varphi)$. In this case

$$z^2 = r^2(\cos 2\varphi + i\sin 2\varphi)$$

and

$$\begin{aligned} z^3 &= z\cdot z^2 \\ &= r(\cos\varphi + i\sin\varphi)\cdot r^2(\cos 2\varphi + i\sin 2\varphi) \\ &= r^3(\cos 3\varphi + i\sin 3\varphi) \end{aligned}$$

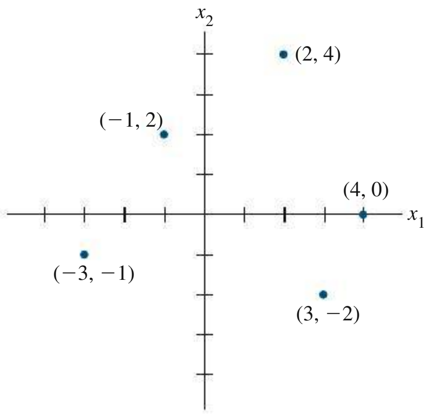
In general, for any positive integer k,

$$z^k = r^k(\cos k\varphi + i\sin k\varphi)$$

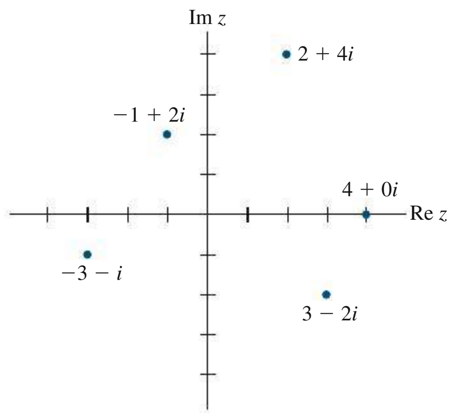
This fact is known as *De Moivre's Theorem*.

Complex Numbers and $\mathbb{R}^2$

Although the elements of $\mathbb{R}^2$ and $\mathbb{C}$ are in one-to-one correspondence, and the operations of addition are essentially the same, there is a logical distinction between $\mathbb{R}^2$ and $\mathbb{C}$. In $\mathbb{R}^2$ we can only multiply a vector by a real scalar, whereas in $\mathbb{C}$ we can multiply any two complex numbers to obtain a third complex number. (The dot product in $\mathbb{R}^2$ doesn't count, because it produces a scalar, not an element of $\mathbb{R}^2$.) We use scalar notation for elements in $\mathbb{C}$ to emphasize this distinction.



The real plane $\mathbb{R}^2$.



The complex plane $\mathbb{C}$.

Credits

Cover

DigiPub/Moment/Getty Images

Chapter 1

1, Hulton Archive/Keystone/Stringer/Getty Images; **54**, Olivier Le Queinec/Shutterstock; **54**, AButyrin 22/Shutterstock; **54**, SasinTipchai/Shutterstock; **58**, Shutterstock; Kurhan/Shutterstock; **86**, Ingredients in the diet as of 1984; nutrient data for ingredients adapted from USDA Agricultural Handbooks No. 8-1 and 8-6, 1976; **90**, Josef Hanus/Shutterstock; **90**, Gary Blakeley/Shutterstock

Chapter 2

97, LANBO/Shutterstock; **98**, Stocktrek Images, Inc/Alamy Stock Photo; **126**, Radu Bercan/Shutterstock; **131**, NASA; **145**, Fotokostic/Shutterstock; **145**, Rainer Plendl/Shutterstock; **145**, Designpics/123RF; **145**, Dotshock/Shutterstock; **146**, Wassily W. Leontief, "The Structure of the U.S. Economy," *Scientific American*, April 1965, pp. 30–32; **150**, Alexander Raths/Shutterstock

Chapter 3

171, Akkaradet Bangchun/Shutterstock; **171**, Zhengzaishuru/Shutterstock

Chapter 4

201, Interneteable/Shutterstock; **236**, Ifong/Shutterstock

Chapter 5

273, Digitalmedia.fws.gov; **336**, Archana bhartia/Shutterstock; **336**, Noah Strycker/Shutterstock

Chapter 6

349, Courtesy of Lester McDonald; **397**, Valerio Pardi/123RF; **397**, Jess Alford/Photodisc/Getty Images

Chapter 7

417, Landsat Data/U.S. Geological Survey; **418**, (Left): Landsat Data/U.S. Geological Survey; (Middle): Landsat Data/U.S. Geological Survey; (Right): Landsat Data/U.S. Geological Survey; (Left 2): Landsat Data/U.S. Geological Survey; (Middle 2): Landsat Data/U.S. Geological Survey; (Right 2): Landsat Data/U.S. Geological Survey; **436**, Onixxino/Getty Images; **436**, Dmitry Kalinovsky/Shutterstock; **436**, Viki2win/Shutterstock

Chapter 8

459, Canali/Index/Heritage Image Partnership Ltd/Alamy Stock Photo; **459**, City of Athens in 387 B.C., the Greek philosopher Plato

Chapter 9

521, CSU Archives/Everett Collection Inc/Alamy Stock Photo

Chapter 10 **(online)**

C-1, Dpa picture alliance/Alamy Stock Photo

Answers to Odd-Numbered Exercises

Chapter 1

Section 1.1, page 10

1. The solution is $(x_1, x_2) = (-8, 3)$, or simply $(-8, 3)$.

3. $(4/7, 9/7)$

5. Replace row 2 by its sum with 3 times row 3, and then replace row 1 by its sum with -5 times row 3.

7. The solution set is empty.

9. No solutions **11.** $(19, -8, 1)$

13. $(5, 3, -1)$

15.
$$\begin{array}{rcrcrcr} & & -8 & + & 4(1) & = & -4 \\ 19 & + & 3(-8) & + & 3(1) & = & -2 \\ 3(19) & + & 7(-8) & + & 5(1) & = & 6 \end{array}$$

17.
$$\begin{array}{rcrcrcr} (5) & & & - & 3(-1) & = & 8 \\ 2(5) & + & 2(3) & + & 9(-1) & = & 7 \\ & & (3) & + & 5(-1) & = & -2 \end{array}$$

19. Consistent

21. The three lines have one point in common.

23. $h \neq 2$ **25.** All h

27–33. Mark a statement True only if the statement is *always* true. Giving you the answers here would defeat the purpose of the true-false questions, which is to help you learn to read the text carefully. The *Study Guide* will tell you where to look for the answers, but you should not consult it until you have made an honest attempt to find the answers yourself.

35. $k + 2g + h = 0$

37. The row reduction of $\begin{bmatrix} 1 & 3 & f \\ c & d & g \end{bmatrix}$ to $\begin{bmatrix} 1 & 3 & f \\ 0 & d - 3c & g - cf \end{bmatrix}$ shows that $d - 3c$ must be nonzero, since f and g are arbitrary. Otherwise, for some choices of f and g the second row could correspond to an equation of the form $0 = b$, where b is nonzero. Thus $d \neq 3c$.

39. Swap row 1 and row 2; swap row 1 and row 2.

41. Replace row 3 by row 3 + (-4) row 1; replace row 3 by row 3 + (4) row 1.

43.
$$\begin{array}{rcrcrcrcl} 4T_1 & - & T_2 & & & - & T_4 & = & 30 \\ -T_1 & + & 4T_2 & - & T_3 & & & = & 60 \\ & & -T_2 & + & 4T_3 & - & T_4 & = & 70 \\ -T_1 & & & - & T_3 & + & 4T_4 & = & 40 \end{array}$$

Section 1.2, page 23

1. Reduced echelon form: a and c. Echelon form: b and d.

3. $\begin{bmatrix} \boxed{1} & 0 & -1 & -2 \\ 0 & \boxed{1} & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Pivot cols 1 and 2: $\begin{bmatrix} \boxed{1} & 2 & 3 & 4 \\ 4 & \boxed{5} & 6 & 7 \\ 6 & 7 & 8 & 9 \end{bmatrix}$.

5. $\begin{bmatrix} \blacksquare & * \\ 0 & \blacksquare \end{bmatrix}$, $\begin{bmatrix} \blacksquare & * \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & \blacksquare \\ 0 & 0 \end{bmatrix}$

7. $\begin{cases} x_1 = -5 - 3x_2 \\ x_2 \text{ is free} \\ x_3 = 3 \end{cases}$ **9.** $\begin{cases} x_1 = 6 + 5x_3 \\ x_2 = 5 + 6x_3 \\ x_3 \text{ is free} \end{cases}$

11. $\begin{cases} x_1 = \frac{4}{3}x_2 - \frac{2}{3}x_3 \\ x_2 \text{ is free} \\ x_3 \text{ is free} \end{cases}$

13. $\begin{cases} x_1 = -3 + 3x_5 \\ x_2 = 1 + 4x_5 \\ x_3 \text{ is free} \\ x_4 = -4 - 9x_5 \\ x_5 \text{ is free} \end{cases}$

Note: The *Study Guide* discusses the common mistake $x_3 = 0$.

15. $\begin{array}{rcrcrcr} & & x_2 & - & 6x_3 & = & 5 \\ x_1 & - & 2x_2 & + & 7x_3 & = & -4 \end{array}$, verify

$$\begin{array}{rcrcrcr} & & 5+6x_3 & - & 6x_3 & = & 5 \\ (6+5x_3) & - & 2(5+6x_3) & + & 7x_3 & = & -4 \end{array}$$

17. $\begin{array}{rcrcrcr} 3x_1 & - & 4x_2 & + & 2x_3 & = & 0 \\ -9x_1 & + & 12x_2 & - & 6x_3 & = & 0 \\ -6x_1 & + & 8x_2 & - & 4x_3 & = & 0 \end{array}$, verify

$$\begin{array}{rcrcrcr} 3\left(\frac{4}{3}x_2 - \frac{2}{3}x_3\right) & - & 4x_2 & + & 2x_3 & = & 0 \\ -9\left(\frac{4}{3}x_2 - \frac{2}{3}x_3\right) & + & 12x_2 & - & 6x_3 & = & 0 \\ -6\left(\frac{4}{3}x_2 - \frac{2}{3}x_3\right) & + & 8x_2 & - & 4x_3 & = & 0 \end{array}$$

19. **a.** Consistent, with a unique solution

b. Inconsistent

21. $h = 7/2$

23. **a.** Inconsistent when $h = 2$ and $k \neq 8$

b. A unique solution when $h \neq 2$

c. Many solutions when $h = 2$ and $k = 8$

25–33. Read the text carefully, and write your answers before you consult the *Study Guide*. Remember, a statement is true only if it is true in all cases.

35. Yes. The system is consistent because with three pivots, there must be a pivot in the third (bottom) row of the coefficient matrix. The reduced echelon form cannot contain a row of the form $[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1]$.

37. If the coefficient matrix has a pivot position in every row, then there is a pivot position in the bottom row, and there is no room for a pivot in the augmented column. So, the system is consistent, by Theorem 2.

39. If a linear system is consistent, then the solution is unique if and only if *every column in the coefficient matrix is a pivot column; otherwise, there are infinitely many solutions.*

41. An underdetermined system always has more variables than equations. There cannot be more basic variables than there are equations, so there must be at least one free variable. Such a variable may be assigned infinitely many different values. If the system is consistent, each different value of a free variable will produce a different solution.

43. Yes, a system of linear equations with more equations than unknowns can be consistent. The following system has a solution ($x_1 = x_2 = 1$):

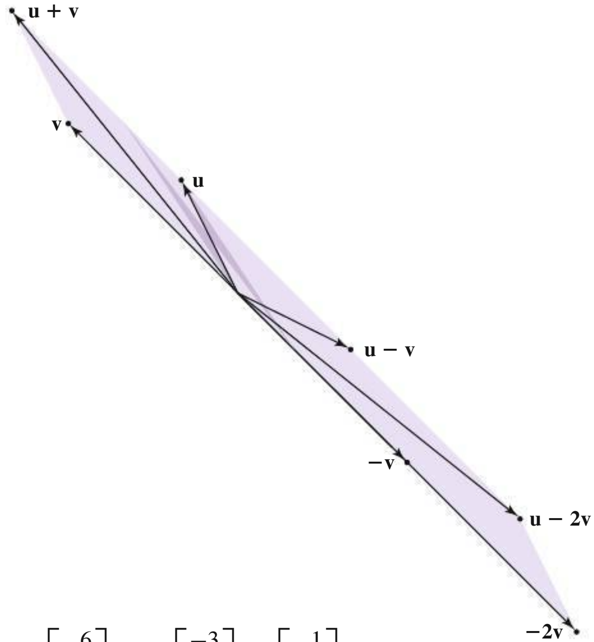
$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 0 \\ 3x_1 + 2x_2 &= 5 \end{aligned}$$

45. $p(t) = 7 + 6t - t^2$

Section 1.3, page 34

1. $\begin{bmatrix} -4 \\ 5 \end{bmatrix}, \begin{bmatrix} 5 \\ -4 \end{bmatrix}$

3.



5. $x_1 \begin{bmatrix} 6 \\ -1 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ -5 \end{bmatrix}$,

$\begin{bmatrix} 6x_1 \\ -x_1 \\ 5x_1 \end{bmatrix} + \begin{bmatrix} -3x_2 \\ 4x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ -5 \end{bmatrix}, \begin{bmatrix} 6x_1 - 3x_2 \\ -x_1 + 4x_2 \\ 5x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ -5 \end{bmatrix}$

$$\begin{array}{rcrcr} 6x_1 & - & 3x_2 & = & 1 \\ -x_1 & + & 4x_2 & = & -7 \\ 5x_1 & & & = & -5 \end{array}$$

Usually the intermediate steps are not displayed.

7. $\mathbf{a} = \mathbf{u} - 2\mathbf{v}$, $\mathbf{b} = 2\mathbf{u} - 2\mathbf{v}$, $\mathbf{c} = 2\mathbf{u} - 3.5\mathbf{v}$, $\mathbf{d} = 3\mathbf{u} - 4\mathbf{v}$

9. $x_1 \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 6 \\ 3 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ -1 \\ -8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

11. Yes, $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

13. No, $\mathbf{b}$ is *not* a linear combination of the columns of A.

15. Noninteger weights are acceptable, of course, but some simple choices are $0 \cdot \mathbf{v}_1 + 0 \cdot \mathbf{v}_2 = \mathbf{0}$, and

$$1 \cdot \mathbf{v}_1 + 0 \cdot \mathbf{v}_2 = \begin{bmatrix} 7 \\ 1 \\ -6 \end{bmatrix}, 0 \cdot \mathbf{v}_1 + 1 \cdot \mathbf{v}_2 = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}$$

$$1 \cdot \mathbf{v}_1 + 1 \cdot \mathbf{v}_2 = \begin{bmatrix} 2 \\ 4 \\ -6 \end{bmatrix}, 1 \cdot \mathbf{v}_1 - 1 \cdot \mathbf{v}_2 = \begin{bmatrix} 12 \\ -2 \\ -6 \end{bmatrix}$$

17. $h = -17$

19. Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ is the set of points on the line through $\mathbf{v}_1$ and $\mathbf{0}$.

21. *Hint:* Show that $\begin{bmatrix} 2 & 2 & h \\ -1 & 1 & k \end{bmatrix}$ is consistent for all h and k. Explain what this calculation shows about Span $\{\mathbf{u}, \mathbf{v}\}$.

23–31. Before you consult your *Study Guide*, read the entire section carefully. Pay special attention to definitions and theorem statements, and note any remarks that precede or follow them.

33. **a.** No, three **b.** Yes, infinitely many
c. $\mathbf{a}_1 = 1 \cdot \mathbf{a}_1 + 0 \cdot \mathbf{a}_2 + 0 \cdot \mathbf{a}_3$

35. **a.** $5\mathbf{v}_1$ is the output of 5 day's operation of mine #1.
b. The total output is $x_1\mathbf{v}_1 + x_2\mathbf{v}_2$, so x_1 and x_2 should satisfy $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \begin{bmatrix} 150 \\ 2825 \end{bmatrix}$.
c. 1.5 days for mine #1 and 4 days for mine #2

37. $(1.3, .9, 0)$

39. **a.** $\begin{bmatrix} 10/3 \\ 2 \end{bmatrix}$
b. Add 3.5 g at $(0, 1)$, add .5 g at $(8, 1)$, and add 2 g at $(2, 4)$.

41. Review Practice Problem 1 and then *write* a solution. The *Study Guide* has a solution.

Section 1.4, page 42

1. The product is not defined because the number of columns (2) in the 3×2 matrix does not match the number of entries (3) in the vector.

3. $$A\mathbf{x} = \begin{bmatrix} 6 & 5 \\ -4 & -3 \\ 7 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix} = 1\begin{bmatrix} 6 \\ -4 \\ 7 \end{bmatrix} - 3\begin{bmatrix} 5 \\ -3 \\ 6 \end{bmatrix}$$
$$= \begin{bmatrix} 6 \\ -4 \\ 7 \end{bmatrix} + \begin{bmatrix} -15 \\ 9 \\ -18 \end{bmatrix} = \begin{bmatrix} -9 \\ 5 \\ -11 \end{bmatrix}, \text{ and}$$
$$A\mathbf{x} = \begin{bmatrix} 6 & 5 \\ -4 & -3 \\ 7 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 6(1) + 5(-3) \\ (-4)(1) + (-3)(-3) \\ 7(1) + 6(-3) \end{bmatrix}$$
$= \begin{bmatrix} -9 \\ 5 \\ -11 \end{bmatrix}$. Show your work here and for Exercises 4–6, but thereafter perform the calculations mentally.

5. $5\begin{bmatrix} 5 \\ -2 \end{bmatrix} - 1\begin{bmatrix} 1 \\ -7 \end{bmatrix} + 3\begin{bmatrix} -8 \\ 3 \end{bmatrix} - 2\begin{bmatrix} 4 \\ -5 \end{bmatrix} = \begin{bmatrix} -8 \\ 16 \end{bmatrix}$

7. $$\begin{bmatrix} 4 & -5 & 7 \\ -1 & 3 & -8 \\ 7 & -5 & 0 \\ -4 & 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -8 \\ 0 \\ -7 \end{bmatrix}$$

9. $x_1\begin{bmatrix} 3 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} -5 \\ 4 \end{bmatrix} = \begin{bmatrix} 9 \\ 0 \end{bmatrix}$ and
$$\begin{bmatrix} 3 & 1 & -5 \\ 0 & 1 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 0 \end{bmatrix}$$

11. $$\begin{bmatrix} 1 & 2 & 4 & -2 \\ 0 & 1 & 5 & 2 \\ -2 & -4 & -3 & 9 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix}$$

13. Yes. (Justify your answer.)

u are here

15. The equation $A\mathbf{x} = \mathbf{b}$ is not consistent when $3b_1 + b_2$ is nonzero. (Show your work.) The set of $\mathbf{b}$ for which the equation *is* consistent is a line through the origin—the set of all points (b_1, b_2) satisfying $b_2 = -3b_1$.

17. Only three rows contain a pivot position. The equation $A\mathbf{x} = \mathbf{b}$ does *not* have a solution for each $\mathbf{b}$ in $\mathbb{R}^4$, by Theorem 4.

19. The work in Exercise 17 shows that statement (d) in Theorem 4 is false. So all four statements in Theorem 4 are false. Thus, not all vectors in $\mathbb{R}^4$ can be written as a linear combination of the columns of A. Also, the columns of A do *not* span $\mathbb{R}^4$.

21. The matrix $[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]$ does not have a pivot in each row, so the columns of the matrix do not span $\mathbb{R}^4$, by Theorem 4. That is, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ does not span $\mathbb{R}^4$.

23–33. Read the text carefully and try to mark each exercise statement True or False before you consult the *Study Guide*. Several parts of Exercises 23–24 are *implications* of the form

"If ⟨statement 1⟩, then ⟨statement 2⟩"

or equivalently,

"⟨statement 2⟩, if ⟨statement 1⟩"

Mark such an implication as True if ⟨statement 2⟩ is true in all cases when ⟨statement 1⟩ is true.

35. $c_1 = -3, c_2 = -1, c_3 = 2$

37. $Q\mathbf{x} = \mathbf{v}$, where $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3]$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

Note: If your answer is the equation $A\mathbf{x} = \mathbf{b}$, you must specify what A and $\mathbf{b}$ are.

39. *Hint:* Start with any 3×3 matrix B in echelon form that has three pivot positions.

41. *Write* your solution before you check the *Study Guide*.

43. *Hint:* How many pivot columns does A have? Why?

45. Given $A\mathbf{x}_1 = \mathbf{y}_1$ and $A\mathbf{x}_2 = \mathbf{y}_2$, you are asked to show that the equation $A\mathbf{x} = \mathbf{w}$ has a solution, where $\mathbf{w} = \mathbf{y}_1 + \mathbf{y}_2$. Observe that $\mathbf{w} = A\mathbf{x}_1 + A\mathbf{x}_2$ and use Theorem 5(a) with $\mathbf{x}_1$ and $\mathbf{x}_2$ in place of $\mathbf{u}$ and $\mathbf{v}$, respectively. That is, $\mathbf{w} = A\mathbf{x}_1 + A\mathbf{x}_2 = A(\mathbf{x}_1 + \mathbf{x}_2)$. So the vector $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ is a solution of $\mathbf{w} = A\mathbf{x}$.

47. The columns do not span $\mathbb{R}^4$.

49. The columns span $\mathbb{R}^4$.

51. Delete column 4 of the matrix in Exercise 49. It is also possible to delete column 3 instead of column 4.

Section 1.5, page 51

1. The system has a nontrivial solution because there is a free variable, x_3.

3. The system has a nontrivial solution because there is a free variable, x_3.

5. $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$

7. $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -9 \\ 4 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 8 \\ -5 \\ 0 \\ 1 \end{bmatrix}$

9. $\mathbf{x} = x_2 \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$

11. *Hint:* The system derived from the *reduced* echelon form is

$$\begin{aligned} x_1 - 4x_2 \qquad + 5x_6 &= 0 \\ x_3 \quad - \; x_6 &= 0 \\ x_5 - 4x_6 &= 0 \\ 0 &= 0 \end{aligned}$$

The basic variables are x_1, x_3, and x_5. The remaining variables are free. The *Study Guide* discusses two mistakes that are often made on this type of problem.

13. $\begin{bmatrix} 3 & -9 & 6 \\ -1 & 3 & -2 \end{bmatrix} \left(x_2 \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right) =$

$x_2 \begin{bmatrix} 3 & -9 & 6 \\ -1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} +$

$x_3 \begin{bmatrix} 3 & -9 & 6 \\ -1 & 3 & -2 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$

$= x_2 \begin{bmatrix} 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

15. $\begin{bmatrix} 1 & -4 & -2 & 0 & 3 & -5 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \left(x_2 \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} -5 \\ 0 \\ 1 \\ 0 \\ 4 \\ 1 \end{bmatrix} \right)$

$= x_2 \begin{bmatrix} 1 & -4 & -2 & 0 & 3 & -5 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} +$

$x_4 \begin{bmatrix} 1 & -4 & -2 & 0 & 3 & -5 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} +$

$x_6 \begin{bmatrix} 1 & -4 & -2 & 0 & 3 & -5 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -5 \\ 0 \\ 1 \\ 0 \\ 4 \\ 1 \end{bmatrix}$

$= x_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

17. $\mathbf{x} = \begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix} = \mathbf{p} + x_3\mathbf{q}$. Geometrically, the solution set is the line through $\begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix}$ parallel to $\begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix}$.

19. $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$. The solution set is the line through $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$, parallel to the line that is the solution set of the homogeneous system in Exercise 5.

21. Let $\mathbf{u} = \begin{bmatrix} -9 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}$. The solution of the homogeneous equation is $\mathbf{x} = x_2\mathbf{u} + x_3\mathbf{v}$, the plane through the origin spanned by $\mathbf{u}$ and $\mathbf{v}$. The solution set of the nonhomogeneous system is $\mathbf{x} = \mathbf{p} + x_2\mathbf{u} + x_3\mathbf{v}$, the plane through $\mathbf{p}$ parallel to the solution set of the homogeneous equation.

23. $\mathbf{x} = \mathbf{a} + t\mathbf{b}$, where t represents a parameter, or $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix} + t \begin{bmatrix} -5 \\ 3 \end{bmatrix}$, or $\begin{cases} x_1 = -2 - 5t \\ x_2 = 3t \end{cases}$

25. $\mathbf{x} = \mathbf{p} + t(\mathbf{q} - \mathbf{p}) = \begin{bmatrix} 2 \\ -5 \end{bmatrix} + t\begin{bmatrix} -5 \\ 6 \end{bmatrix}$

27–35. It is important to read the text carefully and write your answers. After that, check the *Study Guide*, if necessary.

37. $A\mathbf{v}_h = A(\mathbf{w} - \mathbf{p}) = A\mathbf{w} - A\mathbf{p} = \mathbf{b} - \mathbf{b} = \mathbf{0}$

39. When A is the 3×3 zero matrix, *every* $\mathbf{x}$ in $\mathbb{R}^3$ satisfies $A\mathbf{x} = \mathbf{0}$. So the solution set is all vectors in $\mathbb{R}^3$.

41. a. When A is a 3×3 matrix with three pivot positions, the equation $A\mathbf{x} = \mathbf{0}$ has no free variables and hence has no nontrivial solution.

b. With three pivot positions, A has a pivot position in each of its three rows. By Theorem 4 in Section 1.4, the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every possible $\mathbf{b}$. The word "possible" in the exercise means that the only vectors considered in this case are those in $\mathbb{R}^3$, because A has three rows.

43. a. When A is a 3×2 matrix with two pivot positions, each column is a pivot column. So the equation $A\mathbf{x} = \mathbf{0}$ has no free variables and hence no nontrivial solution.

b. With two pivot positions and three rows, A cannot have a pivot in every row. So the equation $A\mathbf{x} = \mathbf{b}$ cannot have a solution for every possible $\mathbf{b}$ (in $\mathbb{R}^3$), by Theorem 4 in Section 1.4.

45. One answer: $\mathbf{x} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$

47. Your example should have the property that the sum of the entries in each row is zero. Why?

49. One answer is $A = \begin{bmatrix} 1 & -4 \\ 1 & -4 \end{bmatrix}$. The *Study Guide* shows how to analyze the problem in order to construct A. If $\mathbf{b}$ is any vector *not* a multiple of the first column of A, then the solution set of $A\mathbf{x} = \mathbf{b}$ is empty and thus cannot be formed by translating the solution set of $A\mathbf{x} = \mathbf{b}$. This does not contradict Theorem 6, because that theorem applies when the equation $A\mathbf{x} = \mathbf{b}$ has a nonempty solution set.

51. If c is a scalar, then $A(c\mathbf{u}) = cA\mathbf{u}$, by Theorem 5(b) in Section 1.4. If $\mathbf{u}$ satisfies $A\mathbf{x} = \mathbf{0}$, then $A\mathbf{u} = \mathbf{0}$, $cA\mathbf{u} = c \cdot \mathbf{0} = \mathbf{0}$, and so $A(c\mathbf{u}) = \mathbf{0}$.

Section 1.6, page 58

1. The general solution is $p_{\text{Goods}} = .875p_{\text{Services}}$, with p_{Services} free. One equilibrium solution is $p_{\text{Services}} = 1000$ and $p_{\text{Goods}} = 875$. Using fractions, the general solution could be written $p_{\text{Goods}} = (7/8)p_{\text{Services}}$, and a natural choice of prices might be $p_{\text{Services}} = 80$ and $p_{\text{Goods}} = 70$. Only the *ratio* of the prices is important. The economic equilibrium is unaffected by a proportional change in prices.

3. a.

	Distribution of Output From				
	C&M	F&P	Mach.		
Output	↓	↓	↓	Input	Purchased By
	.2	.8	.4	→	C&M
	.3	.1	.4	→	F&P
	.5	.1	.2	→	Mach.

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

c. $p_{\text{Chemicals}} = 141.7$, $p_{\text{Fuels}} = 91.7$, $p_{\text{Machinery}} = 100$. To two significant figures, $p_{\text{Chemicals}} = 140$, $p_{\text{Fuels}} = 92$, $p_{\text{Machinery}} = 100$.

5. $B_2S_3 + 6H_2O \rightarrow 2H_3BO_3 + 3H_2S$

7. $3NaHCO_3 + H_3C_6H_5O_7 \rightarrow Na_3C_6H_5O_7 + 3H_2O + 3CO_2$

9. $15PbN_6 + 44CrMn_2O_8 \rightarrow$ $5Pb_3O_4 + 22Cr_2O_3 + 88MnO_2 + 90NO$

11. $\begin{cases} x_1 = 20 - x_3 \\ x_2 = 60 + x_3 \\ x_3 \text{ is free} \\ x_4 = 60 \end{cases}$ The largest value of x_3 is 20.

13. a. $\begin{cases} x_1 = x_3 - 40 \\ x_2 = x_3 + 10 \\ x_3 \text{ is free} \\ x_4 = x_6 + 50 \\ x_5 = x_6 + 60 \\ x_6 \text{ is free} \end{cases}$ **b.** $\begin{cases} x_2 = 50 \\ x_3 = 40 \\ x_4 = 50 \\ x_5 = 60 \end{cases}$

Section 1.7, page 65

Justify your answers to Exercises 1–22.

1. Lin. dep. **3.** Lin. indep.

5. Lin. indep. **7.** Lin. depen.

9. a. $h = 4$ **b.** $h = 4$

11. $h = 6$ **13.** All h

15. Lin. depen. **17.** Lin. depen. **19.** Lin. indep.

21–27. Read through the definitions, examples, and theorems for this section before you consult the *Study Guide*.

29. $\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & \blacksquare \end{bmatrix}$ **31.** $\begin{bmatrix} \blacksquare & * \\ 0 & \blacksquare \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & \blacksquare \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

33. All five columns of the 7×5 matrix A must be pivot columns. Otherwise, the equation $A\mathbf{x} = \mathbf{0}$ would have a free variable, in which case the columns of A would be linearly dependent.

35. A: Any 3×2 matrix with two nonzero columns such that neither column is a multiple of the other. In this case, the columns are linearly independent, and so the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
B: Any 3×2 matrix with one column a multiple of the other.

37. $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

39. True, by Theorem 7. (The *Study Guide* adds another justification.)

41. False. The vector $\mathbf{v}_1$ could be the zero vector.

43. True. A linear dependence relation among $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ may be extended to a linear dependence relation among $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ by placing a zero weight on $\mathbf{v}_4$.

45. You should be able to work this important problem without help. *Write* your solution before you consult the *Study Guide*.

47. $B = \begin{bmatrix} 8 & -3 & 2 \\ -9 & 4 & -7 \\ 6 & -2 & 4 \\ 5 & -1 & 10 \end{bmatrix}$. Other choices are possible.

49. Each column of A that is not a column of B is in the set spanned by the columns of B.

Section 1.8, page 73

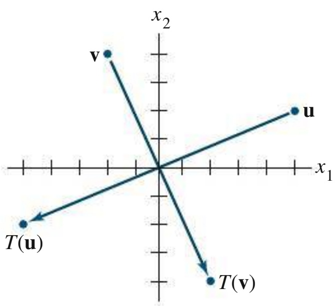
1. $\begin{bmatrix} 2 \\ -6 \end{bmatrix}, \begin{bmatrix} 2a \\ 2b \end{bmatrix}$ **3.** $\mathbf{x} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$, unique solution

5. $\mathbf{x} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$, not unique **7.** $a = 5, b = 6$

9. $\mathbf{x} = x_3 \begin{bmatrix} 9 \\ 4 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -7 \\ -3 \\ 0 \\ 1 \end{bmatrix}$

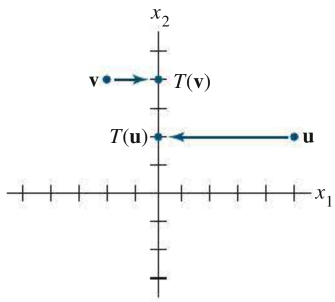
11. Yes, because the system represented by $[A \quad \mathbf{b}]$ is consistent.

13.



A reflection through the origin

15.

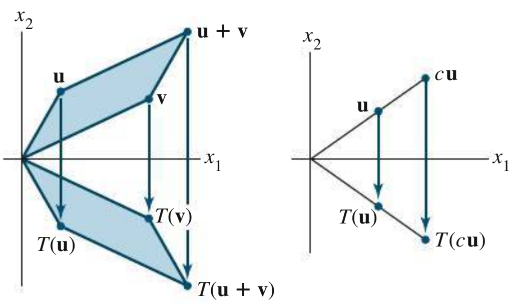


A projection onto the x_2-axis.

17. $\begin{bmatrix} 6 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ 9 \end{bmatrix}$ **19.** $\begin{bmatrix} 13 \\ 7 \end{bmatrix}, \begin{bmatrix} 2x_1 - x_2 \\ 5x_1 + 6x_2 \end{bmatrix}$

21–29. If you consult your *Study Guide* before you make a good effort to answer the true-false questions, you will destroy most of their value.

31.



33. *Hint:* Show that the image of a line (that is, the set of images of all points on a line) can be represented by the parametric equation of a line.

35. **a.** The line through $\mathbf{p}$ and $\mathbf{q}$ is parallel to $\mathbf{q} - \mathbf{p}$. (See Exercises 25 and 26 in Section 1.5.) Since $\mathbf{p}$ is on the line, the equation of the line is $\mathbf{x} = \mathbf{p} + t(\mathbf{q} - \mathbf{p})$. Rewrite this as $\mathbf{x} = \mathbf{p} - t\mathbf{p} + t\mathbf{q}$ and $\mathbf{x} = (1-t)\mathbf{p} + t\mathbf{q}$.

b. Consider $\mathbf{x} = (1-t)\mathbf{p} + t\mathbf{q}$ for t such that $0 \le t \le 1$. Then, by linearity of T, for $0 \le t \le 1$

$$T(\mathbf{x}) = T((1-t)\mathbf{p} + t\mathbf{q}) = (1-t)T(\mathbf{p}) + tT(\mathbf{q}) \quad (*)$$

If $T(\mathbf{p})$ and $T(\mathbf{q})$ are distinct, then (*) is the equation for the line segment between $T(\mathbf{p})$ and $T(\mathbf{q})$, as shown in part (a). Otherwise, the set of images is just the single point $T(\mathbf{p})$, because

$$(1-t)T(\mathbf{p}) + tT(\mathbf{q}) = (1-t)T(\mathbf{p}) + tT(\mathbf{p}) = T(\mathbf{p})$$

37. **a.** When $b = 0$, $f(x) = mx$. In this case, for all x, y in $\mathbb{R}$ and all scalars c and d,

$$\begin{aligned} f(cx + dy) &= m(cx + dy) = mcx + mdy \\ &= c(mx) + d(my) = c \cdot f(x) + d \cdot f(y) \end{aligned}$$

This shows that f is linear.

b. When $f(x) = mx + b$, with b nonzero, $f(0) = m(0) + b = b \neq 0$.

c. In calculus, f is called a "linear function" because the graph of f is a line.

39. *Hint:* Since $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent, you can write a certain equation and work with it.

41. One possibility is to show that T does not map the zero vector into the zero vector, something that every linear transformation *does* do: $T(0, 0) = (0, 4, 0)$.

43. Take $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$ and let c and d be scalars. Then

$$c\mathbf{u} + d\mathbf{v} = (cu_1 + dv_1, cu_2 + dv_2, cu_3 + dv_3)$$

The transformation T is linear because

$$\begin{aligned} T(c\mathbf{u} + d\mathbf{v}) &= (cu_1 + dv_1, cu_2 + dv_2, -(cu_3 + dv_3)) \\ &= (cu_1 + dv_1, cu_2 + dv_2, -cu_3 - dv_3) \\ &= (cu_1, cu_2, -cu_3) + (dv_1, dv_2, -dv_3) \\ &= c(u_1, u_2, -u_3) + d(v_1, v_2, -v_3) \\ &= cT(\mathbf{u}) + dT(\mathbf{v}) \end{aligned}$$

45. All multiples of $(7, 9, 0, 2)$

47. Yes. One choice for $\mathbf{x}$ is $(4, 7, 1, 0)$.

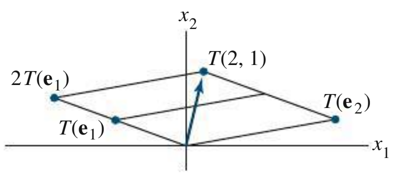
Section 1.9, page 82

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

7. $\begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$ **9.** $\begin{bmatrix} 0 & -1 \\ -1 & 3 \end{bmatrix}$

11. The described transformation T maps $\mathbf{e}_1$ into $-\mathbf{e}_1$ and maps $\mathbf{e}_2$ into $-\mathbf{e}_2$. A rotation through π radians also maps $\mathbf{e}_1$ into $-\mathbf{e}_1$ and maps $\mathbf{e}_2$ into $-\mathbf{e}_2$. Since a linear transformation is completely determined by what it does to the columns of the identity matrix, the rotation transformation has the same effect as T on every vector in $\mathbb{R}^2$.

13.



15. $\begin{bmatrix} 2 & 0 & -3 \\ 4 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix}$ **17.** $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

19. $\begin{bmatrix} 1 & -5 & 4 \\ 0 & 1 & -6 \end{bmatrix}$ **21.** $\mathbf{x} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$

23–31. Read the text carefully, and write your answers before you consult the *Study Guide*. Remember, a statement is true only if it is true in all cases.

Justify your answers to Exercises 33–35.

33. Not one-to-one and does not map $\mathbb{R}^4$ onto $\mathbb{R}^4$

35. Not one-to-one but maps $\mathbb{R}^3$ onto $\mathbb{R}^2$

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

39. n. (Explain why, and then check the *Study Guide*).

41. *Hint:* If $\mathbf{e}_j$ is the jth column of I_n, then $B\mathbf{e}_j$ is the jth column of B.

43. *Hint:* Is it possible that $m > n$? What about $m < n$?

45. No. (Explain why.)

47. No. (Explain why.)

Section 1.10, page 91

1. a. $x_1 \begin{bmatrix} 110 \\ 4 \\ 20 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 130 \\ 3 \\ 18 \\ 5 \end{bmatrix} = \begin{bmatrix} 295 \\ 9 \\ 48 \\ 8 \end{bmatrix}$, where x_1 is the number of servings of Cheerios and x_2 is the number of servings of 100% Natural Cereal.

b. $\begin{bmatrix} 110 & 130 \\ 4 & 3 \\ 20 & 18 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 295 \\ 9 \\ 48 \\ 8 \end{bmatrix}$. Mix 1.5 servings of Cheerios together with 1 serving of 100% Natural Cereal.

3. a. She should mix .99 serving of Mac and Cheese, 1.54 servings of broccoli, and .79 serving of chicken to get her desired nutritional content.

b. She should mix 1.09 servings of shells and white cheddar, .88 serving of broccoli, and 1.03 servings of chicken to get her desired nutritional content. Notice that this mix contains significantly less broccoli, so she should like it better.

5. $R\mathbf{i} = \mathbf{v}$, $\begin{bmatrix} 11 & -5 & 0 & 0 \\ -5 & 10 & -1 & 0 \\ 0 & -1 & 9 & -2 \\ 0 & 0 & -2 & 10 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 50 \\ -40 \\ 30 \\ -30 \end{bmatrix}$

$\mathbf{i} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 3.68 \\ -1.90 \\ 2.57 \\ -2.49 \end{bmatrix}$

7. $R\mathbf{i} = \mathbf{v}$, $\begin{bmatrix} 12 & -7 & 0 & -4 \\ -7 & 15 & -6 & 0 \\ 0 & -6 & 14 & -5 \\ -4 & 0 & -5 & 13 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 40 \\ 30 \\ 20 \\ -10 \end{bmatrix}$

$\mathbf{i} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 11.43 \\ 10.55 \\ 8.04 \\ 5.84 \end{bmatrix}$

9. $\mathbf{x}_{k+1} = M\mathbf{x}_k$ for $k = 0, 1, 2, \ldots,$ where

$$M = \begin{bmatrix} .93 & .05 \\ .07 & .95 \end{bmatrix} \text{ and } \mathbf{x}_0 = \begin{bmatrix} 800{,}000 \\ 500{,}000 \end{bmatrix}.$$

The population in 2022 (for $k = 2$) is $\mathbf{x}_2 = \begin{bmatrix} 741{,}720 \\ 558{,}280 \end{bmatrix}$.

11. 32 in Pullman, 76 in Spokane, and 212 in Seattle.

13. a. The population of the city decreases. After 7 years, the populations are about equal, but the city population continues to decline. After 20 years, there are only 417,000 persons in the city (417,456 rounded off). However, the changes in population seem to grow smaller each year.

b. The city population is increasing slowly, and the suburban population is decreasing. After 20 years, the city population has grown from 350,000 to about 370,000.

Chapter 1 Supplementary Exercises, page 93

1. F	**2.** F	**3.** T	**4.** F	**5.** T
6. T	**7.** F	**8.** F	**9.** T	**10.** F
11. T	**12.** F	**13.** T	**14.** T	**15.** T
16. T	**17.** F	**18.** T	**19.** F	**20.** T
21. F	**22.** F	**23.** F	**24.** T	**25.** T

27. a. Any consistent linear system whose echelon form is

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

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

b. Any consistent linear system whose reduced echelon form is I_3.

c. Any inconsistent linear system of three equations in three variables.

29. a. The solution set: (i) is empty if $h = 12$ and $k \neq 2$; (ii) contains a unique solution if $h \neq 12$; (iii) contains infinitely many solutions if $h = 12$ and $k = 2$.

b. The solution set is empty if $k + 3h = 0$; otherwise, the solution set contains a unique solution.

31. a. Set $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -5 \\ 7 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ -5 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ -3 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. "Determine if $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ span $\mathbb{R}^3$."
Solution: No.

b. Set $A = \begin{bmatrix} 2 & -4 & -2 \\ -5 & 1 & 1 \\ 7 & -5 & -3 \end{bmatrix}$. "Determine if the columns of A span $\mathbb{R}^3$."

c. Define $T(\mathbf{x}) = A\mathbf{x}$. "Determine if T maps $\mathbb{R}^3$ onto $\mathbb{R}^3$."

33. $\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \frac{4}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{7}{3}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ or $\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 8/3 \\ 4/3 \end{bmatrix} + \begin{bmatrix} 7/3 \\ 14/3 \end{bmatrix}$

34. *Hint:* Construct a "grid" on the x_1x_2-plane determined by $\mathbf{a}_1$ and $\mathbf{a}_2$.

35. A solution set is a line when the system has one free variable. If the coefficient matrix is 2×3, then two of the columns should be pivot columns. For instance, take $\begin{bmatrix} 1 & 2 & * \\ 0 & 3 & * \end{bmatrix}$. Put anything in column 3. The resulting matrix will be in echelon form. Make one row replacement operation on the second row to create a matrix *not* in echelon form, such as $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 1 \\ 1 & 5 & 2 \end{bmatrix}$.

36. *Hint:* How many free variables are in the equation $A\mathbf{x} = \mathbf{0}$?

37. $E = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$

39. a. If the three vectors are linearly independent, then a, c, and f must all be nonzero.

b. The numbers $a, \ldots, f$ can have any values.

40. *Hint:* List the columns from right to left as $\mathbf{v}_1, \ldots, \mathbf{v}_4$.

41. *Hint:* Use Theorem 7.

43. Let M be the line through the origin that is parallel to the line through $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Then $\mathbf{v}_2 - \mathbf{v}_1$ and $\mathbf{v}_3 - \mathbf{v}_1$ are both on M. So one of these two vectors is a multiple of the other, say $\mathbf{v}_2 - \mathbf{v}_1 = k(\mathbf{v}_3 - \mathbf{v}_1)$. This equation produces a linear dependence relation: $(k-1)\mathbf{v}_1 + \mathbf{v}_2 - k\mathbf{v}_3 = \mathbf{0}$.

A second solution: A parametric equation of the line is $\mathbf{x} = \mathbf{v}_1 + t(\mathbf{v}_2 - \mathbf{v}_1)$. Since $\mathbf{v}_3$ is on the line, there is some t_0 such that $\mathbf{v}_3 = \mathbf{v}_1 + t_0(\mathbf{v}_2 - \mathbf{v}_1) = (1 - t_0)\mathbf{v}_1 + t_0\mathbf{v}_2$. So $\mathbf{v}_3$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

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

47. $a = 4/5$ and $b = -3/5$

49. a. The vector lists the number of three-, two-, and one-bedroom apartments provided when x_1 floors of plan A are constructed.

b. $x_1 \begin{bmatrix} 3 \\ 7 \\ 8 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ 3 \\ 9 \end{bmatrix}$

c. Use 2 floors of plan A and 15 floors of plan B. Or, use 6 floors of plan A, 2 floors of plan B, and 8 floors of plan C. These are the only feasible solutions. There are other mathematical solutions, but they require a negative number of floors of one or two of the plans, which makes no physical sense.

Chapter 2

Section 2.1, page 108

1. $\begin{bmatrix} -4 & 0 & 2 \\ -8 & 6 & -4 \end{bmatrix}$, $\begin{bmatrix} 3 & -5 & 3 \\ -7 & 2 & -7 \end{bmatrix}$, not defined, $\begin{bmatrix} 1 & 13 \\ -7 & -6 \end{bmatrix}$

3. $\begin{bmatrix} -1 & 1 \\ -5 & 5 \end{bmatrix}$, $\begin{bmatrix} 12 & -3 \\ 15 & -6 \end{bmatrix}$

5. a. $A\mathbf{b}_1 = \begin{bmatrix} -7 \\ 7 \\ 12 \end{bmatrix}$, $A\mathbf{b}_2 = \begin{bmatrix} 6 \\ -16 \\ -11 \end{bmatrix}$,

$$AB = \begin{bmatrix} -7 & 6 \\ 7 & -16 \\ 12 & -11 \end{bmatrix}$$

b. $$AB = \begin{bmatrix} -1(3)+2(-2) & -1(-4)+2(1) \\ 5(3)+4(-2) & 5(-4)+4(1) \\ 2(3)-3(-2) & 2(-4)-3(1) \end{bmatrix} = \begin{bmatrix} -7 & 6 \\ 7 & -16 \\ 12 & -11 \end{bmatrix}$$

7. 3×7 **9.** $k = 5$

11. $AD = \begin{bmatrix} 2 & 3 & 5 \\ 2 & 6 & 15 \\ 2 & 12 & 25 \end{bmatrix}$, $DA = \begin{bmatrix} 2 & 2 & 2 \\ 3 & 6 & 9 \\ 5 & 20 & 25 \end{bmatrix}$

Right-multiplication (that is, multiplication on the right) by D multiplies each *column* of A by the corresponding diagonal entry of D. Left-multiplication by D multiplies each *row* of A by the corresponding diagonal entry of D. The *Study Guide* tells how to make $AB = BA$, but you should try this yourself before looking there.

13. *Hint:* One of the two matrices is Q.

15–23. Answer the questions before looking in the *Study Guide.*

25. $\mathbf{b}_1 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -8 \\ -5 \end{bmatrix}$

27. The third column of AB is the sum of the first two columns of AB. Here's why. Write $B = [\,\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3\,]$. By definition, the third column of AB is $A\mathbf{b}_3$. If $\mathbf{b}_3 = \mathbf{b}_1 + \mathbf{b}_2$, then $A\mathbf{b}_3 = A(\mathbf{b}_1 + \mathbf{b}_2) = A\mathbf{b}_1 + A\mathbf{b}_2$, by a property of matrix-vector multiplication.

29. The columns of A are linearly dependent. Why?

31. *Hint:* Suppose $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{0}$, and show that $\mathbf{x}$ must be $\mathbf{0}$.

33. *Hint:* Use the results of Exercises 31 and 32, and apply the associative law of multiplication to the product CAD.

35. $\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u} = -2a + 3b - 4c$,

$$\mathbf{u}\mathbf{v}^T = \begin{bmatrix} -2a & -2b & -2c \\ 3a & 3b & 3c \\ -4a & -4b & -4c \end{bmatrix},$$

$$\mathbf{v}\mathbf{u}^T = \begin{bmatrix} -2a & 3a & -4a \\ -2b & 3b & -4b \\ -2c & 3c & -4c \end{bmatrix}$$

37. *Hint:* For Theorem 2(b), show that the (i, j)-entry of $A(B + C)$ equals the (i, j)-entry of $AB + AC$.

39. *Hint:* Use the definition of the product $I_m A$ and the fact that $I_m\mathbf{x} = \mathbf{x}$ for $\mathbf{x}$ in $\mathbb{R}^m$.

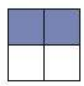
41. *Hint:* First write the (i, j)-entry of $(AB)^T$, which is the (j, i)-entry of AB. Then, to compute the (i, j)-entry in B^TA^T, use the facts that the entries in row i of B^T are $b_{1i}, \ldots, b_{ni}$, because they come from column i of B, and the entries in column j of A^T are $a_{j1}, \ldots, a_{jn}$, because they come from row j of A.

43. The answer here depends on the choice of matrix program. For MATLAB, use the `help` command to read about `zeros`, `ones`, `eye`, and `diag`.

45. Display your results and report your conclusions.

47. The matrix S "shifts" the entries in a vector (a, b, c, d, e) to yield $(b, c, d, e, 0)$. S^5 is the 5×5 zero matrix. So is S^6.

49. $x = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$



51. $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \\ 2 & 3 & 16 & 24 & 25 & 26 & 6 & 7 & 19 & 26 \end{bmatrix}$

Section 2.2, page 118

1. $\begin{bmatrix} 2 & -3 \\ -5 & 8 \end{bmatrix}$ **3.** $\frac{1}{3}\begin{bmatrix} 3 & 3 \\ -7 & -8 \end{bmatrix}$ or $\begin{bmatrix} 1 & 1 \\ -7/3 & -8/3 \end{bmatrix}$

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

7. $x_1 = 7$ and $x_2 = -18$

9. a and b: $\begin{bmatrix} -9 \\ 4 \end{bmatrix}$, $\begin{bmatrix} 11 \\ -5 \end{bmatrix}$, $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$, and $\begin{bmatrix} 13 \\ -5 \end{bmatrix}$

11–19. Write out your answers before checking the *Study Guide.*

21. The proof can be modeled after the proof of Theorem 5.

23. $AB = AC \Rightarrow A^{-1}AB = A^{-1}AC \Rightarrow IB = IC \Rightarrow B = C$. No, in general, B and C can be different when A is not invertible. See Exercise 10 in Section 2.1.

25. $D = C^{-1}B^{-1}A^{-1}$. Show that D works.

27. $A = BCB^{-1}$

29. After you find $X = CB - A$, show that X is a solution.

31. *Hint:* Consider the equation $A\mathbf{x} = \mathbf{0}$.

33. *Hint:* If $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then there are no free variables in the equation $A\mathbf{x} = \mathbf{0}$, and each column of A is a pivot column.

35. *Hint:* Consider the case $a = b = 0$. Then consider the vector $\begin{bmatrix} -b \\ a \end{bmatrix}$, and use the fact that $ad - bc = 0$.

37. *Hint:* For part (a), interchange A and B in the box following Example 6 in Section 2.1, and then replace B by the identity matrix. For parts (b) and (c), begin by writing

$$A = \begin{bmatrix} \text{row}_1(A) \\ \text{row}_2(A) \\ \text{row}_3(A) \end{bmatrix}$$

39. $\begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix}$ **41.** $\begin{bmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ 7/2 & 3/2 & 1/2 \end{bmatrix}$

43. $A^{-1} = B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$. *Hint:* For $j = 1, \ldots, n$, let $\mathbf{a}_j$, $\mathbf{b}_j$, and $\mathbf{e}_j$ denote the jth columns of A, B, and I, respectively. Use the facts that $\mathbf{a}_j - \mathbf{a}_{j+1} = \mathbf{e}_j$ and $\mathbf{b}_j = \mathbf{e}_j - \mathbf{e}_{j+1}$ for $j = 1, \ldots, n-1$, and $\mathbf{a}_n = \mathbf{b}_n = \mathbf{e}_n$.

45. $\begin{bmatrix} 3 \\ -6 \\ 4 \end{bmatrix}$. Find this by row reducing $[\, A \quad \mathbf{e}_3 \,]$.

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

49. .27, .30, and .23 inch, respectively

51. 12, 1.5, 21.5, and 12 newtons, respectively

Section 2.3, page 124

The abbreviation IMT (here and in the *Study Guide*) denotes the Invertible Matrix Theorem (Theorem 8).

1. Invertible, by the IMT. Neither column of the matrix is a multiple of the other column, so they are linearly independent. Also, the matrix is invertible by Theorem 4 in Section 2.2 because the determinant is nonzero.

3. Invertible, by the IMT. The matrix row reduces to $\begin{bmatrix} 5 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ and has 3 pivot positions.

5. Not invertible, by the IMT. The matrix row reduces to $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & -5 \\ 0 & 0 & 0 \end{bmatrix}$ and is not row equivalent to I_3.

7. Invertible, by the IMT. The matrix row reduces to $\begin{bmatrix} -1 & -3 & 0 & 1 \\ 0 & -4 & 8 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and has four pivot positions.

9. The 4×4 matrix has four pivot positions, so it is invertible by the IMT.

11–19. The *Study Guide* will help, but first try to answer the questions based on your careful reading of the text.

21. A square upper triangular matrix is invertible if and only if all the entries on the diagonal are nonzero. Why?

Note: The answers below for Exercises 15–29 mention the IMT. In many cases, part or all of an acceptable answer could also be based on results that were used to establish the IMT.

23. If A has two identical columns then its columns are linearly dependent. Part (e) of the IMT shows that A cannot be invertible.

25. If A is invertible, so is A^{-1}, by Theorem 6 in Section 2.2. By (e) of the IMT applied to A^{-1}, the columns of A^{-1} are linearly independent.

27. By (e) of the IMT, D is invertible. Thus the equation $D\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^7$, by (g) of the IMT. Can you say more?

29. The matrix G cannot be invertible, by Theorem 5 in Section 2.2 or by the paragraph following the IMT. So (g) of the IMT is false and so is (h). The columns of G do not span $\mathbb{R}^n$.

31. Statement (b) of the IMT is false for K, so statements (e) and (h) are also false. That is, the columns of K are linearly *dependent* and the columns do *not* span $\mathbb{R}^n$.

33. *Hint:* Use the IMT first.

35. Let W be the inverse of AB. Then $ABW = I$ and $A(BW) = I$. Unfortunately, this equation by itself does not prove that A is invertible. Why not? Finish the proof before you check the *Study Guide*.

37. Since the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is not one-to-one, statement (f) of the IMT is false. Then (i) is also false and the transformation $\mathbf{x} \mapsto A\mathbf{x}$ does not map $\mathbb{R}^n$ onto $\mathbb{R}^n$. Also, A is not invertible, which implies that the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is not invertible, by Theorem 9.

39. *Hint:* If the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$, then A has a pivot in each row (Theorem 4 in Section 1.4). Could there be free variables in an equation $A\mathbf{x} = \mathbf{b}$?

41. *Hint:* First show that the standard matrix of T is invertible. Then use a theorem or theorems to show that $T^{-1}(\mathbf{x}) = B\mathbf{x}$, where $B = \begin{bmatrix} 7 & 9 \\ 4 & 5 \end{bmatrix}$.

43. *Hint:* To show that T is one-to-one, suppose that $T(\mathbf{u}) = T(\mathbf{v})$ for some vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$. Deduce that $\mathbf{u} = \mathbf{v}$. To show that T is onto, suppose $\mathbf{y}$ represents an arbitrary vector in $\mathbb{R}^n$ and use the inverse S to produce an $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{y}$. A second proof can be given using Theorem 9 together with a theorem from Section 1.9.

45. *Hint:* Consider the standard matrices of T and U.

47. Given any $\mathbf{v}$ in $\mathbb{R}^n$, we may write $\mathbf{v} = T(\mathbf{x})$ for some $\mathbf{x}$, because T is an onto mapping. Then, the assumed properties of S and U show that $S(\mathbf{v}) = S(T(\mathbf{x})) = \mathbf{x}$ and $U(\mathbf{v}) = U(T(\mathbf{x})) = \mathbf{x}$. So $S(\mathbf{v})$ and $U(\mathbf{v})$ are equal for each $\mathbf{v}$. That is, S and U are the same function from $\mathbb{R}^n$ into $\mathbb{R}^n$.

49. a. The exact solution of (3) is $x_1 = 3.94$ and $x_2 = .49$. The exact solution of (4) is $x_1 = 2.90$ and $x_2 = 2.00$.

b. When the solution of (4) is used as an approximation for the solution in (3), the error in using the value of 2.90 for x_1 is about 26%, and the error in using 2.0 for x_2 is about 308%.

c. The condition number of the coefficient matrix is 3363. The percentage change in the solution from (3) to (4) is about 7700 times the percentage change in the right side of the equation. This is the same order of magnitude as the condition number. The condition number gives a rough measure of how sensitive the solution of $A\mathbf{x} = \mathbf{b}$ can be to changes in $\mathbf{b}$. Further information about the condition number is given at the end of Chapter 6 and in Chapter 7.

51. cond$(A) \approx 69{,}000$, which is between 10^4 and 10^5. So about 4 or 5 digits of accuracy may be lost. Several experiments with MATLAB should verify that $\mathbf{x}$ and $\mathbf{x}_1$ agree to 11 or 12 digits.

53. Some versions of MATLAB issue a warning when asked to invert a Hilbert matrix of order about 12 or larger using floating-point arithmetic. The product AA^{-1} should have several off-diagonal entries that are far from being zero. If not, try a larger matrix.

Section 2.4, page 130

1. $\begin{bmatrix} A & B \\ EA + C & EB + D \end{bmatrix}$ **3.** $\begin{bmatrix} Y & Z \\ W & X \end{bmatrix}$

5. $Y = B^{-1}$ (explain why), $X = -B^{-1}A$, $Z = C$

7. $X = A^{-1}$ (why?), $Y = -BA^{-1}$, $Z = 0$ (why?)

9. $X = -A_{21}A_{11}^{-1}$, $Y = -A_{31}A_{11}^{-1}$, $B_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12}$

11–13. You can check your answers in the *Study Guide*.

15. *Hint:* Suppose A is invertible, and let $A^{-1} = \begin{bmatrix} D & E \\ F & G \end{bmatrix}$. Show that $BD = I$ and $CG = I$. This implies that B and C are invertible. (Explain why!) Conversely, suppose B and C are invertible. To prove that A is invertible, guess what A^{-1} must be and check that it works.

17. $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} =$
$$\begin{bmatrix} I & 0 \\ A_{21}A_{11}^{-1} & I \end{bmatrix}\begin{bmatrix} A_{11} & 0 \\ 0 & S \end{bmatrix}\begin{bmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix}$$
with $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$.

19. $G_{k+1} = \begin{bmatrix} X_k & \mathbf{x}_{k+1} \end{bmatrix}\begin{bmatrix} X_k^T \\ \mathbf{x}_{k+1}^T \end{bmatrix} = X_kX_k^T + \mathbf{x}_{k+1}\mathbf{x}_{k+1}^T$
$= G_k + \mathbf{x}_{k+1}\mathbf{x}_{k+1}^T$
Only the outer product matrix $\mathbf{x}_{k+1}\mathbf{x}_{k+1}^T$ needs to be computed (and then added to G_k).

21. $W(s) = I_m - C(A - sI_n)^{-1}B$. This is the Schur complement of $A - sI_n$ in the system matrix.

23. a. $A^2 = \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}$
$= \begin{bmatrix} 1+0 & 0+0 \\ 3-3 & 0+(-1)^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

b. $M^2 = \begin{bmatrix} A & 0 \\ I & -A \end{bmatrix}\begin{bmatrix} A & 0 \\ I & -A \end{bmatrix}$
$= \begin{bmatrix} A^2+0 & 0+0 \\ A-A & 0+(-A)^2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$

25. If A_1 and B_1 are $(k+1)\times(k+1)$ and lower triangular, then we can write $A_1 = \begin{bmatrix} a & \mathbf{0}^T \\ \mathbf{v} & A \end{bmatrix}$ and $B_1 = \begin{bmatrix} b & \mathbf{0}^T \\ \mathbf{w} & B \end{bmatrix}$, where A and B are $k \times k$ and lower triangular, $\mathbf{v}$ and $\mathbf{w}$ are in $\mathbb{R}^k$, and a and b are suitable scalars. Assume that the product of $k \times k$ lower triangular matrices is lower triangular, and compute the product A_1B_1. What do you conclude?

27. Use Example 5 to find the inverse of a matrix of the form $B = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix}$, where B_{11} is $p \times p$, B_{22} is $q \times q$ and B is invertible. Partition the matrix A, and apply your result twice to find that
$$A^{-1} = \begin{bmatrix} -5 & 2 & 0 & 0 & 0 \\ 3 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 3 & -4 \\ 0 & 0 & 0 & -5/2 & 7/2 \end{bmatrix}$$

29. a, b. The commands to be used in these exercises will depend on the matrix program.

c. The algebra needed comes from the block matrix equation
$$\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$$
where $\mathbf{x}_1$ and $\mathbf{b}_1$ are in $\mathbb{R}^{20}$ and $\mathbf{x}_2$ and $\mathbf{b}_2$ are in $\mathbb{R}^{30}$. Then $A_{11}\mathbf{x}_1 = \mathbf{b}_1$, which can be solved to produce $\mathbf{x}_1$. The equation $A_{21}\mathbf{x}_1 + A_{22}\mathbf{x}_2 = \mathbf{b}_2$ yields $A_{22}\mathbf{x}_2 = \mathbf{b}_2 - A_{21}\mathbf{x}_1$, which can be solved for $\mathbf{x}_2$ by row reducing the matrix $[A_{22} \quad \mathbf{c}]$, where $\mathbf{c} = \mathbf{b}_2 - A_{21}\mathbf{x}_1$.

Section 2.5, page 138

1. $L\mathbf{y} = \mathbf{b} \Rightarrow \mathbf{y} = \begin{bmatrix} -7 \\ -2 \\ 6 \end{bmatrix}$, $U\mathbf{x} = \mathbf{y} \Rightarrow \mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ -6 \end{bmatrix}$

3. $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -1 \\ 3 \\ 3 \end{bmatrix}$ **5.** $\mathbf{y} = \begin{bmatrix} 1 \\ 5 \\ 1 \\ -3 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -2 \\ -1 \\ 2 \\ -3 \end{bmatrix}$

7. $LU = \begin{bmatrix} 1 & 0 \\ -3/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 0 & 7/2 \end{bmatrix}$

9. $\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & 2/3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 & 2 \\ 0 & -3 & 12 \\ 0 & 0 & -8 \end{bmatrix}$

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

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

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

17. $U^{-1} = \begin{bmatrix} 1/4 & 3/8 & 1/4 \\ 0 & -1/2 & 1/2 \\ 0 & 0 & 1/2 \end{bmatrix}$,

$L^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$,

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

19. *Hint:* Think about row reducing $\begin{bmatrix} A & I \end{bmatrix}$.

21. *Hint:* Represent the row operations by a sequence of elementary matrices.

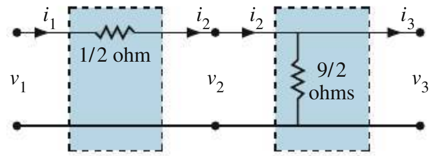
23. a. Denote the rows of D as transposes of column vectors. Then partitioned matrix multiplication yields

$$A = CD = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_4 \end{bmatrix} \begin{bmatrix} \mathbf{d}_1^T \\ \vdots \\ \mathbf{d}_4^T \end{bmatrix} = \mathbf{c}_1\mathbf{d}_1^T + \cdots + \mathbf{c}_4\mathbf{d}_4^T$$

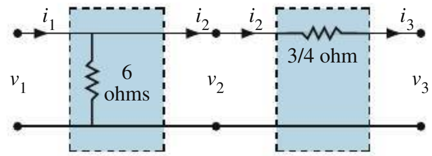
b. A has 40,000 entries. Since C has 1600 entries and D has 400 entries, together they occupy only 5% of the memory needed to store A.

25. Explain why U, D, and V^T are invertible. Then use a theorem on the inverse of a product of invertible matrices.

27. a.

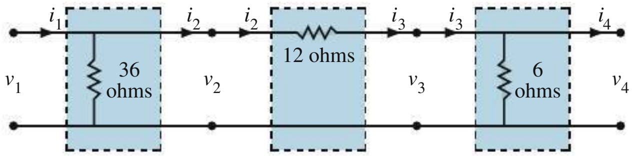


b.



29. a. $\begin{bmatrix} 1 + R_2/R_1 & -R_2 \\ -1/R_1 - R_2/(R_1R_3) - 1/R_3 & 1 + R_2/R_3 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & 0 \\ -1/6 & 1 \end{bmatrix} \begin{bmatrix} 1 & -12 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/36 & 1 \end{bmatrix}$



31. a. $L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.25 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.25 & -.0667 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -.2667 & -.2857 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -.2679 & -.0833 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -.2917 & -.2921 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -.2697 & -.0861 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -.2948 & -.2931 & 1 \end{bmatrix}$

$U = \begin{bmatrix} 4 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3.75 & -.25 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3.7333 & -1.0667 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3.4286 & -.2857 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.7083 & -1.0833 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.3919 & -.2921 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3.7052 & -1.0861 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.3868 \end{bmatrix}$

b. $\mathbf{x} = (3.9569, 6.5885, 4.2392, 7.3971, 5.6029, 8.7608, 9.4115, 12.0431)$

c. $A^{-1} = \begin{bmatrix} .2953 & .0866 & .0945 & .0509 & .0318 & .0227 & .0100 & .0082 \\ .0866 & .2953 & .0509 & .0945 & .0227 & .0318 & .0082 & .0100 \\ .0945 & .0509 & .3271 & .1093 & .1045 & .0591 & .0318 & .0227 \\ .0509 & .0945 & .1093 & .3271 & .0591 & .1045 & .0227 & .0318 \\ .0318 & .0227 & .1045 & .0591 & .3271 & .1093 & .0945 & .0509 \\ .0227 & .0318 & .0591 & .1045 & .1093 & .3271 & .0509 & .0945 \\ .0100 & .0082 & .0318 & .0227 & .0945 & .0509 & .2953 & .0866 \\ .0082 & .0100 & .0227 & .0318 & .0509 & .0945 & .0866 & .2953 \end{bmatrix}$

Obtain A^{-1} directly and then compute $A^{-1} - U^{-1}L^{-1}$ to compare the two methods for inverting a matrix.

Section 2.6, page 145

1. $C = \begin{bmatrix} .10 & .60 & .60 \\ .30 & .20 & 0 \\ .30 & .10 & .10 \end{bmatrix}$, $\begin{Bmatrix} \text{intermediate} \\ \text{demand} \end{Bmatrix} = \begin{bmatrix} 60 \\ 20 \\ 10 \end{bmatrix}$

3. $\mathbf{x} = \begin{bmatrix} 40 \\ 15 \\ 15 \end{bmatrix}$ **5.** $\mathbf{x} = \begin{bmatrix} 110 \\ 120 \end{bmatrix}$

7. a. $\begin{bmatrix} 1.6 \\ 1.2 \end{bmatrix}$ **b.** $\begin{bmatrix} 111.6 \\ 121.2 \end{bmatrix}$

9. $\mathbf{x} = \begin{bmatrix} 82.8 \\ 131.0 \\ 110.3 \end{bmatrix}$

11. *Hint:* Use properties of transposes to obtain $\mathbf{p}^T = \mathbf{p}^T C + \mathbf{v}^T$, so that $\mathbf{p}^T\mathbf{x} = (\mathbf{p}^T C + \mathbf{v}^T)\mathbf{x} = \mathbf{p}^T C\mathbf{x} + \mathbf{v}^T\mathbf{x}$. Now compute $\mathbf{p}^T\mathbf{x}$ from the production equation.

13. $\mathbf{x} = (99576, 97703, 51231, 131570, 49488, 329554, 13835)$. The entries in $\mathbf{x}$ suggest more precision in the answer than is warranted by the entries in $\mathbf{d}$, which appear to be accurate only to perhaps the nearest thousand. So a more realistic answer for $\mathbf{x}$ might be $\mathbf{x} = 1000 \times (100, 98, 51, 132, 49, 330, 14)$.

15. $\mathbf{x}^{(12)}$ is the first vector whose entries are accurate to the nearest thousand. The calculation of $\mathbf{x}^{(12)}$ takes about 1260 flops, while row reduction of $\begin{bmatrix} (I - C) & \mathbf{d} \end{bmatrix}$ takes only about 550 flops. If C is larger than 20×20, then fewer flops are needed to compute $\mathbf{x}^{(12)}$ by iteration than to compute the equilibrium vector $\mathbf{x}$ by row reduction. As the size of C grows, the advantage of the iterative method increases. Also, because C becomes more sparse for larger models of the economy, fewer iterations are needed for reasonable accuracy.

Section 2.7, page 153

1. $\begin{bmatrix} 1 & .25 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **3.** $\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & \sqrt{2} \\ \sqrt{2}/2 & \sqrt{2}/2 & 2\sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$

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

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

See the Practice Problem.

9. $A(BD)$ requires 1600 multiplications. $(AB)D$ requires 808 multiplications. The first method uses about twice as many multiplications. If D had 20,000 columns, the counts would be 160,000 and 80,008, respectively.

11. Use the fact that

$$\sec\varphi - \tan\varphi \sin\varphi = \frac{1}{\cos\varphi} - \frac{\sin^2\varphi}{\cos\varphi} = \cos\varphi$$

13. $\begin{bmatrix} A & \mathbf{p} \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} I & \mathbf{p} \\ \mathbf{0}^T & 1 \end{bmatrix}\begin{bmatrix} A & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}$. First apply the linear transformation A, and then translate by $\mathbf{p}$.

15. $(12, -6, 3)$ **17.** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & -\sqrt{3}/2 & 0 \\ 0 & \sqrt{3}/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

19. The triangle with vertices at $(7, 2, 0)$, $(7.5, 5, 0)$, $(5, 5, 0)$

21. $\begin{bmatrix} 2.2586 & -1.0395 & -.3473 \\ -1.3495 & 2.3441 & .0696 \\ .0910 & -.3046 & 1.2777 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix}$

Section 2.8, page 160

1. The set is closed under sums but not under multiplication by a negative scalar. (Sketch an example.)

3. The set is not closed under sums or scalar multiples. The subset consisting of the points on the line $x_2 = x_1$ is a subspace, so any "counterexample" must use at least one point not on this line.

5. No. The system corresponding to $[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{w}\,]$ is inconsistent.

7. **a.** The three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$

b. Infinitely many vectors

c. Yes, because $A\mathbf{x} = \mathbf{p}$ has a solution.

9. No, because $A\mathbf{p} \neq \mathbf{0}$.

11. $p = 4$ and $q = 3$. Nul A is a subspace of $\mathbb{R}^4$ because solutions of $A\mathbf{x} = \mathbf{0}$ must have four entries, to match the columns of A. Col A is a subspace of $\mathbb{R}^3$ because each column vector has three entries.

13. For Nul A, choose $(1, -2, 1, 0)$ or $(-1, 4, 0, 1)$, for example. For Col A, select any column of A.

15. Yes. Let A be the matrix whose columns are the vectors given. Then A is invertible because its determinant is nonzero, and so its columns form a basis for $\mathbb{R}^2$, by the IMT (or by Example 5). (Other reasons for the invertibility of A could be given.)

17. Yes. Let A be the matrix whose columns are the vectors given. Row reduction shows three pivots, so A is invertible. By the IMT, the columns of A form a basis for $\mathbb{R}^3$.

19. No. Let A be the 3×2 matrix whose columns are the vectors given. The columns of A cannot possibly span $\mathbb{R}^3$ because A cannot have a pivot in every row. So the columns are not a basis for $\mathbb{R}^3$. (They are a basis for a plane in $\mathbb{R}^3$.)

21–29. Read the section carefully, and write your answers before checking the *Study Guide*. This section has terms and key concepts that you must learn now before going on.

31. Basis for Col A: $\begin{bmatrix} 4 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \\ 4 \end{bmatrix}$

Basis for Nul A: $\begin{bmatrix} 4 \\ -5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -7 \\ 6 \\ 0 \\ 1 \end{bmatrix}$

33. Basis for Col A: $\begin{bmatrix} 1 \\ -1 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 2 \\ 6 \end{bmatrix}, \begin{bmatrix} -3 \\ 3 \\ 5 \\ -5 \end{bmatrix}$

Basis for Nul A: $\begin{bmatrix} 2 \\ -2.5 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -7 \\ .5 \\ 0 \\ -4 \\ 1 \end{bmatrix}$

35. Construct a nonzero 3×3 matrix A, and construct $\mathbf{b}$ to be almost any convenient linear combination of the columns of A.

37. *Hint:* You need a nonzero matrix whose columns are linearly dependent.

39. If Col $F \neq \mathbb{R}^5$, then the columns of F do not span $\mathbb{R}^5$. Since F is square, the IMT shows that F is not invertible and the equation $F\mathbf{x} = \mathbf{0}$ has a nontrivial solution. That is, Nul F contains a nonzero vector. Another way to describe this is to write Nul $F \neq \{\mathbf{0}\}$.

41. If Col $Q = \mathbb{R}^4$, then the columns of Q span $\mathbb{R}^4$. Since Q is square, the IMT shows that Q is invertible and the equation $Q\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^4$. Also, each solution is unique, by Theorem 5 in Section 2.2.

43. If the columns of B are linearly independent, then the equation $B\mathbf{x} = \mathbf{0}$ has only the trivial (zero) solution. That is, Nul $B = \{\mathbf{0}\}$.

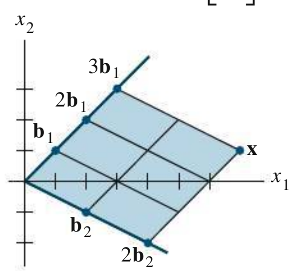
45. Display the reduced echelon form of A, and select the pivot columns of A as a basis for Col A. For Nul A, write the solution of $A\mathbf{x} = \mathbf{0}$ in parametric vector form.

Basis for Col A : $\begin{bmatrix} 3 \\ -7 \\ -5 \\ 3 \end{bmatrix}, \begin{bmatrix} -5 \\ 9 \\ 7 \\ -7 \end{bmatrix}$

Basis for Nul A : $\begin{bmatrix} -2.5 \\ -1.5 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4.5 \\ 2.5 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3.5 \\ -1.5 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

Section 2.9, page 166

1. $\mathbf{x} = 3\mathbf{b}_1 + 2\mathbf{b}_2 = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$



3. $\begin{bmatrix} 7 \\ 5 \end{bmatrix}$ **5.** $\begin{bmatrix} 1/4 \\ -5/4 \end{bmatrix}$

7. $[\mathbf{w}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 1.5 \\ .5 \end{bmatrix}$

9. Basis for Col A: $\begin{bmatrix} 1 \\ -3 \\ 2 \\ -4 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 5 \\ -3 \\ 7 \end{bmatrix}$; dim Col $A = 3$

Basis for Nul A: $\begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}$; dim Nul $A = 1$

11. Basis for Col A: $\begin{bmatrix} 1 \\ 2 \\ -3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ -9 \\ 10 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ -7 \\ 11 \end{bmatrix}$;

dim Col $A = 3$; Basis for Nul A: $\begin{bmatrix} 9 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 3 \\ 0 \\ -2 \\ 1 \end{bmatrix}$;

dim Nul $A = 2$

13. Columns 1, 3, and 4 of the original matrix form a basis for H, so dim $H = 3$.

15. Col $A = \mathbb{R}^3$, because A has a pivot in each row, and so the columns of A span $\mathbb{R}^3$. Nul A *cannot* equal $\mathbb{R}^2$, because Nul A is a subspace of $\mathbb{R}^5$. It is true, however, that Nul A is two-dimensional. Reason: The equation $A\mathbf{x} = \mathbf{0}$ has two free variables, because A has five columns and only three of them are pivot columns.

17–25. See the *Study Guide* after you write your justifications.

27. The fact that the solution space of $A\mathbf{x} = \mathbf{0}$ has a basis of three vectors means that dim Nul $A = 3$. Since a 5×7 matrix A has seven columns, the Rank Theorem shows that rank $A = 7 - \dim \text{Nul } A = 4$. See the *Study Guide* for a justification that does not explicitly mention the Rank Theorem.

29. A 7×6 matrix has six columns. By the Rank Theorem, dim Nul $A = 6 - \text{rank } A$. Since the rank is four, dim Nul $A = 2$. That is, the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$ is two.

31. A 3×4 matrix A with a two-dimensional column space has two pivot columns. The remaining two columns will correspond to free variables in the equation $A\mathbf{x} = \mathbf{0}$. So the desired construction is possible. There are six possible locations for the two pivot columns, one of which is $\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 \end{bmatrix}$. A simple construction is to take two vectors in $\mathbb{R}^3$ that are obviously not linearly dependent and place them in a matrix along with a copy of each vector, in any order. The resulting matrix will obviously have a two-dimensional column space. There is no need to worry about whether Nul A has the correct dimension, since this is guaranteed by the Rank Theorem: dim Nul $A = 4 - \text{rank } A$.

33. The p columns of A span Col A by definition. If dim Col $A = p$, then the spanning set of p columns is automatically a basis for Col A, by the Basis Theorem. In particular, the columns are linearly independent.

35. **a.** *Hint:* The columns of B span W, and each vector $\mathbf{a}_j$ is in W. The vector $\mathbf{c}_j$ is in $\mathbb{R}^p$ because B has p columns.

b. *Hint:* What is the size of C?

c. *Hint:* How are B and C related to A?

37. Your calculations should show that the matrix $[\ \mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{x}\ ]$ corresponds to a consistent system. The $\mathcal{B}$-coordinate vector of $\mathbf{x}$ is $(-5/3, 8/3)$.

Chapter 2 Supplementary Exercises, page 169

1. T **2.** F **3.** T **4.** F
5. F **6.** F **7.** T **8.** T
9. T **10.** F **11.** T **12.** F
13. F **14.** T **15.** F

17. I

19. $A^2 = 2A - I$. Multiply by A: $A^3 = 2A^2 - A$.
Substitute $A^2 = 2A - I$: $A^3 = 2(2A - I) - A = 3A - 2I$.
Multiply by A again: $A^4 = A(3A - 2I) = 3A^2 - 2A$.
Substitute the identity $A^2 = 2A - I$ again:
$A^4 = 3(2A - I) - 2A = 4A - 3I$.

21. $\begin{bmatrix} 10 & -1 \\ 9 & 10 \\ -5 & -3 \end{bmatrix}$ **23.** $\begin{bmatrix} -3 & 13 \\ -8 & 27 \end{bmatrix}$

25. **a.**
$$\begin{aligned} p(x_i) &= c_0 + c_1 x_i + \cdots + c_{n-1} x_i^{n-1} \\ &= \text{row}_i(V)\begin{bmatrix} c_0 \\ \vdots \\ c_{n-1} \end{bmatrix} = \text{row}_i(V\mathbf{c}) = y_i \end{aligned}$$

b. Suppose $x_1, \ldots, x_n$ are distinct, and suppose $V\mathbf{c} = \mathbf{0}$ for some vector $\mathbf{c}$. Then the entries in $\mathbf{c}$ are the coefficients of a polynomial whose value is zero at the distinct points $x_1, \ldots, x_n$. However, a nonzero polynomial of degree $n-1$ cannot have n zeros, so the polynomial must be identically zero. That is, the entries in $\mathbf{c}$ must all be zero. This shows that the columns of V are linearly independent.

c. *Hint:* When $x_1, \ldots, x_n$ are distinct, there is a vector $\mathbf{c}$ such that $V\mathbf{c} = \mathbf{y}$. Why?

27. **a.** $P^2 = (\mathbf{u}\mathbf{u}^T)(\mathbf{u}\mathbf{u}^T) = \mathbf{u}(\mathbf{u}^T\mathbf{u})\mathbf{u}^T = \mathbf{u}(1)\mathbf{u}^T = P$

b. $P^T = (\mathbf{u}\mathbf{u}^T)^T = \mathbf{u}^{TT}\mathbf{u}^T = \mathbf{u}\mathbf{u}^T = P$

c.
$$\begin{aligned} Q^2 &= (I - 2P)(I - 2P) \\ &= I - I(2P) - 2PI + 2P(2P) \\ &= I - 4P + 4P^2 = I, \text{ because of part (a).} \end{aligned}$$

29. Left-multiplication by an elementary matrix produces an elementary row operation:

$$B \sim E_1 B \sim E_2 E_1 B \sim E_3 E_2 E_1 B = C$$

So B is row equivalent to C. Since row operations are reversible, C is row equivalent to B. (Alternatively, show C being changed into B by row operations using the inverses of the E_i.)

31. Since B is 4×6 (with more columns than rows), its six columns are linearly dependent and there is a nonzero $\mathbf{x}$ such that $B\mathbf{x} = \mathbf{0}$. Thus $AB\mathbf{x} = A\mathbf{0} = \mathbf{0}$, which shows that the matrix AB is not invertible, by the Invertible Matrix Theorem.

33. To four decimal places, as k increases,

$$A^k \to \begin{bmatrix} .2857 & .2857 & .2857 \\ .4286 & .4286 & .4286 \\ .2857 & .2857 & .2857 \end{bmatrix} \quad \text{and}$$

$$B^k \to \begin{bmatrix} .2022 & .2022 & .2022 \\ .3708 & .3708 & .3708 \\ .4270 & .4270 & .4270 \end{bmatrix}$$

or, in rational format,

$$A^k \to \begin{bmatrix} 2/7 & 2/7 & 2/7 \\ 3/7 & 3/7 & 3/7 \\ 2/7 & 2/7 & 2/7 \end{bmatrix} \quad \text{and}$$

$$B^k \to \begin{bmatrix} 18/89 & 18/89 & 18/89 \\ 33/89 & 33/89 & 33/89 \\ 38/89 & 38/89 & 38/89 \end{bmatrix}$$

Chapter 3

Section 3.1, page 177

1. 1 **3.** 0 **5.** -24 **7.** 4

9. 15. Start with row 3.

11. -18. Start with column 1 or row 4.

13. 6. Start with row 2 or column 2.

15. 24 **17.** -10

19. $ad - bc$, $cb - da$. Interchanging two rows changes the sign of the determinant.

21. 2; $3(4 + 2k) - 2(5 + 3k) = 12 + 6k - 10 - 6k$. Row replacement does not change a determinant.

23. $7a - 14b + 7c$. $-7a + 14b - 7c$. Interchanging two rows changes the sign of the determinant.

25. 1 **27.** 1 **29.** k

31. 1. The matrix is upper or lower triangular, with only 1's on the diagonal. The determinant is 1, the product of the diagonal entries.

33.
$$\begin{aligned} \det EA &= \det\begin{bmatrix} a + kc & b + kd \\ c & d \end{bmatrix} \\ &= (a + kc)d - (b + kd)c \\ &= ad + kcd - bc - kdc = (+1)(ad - bc) \\ &= (\det E)(\det A) \end{aligned}$$

35.
$$\begin{aligned} \det EA &= \det\begin{bmatrix} c & d \\ a & b \end{bmatrix} = cb - ad = (-1)(ad - bc) \\ &= (\det E)(\det A) \end{aligned}$$

37. $5A = \begin{bmatrix} 15 & 5 \\ 20 & 10 \end{bmatrix}$; no

39–41. Hints are in the *Study Guide*.

43. The area of the parallelogram and the determinant of $\begin{bmatrix} \mathbf{u} & \mathbf{v} \end{bmatrix}$ both equal 6. If $\mathbf{v} = \begin{bmatrix} x \\ 2 \end{bmatrix}$ for any x, the area is still 6. In each case the base of the parallelogram is unchanged, and the altitude remains 2 because the second coordinate of $\mathbf{v}$ is always 2.

45. **a.** yes **b.** no **c.** yes **d.** no

47. In general, $\det A^{-1} = 1/\det A$ as long as $\det A$ is nonzero.

49. You can check your conjectures when you get to Section 3.2.

51. **b.** $llll; rllr; lrlr; llrr;$

Section 3.2, page 185

1. Interchanging two rows reverses the sign of the determinant.

3. A row replacement operation does not change the determinant.

5. -3 **7.** 0 **9.** -28 **11.** -48

13. 6 **15.** 21 **17.** 7 **19.** 14

21. Not invertible **23.** Invertible

25. Linearly independent

27–33. See the *Study Guide*.

35. 16

37. *Hint:* Show that $(\det A)(\det A^{-1}) = 1$.

39. *Hint:* Use Theorem 6.

41. *Hint:* Use Theorem 6 and another theorem.

43. $\det AB = \det\begin{bmatrix} 6 & 0 \\ 17 & 4 \end{bmatrix} = 24$; $(\det A)(\det B) = 3 \cdot 8 = 24$

45. **a.** -6 **b.** -250 **c.** 3 **d.** $-1/2$ **e.** -8

47.
$$\begin{aligned} \det A &= (a+e)d - (b+f)c = ad + ed - bc - fc \\ &= (ad - bc) + (ed - fc) = \det B + \det C \end{aligned}$$

49. *Hint:* Compute $\det A$ by a cofactor expansion down column 3.

51. No. $\det A \det A^{-1}$ should equal 1.

53. See the *Study Guide* after you have made a conjecture about A^TA and AA^T.

Section 3.3, page 195

1. $\begin{bmatrix} 5/6 \\ -1/6 \end{bmatrix}$ **3.** $\begin{bmatrix} 4/5 \\ -3/10 \end{bmatrix}$ **5.** $\begin{bmatrix} 1/4 \\ 11/4 \\ 3/8 \end{bmatrix}$

7. $s \neq \pm\sqrt{3}$; $x_1 = \dfrac{5s+4}{6(s^2-3)}$, $x_2 = \dfrac{-4s-15}{4(s^2-3)}$

9. $s \neq 0, 1$; $x_1 = \dfrac{-7}{3(s-1)}$, $x_2 = \dfrac{4s+3}{6s(s-1)}$

11. $\text{adj}\, A = \begin{bmatrix} 0 & 1 & 0 \\ -5 & -1 & -5 \\ 5 & 2 & 10 \end{bmatrix}$, $A^{-1} = \dfrac{1}{5}\begin{bmatrix} 0 & 1 & 0 \\ -5 & -1 & -5 \\ 5 & 2 & 10 \end{bmatrix}$

13. $\text{adj}\, A = \begin{bmatrix} -1 & -1 & 5 \\ 1 & -5 & 1 \\ 1 & 7 & -5 \end{bmatrix}$, $A^{-1} = \dfrac{1}{6}\begin{bmatrix} -1 & -1 & 5 \\ 1 & -5 & 1 \\ 1 & 7 & -5 \end{bmatrix}$

15. $\text{adj}\, A = \begin{bmatrix} -1 & 0 & 0 \\ -1 & -5 & 0 \\ -1 & -15 & 5 \end{bmatrix}$, $A^{-1} = \dfrac{-1}{5}\begin{bmatrix} -1 & 0 & 0 \\ -1 & -5 & 0 \\ -1 & -15 & 5 \end{bmatrix}$

17. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $C_{11} = d$, $C_{12} = -c$, $C_{21} = -b$, $C_{22} = a$. The adjugate matrix is the transpose of cofactors:
$$\text{adj}\, A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$
Following Theorem 8, we divide by $\det A$; this produces the formula from Section 2.2.

19. 8 **21.** 3 **23.** 23

25. A 3×3 matrix A is not invertible if and only if its columns are linearly dependent (by the Invertible Matrix Theorem). This happens if and only if one of the columns is in the plane spanned by the other two columns, which is equivalent to the condition that the parallelepiped determined by these columns has zero volume, which in turn is equivalent to the condition that $\det A = 0$.

27. 12 **29.** $\frac{1}{2}|\det\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}|$

31. **a.** See Example 5. **b.** $4\pi abc/3$

33. I.

35–37. By now you know to try these before you look in the *Study Guide*.

39. In MATLAB, the entries in $B - \text{inv}(A)$ are approximately 10^{-15} or smaller. See the *Study Guide* for suggestions that may save you keystrokes as you work.

41. MATLAB Student Version 4.0 uses 57,771 flops for $\text{inv}(A)$, and 14,269,045 flops for the inverse formula. The `inv(A)` command requires only about 0.4% of the operations for the inverse formula. The *Study Guide* shows how to use the `flops` command.

Chapter 3 Supplementary Exercises, page 197

1. T **2.** T **3.** F **4.** F

5. F **6.** F **7.** T **8.** T

9. F **10.** F **11.** T **12.** F

13. F **14.** T **15.** F

The solution for Exercise 17 is based on the fact that if a matrix contains two rows (or two columns) that are multiples of each other, then the determinant of the matrix is zero, by Theorem.

17. Make two row replacement operations, and then factor out a common multiple in row 2 and a common multiple in row 3.

$$\begin{vmatrix} 1 & a & b+c \\ 1 & b & a+c \\ 1 & c & a+b \end{vmatrix} = \begin{vmatrix} 1 & a & b+c \\ 0 & b-a & a-b \\ 0 & c-a & a-c \end{vmatrix}$$

$$= (b-a)(c-a)\begin{vmatrix} 1 & a & b+c \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{vmatrix}$$

$$= 0$$

19. -12

21. When the determinant is expanded by cofactors of the first row, the equation has the form $ax + by + c = 0$, where at least one of a and b is not zero. This is the equation of a line. It is clear that (x_1, y_1) and (x_2, y_2) are on the line, because when the coordinates of one of the points are substituted for x and y, two rows of the matrix are equal and so the determinant is zero.

23. $T \sim \begin{bmatrix} 1 & a & a^2 \\ 0 & b-a & b^2-a^2 \\ 0 & c-a & c^2-a^2 \end{bmatrix}$. Thus, by Theorem 3,

$$\begin{aligned} \det T &= (b-a)(c-a)\det\begin{bmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 1 & c+a \end{bmatrix} \\ &= (b-a)(c-a)\det\begin{bmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & c-b \end{bmatrix} \\ &= (b-a)(c-a)(c-b) \end{aligned}$$

25. Area $= 12$. If one vertex is subtracted from all four vertices, and if the new vertices are $\mathbf{0}$, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, then the translated figure (and hence the original figure) will be a parallelogram if and only if one of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ is the sum of the other two vectors.

27. By the Inverse Formula, $(\text{adj } A) \cdot \dfrac{1}{\det A} A = A^{-1}A = I$. By the Invertible Matrix Theorem, adj A is invertible $(\text{adj } A)^{-1} = \dfrac{1}{\det A} A$.

29. **a.** $X = CA^{-1}$, $Y = D - CA^{-1}B$. Now use Exercise 28(c).

b. From part (a), and the property of determinants,

$$\begin{aligned} \det\begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \det[A(D - CA^{-1}B)] \\ &= \det[AD - ACA^{-1}B] \\ &= \det[AD - CAA^{-1}B] \\ &= \det[AD - CB] \end{aligned}$$

where the equality $AC = CA$ was used in the third step.

31. First consider the case $n = 2$, and prove that the result holds by directly computing the determinants of B and C. Now assume that the formula holds for all $(k-1) \times (k-1)$ matrices, and let A, B, and C be $k \times k$ matrices. Use a cofactor expansion along the first column and the inductive hypothesis to find det B. Use row replacement operations on C to create zeros below the first pivot and produce a triangular matrix. Find the determinant of this matrix and add to det B to get the result.

33. Compute:

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{vmatrix} = 1, \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{vmatrix} = 1,$$

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

Conjecture:

$$\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2 & & 2 \\ 1 & 2 & 3 & & 3 \\ \vdots & & & \ddots & \vdots \\ 1 & 2 & 3 & \dots & n \end{vmatrix} = 1$$

To confirm the conjecture, use row replacement operations to create zeros below the first pivot, then the second pivot, and so on. The resulting matrix is

$$\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & & 1 \\ 0 & 0 & 1 & & 1 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}$$

which is an upper triangular matrix with determinant 1.

Chapter 4

Section 4.1, page 208

1. **a.** $\mathbf{u} + \mathbf{v}$ is in V because its entries will both be nonnegative.

b. *Example:* If $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $c = -1$, then $\mathbf{u}$ is in V, but $c\mathbf{u}$ is not in V.

3. *Example:* If $\mathbf{u} = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$ and $c = 4$, then $\mathbf{u}$ is in H, but $c\mathbf{u}$ is not in H.

5. Yes, by Theorem 1, because the set is Span $\{t^2\}$.

7. No, the set is not closed under multiplication by scalars that are not integers.

9. $H = \text{Span}\,\{\mathbf{v}\}$, where $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$. By Theorem 1, H is a subspace of $\mathbb{R}^3$.

11. $W = \text{Span}\{\mathbf{u}, \mathbf{v}\}$, where $\mathbf{u} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$. By Theorem 1, W is a subspace of $\mathbb{R}^3$.

13. **a.** There are only three vectors in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, and $\mathbf{w}$ is not one of them.

b. There are infinitely many vectors in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

c. $\mathbf{w}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

15. Not a vector space because the zero vector is not in W

17. $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\}$

19. *Hint:* Use Theorem 1.

Warning: Although the *Study Guide* has complete solutions for *every* odd-numbered exercise whose answer here is only a "Hint," you *must* really try to work the solution yourself. Otherwise, you will not benefit from the exercise.

21. Yes. The conditions for a subspace are obviously satisfied: The zero matrix is in H, the sum of two upper triangular matrices is upper triangular, and any scalar multiple of an upper triangular matrix is again upper triangular.

23–31. See the *Study Guide* after you have written your answers.

33. 4 **35. a.** 8 **b.** 3 **c.** 5 **d.** 4

37.
$$\begin{aligned} \mathbf{u} + (-1)\mathbf{u} &= 1\mathbf{u} + (-1)\mathbf{u} && \text{Axiom 10} \\ &= [1 + (-1)]\mathbf{u} && \text{Axiom 8} \\ &= 0\mathbf{u} = \mathbf{0} && \text{Exercise 35} \end{aligned}$$
From Exercise 34, it follows that $(-1)\mathbf{u} = -\mathbf{u}$.

39. Any subspace H that contains $\mathbf{u}$ and $\mathbf{v}$ must also contain all scalar multiples of $\mathbf{u}$ and $\mathbf{v}$ and hence must contain all sums of scalar multiples of $\mathbf{u}$ and $\mathbf{v}$. Thus H must contain Span $\{\mathbf{u}, \mathbf{v}\}$.

41. *Hint:* For part of the solution, consider $\mathbf{w}_1$ and $\mathbf{w}_2$ in $H + K$, and write $\mathbf{w}_1$ and $\mathbf{w}_2$ in the form $\mathbf{w}_1 = \mathbf{u}_1 + \mathbf{v}_1$ and $\mathbf{w}_2 = \mathbf{u}_2 + \mathbf{v}_2$, where $\mathbf{u}_1$ and $\mathbf{u}_2$ are in H, and $\mathbf{v}_1$ and $\mathbf{v}_2$ are in K.

43. The reduced echelon form of $[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3 \quad \mathbf{w}\,]$ shows that $\mathbf{w} = \mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3$.

45. The functions are $\cos 4t$ and $\cos 6t$. See Exercise 54 in Section 4.5.

Section 4.2, page 219

1. $\begin{bmatrix} 3 & -5 & -3 \\ 6 & -2 & 0 \\ -8 & 4 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ -4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, so $\mathbf{w}$ is in Nul A.

3. $\begin{bmatrix} 7 \\ -4 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -6 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ **5.** $\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 9 \\ 1 \\ 0 \end{bmatrix}$

7. W is not a subspace of $\mathbb{R}^3$ because the zero vector $(0, 0, 0)$ is not in W.

9. W is a subspace of $\mathbb{R}^4$, by Theorem 2, because W is the set of solutions of the homogeneous system
$$\begin{aligned} a - 2b - 4c \phantom{{}- 3d} &= 0 \\ 2a \phantom{{}- 2b} - c - 3d &= 0 \end{aligned}$$

11. W is not a subspace because $\mathbf{0}$ is not in W. *Justification:* If a typical element $(b - 2d, 5 + d, b + 3d, d)$ were zero, then $5 + d = 0$ and $d = 0$, which is impossible.

13. $W = \text{Col}\, A$ for $A = \begin{bmatrix} 1 & -6 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$, so W is a vector space by Theorem 3.

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

17. a. 2 **b.** 4 **19. a.** 5 **b.** 2

21. $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ in Nul A, $\begin{bmatrix} 2 \\ -1 \\ -4 \\ 3 \end{bmatrix}$ in Col A, and $[2 \;\; -6]$ is in Row A. Other answers possible.

23. $\mathbf{w}$ is in both Nul A and Col A.

25–37. See the *Study Guide*. By now you should know how to use it properly.

39. Let $\mathbf{x} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & -3 & -3 \\ -2 & 4 & 2 \\ -1 & 5 & 7 \end{bmatrix}$. Then $\mathbf{x}$ is in Nul A. Since Nul A is a subspace of $\mathbb{R}^3$, $10\mathbf{x}$ is in Nul A.

41. **a.** $A\mathbf{0} = \mathbf{0}$, so the zero vector is in Col A.

b. By a property of matrix multiplication, $A\mathbf{x} + A\mathbf{w} = A(\mathbf{x} + \mathbf{w})$, which shows that $A\mathbf{x} + A\mathbf{w}$ is a linear combination of the columns of A and hence is in Col A.

c. $c(A\mathbf{x}) = A(c\mathbf{x})$, which shows that $c(A\mathbf{x})$ is in Col A for all scalars c.

43. **a.** For arbitrary polynomials $\mathbf{p}$, $\mathbf{q}$ in $\mathbb{P}_2$ and any scalar c,
$$\begin{aligned} T(\mathbf{p} + \mathbf{q}) &= \begin{bmatrix} (\mathbf{p} + \mathbf{q})(0) \\ (\mathbf{p} + \mathbf{q})(1) \end{bmatrix} = \begin{bmatrix} \mathbf{p}(0) + \mathbf{q}(0) \\ \mathbf{p}(1) + \mathbf{q}(1) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(1) \end{bmatrix} + \begin{bmatrix} \mathbf{q}(0) \\ \mathbf{q}(1) \end{bmatrix} = T(\mathbf{p}) + T(\mathbf{q}) \\ T(c\mathbf{p}) &= \begin{bmatrix} c\mathbf{p}(0) \\ c\mathbf{p}(1) \end{bmatrix} = c\begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(1) \end{bmatrix} = cT(\mathbf{p}) \end{aligned}$$

So T is a linear transformation from $\mathbb{P}_2$ into $\mathbb{P}_2$.

b. Any quadratic polynomial that vanishes at 0 and 1 must be a multiple of $\mathbf{p}(t) = t(t - 1)$. The range of T is $\mathbb{R}^2$.

45. a. For A, B in $M_{2\times2}$ and any scalar c,

$$\begin{aligned} T(A+B) &= (A+B)+(A+B)^T \\ &= A+B+A^T+B^T \quad \text{Transpose property} \\ &= (A+A^T)+(B+B^T) = T(A)+T(B) \\ T(cA) &= (cA)+(cA)^T = cA+cA^T \\ &= c(A+A^T) = cT(A) \end{aligned}$$

So T is a linear transformation from $M_{2\times2}$ into $M_{2\times2}$.

b. If B is any element in $M_{2\times2}$ with the property that $B^T = B$, and if $A = \frac{1}{2}B$, then

$$T(A) = \tfrac{1}{2}B + \left(\tfrac{1}{2}B\right)^T = \tfrac{1}{2}B + \tfrac{1}{2}B = B$$

c. Part (b) showed that the range of T contains all B such that $B^T = B$. So it suffices to show that any B in the range of T has this property. If $B = T(A)$, then by properties of transposes,

$$B^T = (A+A^T)^T = A^T + A^{TT} = A^T + A = B$$

d. The kernel of T is $\left\{ \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix} : b \text{ real} \right\}$.

47. *Hint:* Check the three conditions for a subspace. Typical elements of $T(U)$ have the form $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$, where $\mathbf{u}_1$ and $\mathbf{u}_2$ are in U.

49. $\mathbf{w}$ is in Col A but not in Nul A. (Explain why.)

51. The reduced echelon form of A is

$$\begin{bmatrix} 1 & 0 & 1/3 & 0 & 10/3 \\ 0 & 1 & 1/3 & 0 & -26/3 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Section 4.3, page 228

1. Yes, the 3×3 matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ has 3 pivot positions. By the Invertible Matrix Theorem, A is invertible and its columns form a basis for $\mathbb{R}^3$. (See Example 3.)

3. No, the vectors are linearly dependent and do not span $\mathbb{R}^3$.

5. No, the set is linearly dependent because the zero vector is in the set. However,

$$\begin{bmatrix} 1 & -2 & 0 & 0 \\ -3 & 9 & 0 & -3 \\ 0 & 0 & 0 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 & 0 \\ 0 & 3 & 0 & -3 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

The matrix has pivots in each row and hence its columns span $\mathbb{R}^3$.

7. No, the vectors are linearly independent because they are not multiples. (More precisely, neither vector is a multiple of the other.) However, the vectors do not span $\mathbb{R}^3$. The matrix $\begin{bmatrix} -2 & 6 \\ 3 & -1 \\ 0 & 5 \end{bmatrix}$ can have at most two pivots since it has only two columns. So there will not be a pivot in each row.

9. $\begin{bmatrix} 3 \\ 5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \\ 0 \\ 1 \end{bmatrix}$ **11.** $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

13. Basis for Nul A: $\begin{bmatrix} -6 \\ -5/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -3/2 \\ 0 \\ 1 \end{bmatrix}$

Basis for Col A: $\begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix}$

Basis for Row A: $\begin{bmatrix} 1 & 0 & 6 & 5 \end{bmatrix}, \begin{bmatrix} 0 & 2 & 5 & 3 \end{bmatrix}$

15. $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ **17.** $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$

19. The three simplest answers are $\{\mathbf{v}_1, \mathbf{v}_2\}$ or $\{\mathbf{v}_1, \mathbf{v}_3\}$ or $\{\mathbf{v}_2, \mathbf{v}_3\}$. Other answers are possible.

21–31. See the *Study Guide* for hints.

33. *Hint:* Use the Invertible Matrix Theorem.

35. No. (Why is the set not a basis for H?)

37. $\{\cos \omega t, \sin \omega t\}$

39. Let A be the $n \times k$ matrix $[\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_k\,]$. Since A has fewer columns than rows, there cannot be a pivot position in each row of A. By Theorem 4 in Section 1.4, the columns of A do not span $\mathbb{R}^n$ and hence are not a basis for $\mathbb{R}^n$.

41. *Hint:* If $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly dependent, then there exist $c_1, \dots, c_p$, not all zero, such that $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$. Use this equation.

43. Neither polynomial is a multiple of the other polynomial, so $\{\mathbf{p}_1, \mathbf{p}_2\}$ is a linearly independent set in $\mathbb{P}_3$.

45. Let $\{\mathbf{v}_1, \mathbf{v}_3\}$ be any linearly independent set in the vector space V, and let $\mathbf{v}_2$ and $\mathbf{v}_4$ be linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_3$. Then $\{\mathbf{v}_1, \mathbf{v}_3\}$ is a basis for Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

47. You could be clever and find special values of t that produce several zeros in (5), and thereby create a system of equations that can be solved easily by hand. Or, you could use values of t such as $t = 0, .1, .2, \dots$ to create a system of equations that you can solve with a matrix program.

Section 4.4, page 238

1. $\begin{bmatrix} 3 \\ -7 \end{bmatrix}$ **3.** $\begin{bmatrix} -1 \\ -5 \\ 9 \end{bmatrix}$ **5.** $\begin{bmatrix} 8 \\ -5 \end{bmatrix}$ **7.** $\begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}$

9. $\begin{bmatrix} 2 & 1 \\ -9 & 8 \end{bmatrix}$ **11.** $\begin{bmatrix} 6 \\ 4 \end{bmatrix}$ **13.** $\begin{bmatrix} 2 \\ 6 \\ -1 \end{bmatrix}$

15–19. The *Study Guide* has hints.

21. $\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 5\mathbf{v}_1 - 2\mathbf{v}_2 = 10\mathbf{v}_1 - 3\mathbf{v}_2 + \mathbf{v}_3$ (infinitely many answers)

23. *Hint:* By hypothesis, the zero vector has a unique representation as a linear combination of elements of S.

25. $\begin{bmatrix} 9 & 2 \\ 4 & 1 \end{bmatrix}$

27. *Hint:* Suppose that $[\mathbf{u}]_{\mathcal{B}} = [\mathbf{w}]_{\mathcal{B}}$ for some $\mathbf{u}$ and $\mathbf{w}$ in V, and denote the entries in $[\mathbf{u}]_{\mathcal{B}}$ by $c_1, \dots, c_n$. Use the definition of $[\mathbf{u}]_{\mathcal{B}}$.

29. One possible approach: First, show that if $\mathbf{u}_1, \dots, \mathbf{u}_p$ are linearly *dependent*, then $[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_p]_{\mathcal{B}}$ are linearly dependent. Second, show that if $[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_p]_{\mathcal{B}}$ are linearly dependent, then $\mathbf{u}_1, \dots, \mathbf{u}_p$ are linearly *dependent*. Use the two equations displayed in the exercise. A slightly different proof is given in the *Study Guide*.

31. Linearly independent. (Justify answers to Exercises 31–38.)

33. Linearly dependent

35. a. The coordinate vectors $\begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 5 \\ -7 \end{bmatrix}, \begin{bmatrix} -4 \\ 5 \\ -6 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ do not span $\mathbb{R}^3$. Because of the isomorphism between $\mathbb{R}^3$ and $\mathbb{P}_2$, the corresponding polynomials do not span $\mathbb{P}_2$.

b. The coordinate vectors $\begin{bmatrix} 0 \\ 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -8 \\ -2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}$ span $\mathbb{R}^3$. Because of the isomorphism between $\mathbb{R}^3$ and $\mathbb{P}_2$, the corresponding polynomials span $\mathbb{P}_2$.

37. The coordinate vectors $\begin{bmatrix} 3 \\ 7 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 16 \\ -6 \\ 2 \end{bmatrix}$ are a linearly dependent subset of $\mathbb{R}^4$. Because of the isomorphism between $\mathbb{R}^4$ and $\mathbb{P}_3$, the corresponding polynomials form a linearly dependent subset of $\mathbb{P}_3$, and thus cannot be a basis for $\mathbb{P}_3$.

39. $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -5/3 \\ 8/3 \end{bmatrix}$ **41.** $\begin{bmatrix} 1.3 \\ 0 \\ 0.8 \end{bmatrix}$

Section 4.5, page 247

1. $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}$; dim is 2

3. $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -3 \\ 0 \end{bmatrix}$; dim is 3

5. $\begin{bmatrix} 1 \\ 2 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} -4 \\ 5 \\ 0 \\ 7 \end{bmatrix}$; dim is 2

7. No basis; dim is 0 **9.** 2 **11.** 2, 3, 3

13. 2, 2, 2 **15.** 0, 3, 3

17–25. See the *Study Guide*.

27. *Hint:* You need only show that the first four Hermite polynomials are linearly independent. Why?

29. $[\mathbf{p}]_{\mathcal{B}} = \left(3, 3, -2, \frac{3}{2}\right)$

31. *Hint:* Suppose S does span V, and use the Spanning Set Theorem. This leads to a contradiction, which shows that the spanning hypothesis is false.

33. 5, 3, 3

35. Yes; no. Since Col A is a four-dimensional subspace of $\mathbb{R}^4$, it coincides with $\mathbb{R}^4$. The null space cannot be $\mathbb{R}^3$, because the vectors in Nul A have 7 entries. Nul A is a three-dimensional subspace of $\mathbb{R}^7$, by the Rank Theorem.

37. 2

39. 5, 5. In both cases, the number of pivots cannot exceed the number of columns or the number of rows.

41. The functions $\{1, x, x^2, \dots\}$ are a linearly independent set with infinitely many vectors.

43–47. Consult the *Study Guide.*

49. $\dim \operatorname{Row} A = \dim \operatorname{Col} A = \operatorname{rank} A$, so the result follows from the Rank Theorem.

51. *Hint:* Since H is a nonzero subspace of a finite-dimensional space, H is finite-dimensional and has a basis, say, $\mathbf{v}_1, \dots, \mathbf{v}_p$. First show that $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)\}$ spans $T(H)$.

53. a. One basis is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_2, \mathbf{e}_3\}$. In fact, any two of the vectors $\mathbf{e}_2, \dots, \mathbf{e}_5$ will extend $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ to a basis of $\mathbb{R}^5$.

Section 4.6, page 253

1. a. $\begin{bmatrix} 6 & 9 \\ -2 & -4 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 \\ -2 \end{bmatrix}$ **3.** (ii)

5. a. $\begin{bmatrix} 4 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 1 & -2 \end{bmatrix}$ **b.** $\begin{bmatrix} 8 \\ 2 \\ 2 \end{bmatrix}$

7. $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} -3 & 1 \\ -5 & 2 \end{bmatrix}$, $\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} -2 & 1 \\ -5 & 3 \end{bmatrix}$

9. $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} 9 & -2 \\ -4 & 1 \end{bmatrix}$, $\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} 1 & 2 \\ 4 & 9 \end{bmatrix}$

11–13. See the *Study Guide*.

15. $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} 1 & 3 & 0 \\ -2 & -5 & 2 \\ 1 & 4 & 3 \end{bmatrix}$, $[-1 + 2t]_{\mathcal{B}} = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$

17. a. $\mathcal{B}$ is a basis for V.

b. The coordinate mapping is a linear transformation.

c. The product of a matrix and a vector

d. The coordinate vector of $\mathbf{v}$ relative to $\mathcal{B}$

19. a. $P^{-1} = \frac{1}{32}\begin{bmatrix} 32 & 0 & 16 & 0 & 12 & 0 & 10 \\ 0 & 32 & 0 & 24 & 0 & 20 & 0 \\ 0 & 0 & 16 & 0 & 16 & 0 & 15 \\ 0 & 0 & 0 & 8 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

b. P is the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$. So P^{-1} is the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$, by equation (5), and the columns of this matrix are the $\mathcal{C}$-coordinate vectors of the basis vectors in $\mathcal{B}$, by Theorem 15.

21. *Hint:* Let $\mathcal{C}$ be the basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Then the columns of P are $[\mathbf{u}_1]_\mathcal{C}$, $[\mathbf{u}_2]_\mathcal{C}$, and $[\mathbf{u}_3]_\mathcal{C}$. Use the definition of $\mathcal{C}$-coordinate vectors and matrix algebra to compute $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$. The solution method is discussed in the *Study Guide*. Here are the numerical answers:

a. $\mathbf{u}_1 = \begin{bmatrix} -6 \\ -5 \\ 21 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -6 \\ -9 \\ 32 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} -5 \\ 0 \\ 3 \end{bmatrix}$

b. $\mathbf{w}_1 = \begin{bmatrix} 28 \\ -9 \\ -3 \end{bmatrix}$, $\mathbf{w}_2 = \begin{bmatrix} 38 \\ -13 \\ 2 \end{bmatrix}$, $\mathbf{w}_3 = \begin{bmatrix} 21 \\ -7 \\ 3 \end{bmatrix}$

Section 4.7, Page 260

1. $(\ldots, 0, 2, 0, 2, 0, 2, 0, \ldots)$

3. $(\ldots, -2, 2, -2, 3, -1, 3, -1, \ldots)$

5. α

7. ϵ_c

9. Verify that the three properties in the definition of a LTI transformation are satisfied.

11. χ

13. Apply T to any signal to get a signal in the range of T.

15–21. See the *Study Guide*.

23. $I - \frac{3}{4}S$.

25. Show that W satisfies the three properties of a subspace.

27. $\{\chi - \alpha\}, 1$

29. Show that W satisfies the three properties of a subspace.

31. $\{S^{2m-1}(\delta)|$ where m is any integer$\}$. Yes W is an infinite dimensional subspace. Justify your answer.

Section 4.8, page 268

1. If $y_k = 2^k$, then $y_{k+1} = 2^{k+1}$ and $y_{k+2} = 2^{k+2}$. Substituting these formulas into the left side of the equation gives

$$\begin{aligned} y_{k+2} + 2y_{k+1} - 8y_k &= 2^{k+2} + 2 \cdot 2^{k+1} - 8 \cdot 2^k \\ &= 2^k(2^2 + 2 \cdot 2 - 8) \\ &= 2^k(0) = 0 \quad \text{for all } k \end{aligned}$$

Since the difference equation holds for all k, 2^k is a solution. A similar calculation works for $y_k = (-4)^k$.

3. The signals 2^k and $(-4)^k$ are linearly independent because neither is a multiple of the other. For instance, there is no scalar c such that $2^k = c(-4)^k$ *for all* k. By Theorem 17, the solution set H of the difference equation in Exercise 1 is two-dimensional. By the Basis Theorem in Section 4.5, the two linearly independent signals 2^k and $(-4)^k$ form a basis for H.

5. If $y_k = (-3)^k$, then

$$\begin{aligned} y_{k+2} + 6y_{k+1} + 9y_k &= (-3)^{k+2} + 6(-3)^{k+1} + 9(-3)^k \\ &= (-3)^k[(-3)^2 + 6(-3) + 9] \\ &= (-3)^k(0) = 0 \quad \text{for all } k \end{aligned}$$

Similarly, if $y_k = k(-3)^k$, then

$$\begin{aligned} &y_{k+2} + 6y_{k+1} + 9y_k \\ &\quad = (k+2)(-3)^{k+2} + 6(k+1)(-3)^{k+1} + 9k(-3)^k \\ &\quad = (-3)^k[(k+2)(-3)^2 + 6(k+1)(-3) + 9k] \\ &\quad = (-3)^k[9k + 18 - 18k - 18 + 9k] \\ &\quad = (-3)^k(0) \quad \text{for all } k \end{aligned}$$

Thus both $(-3)^k$ and $k(-3)^k$ are in the solution space H of the difference equation. Also, there is no scalar c such that $k(-3)^k = c(-3)^k$ *for all* k, because c must be chosen independently of k. Likewise,
there is no scalar c such that $(-3)^k = ck(-3)^k$ for all k. So the two signals are linearly independent. Since $\dim H = 2$, the signals form a basis for H, by the Basis Theorem.

7. Yes **9.** Yes

11. No, two
signals cannot span the three-dimensional solution space.

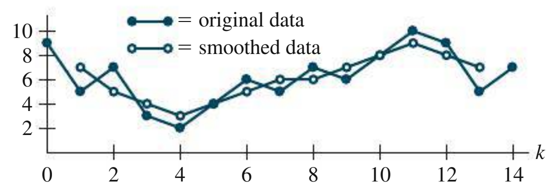
13. $\left(\frac{1}{3}\right)^k, \left(\frac{2}{3}\right)^k$ **15.** $5^k, (-5)^k$

17. $y_k = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^k$

19. $Y_k = c_1(.8)^k + c_2(.5)^k + 10 \to 10 \quad \text{as } k \to \infty$

21. $y_k = c_1(-2 + \sqrt{3})^k + c_2(-2 - \sqrt{3})^k$

23. 7, 5, 4, 3, 4, 5, 6, 6, 7, 8, 9, 8, 7; see figure below.



25. a. $y_{k+1} - 1.01y_k = -450$, $y_0 = 10{,}000$

b. MATLAB code:

```
pay = 450, y = 10000, m = 0
table = [0 ; y]
while y > 450
      y = 1.01*y - pay
      m = m + 1
      table = [table [m ; y] ]
                %append new column
end
m, y
```

c. At month 26, the last payment is \$114.88. The total paid by the borrower is \$11,364.88.

27. $k^2 + c_1 \cdot (-4)^k + c_2$ **29.** $2 - 2k + c_1 \cdot 4^k + c_2 \cdot 2^{-k}$

31. $\mathbf{x}_{k+1} = A\mathbf{x}_k$, where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 9 & -6 & -8 & 6 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} y_k \\ y_{k+1} \\ y_{k+2} \\ y_{k+3} \end{bmatrix}$$

33. The equation holds for all k, so it holds with k replaced by $k-1$, which transforms the equation into

$$y_{k+2} + 5y_{k+1} + 6y_k = 0 \quad \text{for all } k$$

The equation is of order 2.

35. For all k, the Casorati matrix $C(k)$ is not invertible. In this case, the Casorati matrix gives no information about the linear independence/dependence of the set of signals. In fact, neither signal is a multiple of the other, so they are linearly independent.

Chapter 4 Supplementary Exercises, page 271

1. T **2.** T **3.** F **4.** F **5.** T **6.** T
7. F **8.** F **9.** T **10.** F **11.** F **12.** F
13. T **14.** F **15.** T **16.** T **17.** F **18.** T
19. T

21. The set of all (b_1, b_2, b_3) satisfying $b_1 + 2b_2 + b_3 = 0$.

23. The vector $\mathbf{p}_1$ is not zero and $\mathbf{p}_2$ is not a multiple of $\mathbf{p}_1$, so keep both of these vectors. Since $\mathbf{p}_3 = 2\mathbf{p}_1 + 2\mathbf{p}_2$, discard $\mathbf{p}_3$. Since $\mathbf{p}_4$ has a t^2 term, it cannot be a linear combination of $\mathbf{p}_1$ and $\mathbf{p}_2$, so keep $\mathbf{p}_4$. Finally, $\mathbf{p}_5 = \mathbf{p}_1 + \mathbf{p}_4$, so discard $\mathbf{p}_5$. The resulting basis is $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_4\}$.

25. You would have to know that the solution set of the homogeneous system is spanned by two solutions. In this case, the null space of the 18×20 coefficient matrix A is at most two-dimensional. By the Rank Theorem, $\dim \text{Col}\, A \geq 20 - 2 = 18$, which means that $\text{Col}\, A = \mathbb{R}^{18}$, because A has 18 rows, and every equation $A\mathbf{x} = \mathbf{b}$ is consistent.

27. Let A be the standard $m \times n$ matrix of the transformation T.

a. If T is one-to-one, then the columns of A are linearly independent (Theorem 12 in Section 1.9), so $\dim \text{Nul}\, A = 0$. By the Rank Theorem, $\dim \text{Col}\, A = \text{rank}\, A = n$. Since the range of T is $\text{Col}\, A$, the dimension of the range of T is n.

b. If T is onto, then the columns of A span $\mathbb{R}^m$ (Theorem 12 in Section 1.9), so $\dim \text{Col}\, A = m$. By the Rank Theorem, $\dim \text{Nul}\, A = n - \dim \text{Col}\, A = n - m$. Since the kernel of T is $\text{Nul}\, A$, the dimension of the kernel of T is $n - m$.

29. If S is a finite spanning set for V, then a subset of S—say S'—is a basis for V. Since S' must span V, S' cannot be a proper subset of S because of the minimality of S. Thus $S' = S$, which proves that S is a basis for V.

30. a. *Hint:* Any $\mathbf{y}$ in $\text{Col}\, AB$ has the form $\mathbf{y} = AB\mathbf{x}$ for some $\mathbf{x}$.

31. By Exercise 12, $\text{rank}\, PA \leq \text{rank}\, A$, and $\text{rank}\, A = \text{rank}\, P^{-1}PA \leq \text{rank}\, PA$. Thus $\text{rank}\, PA = \text{rank}\, A$.

33. The equation $AB = 0$ shows that each column of B is in $\text{Nul}\, A$. Since $\text{Nul}\, A$ is a subspace, all linear combinations of the columns of B are in $\text{Nul}\, A$, so $\text{Col}\, B$ is a subspace of $\text{Nul}\, A$. By Theorem 12 in Section 4.5, $\dim \text{Col}\, B \leq \dim \text{Nul}\, A$. Applying the Rank Theorem, we find that

$$n = \text{rank}\, A + \dim \text{Nul}\, A \geq \text{rank}\, A + \text{rank}\, B$$

35. a. Let A_1 consist of the r pivot columns in A. The columns of A_1 are linearly independent. So A_1 is an $m \times r$ with rank r.

b. By the Rank Theorem applied to A_1, the dimension of $\text{Row}\, A$ is r, so A_1 has r linearly independent rows. Use them to form A_2. Then A_2 is $r \times r$ with linearly independent rows. By the Invertible Matrix Theorem, A_2 is invertible.

37. $$[\,B \quad AB \quad A^2B\,] = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -.9 & .81 \\ 1 & .5 & .25 \end{bmatrix} \sim \begin{bmatrix} 1 & -.9 & .81 \\ 0 & 1 & 0 \\ 0 & 0 & -.56 \end{bmatrix}$$

This matrix has rank 3, so the pair (A, B) is controllable.

39. $\text{rank}\,[\,B \quad AB \quad A^2B \quad A^3B\,] = 3$. The pair (A, B) is not controllable.

Chapter 5

Section 5.1, page 280

1. Yes **3.** No **5.** Yes, $\lambda = 0$ **7.** Yes, $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

9. $\lambda = 1$: $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$; $\lambda = 5$: $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ **11.** $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$

13. $\lambda = 1$: $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$; $\lambda = 2$: $\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$; $\lambda = 3$: $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

15. $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$ **17.** $0, 2, -1$

19. 0. Justify your answer.

21–29. See the *Study Guide*, after you have written your answers.

31. *Hint:* Use Theorem 2.

33. *Hint:* Use the equation $A\mathbf{x} = \lambda\mathbf{x}$ to find an equation involving A^{-1}.

35. *Hint:* For any λ, $(A - \lambda I)^T = A^T - \lambda I$. By a theorem (which one?), $A^T - \lambda I$ is invertible if and only if $A - \lambda I$ is invertible.

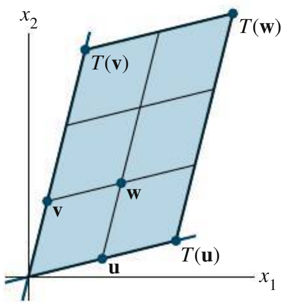
37. Let $\mathbf{v}$ be the vector in $\mathbb{R}^n$ whose entries are all 1's. Then $A\mathbf{v} = s\mathbf{v}$.

39. *Hint:* If A is the standard matrix of T, look for a nonzero vector $\mathbf{v}$ (a point in the plane) such that $A\mathbf{v} = \mathbf{v}$.

41. **a.** $\mathbf{x}_{k+1} = c_1\lambda^{k+1}\mathbf{u} + c_2\mu^{k+1}\mathbf{v}$

b.
$$\begin{aligned} A\mathbf{x}_k &= A(c_1\lambda^k\mathbf{u} + c_2\mu^k\mathbf{v}) \\ &= c_1\lambda^k A\mathbf{u} + c_2\mu^k A\mathbf{v} && \text{Linearity} \\ &= c_1\lambda^k\lambda\mathbf{u} + c_2\mu^k\mu\mathbf{v} && \mathbf{u}\text{ and }\mathbf{v}\text{ are eigenvectors.} \\ &= \mathbf{x}_{k+1} \end{aligned}$$

43.



45. $\lambda = 3$: $\begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$; $\lambda = 13$: $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. You can speed up your calculations with the program `nulbasis` discussed in the *Study Guide*.

47. $\lambda = -2$: $\begin{bmatrix} -2 \\ 7 \\ -5 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 7 \\ -5 \\ 0 \\ 5 \end{bmatrix}$;

$\lambda = 5$: $\begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

Section 5.2, page 288

1. $\lambda^2 - 4\lambda - 45$; $9, -5$ **3.** $\lambda^2 - 2\lambda - 1$; $1 \pm \sqrt{2}$

5. $\lambda^2 - 6\lambda + 9$; 3 **7.** $\lambda^2 - 9\lambda + 32$; no real eigenvalues

9. $-\lambda^3 + 4\lambda^2 - 9\lambda - 6$ **11.** $-\lambda^3 + 9\lambda^2 - 26\lambda + 24$

13. $-\lambda^3 + 18\lambda^2 - 95\lambda + 150$ **15.** 4, 3, 3, 1

17. 3, 3, 1, 1, 0

19. *Hint:* The equation given holds for all λ.

21–29. The *Study Guide* has hints.

31. *Hint:* Find an invertible matrix P so that $RQ = P^{-1}AP$.

33. In general, the eigenvectors of A are not the same as the eigenvectors of A^T, unless, of course, $A^T = A$.

35. $a = 32$: $\lambda = 1, 1, 2$
$a = 31.9$: $\lambda = .2958, 1, 2.7042$
$a = 31.8$: $\lambda = -.1279, 1, 3.1279$
$a = 32.1$: $\lambda = 1, 1.5 \pm .9747i$
$a = 32.2$: $\lambda = 1; 1.5 \pm 1.4663i$

Section 5.3, page 295

1. $\begin{bmatrix} 226 & -525 \\ 90 & -209 \end{bmatrix}$ **3.** $\begin{bmatrix} a^k & 0 \\ 3(a^k - b^k) & b^k \end{bmatrix}$

5. $\lambda = 5$: $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$; $\lambda = 1$: $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$

When an answer involves a diagonalization, $A = PDP^{-1}$, the factors P and D are not unique, so your answer may differ from that given here.

7. $P = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ **9.** Not diagonalizable

11. $P = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 3 & 1 \\ 4 & 3 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

13. $P = \begin{bmatrix} -1 & 2 & 1 \\ -1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

15. $P = \begin{bmatrix} -1 & -4 & -2 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

17. Not diagonalizable

19. $P = \begin{bmatrix} 1 & 3 & -1 & -1 \\ 0 & 2 & -1 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

21–27. See the *Study Guide*.

29. Yes. (Explain why.)

31. No, A must be diagonalizable. (Explain why.)

33. *Hint:* Write $A = PDP^{-1}$. Since A is invertible, 0 is not an eigenvalue of A, so D has nonzero entries on its diagonal.

35. One answer is $P_1 = \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}$, whose columns are eigenvectors corresponding to the eigenvalues in D_1.

37. *Hint:* Construct a suitable 2×2 triangular matrix.

39. $P = \begin{bmatrix} 2 & 2 & 1 & 6 \\ 1 & -1 & 1 & -3 \\ -1 & -7 & 1 & 0 \\ 2 & 2 & 0 & 4 \end{bmatrix}$,
$D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}$

41. $P = \begin{bmatrix} 6 & 3 & 2 & 4 & 3 \\ -1 & -1 & -1 & -3 & -1 \\ -3 & -3 & -4 & -2 & -4 \\ 3 & 0 & -1 & 5 & 0 \\ 0 & 3 & 4 & 0 & 5 \end{bmatrix}$,
$D = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Section 5.4, page 303

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

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

5. $24\mathbf{b}_1 - 20\mathbf{b}_2 + 11\mathbf{b}_3$

7. $\begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}$ **9.** $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

11. $\mathbf{b}_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

13. a. $A\mathbf{b}_1 = 2\mathbf{b}_1$, so $\mathbf{b}_1$ is an eigenvector of A. However, A has only one eigenvalue, $\lambda = 2$, and the eigenspace is only one-dimensional, so A is not diagonalizable.

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

15. a. $T(\mathbf{p}) = 3 + 3t + 3t^2 = 3\mathbf{p}$ so $\mathbf{p}$ is an eigenvector of T with eigenvalue 3.

b. $T(\mathbf{p}) = -1 - t - t^2$ so $\mathbf{p}$ is not an eigenvector.

17–19. See the *Study Guide.*

21. By definition, if A is similar to B, there exists an invertible matrix P such that $P^{-1}AP = B$. (See Section 5.2.) Then B is invertible because it is the product of invertible matrices. To show that A^{-1} is similar to B^{-1}, use the equation $P^{-1}AP = B$. See the *Study Guide*.

23. *Hint:* Review Practice Problem 2.

25. *Hint:* Compute $B(P^{-1}\mathbf{x})$.

27. *Hint:* Write $A = PBP^{-1} = (PB)P^{-1}$, and use the trace property.

29. $S(\chi) = \chi$ so χ is an eigenvector of S with eigenvalue 1.

31. $M_2(\alpha) = 0$ so α is an eigenvector of M_2 with eigenvalue 0.

33. $P^{-1}AP = \begin{bmatrix} 8 & 3 & -6 \\ 0 & 1 & 3 \\ 0 & 0 & -3 \end{bmatrix}$

35. $\lambda = 2\colon \mathbf{b}_1 = \begin{bmatrix} 0 \\ -3 \\ 3 \\ 2 \end{bmatrix}; \lambda = 4\colon \mathbf{b}_2 = \begin{bmatrix} -30 \\ -7 \\ 3 \\ 0 \end{bmatrix}$,
$\mathbf{b}_3 = \begin{bmatrix} 39 \\ 5 \\ 0 \\ 3 \end{bmatrix}; \lambda = 5\colon \mathbf{b}_4 = \begin{bmatrix} 11 \\ -3 \\ 4 \\ 4 \end{bmatrix}$;
basis: $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$

Section 5.5, page 310

1. $\lambda = 2 + i, \begin{bmatrix} -1 + i \\ 1 \end{bmatrix}; \quad \lambda = 2 - i, \begin{bmatrix} -1 - i \\ 1 \end{bmatrix}$

3. $\lambda = 2 + 3i, \begin{bmatrix} 1 - 3i \\ 2 \end{bmatrix}; \quad \lambda = 2 - 3i, \begin{bmatrix} 1 + 3i \\ 2 \end{bmatrix}$

5. $\lambda = 2 + 2i, \begin{bmatrix} 1 \\ 2 + 2i \end{bmatrix}; \quad \lambda = 2 - 2i, \begin{bmatrix} 1 \\ 2 - 2i \end{bmatrix}$

7. $\lambda = \sqrt{3} \pm i, \varphi = \pi/6$ radian, $r = 2$

9. $\lambda = -\sqrt{3}/2 \pm (1/2)i, \varphi = -5\pi/6$ radians, $r = 1$

11. $\lambda = .1 \pm .1i, \varphi = -\pi/4$ radian, $r = \sqrt{2}/10$

In Exercises 13–20, other answers are possible. Any P that makes $P^{-1}AP$ equal to the given C or to C^T is a satisfactory answer. First find P; then compute $P^{-1}AP$.

13. $P = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$

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

17. $P = \begin{bmatrix} 2 & -1 \\ 5 & 0 \end{bmatrix}, C = \begin{bmatrix} -.6 & -.8 \\ .8 & -.6 \end{bmatrix}$

19. $P = \begin{bmatrix} 2 & -1 \\ 2 & 0 \end{bmatrix}, C = \begin{bmatrix} .96 & -.28 \\ .28 & .96 \end{bmatrix}$

21. $\mathbf{y} = \begin{bmatrix} 2 \\ -1 + 2i \end{bmatrix} = \dfrac{-1 + 2i}{5} \begin{bmatrix} -2 - 4i \\ 5 \end{bmatrix}$

23–25. See the *Study Guide.*

27. (a) Properties of conjugates and the fact that $\overline{\mathbf{x}}^T = \overline{\mathbf{x}^T}$; (b) $\overline{A\mathbf{x}} = A\overline{\mathbf{x}}$ and A is real; (c) because $\mathbf{x}^T A\overline{\mathbf{x}}$ is a scalar and hence may be viewed as a 1×1 matrix; (d) properties of transposes; (e) $A^T = A$, definition of q

29. *Hint:* First write $\mathbf{x} = \text{Re}\,\mathbf{x} + i(\text{Im}\,\mathbf{x})$.

31. $P = \begin{bmatrix} 1 & -1 & -2 & 0 \\ -4 & 0 & 0 & 2 \\ 0 & 0 & -3 & -1 \\ 2 & 0 & 4 & 0 \end{bmatrix}, C = \begin{bmatrix} .2 & -.5 & 0 & 0 \\ .5 & .2 & 0 & 0 \\ 0 & 0 & .3 & -.1 \\ 0 & 0 & .1 & .3 \end{bmatrix}$

Other choices are possible, but C must equal $P^{-1}AP$.

Section 5.6, page 320

1. **a.** *Hint:* Find c_1, c_2 such that $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. Use this representation and the fact that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A to compute $\mathbf{x}_1 = \begin{bmatrix} 49/3 \\ 41/3 \end{bmatrix}$.

   b. In general, $\mathbf{x}_k = 5(3)^k\mathbf{v}_1 - 4(\frac{1}{3})^k\mathbf{v}_2 \quad \text{for } k \geq 0$.

3. When $p = .2$, the eigenvalues of A are .9 and .7, and

$$\mathbf{x}_k = c_1(.9)^k \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2(.7)^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} \to \mathbf{0} \quad \text{as } k \to \infty$$

The higher predation rate cuts down the owls' food supply, and eventually both predator and prey populations perish.

5. If $p = .325$, the eigenvalues are 1.05 and .55. Since $1.05 > 1$, both populations will grow at 5% per year. An eigenvector for 1.05 is $(6, 13)$, so eventually there will be approximately 6 spotted owls to every 13 (thousand) flying squirrels.

7. **a.** The origin is a saddle point because A has one eigenvalue larger than 1 and one smaller than 1 (in absolute value).

   b. The direction of greatest attraction is given by the eigenvector corresponding to the eigenvalue $1/3$, namely, $\mathbf{v}_2$. All vectors that are multiples of $\mathbf{v}_2$ are attracted to the origin. The direction of greatest repulsion is given by the eigenvector $\mathbf{v}_1$. All multiples of $\mathbf{v}_1$ are repelled.

   c. See the *Study Guide*.

9. Saddle point; eigenvalues: 2, .5; direction of greatest repulsion: the line through $(0, 0)$ and $(-1, 1)$; direction of greatest attraction: the line through $(0, 0)$ and $(1, 4)$

11. Attractor; eigenvalues: .9, .8; greatest attraction: line through $(0, 0)$ and $(5, 4)$

13. Repellor; eigenvalues: 1.2, 1.1; greatest repulsion: line through $(0, 0)$ and $(3, 4)$

15. $\mathbf{x}_k = \mathbf{v}_1 + .1(.5)^k \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} + .3(.2)^k \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \to \mathbf{v}_1$ as $k \to \infty$

17. **a.** $A = \begin{bmatrix} 0 & 1.6 \\ .3 & .8 \end{bmatrix}$

   b. The population is growing because the largest eigenvalue of A is 1.2, which is larger than 1 in magnitude. The eventual growth rate is 1.2, which is 20% per year. The eigenvector $(4, 3)$ for $\lambda_1 = 1.2$ shows that there will be 4 juveniles for every 3 adults.

   c. The juvenile–adult ratio seems to stabilize after about 5 or 6 years. The *Study Guide* describes how to construct a matrix program to generate a data matrix whose columns list the numbers of juveniles and adults each year. Graphing the data is also discussed.

Section 5.7, page 328

1. $\mathbf{x}(t) = \frac{5}{2}\begin{bmatrix} -3 \\ 1 \end{bmatrix}e^{4t} - \frac{3}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix}e^{2t}$

3. $-\frac{5}{2}\begin{bmatrix} -3 \\ 1 \end{bmatrix}e^{t} + \frac{9}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix}e^{-t}$. The origin is a saddle point. The direction of greatest attraction is the line through $(-1, 1)$ and the origin. The direction of greatest repulsion is the line through $(-3, 1)$ and the origin.

5. $-\frac{1}{2}\begin{bmatrix} 1 \\ 3 \end{bmatrix}e^{4t} + \frac{7}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}e^{6t}$. The origin is a repellor. The direction of greatest repulsion is the line through $(1, 1)$ and the origin.

7. Set $P = \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 4 & 0 \\ 0 & 6 \end{bmatrix}$. Then $A = PDP^{-1}$. Substituting $\mathbf{x} = P\mathbf{y}$ into $\mathbf{x}' = A\mathbf{x}$, we have

$$\begin{aligned} \frac{d}{dt}(P\mathbf{y}) &= A(P\mathbf{y}) \\ P\mathbf{y}' &= PDP^{-1}(P\mathbf{y}) = PD\mathbf{y} \end{aligned}$$

Left-multiplying by P^{-1} gives

$$\mathbf{y}' = D\mathbf{y}, \quad \text{or} \quad \begin{bmatrix} y_1'(t) \\ y_2'(t) \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 6 \end{bmatrix}\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$$

9. (complex solution):

$$c_1\begin{bmatrix} 1-i \\ 1 \end{bmatrix}e^{(-2+i)t} + c_2\begin{bmatrix} 1+i \\ 1 \end{bmatrix}e^{(-2-i)t}$$

(real solution):

$$c_1\begin{bmatrix} \cos t + \sin t \\ \cos t \end{bmatrix}e^{-2t} + c_2\begin{bmatrix} \sin t - \cos t \\ \sin t \end{bmatrix}e^{-2t}$$

The trajectories spiral in toward the origin.

11. (complex): $c_1\begin{bmatrix} -3+3i \\ 2 \end{bmatrix}e^{3it} + c_2\begin{bmatrix} -3-3i \\ 2 \end{bmatrix}e^{-3it}$

(real):

$$c_1\begin{bmatrix} -3\cos 3t - 3\sin 3t \\ 2\cos 3t \end{bmatrix} + c_2\begin{bmatrix} -3\sin 3t + 3\cos 3t \\ 2\sin 3t \end{bmatrix}$$

The trajectories are ellipses about the origin.

13. (complex): $c_1\begin{bmatrix} 1+i \\ 2 \end{bmatrix}e^{(1+3i)t} + c_2\begin{bmatrix} 1-i \\ 2 \end{bmatrix}e^{(1-3i)t}$

(real): $c_1\begin{bmatrix} \cos 3t - \sin 3t \\ 2\cos 3t \end{bmatrix}e^{t} + c_2\begin{bmatrix} \sin 3t + \cos 3t \\ 2\sin 3t \end{bmatrix}e^{t}$

The trajectories spiral out, away from the origin.

15. $\mathbf{x}(t) = c_1\begin{bmatrix}-1\\0\\1\end{bmatrix}e^{-2t} + c_2\begin{bmatrix}-6\\1\\5\end{bmatrix}e^{-t} + c_3\begin{bmatrix}-4\\1\\4\end{bmatrix}e^{t}$
The origin is a saddle point. A solution with $c_3 = 0$ is attracted to the origin. A solution with $c_1 = c_2 = 0$ is repelled.

17. (complex):
$c_1\begin{bmatrix}-3\\1\\1\end{bmatrix}e^{t} + c_2\begin{bmatrix}23-34i\\-9+14i\\3\end{bmatrix}e^{(5+2i)t} +$
$c_3\begin{bmatrix}23+34i\\-9-14i\\3\end{bmatrix}e^{(5-2i)t}$
(real): $c_1\begin{bmatrix}-3\\1\\1\end{bmatrix}e^{t} + c_2\begin{bmatrix}23\cos 2t + 34\sin 2t\\-9\cos 2t - 14\sin 2t\\3\cos 2t\end{bmatrix}e^{5t} +$
$c_3\begin{bmatrix}23\sin 2t - 34\cos 2t\\-9\sin 2t + 14\cos 2t\\3\sin 2t\end{bmatrix}e^{5t}$
The origin is a repellor. The trajectories spiral outward, away from the origin.

19. $A = \begin{bmatrix}-2 & 3/4\\1 & -1\end{bmatrix}$,
$\begin{bmatrix}v_1(t)\\v_2(t)\end{bmatrix} = \frac{5}{2}\begin{bmatrix}1\\2\end{bmatrix}e^{-.5t} - \frac{1}{2}\begin{bmatrix}-3\\2\end{bmatrix}e^{-2.5t}$

21. $A = \begin{bmatrix}-1 & -8\\5 & -5\end{bmatrix}$,
$\begin{bmatrix}i_L(t)\\v_C(t)\end{bmatrix} = \begin{bmatrix}-20\sin 6t\\15\cos 6t - 5\sin 6t\end{bmatrix}e^{-3t}$

Section 5.8, page 335

1. Eigenvector: $\mathbf{x}_4 = \begin{bmatrix}1\\.3326\end{bmatrix}$, or $A\mathbf{x}_4 = \begin{bmatrix}4.9978\\1.6652\end{bmatrix}$;
$\lambda \approx 4.9978$

3. Eigenvector: $\mathbf{x}_4 = \begin{bmatrix}.5188\\1\end{bmatrix}$, or $A\mathbf{x}_4 = \begin{bmatrix}.4594\\.9075\end{bmatrix}$;
$\lambda \approx .9075$

5. $\mathbf{x} = \begin{bmatrix}-.7999\\1\end{bmatrix}$, $A\mathbf{x} = \begin{bmatrix}4.0015\\-5.0020\end{bmatrix}$; estimated
$\lambda = -5.0020$

7. $\mathbf{x}_k$: $\begin{bmatrix}.75\\1\end{bmatrix}, \begin{bmatrix}1\\.9565\end{bmatrix}, \begin{bmatrix}.9932\\1\end{bmatrix}, \begin{bmatrix}1\\.9990\end{bmatrix}, \begin{bmatrix}.9998\\1\end{bmatrix}$
μ_k: 11.5, 12.78, 12.96, 12.9948, 12.9990

9. $\mu_5 = 8.4233$, $\mu_6 = 8.4246$; actual value: 8.42443 (accurate to 5 places)

11. μ_k: 5.8000, 5.9655, 5.9942, 5.9990 $(k = 1, 2, 3, 4)$;
$R(\mathbf{x}_k)$: 5.9655, 5.9990, 5.99997, 5.9999993

13. Yes, but the sequences may converge very slowly.

15. *Hint:* Write $A\mathbf{x} - \alpha\mathbf{x} = (A - \alpha I)\mathbf{x}$, and use the fact that $(A - \alpha I)$ is invertible when α is *not* an eigenvalue of A.

17. $\nu_0 = 3.3384$, $\nu_1 = 3.32119$ (accurate to 4 places with rounding), $\nu_2 = 3.3212209$. Actual value: 3.3212201 (accurate to 7 places)

19. **a.** $\mu_6 = 30.2887 = \mu_7$ to four decimal places. To six places, the largest eigenvalue is 30.288685, with eigenvector (.957629, .688937, 1, .943782).

b. The inverse power method (with $\alpha = 0$) produces $\mu_1^{-1} = .010141$, $\mu_2^{-1} = .010150$. To seven places, the smallest eigenvalue is .0101500, with eigenvector (−.603972, 1, −.251135, .148953). The reason for the rapid convergence is that the next-to-smallest eigenvalue is near .85.

21. **a.** If the eigenvalues of A are all less than 1 in magnitude, and if $\mathbf{x} \neq \mathbf{0}$, then $A^k\mathbf{x}$ is approximately an eigenvector for large k.

b. If the strictly dominant eigenvalue is 1, and if $\mathbf{x}$ has a component in the direction of the corresponding eigenvector, then $\{A^k\mathbf{x}\}$ will converge to a multiple of that eigenvector.

c. If the eigenvalues of A are all greater than 1 in magnitude, and if $\mathbf{x}$ is not an eigenvector, then the distance from $A^k\mathbf{x}$ to the nearest eigenvector will *increase* as $k \to \infty$.

Section 5.9, page 344

1. **a.** $\begin{array}{cc} \text{From} & \\ \begin{array}{cc}\text{N} & \text{M}\end{array} & \text{To} \\ \begin{bmatrix}.7 & .6\\.3 & .4\end{bmatrix} & \begin{array}{l}\text{News}\\\text{Music}\end{array}\end{array}$ **b.** $\begin{bmatrix}1\\0\end{bmatrix}$ **c.** 33%

3. **a.** $\begin{array}{cc} \text{From} & \\ \begin{array}{cc}\text{H} & \text{I}\end{array} & \text{To} \\ \begin{bmatrix}.95 & .45\\.05 & .55\end{bmatrix} & \begin{array}{l}\text{Healthy}\\\text{Ill}\end{array}\end{array}$ **b.** 15%, 12.5%

c. .925; use $\mathbf{x}_0 = \begin{bmatrix}1\\0\end{bmatrix}$.

5. $\begin{bmatrix}.4\\.6\end{bmatrix}$ **7.** $\begin{bmatrix}1/4\\1/2\\1/4\end{bmatrix}$

9. Yes, because P^2 has all positive entries.

11. **a.** $\begin{bmatrix}2/3\\1/3\end{bmatrix}$ **b.** 2/3

13. **a.** $\begin{bmatrix}.9\\.1\end{bmatrix}$ **b.** .10, no

15–19. See the *Study Guide.*

21. No. $\mathbf{q}$ is not a probability vector since its entries do not add to 1.

23. No. $A\mathbf{q}$ does not equal $\mathbf{q}$.

25. 67%

27. a. The entries in a column of P sum to 1. A column in the matrix $P - I$ has the same entries as in P except that one of the entries is decreased by 1. Hence each column sum is 0.

b. By (a), the bottom row of $P - I$ is the negative of the sum of the other rows.

c. By (b) and the Spanning Set Theorem, the bottom row of $P - I$ can be removed and the remaining $(n-1)$ rows will still span the row space. Alternatively, use (a) and the fact that row operations do not change the row space. Let A be the matrix obtained from $P - I$ by adding to the bottom row all the other rows. By (a), the row space is spanned by the first $(n-1)$ rows of A.

d. By the Rank Theorem and (c), the dimension of the column space of $P - I$ is less than n, and hence the null space is nontrivial. Instead of the Rank Theorem, you may use the Invertible Matrix Theorem, since $P - I$ is a square matrix.

29. a. The product $S\mathbf{x}$ equals the sum of the entries in $\mathbf{x}$. For a probability vector, this sum must be 1.

b. $P = [\,\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n\,]$, where the $\mathbf{p}_i$ are probability vectors. By matrix multiplication and part (a),

$$SP = [\,S\mathbf{p}_1 \quad S\mathbf{p}_2 \quad \cdots \quad S\mathbf{p}_n\,] = [\,1 \quad 1 \quad \cdots \quad 1\,] = S$$

c. By part (b), $S(P\mathbf{x}) = (SP)\mathbf{x} = S\mathbf{x} = 1$. Also, the entries in $P\mathbf{x}$ are nonnegative (because P and $\mathbf{x}$ have nonnegative entries). Hence, by (a), $P\mathbf{x}$ is a probability vector.

31. a. To four decimal places,

$$P^4 = P^5 = \begin{bmatrix} .2816 & .2816 & .2816 & .2816 \\ .3355 & .3355 & .3355 & .3355 \\ .1819 & .1819 & .1819 & .1819 \\ .2009 & .2009 & .2009 & .2009 \end{bmatrix},$$

$$\mathbf{q} = \begin{bmatrix} .2816 \\ .3355 \\ .1819 \\ .2009 \end{bmatrix}$$

Note that, due to round-off, the column sums are not 1.

b. To four decimal places,

$$Q^{80} = \begin{bmatrix} .7354 & .7348 & .7351 \\ .0881 & .0887 & .0884 \\ .1764 & .1766 & .1765 \end{bmatrix},$$

$$Q^{116} = Q^{117} = \begin{bmatrix} .7353 & .7353 & .7353 \\ .0882 & .0882 & .0882 \\ .1765 & .1765 & .1765 \end{bmatrix},$$

$$\mathbf{q} = \begin{bmatrix} .7353 \\ .0882 \\ .1765 \end{bmatrix}$$

c. Let P be an $n \times n$ regular stochastic matrix, $\mathbf{q}$ the steady-state vector of P, and $\mathbf{e}_1$ the first column of the identity matrix. Then $P^k\mathbf{e}_1$ is the first column of P^k. By Theorem 11, $P^k\mathbf{e}_1 \to \mathbf{q}$ as $k \to \infty$. Replacing $\mathbf{e}_1$ by the other columns of the identity matrix, we conclude that each column of P^k converges to $\mathbf{q}$ as $k \to \infty$. Thus $P^k \to [\,\mathbf{q} \quad \mathbf{q} \quad \cdots \quad \mathbf{q}\,]$.

Chapter 5 Supplementary Exercises, page 346

1. T	**2.** F	**3.** T	**4.** F	**5.** T
6. T	**7.** F	**8.** T	**9.** F	**10.** T
11. F	**12.** F	**13.** F	**14.** T	**15.** F
16. T	**17.** F	**18.** T	**19.** F	**20.** T
21. T	**22.** T	**23.** F		

25. a. Suppose $A\mathbf{x} = \lambda\mathbf{x}$, with $\mathbf{x} \neq \mathbf{0}$. Then $(5I - A)\mathbf{x} = 5\mathbf{x} - A\mathbf{x} = 5\mathbf{x} - \lambda\mathbf{x} = (5-\lambda)\mathbf{x}$. The eigenvalue is $5 - \lambda$.

b. $(5I - 3A + A^2)\mathbf{x} = 5\mathbf{x} - 3A\mathbf{x} + A(A\mathbf{x}) = 5\mathbf{x} - 3\lambda\mathbf{x} + \lambda^2\mathbf{x} = (5 - 3\lambda + \lambda^2)\mathbf{x}$. The eigenvalue is $5 - 3\lambda + \lambda^2$.

27. Suppose $A\mathbf{x} = \lambda\mathbf{x}$, with $\mathbf{x} \neq \mathbf{0}$. Then

$$\begin{aligned} p(A)\mathbf{x} &= (c_0 I + c_1 A + c_2 A^2 + \cdots + c_n A^n)\mathbf{x} \\ &= c_0\mathbf{x} + c_1 A\mathbf{x} + c_2 A^2\mathbf{x} + \cdots + c_n A^n\mathbf{x} \\ &= c_0\mathbf{x} + c_1\lambda\mathbf{x} + c_2\lambda^2\mathbf{x} + \cdots + c_n\lambda^n\mathbf{x} = p(\lambda)\mathbf{x} \end{aligned}$$

So $p(\lambda)$ is an eigenvalue of the matrix $p(A)$.

29. If $A = PDP^{-1}$, then $p(A) = Pp(D)P^{-1}$, as shown in Exercise 28. If the (j, j) entry in D is λ, then the (j, j) entry in D^k is λ^k, and so the (j, j) entry in $p(D)$ is $p(\lambda)$. If p is the characteristic polynomial of A, then $p(\lambda) = 0$ for each diagonal entry of D, because these entries in D are the eigenvalues of A. Thus $p(D)$ is the zero matrix. Thus $p(A) = P0P^{-1} = 0$.

31. If $I - A$ were not invertible, then the equation $(I - A)\mathbf{x} = \mathbf{0}$ would have a nontrivial solution $\mathbf{x}$. Then $\mathbf{x} - A\mathbf{x} = \mathbf{0}$ and $A\mathbf{x} = 1 \cdot \mathbf{x}$, which shows that A would have 1 as an eigenvalue. This cannot happen if all the eigenvalues are less than 1 in magnitude. So $I - A$ must be invertible.

33. a. Take $\mathbf{x}$ in H. Then $\mathbf{x} = c\mathbf{u}$ for some scalar c. So $A\mathbf{x} = A(c\mathbf{u}) = c(A\mathbf{u}) = c(\lambda\mathbf{u}) = (c\lambda)\mathbf{u}$, which shows that $A\mathbf{x}$ is in H.

b. Let $\mathbf{x}$ be a nonzero vector in K. Since K is one-dimensional, K must be the set of all scalar multiples of $\mathbf{x}$. If K is invariant under A, then $A\mathbf{x}$ is in K and hence $A\mathbf{x}$ is a multiple of $\mathbf{x}$. Thus $\mathbf{x}$ is an eigenvector of A.

35. 1, 3, 7

37. Replace a by $a - \lambda$ in the determinant formula from Exercise 30 in Chapter 3 Supplementary Exercises:

$$\det(A - \lambda I) = (a - b - \lambda)^{n-1}[a - \lambda + (n-1)b]$$

This determinant is zero only if $a - b - \lambda = 0$ or $a - \lambda + (n-1)b = 0$. Thus λ is an eigenvalue of A if and only if $\lambda = a - b$ or $\lambda = a + (n-1)b$. From the formula for $\det(A - \lambda I)$ above, the algebraic multiplicity is $n - 1$ for $a - b$ and 1 for $a + (n - 1)b$.

39. $\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = \lambda^2 - (\operatorname{tr} A)\lambda + \det A$. Use the quadratic formula to solve the characteristic equation:

$$\lambda = \frac{\operatorname{tr} A \pm \sqrt{(\operatorname{tr} A)^2 - 4 \det A}}{2}$$

The eigenvalues are both real if and only if the discriminant is nonnegative, that is, $(\operatorname{tr} A)^2 - 4 \det A \geq 0$. This inequality simplifies to $(\operatorname{tr} A)^2 \geq 4 \det A$ and $\left(\frac{\operatorname{tr} A}{2}\right)^2 \geq \det A$.

41. $C_p = \begin{bmatrix} 0 & 1 \\ -6 & 5 \end{bmatrix}$; $\det(C_p - \lambda I) = 6 - 5\lambda + \lambda^2 = p(\lambda)$

43. If p is a polynomial of order 2, then a calculation such as in Exercise 41 shows that the characteristic polynomial of C_p is $p(\lambda) = (-1)^2 p(\lambda)$, so the result is true for $n = 2$. Suppose the result is true for $n = k$ for some $k \geq 2$, and consider a polynomial p of degree $k + 1$. Then expanding $\det(C_p - \lambda I)$ by cofactors down the first column, the determinant of $C_p - \lambda I$ equals

$$(-\lambda) \det \begin{bmatrix} -\lambda & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & & & 1 \\ -a_1 & -a_2 & \cdots & -a_k - \lambda \end{bmatrix} + (-1)^{k+1} a_0$$

The $k \times k$ matrix shown is $C_q - \lambda I$, where $q(t) = a_1 + a_2 t + \cdots + a_k t^{k-1} + t^k$. By the induction assumption, the determinant of $C_q - \lambda I$ is $(-1)^k q(\lambda)$. Thus

$$\begin{aligned} \det(C_p - \lambda I) &= (-1)^{k+1} a_0 + (-\lambda)(-1)^k q(\lambda) \\ &= (-1)^{k+1}[a_0 + \lambda(a_1 + \cdots + a_k \lambda^{k-1} + \lambda^k)] \\ &= (-1)^{k+1} p(\lambda) \end{aligned}$$

So the formula holds for $n = k + 1$ when it holds for $n = k$. By the principle of induction, the formula for $\det(C_p - \lambda I)$ is true for all $n \geq 2$.

45. From Exercise 44, the columns of the Vandermonde matrix V are eigenvectors of C_p, corresponding to the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ (the roots of the polynomial p). Since these eigenvalues are distinct, the eigenvectors form a linearly independent set, by Theorem 2 in Section 5.1. Thus V has linearly independent columns and hence is invertible, by the Invertible Matrix Theorem. Finally, since the columns of V are eigenvectors of C_p, the Diagonalization Theorem (Theorem 5 in Section 5.3) shows that $V^{-1} C_p V$ is diagonal.

47. If your matrix program computes eigenvalues and eigenvectors by iterative methods rather than symbolic calculations, you may have some difficulties. You should find that $AP - PD$ has extremely small entries and PDP^{-1} is close to A. (This was true just a few years ago, but the situation could change as matrix programs continue to improve.) If you constructed P from the program's eigenvectors, check the condition number of P. This may indicate that you do not really have three linearly independent eigenvectors.

Chapter 6

Section 6.1, page 356

1. $5, 4, \frac{4}{5}$ **3.** $\begin{bmatrix} 3/35 \\ -1/35 \\ -1/7 \end{bmatrix}$ **5.** $\begin{bmatrix} 8/13 \\ 12/13 \end{bmatrix}$

7. $\sqrt{35}$ **9.** $\begin{bmatrix} -.6 \\ .8 \end{bmatrix}$ **11.** $\begin{bmatrix} 7/\sqrt{69} \\ 2/\sqrt{69} \\ 4/\sqrt{69} \end{bmatrix}$

13. $5\sqrt{5}$ **15.** Not orthogonal **17.** Orthogonal

19–27. Refer to the *Study Guide* after you have written your answers.

29. *Hint:* Use Theorems 3 and 2 from Section 2.1.

31. $\mathbf{u} \cdot \mathbf{v} = 0$, $\|\mathbf{u}\|^2 = 30$, $\|\mathbf{v}\|^2 = 101$, $\|\mathbf{u} + \mathbf{v}\|^2 = (-5)^2 + (-9)^2 + 5^2 = 131 = 30 + 101$

33. The set of all multiples of $\begin{bmatrix} -b \\ a \end{bmatrix}$ (when $\mathbf{v} \neq \mathbf{0}$)

35. *Hint:* Use the definition of orthogonality.

37. *Hint:* Consider a typical vector $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ in W.

39. *Hint:* If $\mathbf{x}$ is in $W^{\perp}$, then $\mathbf{x}$ is orthogonal to every vector in W.

41. State your conjecture and verify it algebraically.

Section 6.2, page 364

1. Not orthogonal **3.** Not orthogonal **5.** Orthogonal

7. Show $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$, mention Theorem 4, and observe that two linearly independent vectors in $\mathbb{R}^2$ form a basis. Then obtain

$$\mathbf{x} = \tfrac{39}{13}\begin{bmatrix} 2 \\ -3 \end{bmatrix} + \tfrac{26}{52}\begin{bmatrix} 6 \\ 4 \end{bmatrix} = 3\begin{bmatrix} 2 \\ -3 \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

9. Show $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$, $\mathbf{u}_1 \cdot \mathbf{u}_3 = 0$, and $\mathbf{u}_2 \cdot \mathbf{u}_3 = 0$. Mention Theorem 4, and observe that three linearly independent vectors in $\mathbb{R}^3$ form a basis. Then obtain

$$\mathbf{x} = \tfrac{5}{2}\mathbf{u}_1 - \tfrac{27}{18}\mathbf{u}_2 + \tfrac{18}{9}\mathbf{u}_3 = \tfrac{5}{2}\mathbf{u}_1 - \tfrac{3}{2}\mathbf{u}_2 + 2\mathbf{u}_3$$

11. $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ **13.** $\mathbf{y} = \begin{bmatrix} -4/5 \\ 7/5 \end{bmatrix} + \begin{bmatrix} 14/5 \\ 8/5 \end{bmatrix}$

15. $\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} .6 \\ -.8 \end{bmatrix}$, distance is 1

17. $\begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$

19. Orthonormal **21.** Orthonormal

23–31. See the *Study Guide*.

33. *Hint:* $\|U\mathbf{x}\|^2 = (U\mathbf{x})^T(U\mathbf{x})$. Also, parts (a) and (c) follow from (b).

35. *Hint:* You need two theorems, one of which applies only to *square* matrices.

37. *Hint:* If you have a candidate for an inverse, you can check to see whether the candidate works.

39. Suppose $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. Replace $\mathbf{u}$ by $c\mathbf{u}$ with $c \neq 0$; then

$$\frac{\mathbf{y}\cdot(c\mathbf{u})}{(c\mathbf{u})\cdot(c\mathbf{u})}(c\mathbf{u}) = \frac{c(\mathbf{y}\cdot\mathbf{u})}{c^2\mathbf{u}\cdot\mathbf{u}}(c)\mathbf{u} = \hat{\mathbf{y}}$$

41. Let $L = \text{Span}\{\mathbf{u}\}$, where $\mathbf{u}$ is nonzero, and let $T(\mathbf{x}) = \text{proj}_L\,\mathbf{x}$. By definition,

$$T(\mathbf{x}) = \frac{\mathbf{x}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = (\mathbf{x}\cdot\mathbf{u})(\mathbf{u}\cdot\mathbf{u})^{-1}\mathbf{u}$$

For $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and any scalars c and d, properties of the inner product (Theorem 1) show that

$$\begin{aligned} T(c\mathbf{x} + d\mathbf{y}) &= [(c\mathbf{x} + d\mathbf{y})\cdot\mathbf{u}](\mathbf{u}\cdot\mathbf{u})^{-1}\mathbf{u} \\ &= [c(\mathbf{x}\cdot\mathbf{u}) + d(\mathbf{y}\cdot\mathbf{u})](\mathbf{u}\cdot\mathbf{u})^{-1}\mathbf{u} \\ &= c(\mathbf{x}\cdot\mathbf{u})(\mathbf{u}\cdot\mathbf{u})^{-1}\mathbf{u} + d(\mathbf{y}\cdot\mathbf{u})(\mathbf{u}\cdot\mathbf{u})^{-1}\mathbf{u} \\ &= cT(\mathbf{x}) + dT(\mathbf{y}) \end{aligned}$$

Thus T is linear.

43. The proof of Theorem 6 shows that the inner products to be checked are actually entries in the matrix product A^TA. A calculation shows that $A^TA = 100I_4$. Since the off-diagonal entries in A^TA are zero, the columns of A are orthogonal.

Section 6.3, page 374

1. $\mathbf{x} = -\frac{8}{9}\mathbf{u}_1 - \frac{2}{9}\mathbf{u}_2 + \frac{2}{3}\mathbf{u}_3 + 2\mathbf{u}_4$; $\mathbf{x} = \begin{bmatrix} 0 \\ -2 \\ 4 \\ -2 \end{bmatrix} + \begin{bmatrix} 10 \\ -6 \\ -2 \\ 2 \end{bmatrix}$

3. $\begin{bmatrix} -1 \\ 4 \\ 0 \end{bmatrix}$ **5.** $\begin{bmatrix} -1 \\ 2 \\ 6 \end{bmatrix} = \mathbf{y}$

7. $\mathbf{y} = \begin{bmatrix} 10/3 \\ 2/3 \\ 8/3 \end{bmatrix} + \begin{bmatrix} -7/3 \\ 7/3 \\ 7/3 \end{bmatrix}$ **9.** $\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ 3 \\ -1 \end{bmatrix}$

11. $\begin{bmatrix} 3 \\ -1 \\ 1 \\ -1 \end{bmatrix}$ **13.** $\begin{bmatrix} -1 \\ -3 \\ -2 \\ 3 \end{bmatrix}$ **15.** $\sqrt{40}$

17. a. $U^TU = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

$$UU^T = \begin{bmatrix} 8/9 & -2/9 & 2/9 \\ -2/9 & 5/9 & 4/9 \\ 2/9 & 4/9 & 5/9 \end{bmatrix}$$

b. $\text{proj}_W\,\mathbf{y} = 6\mathbf{u}_1 + 3\mathbf{u}_2 = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$, and $(UU^T)\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$

19. Any multiple of $\begin{bmatrix} 0 \\ 2/5 \\ 1/5 \end{bmatrix}$, such as $\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$

21–29. Write your answers before checking the *Study Guide*.

31. *Hint:* Use Theorem 3 and the Orthogonal Decomposition Theorem. For the uniqueness, suppose $A\mathbf{p} = \mathbf{b}$ and $A\mathbf{p}_1 = \mathbf{b}$, and consider the equations $\mathbf{p} = \mathbf{p}_1 + (\mathbf{p} - \mathbf{p}_1)$ and $\mathbf{p} = \mathbf{p} + \mathbf{0}$.

33. $\mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ $M = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$

35. $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$;

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

37. U has orthonormal columns, by Theorem 6 in Section 6.2, because $U^TU = I_4$. The closest point to $\mathbf{y}$ in Col U is the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto Col U. From Theorem 10, $\hat{\mathbf{y}} = UU^T\mathbf{y} = (1.2, .4, 1.2, 1.2, .4, 1.2, .4, .4)$

Section 6.4, page 380

1. $\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$ **3.** $\begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3/2 \\ 3/2 \end{bmatrix}$

5. $\begin{bmatrix} 1 \\ -4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ -4 \\ -1 \end{bmatrix}$ **7.** $\begin{bmatrix} 2/\sqrt{30} \\ -5/\sqrt{30} \\ 1/\sqrt{30} \end{bmatrix}, \begin{bmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$

9. $\begin{bmatrix} 3 \\ 1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 1 \\ 3 \end{bmatrix}$ **11.** $\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 3 \\ -3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \\ -2 \end{bmatrix}$

13. $R = \begin{bmatrix} 6 & 12 \\ 0 & 6 \end{bmatrix}$

15. $Q = \begin{bmatrix} 1/\sqrt{5} & 1/2 & 1/2 \\ -1/\sqrt{5} & 0 & 0 \\ -1/\sqrt{5} & 1/2 & 1/2 \\ 1/\sqrt{5} & -1/2 & 1/2 \\ 1/\sqrt{5} & 1/2 & -1/2 \end{bmatrix}$,

$R = \begin{bmatrix} \sqrt{5} & -\sqrt{5} & 4\sqrt{5} \\ 0 & 6 & -2 \\ 0 & 0 & 4 \end{bmatrix}$

17–21. See the *Study Guide*.

23. Suppose $\mathbf{x}$ satisfies $R\mathbf{x} = \mathbf{0}$; then $QR\mathbf{x} = Q\,\mathbf{0} = \mathbf{0}$, and $A\mathbf{x} = \mathbf{0}$. Since the columns of A are linearly independent, $\mathbf{x}$ must be zero. This fact, in turn, shows that the columns of R are linearly independent. Since R is square, it is invertible, by the Invertible Matrix Theorem.

25. Denote the columns of Q by $\mathbf{q}_1, \dots, \mathbf{q}_n$. Note that $n \le m$, because A is $m \times n$ and has linearly independent columns. Use the fact that the columns of Q can be extended to an orthonormal basis for $\mathbb{R}^m$, say, $\{\mathbf{q}_1, \dots, \mathbf{q}_m\}$. (The *Study Guide* describes one method.) Let $Q_0 = \begin{bmatrix} \mathbf{q}_{n+1} & \cdots & \mathbf{q}_m \end{bmatrix}$ and $Q_1 = \begin{bmatrix} Q & Q_0 \end{bmatrix}$. Then, using partitioned matrix multiplication, $Q_1 \begin{bmatrix} R \\ 0 \end{bmatrix} = QR = A$.

27. *Hint:* Partition R as a 2×2 block matrix.

29. The diagonal entries of R are 20, 6, 10.3923, and 7.0711, to four decimal places.

Section 6.5, page 388

1. a. $\begin{bmatrix} 6 & -11 \\ -11 & 22 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -4 \\ 11 \end{bmatrix}$ **b.** $\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

3. a. $\begin{bmatrix} 6 & 6 \\ 6 & 42 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \end{bmatrix}$ **b.** $\hat{\mathbf{x}} = \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix}$

5. $\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ **7.** $2\sqrt{5}$

9. a. $\hat{\mathbf{b}} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ **b.** $\hat{\mathbf{x}} = \begin{bmatrix} 2/7 \\ 1/7 \end{bmatrix}$

11. a. $\hat{\mathbf{b}} = \begin{bmatrix} 3 \\ 1 \\ 4 \\ -1 \end{bmatrix}$ **b.** $\hat{\mathbf{x}} = \begin{bmatrix} 2/3 \\ 0 \\ 1/3 \end{bmatrix}$

13. $A\mathbf{u} = \begin{bmatrix} 11 \\ -11 \\ 11 \end{bmatrix}$, $A\mathbf{v} = \begin{bmatrix} 7 \\ -12 \\ 7 \end{bmatrix}$,

$\mathbf{b} - A\mathbf{u} = \begin{bmatrix} 0 \\ 2 \\ -6 \end{bmatrix}$, $\mathbf{b} - A\mathbf{v} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix}$. No, $\mathbf{u}$ could not possibly be a least-squares solution of $A\mathbf{x} = \mathbf{b}$. Why?

15. $\hat{\mathbf{x}} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$

17–25. See the *Study Guide*.

27. a. If $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$. This shows that Nul A is contained in Nul A^TA.

b. If $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x}^TA^TA\mathbf{x} = \mathbf{x}^T\mathbf{0} = 0$. So $(A\mathbf{x})^T(A\mathbf{x}) = 0$ (which means that $\|A\mathbf{x}\|^2 = 0$), and hence $A\mathbf{x} = \mathbf{0}$. This shows that Nul A^TA is contained in Nul A.

29. *Hint:* For part (a), use an important theorem from Chapter 2.

31. By Theorem 14, $\hat{\mathbf{b}} = A\hat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$. The matrix $A(A^TA)^{-1}A^T$ occurs frequently in statistics, where it is sometimes called the *hat-matrix*.

33. The normal equations are $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$, whose solution is the set of (x, y) such that $x + y = 3$. The solutions correspond to points on the line midway between the lines $x + y = 2$ and $x + y = 4$.

Section 6.6, page 397

1. $y = .9 + .4x$ **3.** $y = 1.1 + 1.3x$

5. 2.5

7. 2.1, a difference of .1 is reasonable.

9. No. A y-value of 20 is quite far from the other y-values.

11. If two data points have different x-coordinates, then the two columns of the design matrix X cannot be multiples of each other and hence are linearly independent. By Theorem 14 in Section 6.5, the normal equations have a unique solution.

13. a. $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y} = \begin{bmatrix} 1.8 \\ 2.7 \\ 3.4 \\ 3.8 \\ 3.9 \end{bmatrix}$, $X = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \\ 5 & 25 \end{bmatrix}$,

$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$

b. $y = 1.76x - .20x^2$

c. $y = 3.36$

15. $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y} = \begin{bmatrix} 7.9 \\ 5.4 \\ -.9 \end{bmatrix}$, $X = \begin{bmatrix} \cos 1 & \sin 1 \\ \cos 2 & \sin 2 \\ \cos 3 & \sin 3 \end{bmatrix}$,

$\boldsymbol{\beta} = \begin{bmatrix} A \\ B \end{bmatrix}$, $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$

17. $\beta = 1.45$ and $e = .811$; the orbit is an ellipse. The equation $r = \beta/(1 - e \cdot \cos\vartheta)$ produces $r = 1.33$ when $\vartheta = 4.6$.

19. a. $y = -.8558 + 4.7025t + 5.5554t^2 - .0274t^3$

b. The velocity function is $v(t) = 4.7025 + 11.1108t - .0822t^2$, and $v(4.5) = 53.0$ ft/sec.

21. *Hint:* Write X and $\mathbf{y}$ as in equation (1), and compute X^TX and $X^T\mathbf{y}$.

23. a. The mean of the x-data is $\bar{x} = 5.5$. The data in mean-deviation form are $(-3.5, 1)$, $(-.5, 2)$, $(1.5, 3)$, and $(2.5, 3)$. The columns of X are orthogonal because the entries in the second column sum to 0.

b. $\begin{bmatrix} 4 & 0 \\ 0 & 21 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 7.5 \end{bmatrix}$,

$y = \frac{9}{4} + \frac{5}{14}x^* = \frac{9}{4} + \frac{5}{14}(x - 5.5)$

25. *Hint:* The equation has a nice geometric interpretation.

Section 6.7, page 406

1. a. $3, \sqrt{105}, 225$ **b.** All multiples of $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$

3. 28 **5.** $5\sqrt{2}, 3\sqrt{3}$ **7.** $\frac{56}{25} + \frac{14}{25}t$

9. a. Constant polynomial, $p(t) = 5$

b. $t^2 - 5$ is orthogonal to p_0 and p_1; values: $(4, -4, -4, 4)$; answer: $q(t) = \frac{1}{4}(t^2 - 5)$

11. $\frac{17}{5}t$

13. Verify each of the four axioms. For instance:

$$\begin{aligned} 1.\quad \langle \mathbf{u}, \mathbf{v} \rangle &= (A\mathbf{u})\cdot(A\mathbf{v}) && \text{Definition} \\ &= (A\mathbf{v})\cdot(A\mathbf{u}) && \text{Property of the dot product} \\ &= \langle \mathbf{v}, \mathbf{u} \rangle && \text{Definition} \end{aligned}$$

15.
$$\begin{aligned} \langle \mathbf{u}, c\mathbf{v} \rangle &= \langle c\mathbf{v}, \mathbf{u} \rangle && \text{Axiom 1} \\ &= c\langle \mathbf{v}, \mathbf{u} \rangle && \text{Axiom 3} \\ &= c\langle \mathbf{u}, \mathbf{v} \rangle && \text{Axiom 1} \end{aligned}$$

17. *Hint:* Compute 4 times the right-hand side.

19–23. See the Study Guide.

25. $\langle \mathbf{u}, \mathbf{v} \rangle = \sqrt{a}\sqrt{b} + \sqrt{b}\sqrt{a} = 2\sqrt{ab}$, $\|\mathbf{u}\|^2 = (\sqrt{a})^2 + (\sqrt{b})^2 = a + b$. Since a and b are nonnegative, $\|\mathbf{u}\| = \sqrt{a+b}$. Similarly, $\|\mathbf{v}\| = \sqrt{b+a}$. By Cauchy–Schwarz, $2\sqrt{ab} \le \sqrt{a+b}\sqrt{b+a} = a + b$. Hence, $\sqrt{ab} \le \dfrac{a+b}{2}$.

27. 0 **29.** $2/\sqrt{5}$ **31.** $1, t, 3t^2 - 1$

33. The new orthogonal polynomials are multiples of $-17t + 5t^3$ and $72 - 155t^2 + 35t^4$. Scale these polynomials so their values at $-2, -1, 0, 1,$ and 2 are small integers.

Section 6.8, page 412

1. $y = 2 + \frac{3}{2}t$

3.
$$\begin{aligned} p(t) &= 4p_0 - .1p_1 - .5p_2 + .2p_3 \\ &= 4 - .1t - .5(t^2 - 2) + .2\left(\tfrac{5}{6}t^3 - \tfrac{17}{6}t\right) \end{aligned}$$
(This polynomial happens to fit the data exactly.)

5. Use the identity

$\sin mt \sin nt = \frac{1}{2}[\cos(mt - nt) - \cos(mt + nt)]$

7. Use the identity $\cos^2 kt = \dfrac{1 + \cos 2kt}{2}$.

9. $\pi + 2\sin t + \sin 2t + \frac{2}{3}\sin 3t$ [*Hint:* Save time by using the results from Example 4.]

11. $\frac{1}{2} - \frac{1}{2}\cos 2t$ (Why?)

13. *Hint:* Take functions f and g in $C[0, 2\pi]$, and fix an integer $m \ge 0$. Write the Fourier coefficient of $f + g$ that involves $\cos mt$, and write the Fourier coefficient that involves $\sin mt\,(m > 0)$.

15. The cubic curve is the graph of $g(t) = -.2685 + 3.6095t + 5.8576t^2 - .0477t^3$. The velocity at $t = 4.5$ seconds is $g'(4.5) = 53.4$ ft/sec. This is about .7% faster than the estimate obtained in Exercise 19 in Section 6.6.

Chapter 6 Supplementary Exercises, page 414

1. F **2.** T **3.** T **4.** F **5.** F **6.** T

7. T **8.** T **9.** F **10.** T **11.** T **12.** F

13. T **14.** F **15.** F **16.** T **17.** T **18.** F

19. F

20. *Hint:* If $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthonormal set and $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, then the vectors $c_1\mathbf{v}_1$ and $c_2\mathbf{v}_2$ are orthogonal, and

$$\begin{aligned} \|\mathbf{x}\|^2 &= \|c_1\mathbf{v}_1 + c_2\mathbf{v}_2\|^2 = \|c_1\mathbf{v}_1\|^2 + \|c_2\mathbf{v}_2\|^2 \\ &= (|c_1|\|\mathbf{v}_1\|)^2 + (|c_2|\|\mathbf{v}_2\|)^2 = |c_1|^2 + |c_2|^2 \end{aligned}$$

(Explain why.) So the stated equality holds for $p = 2$. Suppose that the equality holds for $p = k$, with $k \ge 2$, let $\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$ be an orthonormal set, and consider $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k + c_{k+1}\mathbf{v}_{k+1} = \mathbf{u}_k + c_{k+1}\mathbf{v}_{k+1}$, where $\mathbf{u}_k = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$.

21. Given $\mathbf{x}$ and an orthonormal set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^n$, let $\hat{\mathbf{x}}$ be the orthogonal projection of $\mathbf{x}$ onto the subspace spanned by $\mathbf{v}_1, \dots, \mathbf{v}_p$. By Theorem 10 in Section 6.3,

$$\hat{\mathbf{x}} = (\mathbf{x}\cdot\mathbf{v}_1)\mathbf{v}_1 + \cdots + (\mathbf{x}\cdot\mathbf{v}_p)\mathbf{v}_p$$

By Exercise 20, $\|\hat{\mathbf{x}}\|^2 = |\mathbf{x}\cdot\mathbf{v}_1|^2 + \cdots + |\mathbf{x}\cdot\mathbf{v}_p|^2$. Bessel's inequality follows from the fact that $\|\hat{\mathbf{x}}\|^2 \le \|\mathbf{x}\|^2$, noted before the statement of the Cauchy–Schwarz inequality, in Section 6.7.

23. Suppose $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ for all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$, and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the standard basis for $\mathbb{R}^n$. For $j = 1, \dots, n$, $U\mathbf{e}_j$ is the jth column of U. Since $\|U\mathbf{e}_j\|^2 = (U\mathbf{e}_j)\cdot(U\mathbf{e}_j) = \mathbf{e}_j\cdot\mathbf{e}_j = 1$, the columns of U are unit vectors; since $(U\mathbf{e}_j)\cdot(U\mathbf{e}_k) = \mathbf{e}_j\cdot\mathbf{e}_k = 0$ for $j \ne k$, the columns are pairwise orthogonal.

25. *Hint:* Compute Q^TQ, using the fact that $(\mathbf{u}\mathbf{u}^T)^T = \mathbf{u}^{TT}\mathbf{u}^T = \mathbf{u}\mathbf{u}^T$.

27. Let $W = \text{Span}\{\mathbf{u}, \mathbf{v}\}$. Given $\mathbf{z}$ in $\mathbb{R}^n$, let $\hat{\mathbf{z}} = \text{proj}_W \mathbf{z}$. Then $\hat{\mathbf{z}}$ is in Col A, where $A = \begin{bmatrix} \mathbf{u} & \mathbf{v} \end{bmatrix}$, say, $\hat{\mathbf{z}} = A\hat{\mathbf{x}}$ for some $\hat{\mathbf{x}}$ in $\mathbb{R}^2$. So $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{z}$. The normal equations can be solved to produce $\hat{\mathbf{x}}$, and then $\hat{\mathbf{z}}$ is found by computing $A\hat{\mathbf{x}}$.

29. *Hint:* Let $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix}$, and $A = \begin{bmatrix} \mathbf{v}^T \\ \mathbf{v}^T \\ \mathbf{v}^T \end{bmatrix} = \begin{bmatrix} 1 & -2 & 5 \\ 1 & -2 & 5 \\ 1 & -2 & 5 \end{bmatrix}$. The given set of equations is $A\mathbf{x} = \mathbf{b}$, and the set of all least-squares solutions coincides with the set of solutions of $A^TA\mathbf{x} = A^T\mathbf{b}$ (Theorem 13 in Section 6.5). Study this equation, and use the fact that $(\mathbf{v}\mathbf{v}^T)\mathbf{x} = \mathbf{v}(\mathbf{v}^T\mathbf{x}) = (\mathbf{v}^T\mathbf{x})\mathbf{v}$, because $\mathbf{v}^T\mathbf{x}$ is a scalar.

31. a. The row–column calculation of $A\mathbf{u}$ shows that each row of A is orthogonal to every $\mathbf{u}$ in Nul A. So each row of A is in $(\text{Nul } A)^\perp$. Since $(\text{Nul } A)^\perp$ is a subspace, it must contain all linear combinations of the rows of A; hence $(\text{Nul } A)^\perp$ contains Row A.

b. If rank $A = r$, then $\dim \text{Nul } A = n - r$, by the Rank Theorem. By Exercise 32(c) in Section 6.3,

$$\dim \text{Nul } A + \dim(\text{Nul } A)^\perp = n$$

So $\dim(\text{Nul } A)^\perp$ must be r. But Row A is an r-dimensional subspace of $(\text{Nul } A)^\perp$, by the Rank Theorem and part (a). Therefore, Row A must coincide with $(\text{Nul } A)^\perp$.

c. Replace A by A^T in part (b) and conclude that Row A^T coincides with $(\text{Nul } A^T)^\perp$. Since Row A^T = Col A, this proves (c).

33. If $A = URU^T$ with U orthogonal, then A is similar to R (because U is invertible and $U^T = U^{-1}$) and so A has the same eigenvalues as R (by Theorem 4 in Section 5.2), namely, the n real numbers on the diagonal of R.

35. $\dfrac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = .4618$,

$\text{cond}(A) \times \dfrac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = 3363 \times (1.548 \times 10^{-4}) = .5206$.

Observe that $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$ almost equals cond(A) times $\|\Delta\mathbf{b}\|/\|\mathbf{b}\|$.

37. $\dfrac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = 7.178 \times 10^{-8}$, $\dfrac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = 2.832 \times 10^{-4}$. Observe that the relative change in $\mathbf{x}$ is *much* smaller than the relative change in $\mathbf{b}$. In fact, since

$$\text{cond}(A) \times \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = 23{,}683 \times (2.832 \times 10^{-4}) = 6.707$$

the theoretical bound on the relative change in $\mathbf{x}$ is 6.707 (to four significant figures). This exercise shows that even when a condition number is large, the relative error in a solution need not be as large as you might expect.

Chapter 7

Section 7.1, page 423

1. Symmetric **3.** Not symmetric **5.** Symmetric

7. Orthogonal, $\begin{bmatrix} .6 & .8 \\ .8 & -.6 \end{bmatrix}$

9. Orthogonal, $\begin{bmatrix} -4/5 & 3/5 \\ 3/5 & 4/5 \end{bmatrix}$

11. Not orthogonal

13. $P = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$, $D = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$

15. $P = \begin{bmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \\ 0 & 11 \end{bmatrix}$

17. $P = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}$,

$D = \begin{bmatrix} -4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix}$

19. $P = \begin{bmatrix} -1/\sqrt{5} & 4/\sqrt{45} & -2/3 \\ 2/\sqrt{5} & 2/\sqrt{45} & -1/3 \\ 0 & 5/\sqrt{45} & 2/3 \end{bmatrix}$,

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

21. $P = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/2 & 1/2 \\ 0 & -1/\sqrt{2} & 1/2 & 1/2 \\ 1/\sqrt{2} & 0 & -1/2 & 1/2 \\ -1/\sqrt{2} & 0 & -1/2 & 1/2 \end{bmatrix}$,

$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$

23. $P = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}$,

$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

25–31. See the *Study Guide*.

33. $(A\mathbf{x})\cdot\mathbf{y} = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^TA^T\mathbf{y} = \mathbf{x}^TA\mathbf{y} = \mathbf{x}\cdot(A\mathbf{y})$, because $A^T = A$.

35. *Hint:* Use an orthogonal diagonalization of A, or appeal to Theorem 2.

37. The Diagonalization Theorem in Section 5.3 says that the columns of P are (linearly independent) eigenvectors corresponding to the eigenvalues of A listed on the diagonal of D. So P has exactly k columns of eigenvectors corresponding to λ. These k columns form a basis for the eigenspace.

39. $A = 8\mathbf{u}_1\mathbf{u}_1^T + 6\mathbf{u}_2\mathbf{u}_2^T + 3\mathbf{u}_3\mathbf{u}_3^T$

$$= 8\begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + 6\begin{bmatrix} 1/6 & 1/6 & -2/6 \\ 1/6 & 1/6 & -2/6 \\ -2/6 & -2/6 & 4/6 \end{bmatrix} + 3\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

41. *Hint:* $(\mathbf{u}\mathbf{u}^T)\mathbf{x} = \mathbf{u}(\mathbf{u}^T\mathbf{x}) = (\mathbf{u}^T\mathbf{x})\mathbf{u}$, because $\mathbf{u}^T\mathbf{x}$ is a scalar.

43. $P = \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix}$,

$$D = \begin{bmatrix} 19 & 0 & 0 & 0 \\ 0 & 11 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & -11 \end{bmatrix}$$

45. $P = \begin{bmatrix} 1/\sqrt{2} & 3/\sqrt{50} & -2/5 & -2/5 \\ 0 & 4/\sqrt{50} & -1/5 & 4/5 \\ 0 & 4/\sqrt{50} & 4/5 & -1/5 \\ 1/\sqrt{2} & -3/\sqrt{50} & 2/5 & 2/5 \end{bmatrix}$

$$D = \begin{bmatrix} .75 & 0 & 0 & 0 \\ 0 & .75 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1.25 \end{bmatrix}$$

Section 7.2, page 430

1. a. $5x_1^2 + \frac{2}{3}x_1x_2 + x_2^2$ **b.** 185 **c.** 16

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

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

7. $\mathbf{x} = P\mathbf{y}$, where $P = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\mathbf{y}^T D\mathbf{y} = 6y_1^2 - 4y_2^2$

In Exercises 9–14, other answers (change of variables and new quadratic form) are possible.

9. Positive definite; eigenvalues are 6 and 2

Change of variable: $\mathbf{x} = P\mathbf{y}$, with $P = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$

New quadratic form: $6y_1^2 + 2y_2^2$

11. Indefinite; eigenvalues are 3 and -2

Change of variable: $\mathbf{x} = P\mathbf{y}$, with $P = \frac{1}{\sqrt{5}}\begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}$

New quadratic form: $3y_1^2 - 2y_2^2$

13. Positive semidefinite; eigenvalues are 10 and 0

Change of variable: $\mathbf{x} = P\mathbf{y}$, with $P = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}$

New quadratic form: $10y_1^2$

15. Negative definite; eigenvalues are $-13, -9, -7, -1$
Change of variable: $\mathbf{x} = P\mathbf{y}$;

$$P = \begin{bmatrix} 0 & -1/2 & 0 & 3/\sqrt{12} \\ 0 & 1/2 & -2/\sqrt{6} & 1/\sqrt{12} \\ -1/\sqrt{2} & 1/2 & 1/\sqrt{6} & 1/\sqrt{12} \\ 1/\sqrt{2} & 1/2 & 1/\sqrt{6} & 1/\sqrt{12} \end{bmatrix}$$

New quadratic form: $-13y_1^2 - 9y_2^2 - 7y_3^2 - y_4^2$

17. Positive definite; eigenvalues are 1 and 21.
Change of variable: $\mathbf{x} = P\mathbf{y}$;

$$P = \frac{1}{\sqrt{50}}\begin{bmatrix} 4 & 3 & 4 & -3 \\ -5 & 0 & 5 & 0 \\ 3 & -4 & 3 & 4 \\ 0 & 5 & 0 & 5 \end{bmatrix}$$

New quadratic form: $y_1^2 + y_2^2 + 21y_3^2 + 21y_4^2$

19. 8

21–29. See the *Study Guide.*

31. Write the characteristic polynomial in two ways:

$$\det(A - \lambda I) = \det\begin{bmatrix} a-\lambda & b \\ b & d-\lambda \end{bmatrix} = \lambda^2 - (a+d)\lambda + ad - b^2$$

and

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2$$

Equate coefficients to obtain $\lambda_1 + \lambda_2 = a + d$ and $\lambda_1\lambda_2 = ad - b^2 = \det A$.

33. Exercise 34 in Section 7.1 showed that B^TB is symmetric. Also, $\mathbf{x}^TB^TB\mathbf{x} = (B\mathbf{x})^TB\mathbf{x} = \|B\mathbf{x}\|^2 \geq 0$, so the quadratic form is positive semidefinite, and we say that the matrix B^TB is positive semidefinite. *Hint:* To show that B^TB is positive definite when B is square and invertible, suppose that $\mathbf{x}^TB^TB\mathbf{x} = 0$ and deduce that $\mathbf{x} = \mathbf{0}$.

35. *Hint:* Show that $A + B$ is symmetric and the quadratic form $\mathbf{x}^T(A + B)\mathbf{x}$ is positive definite.

Section 7.3, page 438

1. $\mathbf{x} = P\mathbf{y}$, where $P = \begin{bmatrix} 1/3 & 2/3 & -2/3 \\ 2/3 & 1/3 & 2/3 \\ -2/3 & 2/3 & 1/3 \end{bmatrix}$

3. a. 9 **b.** $\pm\begin{bmatrix} 1/3 \\ 2/3 \\ -2/3 \end{bmatrix}$ **c.** 6

5. a. 6 **b.** $\pm\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ **c.** -4

7. $\pm\begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}$ **9.** $5+\sqrt{5}$ **11.** 3

13. *Hint:* If $m = M$, take $\alpha = 0$ in the formula for $\mathbf{x}$. That is, let $\mathbf{x} = \mathbf{u}_n$, and verify that $\mathbf{x}^T A\mathbf{x} = m$. If $m < M$ and if t is a number between m and M, then $0 \le t - m \le M - m$ and $0 \le (t-m)/(M-m) \le 1$. So let $\alpha = (t-m)/(M-m)$. Solve the expression for α to see that $t = (1-\alpha)m + \alpha M$. As α goes from 0 to 1, t goes from m to M. Construct $\mathbf{x}$ as in the statement of the exercise, and verify its properties.

15. a. 9 **b.** $\begin{bmatrix} -2/\sqrt{6} \\ 0 \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$ **c.** 3

17. a. 17 **b.** $\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$ **c.** 13

Section 7.4, page 448

1. 3, 1 **3.** 4, 1

The answers in Exercises 5–13 are not the only possibilities.

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

7. $\begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$
$\times\begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$

9. $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}$
$\times\begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

11. $\begin{bmatrix} -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \end{bmatrix}\begin{bmatrix} \sqrt{90} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$
$\times\begin{bmatrix} 3/\sqrt{10} & -1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{bmatrix}$

13. $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$
$\times\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{18} & 1/\sqrt{18} & -4/\sqrt{18} \\ -2/3 & 2/3 & 1/3 \end{bmatrix}$

15. a. rank $A = 2$

b. Basis for Col A: $\begin{bmatrix} .40 \\ .37 \\ -.84 \end{bmatrix}, \begin{bmatrix} -.78 \\ -.33 \\ -.52 \end{bmatrix}$

Basis for Nul A: $\begin{bmatrix} .58 \\ -.58 \\ .58 \end{bmatrix}$

(Remember that V^T appears in the SVD.)

17. If U is an orthogonal matrix then $\det U = \pm 1$. If $A = U\Sigma V^T$ and A is square, then so are U, Σ, and V. Hence $\det A = \det U \det \Sigma \det V^T$
$= \pm 1 \det \Sigma = \pm\sigma_1 \cdots \sigma_n$

19. *Hint:* Since U and V are orthogonal,

$$A^T A = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T = V(\Sigma^T\Sigma)V^{-1}$$

Thus V diagonalizes $A^T A$. What does this tell you about V?

21. The right singular vector $\mathbf{v}_1$ is an eigenvector for the largest eigenvalue λ_1 of $A^T A$. By Theorem 7 in Section 7.3, the largest eigenvalue, λ_2, is the maximum of $\mathbf{x}^T(A^T A)\mathbf{x}$ over all unit vectors orthogonal to $\mathbf{v}_1$. Since $\mathbf{x}^T(A^T A)\mathbf{x} = ||A\mathbf{x}||^2$, the square root of λ_2, which is the second largest eigenvalue, is the maximum of $||A\mathbf{x}||$ over all unit vectors orthogonal to $\mathbf{v}_1$.

23. *Hint:* Use a column–row expansion of $(U\Sigma)V^T$.

25. *Hint:* Consider the SVD for the standard matrix of T—say, $A = U\Sigma V^T = U\Sigma V^{-1}$. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be bases constructed from the columns of V and U, respectively. Compute the matrix for T relative to $\mathcal{B}$ and $\mathcal{C}$, as in Section 5.4. To do this, you must show that $V^{-1}\mathbf{v}_j = \mathbf{e}_j$, the jth column of I_n.

27. $\begin{bmatrix} -.57 & -.65 & -.42 & .27 \\ .63 & -.24 & -.68 & -.29 \\ .07 & -.63 & .53 & -.56 \\ -.51 & .34 & -.29 & -.73 \end{bmatrix}$
$\times\begin{bmatrix} 16.46 & 0 & 0 & 0 & 0 \\ 0 & 12.16 & 0 & 0 & 0 \\ 0 & 0 & 4.87 & 0 & 0 \\ 0 & 0 & 0 & 4.31 & 0 \end{bmatrix}$
$\times\begin{bmatrix} -.10 & .61 & -.21 & -.52 & .55 \\ -.39 & .29 & .84 & -.14 & -.19 \\ -.74 & -.27 & -.07 & .38 & .49 \\ .41 & -.50 & .45 & -.23 & .58 \\ -.36 & -.48 & -.19 & -.72 & -.29 \end{bmatrix}$

29. 25.9343, 16.7554, 11.2917, 1.0785, .00037793; $\sigma_1/\sigma_5 = 68{,}622$

Section 7.5, page 455

1. $M = \begin{bmatrix} 12 \\ 10 \end{bmatrix}$; $B = \begin{bmatrix} 7 & 10 & -6 & -9 & -10 & 8 \\ 2 & -4 & -1 & 5 & 3 & -5 \end{bmatrix}$;
$S = \begin{bmatrix} 86 & -27 \\ -27 & 16 \end{bmatrix}$

3. $\begin{bmatrix} .95 \\ -.32 \end{bmatrix}$ for $\lambda = 95.2$, $\begin{bmatrix} .32 \\ .95 \end{bmatrix}$ for $\lambda = 6.8$

5. (.130, .874, .468), 75.9% of the variance

7. $y_1 = .95x_1 - .32x_2$; y_1 explains 93.3% of the variance.

9. $c_1 = 1/3$, $c_2 = 2/3$, $c_3 = 2/3$; the variance of y is 9.

11. a. If $\mathbf{w}$ is the vector in $\mathbb{R}^N$ with a 1 in each position, then

$$\begin{bmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_N \end{bmatrix}\mathbf{w} = \mathbf{X}_1 + \cdots + \mathbf{X}_N = \mathbf{0}$$

because the $\mathbf{X}_k$ are in mean-deviation form. Then

$$\begin{aligned} &\begin{bmatrix} \mathbf{Y}_1 & \cdots & \mathbf{Y}_N \end{bmatrix}\mathbf{w} \\ &= \begin{bmatrix} P^T\mathbf{X}_1 & \cdots & P^T\mathbf{X}_N \end{bmatrix}\mathbf{w} \qquad \text{By definition} \\ &= P^T\begin{bmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_N \end{bmatrix}\mathbf{w} = P^T\mathbf{0} = \mathbf{0} \end{aligned}$$

That is, $\mathbf{Y}_1 + \cdots + \mathbf{Y}_N = \mathbf{0}$, so the $\mathbf{Y}_k$ are in mean-deviation form.

b. *Hint:* Because the $\mathbf{X}_j$ are in mean-deviation form, the covariance matrix of the $\mathbf{X}_j$ is

$$1/(N-1)\begin{bmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_N \end{bmatrix}\begin{bmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_N \end{bmatrix}^T$$

Compute the covariance matrix of the $\mathbf{Y}_j$, using part (a).

13. If $B = \begin{bmatrix} \hat{\mathbf{X}}_1 & \cdots & \hat{\mathbf{X}}_N \end{bmatrix}$, then

$$S = \frac{1}{N-1}BB^T = \frac{1}{N-1}\begin{bmatrix} \hat{\mathbf{X}}_1 & \cdots & \hat{\mathbf{X}}_n \end{bmatrix}\begin{bmatrix} \hat{\mathbf{X}}_1^T \\ \vdots \\ \hat{\mathbf{X}}_N^T \end{bmatrix}$$

$$= \frac{1}{N-1}\sum_1^N \hat{\mathbf{X}}_k\hat{\mathbf{X}}_k^T = \frac{1}{N-1}\sum_1^N (\mathbf{X}_k - \mathbf{M})(\mathbf{X}_k - \mathbf{M})^T$$

Chapter 7 Supplementary Exercises, page 457

1. T **2.** F **3.** T **4.** F **5.** F **6.** F

7. F **8.** T **9.** F **10.** F **11.** F **12.** F

13. T **14.** F **15.** T **16.** T **17.** F

19. If rank $A = r$, then dim Nul $A = n - r$, by the Rank Theorem. So 0 is an eigenvalue of multiplicity $n - r$. Hence, of the n terms in the spectral decomposition of A, exactly $n - r$ are zero. The remaining r terms (corresponding to the nonzero eigenvalues) are all rank 1 matrices, as mentioned in the discussion of the spectral decomposition.

21. If $A\mathbf{v} = \lambda\mathbf{v}$ for some nonzero λ, then $\mathbf{v} = \lambda^{-1}A\mathbf{v} = A(\lambda^{-1}\mathbf{v})$, which shows that $\mathbf{v}$ is a linear combination of the columns of A.

23. *Hint:* If $A = R^TR$, where R is invertible, then A is positive definite, by Exercise 33 in Section 7.2. Conversely, suppose that A is positive definite. Then by Exercise 34 in Section 7.2, $A = B^TB$ for some positive definite matrix B. Explain why B admits a QR factorization, and use it to create the Cholesky factorization of A.

25. If A is $m \times n$ and $\mathbf{x}$ is in $\mathbb{R}^n$, then $\mathbf{x}^TA^TA\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = \|A\mathbf{x}\|^2 \geq 0$. Thus A^TA is positive semidefinite. By Exercise 30 in Section 6.5, rank A^TA = rank A.

27. *Hint:* Write an SVD of A in the form $A = U\Sigma V^T = PQ$, where $P = U\Sigma U^T$ and $Q = UV^T$. Show that P is symmetric and has the same eigenvalues as Σ. Explain why Q is an orthogonal matrix.

29. a. If $\mathbf{b} = A\mathbf{x}$, then $\mathbf{x}^+ = A^+\mathbf{b} = A^+A\mathbf{x}$. By Exercise 28(b), $\mathbf{x}^+$ is the orthogonal projection of $\mathbf{x}$ onto Row A.

b. From (a) and then Exercise 28(c), $A\mathbf{x}^+ = A(A^+A\mathbf{x}) = (AA^+A)\mathbf{x} = A\mathbf{x} = \mathbf{b}$.

c. Since $\mathbf{x}^+$ is the orthogonal projection onto Row A, the Pythagorean Theorem shows that $\|\mathbf{u}\|^2 = \|\mathbf{x}^+\|^2 + \|\mathbf{u} - \mathbf{x}^+\|^2$. Part (c) follows immediately.

31. $$A^+ = \frac{1}{40}\cdot\begin{bmatrix} -2 & -14 & 13 & 13 \\ -2 & -14 & 13 & 13 \\ -2 & 6 & -7 & -7 \\ 2 & -6 & 7 & 7 \\ 4 & -12 & -6 & -6 \end{bmatrix}, \quad \hat{\mathbf{x}} = \begin{bmatrix} .7 \\ .7 \\ -.8 \\ .8 \\ .6 \end{bmatrix}$$

The reduced echelon form of $\begin{bmatrix} A \\ \mathbf{x}^T \end{bmatrix}$ is the same as the reduced echelon form of A, except for an extra row of zeros. So adding scalar multiples of the rows of A to $\mathbf{x}^T$ can produce the zero vector, which shows that $\mathbf{x}^T$ is in Row A.

Basis for Nul A: $\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$

Chapter 8

Section 8.1, page 467

1. Some possible answers: $\mathbf{y} = 2\mathbf{v}_1 - 1.5\mathbf{v}_2 + .5\mathbf{v}_3$, $\mathbf{y} = 2\mathbf{v}_1 - 2\mathbf{v}_3 + \mathbf{v}_4$, $\mathbf{y} = 2\mathbf{v}_1 + 3\mathbf{v}_2 - 7\mathbf{v}_3 + 3\mathbf{v}_4$

3. $\mathbf{y} = -3\mathbf{v}_1 + 2\mathbf{v}_2 + 2\mathbf{v}_3$. The weights sum to 1, so this is an affine sum.

5. a. $\mathbf{p}_1 = 3\mathbf{b}_1 - \mathbf{b}_2 - \mathbf{b}_3 \in \text{aff } S$ since the coefficients sum to 1.

b. $\mathbf{p}_2 = 2\mathbf{b}_1 + 0\mathbf{b}_2 + \mathbf{b}_3 \notin \text{aff } S$ since the coefficients do not sum to 1.

c. $\mathbf{p}_3 = -\mathbf{b}_1 + 2\mathbf{b}_2 + 0\mathbf{b}_3 \in \text{aff } S$ since the coefficients sum to 1.

7. a. $\mathbf{p}_1 \in \text{Span } S$, but $\mathbf{p}_1 \notin \text{aff } S$

b. $\mathbf{p}_2 \in \text{Span } S$, and $\mathbf{p}_2 \in \text{aff } S$

c. $\mathbf{p}_3 \notin \text{Span } S$, so $\mathbf{p}_3 \notin \text{aff } S$

9. $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Other answers are possible.

11–19. See the *Study Guide*.

21. Span $\{\mathbf{v}_2 - \mathbf{v}_1, \mathbf{v}_3 - \mathbf{v}_1\}$ is a plane if and only if $\{\mathbf{v}_2 - \mathbf{v}_1, \mathbf{v}_3 - \mathbf{v}_1\}$ is linearly independent. Suppose c_2 and c_3 satisfy $c_2(\mathbf{v}_2 - \mathbf{v}_1) + c_3(\mathbf{v}_3 - \mathbf{v}_1) = \mathbf{0}$. Show that this implies $c_2 = c_3 = 0$.

23. Let $S = \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$. To show that S is affine, it suffices to show that S is a flat, by Theorem 3. Let $W = \{\mathbf{x} : A\mathbf{x} = \mathbf{0}\}$. Then W is a subspace of $\mathbb{R}^n$, by Theorem 2 in Section 4.2 (or Theorem 12 in Section 2.8). Since $S = W + \mathbf{p}$, where $\mathbf{p}$ satisfies $A\mathbf{p} = \mathbf{b}$, by Theorem 6 in Section 1.5, S is a translate of W, and hence S is a flat.

25. A suitable set consists of any three vectors that are not collinear and have 5 as their third entry. If 5 is their third entry, they lie in the plane $z = 5$. If the vectors are not collinear, their affine hull cannot be a line, so it must be the plane.

27. If $\mathbf{p}, \mathbf{q} \in f(S)$, then there exist $\mathbf{r}, \mathbf{s} \in S$ such that $f(\mathbf{r}) = \mathbf{p}$ and $f(\mathbf{s}) = \mathbf{q}$. Given any $t \in \mathbb{R}$, we must show that $\mathbf{z} = (1-t)\mathbf{p} + t\mathbf{q}$ is in $f(S)$. Now use definitions of $\mathbf{p}$ and $\mathbf{q}$, and the fact that f is linear. The complete proof is presented in the *Study Guide*.

29. Since B is affine, Theorem 2 implies that B contains all affine combinations of points of B. Hence B contains all affine combinations of points of A. That is, aff $A \subseteq B$.

31. Since $A \subseteq (A \cup B)$, it follows from Exercise 30 that aff $A \subseteq$ aff $(A \cup B)$. Similarly, aff $B \subseteq$ aff $(A \cup B)$, so [aff $A \cup$ aff B] $\subseteq$ aff $(A \cup B)$.

33. To show that $D \subseteq E \cap F$, show that $D \subseteq E$ and $D \subseteq F$. The complete proof is presented in the *Study Guide*.

Section 8.2, page 477

1. Affinely dependent and $2\mathbf{v}_1 + \mathbf{v}_2 - 3\mathbf{v}_3 = \mathbf{0}$

3. The set is affinely independent. If the points are called $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{v}_4$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for $\mathbb{R}^3$ and $\mathbf{v}_4 = 16\mathbf{v}_1 + 5\mathbf{v}_2 - 3\mathbf{v}_3$, but the weights in the linear combination do not sum to 1.

5. $-4\mathbf{v}_1 + 5\mathbf{v}_2 - 4\mathbf{v}_3 + 3\mathbf{v}_4 = \mathbf{0}$

7. The barycentric coordinates are $(-2, 4, -1)$.

9–17. See the *Study Guide*.

19. When a set of five points is translated by subtracting, say, the first point, the new set of four points must be linearly dependent, by Theorem 8 in Section 1.7, because the four points are in $\mathbb{R}^3$. By Theorem 5, the original set of five points is affinely dependent.

21. If $\{\mathbf{v}_1, \mathbf{v}_2\}$ is affinely dependent, then there exist c_1 and c_2, not both zero, such that $c_1 + c_2 = 0$ and $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$. Show that this implies $\mathbf{v}_1 = \mathbf{v}_2$. For the converse, suppose $\mathbf{v}_1 = \mathbf{v}_2$ and select specific c_1 and c_2 that show their affine dependence. The details are in the *Study Guide*.

23. a. The vectors $\mathbf{v}_2 - \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v}_3 - \mathbf{v}_1 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ are not multiples and hence are linearly independent. By Theorem 5, S is affinely independent.

b. $\mathbf{p}_1 \leftrightarrow \left(-\frac{6}{8}, \frac{9}{8}, \frac{5}{8}\right)$, $\mathbf{p}_2 \leftrightarrow \left(0, \frac{1}{2}, \frac{1}{2}\right)$, $\mathbf{p}_3 \leftrightarrow \left(\frac{14}{8}, -\frac{5}{8}, -\frac{1}{8}\right)$, $\mathbf{p}_4 \leftrightarrow \left(\frac{6}{8}, -\frac{5}{8}, \frac{7}{8}\right)$, $\mathbf{p}_5 \leftrightarrow \left(\frac{1}{4}, \frac{1}{8}, \frac{5}{8}\right)$

c. $\mathbf{p}_6$ is $(-, -, +)$, $\mathbf{p}_7$ is $(0, +, -)$, and $\mathbf{p}_8$ is $(+, +, -)$.

25. Suppose $S = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ is an affinely independent set. Then equation (7) has a solution, because $\mathbf{p}$ is in aff S. Hence equation (8) has a solution. By Theorem 5, the homogeneous forms of the points in S are linearly independent. Thus (8) has a unique solution. Then (7) also has a unique solution, because (8) encodes both equations that appear in (7).

The following argument mimics the proof of Theorem 8 in Section 4.4. If $S = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ is an affinely independent set, then scalars $c_1, \dots, c_k$ exist that satisfy (7), by definition of aff S. Suppose $\mathbf{x}$ also has the representation

$$\mathbf{x} = d_1\mathbf{b}_1 + \cdots + d_k\mathbf{b}_k \quad \text{and} \quad d_1 + \cdots + d_k = 1 \tag{7a}$$

for scalars $d_1, \dots, d_k$. Then subtraction produces the equation

$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (c_1 - d_1)\mathbf{b}_1 + \cdots + (c_k - d_k)\mathbf{b}_k \tag{7b}$$

The weights in (7b) sum to 0 because the c's and the d's separately sum to 1. This is impossible, unless each weight in (8) is 0, because S is an affinely independent set. This proves that $c_i = d_i$ for $i = 1, \dots, k$.

27. If $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ is an affinely dependent set, then there exist scalars c_1, c_2, and c_3, not all zero, such that $c_1\mathbf{p}_1 + c_2\mathbf{p}_2 + c_3\mathbf{p}_3 = \mathbf{0}$ and $c_1 + c_2 + c_3 = 0$. Now use the linearity of f.

29. Let $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$. Then

$\det[\,\tilde{\mathbf{a}}\;\; \tilde{\mathbf{b}}\;\; \tilde{\mathbf{c}}\,] = \det\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ 1 & 1 & 1 \end{bmatrix} =$

$\det\begin{bmatrix} a_1 & a_2 & 1 \\ b_1 & b_2 & 1 \\ c_1 & c_2 & 1 \end{bmatrix}$, by the transpose property of the determinant (Theorem 5 in Section 3.2). By Exercise 30 in Section 3.3, this determinant equals 2 times the area of the triangle with vertices at $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$.

31. If $[\,\tilde{\mathbf{a}}\;\; \tilde{\mathbf{b}}\;\; \tilde{\mathbf{c}}\,]\begin{bmatrix} r \\ s \\ t \end{bmatrix} = \tilde{\mathbf{p}}$, then Cramer's rule gives $r = \det[\,\tilde{\mathbf{p}}\;\; \tilde{\mathbf{b}}\;\; \tilde{\mathbf{c}}\,]/\det[\,\tilde{\mathbf{a}}\;\; \tilde{\mathbf{b}}\;\; \tilde{\mathbf{c}}\,]$. By Exercise 29, the numerator of this quotient is twice the area of $\triangle\mathbf{pbc}$, and the denominator is twice the area of $\triangle\mathbf{abc}$. This proves the formula for r. The other formulas are proved using Cramer's rule for s and t.

33. The intersection point is $\mathbf{x}(4) = -.1\begin{bmatrix} 1 \\ 3 \\ -6 \end{bmatrix} + .6\begin{bmatrix} 7 \\ 3 \\ -5 \end{bmatrix} + .5\begin{bmatrix} 3 \\ 9 \\ -2 \end{bmatrix} = \begin{bmatrix} 5.6 \\ 6.0 \\ -3.4 \end{bmatrix}$. It is not inside the triangle.

Section 8.3, page 484

1. See the *Study Guide*.

3. None are in conv S.

5. $\mathbf{p}_1 = -\frac{1}{6}\mathbf{v}_1 + \frac{1}{3}\mathbf{v}_2 + \frac{2}{3}\mathbf{v}_3 + \frac{1}{6}\mathbf{v}_4$, so $\mathbf{p}_1 \notin \text{conv}\, S$.
$\mathbf{p}_2 = \frac{1}{3}\mathbf{v}_1 + \frac{1}{3}\mathbf{v}_2 + \frac{1}{6}\mathbf{v}_3 + \frac{1}{6}\mathbf{v}_4$, so $\mathbf{p}_2 \in \text{conv}\, S$.

7. a. The barycentric coordinates of $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, and $\mathbf{p}_4$ are, respectively, $\left(\frac{1}{3}, \frac{1}{6}, \frac{1}{2}\right)$, $\left(0, \frac{1}{2}, \frac{1}{2}\right)$, $\left(\frac{1}{2}, -\frac{1}{4}, \frac{3}{4}\right)$, and $\left(\frac{1}{2}, \frac{3}{4}, -\frac{1}{4}\right)$.

b. $\mathbf{p}_3$ and $\mathbf{p}_4$ are outside conv T. $\mathbf{p}_1$ is inside conv T. $\mathbf{p}_2$ is on the edge $\overline{\mathbf{v}_2\mathbf{v}_3}$ of conv T.

9. $\mathbf{p}_1$ and $\mathbf{p}_3$ are outside the tetrahedron conv S. $\mathbf{p}_2$ is on the face containing the vertices $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{v}_4$. $\mathbf{p}_4$ is inside conv S. $\mathbf{p}_5$ is on the edge between $\mathbf{v}_1$ and $\mathbf{v}_3$.

11–15. See the *Study Guide*.

17. If $\mathbf{p}, \mathbf{q} \in f(S)$, then there exist $\mathbf{r}, \mathbf{s} \in S$ such that $f(\mathbf{r}) = \mathbf{p}$ and $f(\mathbf{s}) = \mathbf{q}$. The goal is to show that the line segment $\mathbf{y} = (1-t)\mathbf{p} + t\mathbf{q}$, for $0 \le t \le 1$, is in $f(S)$. Use the linearity of f and the convexity of S to show that $\mathbf{y} = f(\mathbf{w})$ for some $\mathbf{w}$ in S. This will show that $\mathbf{y}$ is in $f(S)$ and that $f(S)$ is convex.

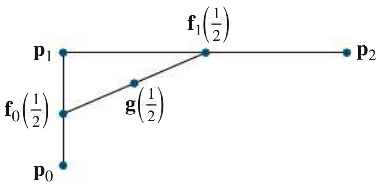
19. $\mathbf{p} = \frac{1}{6}\mathbf{v}_1 + \frac{1}{2}\mathbf{v}_2 + \frac{1}{3}\mathbf{v}_4$ and $\mathbf{p} = \frac{1}{2}\mathbf{v}_1 + \frac{1}{6}\mathbf{v}_2 + \frac{1}{3}\mathbf{v}_3$.

21. Suppose $A \subseteq B$, where B is convex. Then, since B is convex, Theorem 7 implies that B contains all convex combinations of points of B. Hence B contains all convex combinations of points of A. That is, conv $A \subseteq B$.

23. a. Use Exercise 22 to show that conv A and conv B are both subsets of conv $(A \cup B)$. This will imply that their union is also a subset of conv $(A \cup B)$.

b. One possibility is to let A be two adjacent corners of a square and let B be the other two corners. Then what is $(\text{conv}\, A) \cup (\text{conv}\, B)$, and what is conv $(A \cup B)$?

25.



27.
$$\begin{aligned} \mathbf{g}(t) &= (1-t)\mathbf{f}_0(t) + t\mathbf{f}_1(t) \\ &= (1-t)[(1-t)\mathbf{p}_0 + t\mathbf{p}_1] + t[(1-t)\mathbf{p}_1 + t\mathbf{p}_2] \\ &= (1-t)^2\mathbf{p}_0 + 2t(1-t)\mathbf{p}_1 + t^2\mathbf{p}_2. \end{aligned}$$
The sum of the weights in the linear combination for $\mathbf{g}$ is $(1-t)^2 + 2t(1-t) + t^2$, which equals $(1 - 2t + t^2) + (2t - 2t^2) + t^2 = 1$. The weights are each between 0 and 1 when $0 \le t \le 1$, so $\mathbf{g}(t)$ is in conv $\{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2\}$.

Section 8.4, page 493

1. $f(x_1, x_2) = 3x_1 + 4x_2$ and $d = 13$

3. a. Open **b.** Closed **c.** Neither
d. Closed **e.** Closed

5. a. Not compact, convex
b. Compact, convex
c. Not compact, convex
d. Not compact, not convex
e. Not compact, convex

7. a. $\mathbf{n} = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$ or a multiple

b. $f(\mathbf{x}) = 2x_2 + 3x_3$, $d = 11$

9. a. $\mathbf{n} = \begin{bmatrix} 3 \\ -1 \\ 2 \\ 1 \end{bmatrix}$ or a multiple

b. $f(\mathbf{x}) = 3x_1 - x_2 + 2x_3 + x_4$, $d = 5$

11. $\mathbf{v}_2$ is on the same side as $\mathbf{0}$, $\mathbf{v}_1$ is on the other side, and $\mathbf{v}_3$ is in H.

13. One possibility is $\mathbf{p} = \begin{bmatrix} 32 \\ -14 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 10 \\ -7 \\ 1 \\ 0 \end{bmatrix}$,

$\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 1 \end{bmatrix}$.

15. $f(x_1, x_2, x_3, x_4) = x_1 - 3x_2 + 4x_3 - 2x_4$, and $d = 5$

17. $f(x_1, x_2, x_3) = x_1 - 2x_2 + x_3$, and $d = 0$

19. $f(x_1, x_2, x_3) = -5x_1 + 3x_2 + x_3$, and $d = 0$

21–27. See the *Study Guide*.

29. $f(x_1, x_2) = 3x_1 - 2x_2$ with d satisfying $9 < d < 10$ is one possibility.

31. $f(x, y) = 4x + y$. A natural choice for d is 12.75, which equals $f(3, .75)$. The point $(3, .75)$ is three-fourths of the distance between the center of A and the center of B.

33. Exercise 2(a) in Section 8.3 gives one possibility. Or let $S = \{(x, y) : x^2y^2 = 1 \text{ and } y > 0\}$. Then conv S is the upper (open) half-plane.

35. Let $\mathbf{x}, \mathbf{y} \in B(\mathbf{p}, \delta)$ and suppose $\mathbf{z} = (1-t)\mathbf{x} + t\mathbf{y}$, where $0 \le t \le 1$. Then show that
$$\begin{aligned} \|\mathbf{z} - \mathbf{p}\| &= \|[(1-t)\mathbf{x} + t\mathbf{y}] - \mathbf{p}\| \\ &= \|(1-t)(\mathbf{x} - \mathbf{p}) + t(\mathbf{y} - \mathbf{p})\| < \delta. \end{aligned}$$

Section 8.5, page 505

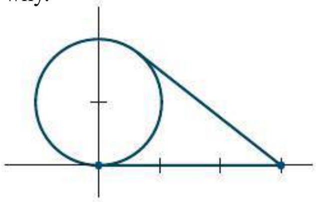
1. a. $m = 1$ at the point $\mathbf{p}_1$ **b.** $m = 5$ at the point $\mathbf{p}_2$
c. $m = 5$ at the point $\mathbf{p}_3$

3. a. $m = -3$ at the point $\mathbf{p}_3$
b. $m = 1$ on the set conv $\{\mathbf{p}_1, \mathbf{p}_3\}$
c. $m = -3$ on the set conv $\{\mathbf{p}_1, \mathbf{p}_2\}$

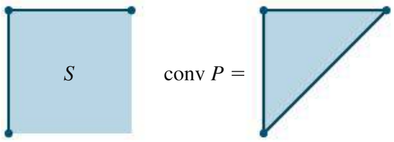
5. $\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \end{bmatrix} \right\}$

7. $\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 6 \end{bmatrix} \right\}$

9. The origin is an extreme point, but it is not a vertex. Explain why.



11. One possibility is to let S be a square that includes part of the boundary but not all of it. For example, include just two adjacent edges. The convex hull of the profile P is a triangular region.



13. a. $f_0(C^5) = 32$, $f_1(C^5) = 80$, $f_2(C^5) = 80$, $f_3(C^5) = 40$, $f_4(C^5) = 10$, and $32 - 80 + 80 - 40 + 10 = 2$.

b.

	f_0	f_1	f_2	f_3	f_4
C^1	2				
C^2	4	4			
C^3	8	12	6		
C^4	16	32	24	8	
C^5	32	80	80	40	10

For a general formula, see the *Study Guide*.

15. a. $f_0(P^n) = f_0(Q) + 1$

b. $f_k(P^n) = f_k(Q) + f_{k-1}(Q)$

c. $f_{n-1}(P^n) = f_{n-2}(Q) + 1$

17–23. See the *Study Guide*.

25. Let S be convex and let $\mathbf{x} \in cS + dS$, where $c > 0$ and $d > 0$. Then there exist $\mathbf{s}_1$ and $\mathbf{s}_2$ in S such that $\mathbf{x} = c\mathbf{s}_1 + d\mathbf{s}_2$. But then

$$\mathbf{x} = c\mathbf{s}_1 + d\mathbf{s}_2 = (c+d)\left(\frac{c}{c+d}\mathbf{s}_1 + \frac{d}{c+d}\mathbf{s}_2\right).$$

Now show that the expression on the right side is a member of $(c+d)S$.

For the converse, pick a typical point in $(c+d)S$ and show it is in $cS + dS$.

27. *Hint:* Suppose A and B are convex. Let $\mathbf{x}, \mathbf{y} \in A + B$. Then there exist $\mathbf{a}, \mathbf{c} \in A$ and $\mathbf{b}, \mathbf{d} \in B$ such that $\mathbf{x} = \mathbf{a} + \mathbf{b}$ and $\mathbf{y} = \mathbf{c} + \mathbf{d}$. For any t such that $0 \le t \le 1$, show that

$$\mathbf{w} = (1-t)\mathbf{x} + t\mathbf{y} = (1-t)(\mathbf{a} + \mathbf{b}) + t(\mathbf{c} + \mathbf{d})$$

represents a point in $A + B$.

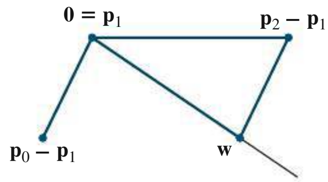
Section 8.6, page 516

1. The control points for $\mathbf{x}(t) + \mathbf{b}$ should be $\mathbf{p}_0 + \mathbf{b}$, $\mathbf{p}_1 + \mathbf{b}$, and $\mathbf{p}_3 + \mathbf{b}$. Write the Bézier curve through these points, and show algebraically that this curve is $\mathbf{x}(t) + \mathbf{b}$. See the *Study Guide*.

3. a. $\mathbf{x}'(t) = (-3 + 6t - 3t^2)\mathbf{p}_0 + (3 - 12t + 9t^2)\mathbf{p}_1 + (6t - 9t^2)\mathbf{p}_2 + 3t^2\mathbf{p}_3$, so
$\mathbf{x}'(0) = -3\mathbf{p}_0 + 3\mathbf{p}_1 = 3(\mathbf{p}_1 - \mathbf{p}_0)$, and
$\mathbf{x}'(1) = -3\mathbf{p}_2 + 3\mathbf{p}_3 = 3(\mathbf{p}_3 - \mathbf{p}_2)$. This shows that the tangent vector $\mathbf{x}'(0)$ points in the direction from $\mathbf{p}_0$ to $\mathbf{p}_1$ and is three times the length of $\mathbf{p}_1 - \mathbf{p}_0$. Likewise, $\mathbf{x}'(1)$ points in the direction from $\mathbf{p}_2$ to $\mathbf{p}_3$ and is three times the length of $\mathbf{p}_3 - \mathbf{p}_2$. In particular, $\mathbf{x}'(1) = \mathbf{0}$ if and only if $\mathbf{p}_3 = \mathbf{p}_2$.

b. $\mathbf{x}''(t) = (6 - 6t)\mathbf{p}_0 + (-12 + 18t)\mathbf{p}_1$
$+(6 - 18t)\mathbf{p}_2 + 6t\mathbf{p}_3$, so that

$\mathbf{x}''(0) = 6\mathbf{p}_0 - 12\mathbf{p}_1 + 6\mathbf{p}_2 = 6(\mathbf{p}_0 - \mathbf{p}_1) + 6(\mathbf{p}_2 - \mathbf{p}_1)$
and
$\mathbf{x}''(1) = 6\mathbf{p}_1 - 12\mathbf{p}_2 + 6\mathbf{p}_3 = 6(\mathbf{p}_1 - \mathbf{p}_2) + 6(\mathbf{p}_3 - \mathbf{p}_2)$
For a picture of $\mathbf{x}''(0)$, construct a coordinate system with the origin at $\mathbf{p}_1$, temporarily, label $\mathbf{p}_0$ as $\mathbf{p}_0 - \mathbf{p}_1$, and label $\mathbf{p}_2$ as $\mathbf{p}_2 - \mathbf{p}_1$. Finally, construct a line from this new origin through the sum of $\mathbf{p}_0 - \mathbf{p}_1$ and $\mathbf{p}_2 - \mathbf{p}_1$, extended out a bit. That line points in the direction of $\mathbf{x}''(0)$.



$\mathbf{w} = (\mathbf{p}_0 - \mathbf{p}_1) + (\mathbf{p}_2 - \mathbf{p}_1) = \frac{1}{6}\mathbf{x}''(0)$

5. a. From Exercise 3(a) or equation (9) in the text,

$\mathbf{x}'(1) = 3(\mathbf{p}_3 - \mathbf{p}_2)$

Use the formula for $\mathbf{x}'(0)$, with the control points from $\mathbf{y}(t)$, and obtain

$\mathbf{y}'(0) = -3\mathbf{p}_3 + 3\mathbf{p}_4 = 3(\mathbf{p}_4 - \mathbf{p}_3)$

For C^1 continuity, $3(\mathbf{p}_3 - \mathbf{p}_2) = 3(\mathbf{p}_4 - \mathbf{p}_3)$, so $\mathbf{p}_3 = (\mathbf{p}_4 + \mathbf{p}_2)/2$, and $\mathbf{p}_3$ is the midpoint of the line segment from $\mathbf{p}_2$ to $\mathbf{p}_4$.

b. If $\mathbf{x}'(1) = \mathbf{y}'(0) = \mathbf{0}$, then $\mathbf{p}_2 = \mathbf{p}_3$ and $\mathbf{p}_3 = \mathbf{p}_4$. Thus, the "line segment" from $\mathbf{p}_2$ to $\mathbf{p}_4$ is just the point $\mathbf{p}_3$. [*Note:* In this case, the combined curve is still C^1 continuous, by definition. However, some choices of the other "control" points, $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_5$, and $\mathbf{p}_6$, can produce a curve with a visible corner at $\mathbf{p}_3$, in which case the curve is not G^1 continuous at $\mathbf{p}_3$.]

7. *Hint:* Use $\mathbf{x}''(t)$ from Exercise 3 and adapt this for the second curve to see that

$\mathbf{y}''(t) = 6(1-t)\mathbf{p}_3 + 6(-2+3t)\mathbf{p}_4 + 6(1-3t)\mathbf{p}_5 + 6t\mathbf{p}_6$

Then set $\mathbf{x}''(1) = \mathbf{y}''(0)$. Since the curve is C^1 continuous at $\mathbf{p}_3$, Exercise 5(a) says that the point $\mathbf{p}_3$ is the midpoint of the segment from $\mathbf{p}_2$ to $\mathbf{p}_4$. This implies that $\mathbf{p}_4 - \mathbf{p}_3 = \mathbf{p}_3 - \mathbf{p}_2$. Use this substitution to show that $\mathbf{p}_4$ and $\mathbf{p}_5$ are uniquely determined by $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$. Only $\mathbf{p}_6$ can be chosen arbitrarily.

9. Write a vector of the polynomial weights for $\mathbf{x}(t)$, expand the polynomial weights, and factor the vector as $M_B\mathbf{u}(t)$:

$$\begin{bmatrix} 1-4t+6t^2-4t^3+t^4 \\ 4t-12t^2+12t^3-4t^4 \\ 6t^2-12t^3+6t^4 \\ 4t^3-4t^4 \\ t^4 \end{bmatrix} = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ 0 & 4 & -12 & 12 & -4 \\ 0 & 0 & 6 & -12 & 6 \\ 0 & 0 & 0 & 4 & -4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ t \\ t^2 \\ t^3 \\ t^4 \end{bmatrix},$$

$$M_B = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ 0 & 4 & -12 & 12 & -4 \\ 0 & 0 & 6 & -12 & 6 \\ 0 & 0 & 0 & 4 & -4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

11–15. See the *Study Guide*.

17. a. *Hint:* Use the fact that $\mathbf{q}_0 = \mathbf{p}_0$.

b. Multiply the first and last parts of equation (13) by $\frac{8}{3}$ and solve for $8\mathbf{q}_2$.

c. Use equation (8) to substitute for $8\mathbf{q}_3$ and then apply part (a).

19. a. From equation (11), $\mathbf{y}'(1) = .5\mathbf{x}'(.5) = \mathbf{z}'(0)$.

b. Observe that $\mathbf{y}'(1) = 3(\mathbf{q}_3 - \mathbf{q}_2)$. This follows from equation (9), with $\mathbf{y}(t)$ and its control points in place of $\mathbf{x}(t)$ and its control points. Similarly, for $\mathbf{z}(t)$ and its control points, $\mathbf{z}'(0) = 3(\mathbf{r}_1 - \mathbf{r}_0)$. By part (a), $3(\mathbf{q}_3 - \mathbf{q}_2) = 3(\mathbf{r}_1 - \mathbf{r}_0)$. Replace $\mathbf{r}_0$ by $\mathbf{q}_3$, and obtain $\mathbf{q}_3 - \mathbf{q}_2 = \mathbf{r}_1 - \mathbf{q}_3$, and hence $\mathbf{q}_3 = (\mathbf{q}_2 + \mathbf{r}_1)/2$.

c. Set $\mathbf{q}_0 = \mathbf{p}_0$ and $\mathbf{r}_3 = \mathbf{p}_3$. Compute $\mathbf{q}_1 = (\mathbf{p}_0 + \mathbf{p}_1)/2$ and $\mathbf{r}_2 = (\mathbf{p}_2 + \mathbf{p}_3)/2$. Compute $\mathbf{m} = (\mathbf{p}_1 + \mathbf{p}_2)/2$. Compute $\mathbf{q}_2 = (\mathbf{q}_1 + \mathbf{m})/2$ and $\mathbf{r}_1 = (\mathbf{m} + \mathbf{r}_2)/2$. Compute $\mathbf{q}_3 = (\mathbf{q}_2 + \mathbf{r}_1)/2$ and set $\mathbf{r}_0 = \mathbf{q}_3$.

21. a. $\mathbf{r}_0 = \mathbf{p}_0$, $\mathbf{r}_1 = \dfrac{\mathbf{p}_0 + 2\mathbf{p}_1}{3}$, $\mathbf{r}_2 = \dfrac{2\mathbf{p}_1 + \mathbf{p}_2}{3}$, $\mathbf{r}_3 = \mathbf{p}_2$

b. *Hint:* Write the standard formula (7) in this section, with $\mathbf{r}_i$ in place of $\mathbf{p}_i$ for $i = 0, \dots, 3$, and then replace $\mathbf{r}_0$ and $\mathbf{r}_3$ by $\mathbf{p}_0$ and $\mathbf{p}_2$, respectively:

$$\begin{aligned} \mathbf{x}(t) = (1 - 3t + 3t^2 - t^3)&\mathbf{p}_0 \\ + (3t - 6t^2 + 3t^3)&\mathbf{r}_1 \\ + (3t^2 - 3t^3)&\mathbf{r}_2 + t^3\mathbf{p}_2 \end{aligned}$$

Use the formulas for $\mathbf{r}_1$ and $\mathbf{r}_2$ from part (a) to examine the second and third terms in this expression for $\mathbf{x}(t)$.

Chapter 8 Supplementary Exercises, pages 519

1. T **2.** T **3.** F **4.** F **5.** T **6.** T

7. T **8.** F **9.** F **10.** F **11.** T **12.** T

13. F **14.** T **15.** T **16.** T **17.** T **18.** T

19. T **20.** F **21.** T

23. Let $\mathbf{y} \in F$. Then $U = F - \mathbf{y}$ and $V = G - \mathbf{y}$ are k-dimensional subspaces with $U \subseteq V$. Let $B = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ be a basis for U. Since dim $V = k$, B is also a basis for V. Hence $U = V$, and $F = U + \mathbf{y} = V + \mathbf{y} = G$.

25. *Hint:* Suppose $F_1 \cap F_2 \neq \varnothing$. Then there exist $\mathbf{v}_1$ and $\mathbf{v}_2$ in V such that $\mathbf{x}_1 + \mathbf{v}_1 = \mathbf{x}_2 + \mathbf{v}_2$. Use this and the properties of a subspace to show that for all $\mathbf{v}$ in V, $\mathbf{x}_1 + \mathbf{v} \in \mathbf{x}_2 + V$ and $\mathbf{x}_2 + \mathbf{v} \in \mathbf{x}_1 + V$.

27. *Hint:* Start with a basis for V and expand it by joining $\mathbf{p}$ to get a basis for $\mathbb{R}^n$.

29. *Hint:* Suppose $\mathbf{x} \in \lambda B(\mathbf{p}, \delta)$. This means that there exists $\mathbf{y} \in B(\mathbf{p}, \delta)$ such that $\mathbf{x} = \lambda\mathbf{y}$. Use the definition of $B(\mathbf{p}, \delta)$ to show that this implies $\mathbf{x} \in B(\lambda\mathbf{p}, \lambda\delta)$. The converse is similar.

31. The positive hull of S is a cone with vertex $(0, 0)$ containing the positive y axis and with sides on the lines $y = \pm x$.

33. *Hint:* It is significant that the set in Exercise 31 consists of exactly two non-collinear points. Explain why this is important.

35. *Hint:* Suppose $\mathbf{x} \in$ pos S. Then $\mathbf{x} = c_1\mathbf{v}_1 + \cdots c_k\mathbf{v}_k$, where $\mathbf{v}_i \in S$ and all $c_i \geq 0$. Let $d = \sum_{i=1}^{k} c_i$. Consider two cases: $d = 0$ and $d \neq 0$.

Chapter 9

Section 9.1, page 533

1. $$\begin{array}{c} \\ \text{d} \\ \text{q} \end{array} \begin{array}{c} \begin{array}{cc} \text{d} & \text{q} \end{array} \\ \begin{bmatrix} -10 & 10 \\ 25 & -25 \end{bmatrix} \end{array}$$

3. $$\begin{array}{l} \\ \text{rock} \\ \text{scissors} \\ \text{paper} \end{array} \begin{array}{c} \begin{array}{ccc} \text{r} & \text{s} & \text{p} \end{array} \\ \begin{bmatrix} 0 & 5 & -5 \\ -5 & 0 & 5 \\ 5 & -5 & 0 \end{bmatrix} \end{array}$$

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

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

9. a. $E(\mathbf{x}, \mathbf{y}) = \frac{13}{12}$, $v(\mathbf{x}) = \min\left\{\frac{5}{6}, 1, \frac{9}{6}\right\} = \frac{5}{6}$, $v(\mathbf{y}) = \max\left\{\frac{3}{4}, \frac{3}{2}, \frac{1}{2}\right\} = \frac{3}{2}$

b. $E(\mathbf{x}, \mathbf{y}) = \frac{9}{8}$, $v(\mathbf{x}) = \min\left\{1, \frac{3}{4}, \frac{7}{4}\right\} = \frac{3}{4}$, $v(\mathbf{y}) = \max\left\{\frac{1}{2}, \frac{5}{4}, \frac{3}{2}\right\} = \frac{3}{2}$

11. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{6} \end{bmatrix}$, $\hat{\mathbf{y}} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$, $v = \frac{1}{2}$

13. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}$, $\hat{\mathbf{y}} = \begin{bmatrix} \frac{4}{5} \\ \frac{1}{5} \end{bmatrix}$, $v = \frac{17}{5}$

15. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$ or $\begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}$ or any convex combination of these row strategies, $\hat{\mathbf{y}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $v = 2$

17. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{5}{7} \\ 0 \\ \frac{2}{7} \\ 0 \end{bmatrix}$, $\hat{\mathbf{y}} = \begin{bmatrix} 0 \\ \frac{5}{7} \\ \frac{2}{7} \\ 0 \\ 0 \end{bmatrix}$, $v = \frac{3}{7}$

19. a. Army: 1/3 river, 2/3 land; guerrillas: 1/3 river, 2/3 land; 2/3 of the supplies get through.

b. Army: 7/11 river, 4/11 land; guerrillas: 7/11 river, 4/11 land; 64/121 of the supplies get through.

21–29. See the *Study Guide*.

31. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{6} \\ 0 \end{bmatrix}$, $\hat{\mathbf{y}} = \begin{bmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$, $v = 0$

33. $\hat{\mathbf{x}} = \left(\dfrac{d-c}{a-b+d-c}, \dfrac{a-b}{a-b+d-c} \right)$,

$\hat{\mathbf{y}} = \left(\dfrac{d-b}{a-b+d-c}, \dfrac{a-c}{a-b+d-c} \right)$,

$v = \dfrac{ad-bc}{a-b+d-c}$

Section 9.2, page 543

1. Let x_1 be the amount invested in mutual funds, x_2 the amount in CDs, and x_3 the amount in savings. Then $\mathbf{b} = \begin{bmatrix} 12{,}000 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} .11 \\ .08 \\ .06 \end{bmatrix}$, and $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 0 & 1 & -2 \end{bmatrix}$.

3. $\mathbf{b} = \begin{bmatrix} 20 \\ -10 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -3 & -5 \end{bmatrix}$

5. $\mathbf{b} = \begin{bmatrix} -35 \\ 20 \\ -20 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} -7 \\ 3 \\ -1 \end{bmatrix}$, $A = \begin{bmatrix} -1 & 4 & 0 \\ 0 & 1 & -2 \\ 0 & -1 & 2 \end{bmatrix}$

7. max = 1360, when $x_1 = \frac{72}{5}$ and $x_2 = \frac{16}{5}$

9. unbounded

11–13. See the *Study Guide*.

15. max profit = \$1250, when $x_1 = 100$ bags of EverGreen and $x_2 = 350$ bags of QuickGreen

17. max profit = \$1180, for 20 widgets and 30 whammies

19. Take any $\mathbf{p}$ and $\mathbf{q}$ in S, with $\mathbf{p} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\mathbf{q} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then $\mathbf{v}^T\mathbf{p} \le c$ and $\mathbf{v}^T\mathbf{q} \le c$. Take any scalar t such that $0 \le t \le 1$. Then, by the linearity of matrix multiplication (or the dot product if $\mathbf{v}^T\mathbf{p}$ is written as $\mathbf{v}\cdot\mathbf{p}$, and so on),

$$\begin{aligned} \mathbf{v}^T[(1-t)\mathbf{p} + t\mathbf{q}] &= (1-t)\mathbf{v}^T\mathbf{p} + t\mathbf{v}^T\mathbf{q} \\ &\le (1-t)c + tc = c \end{aligned}$$

because $(1-t)$ and t are both positive and $\mathbf{p}$ and $\mathbf{q}$ are in S. So the line segment between $\mathbf{p}$ and $\mathbf{q}$ is in S. Since $\mathbf{p}$ and $\mathbf{q}$ were any points in S, the set S is convex.

21. Let $S = \{\mathbf{x} : f(\mathbf{x}) = d\}$, and take $\mathbf{p}$ and $\mathbf{q}$ in S. Also, take t with $0 \le t \le 1$, and let $\mathbf{x} = (1-t)\mathbf{p} + t\mathbf{q}$. Then

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{c}^T\mathbf{x} = \mathbf{c}^T[(1-t)\mathbf{p} + t\mathbf{q}] \\ &= (1-t)\mathbf{c}^T\mathbf{p} + t\mathbf{c}^T\mathbf{q} = (1-t)d + td = d \end{aligned}$$

Thus, $\mathbf{x}$ is in S. This shows that S is convex.

Section 9.3, page 559

1.
$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & M & \\ \end{array}$$
$$\left[\begin{array}{cccccc|c} 2 & 7 & 10 & 1 & 0 & 0 & 20 \\ 3 & 4 & 18 & 0 & 1 & 0 & 25 \\ \hline -21 & -25 & -15 & 0 & 0 & 1 & 0 \end{array}\right]$$

3. a. x_2

b.
$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & M & \\ \end{array}$$
$$\left[\begin{array}{ccccc|c} \frac{7}{2} & 0 & 1 & -\frac{1}{2} & 0 & 5 \\ \frac{3}{2} & 1 & 0 & \frac{1}{2} & 0 & 15 \\ \hline 11 & 0 & 0 & 5 & 1 & 150 \end{array}\right]$$

c. $x_1 = 0$, $x_2 = 15$, $x_3 = 5$, $x_4 = 0$, $M = 150$

d. optimal

5. a. x_1

b.
$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & M & \\ \end{array}$$
$$\left[\begin{array}{ccccc|c} 0 & 2 & 1 & -1 & 0 & 4 \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 8 \\ \hline 0 & -2 & 0 & 3 & 1 & 48 \end{array}\right]$$

c. $x_1 = 8$, $x_2 = 0$, $x_3 = 4$, $x_4 = 0$, $M = 48$

d. not optimal

7–11. See the *Study Guide*.

13. The maximum is 150, when $x_1 = 3$ and $x_2 = 10$.

15. The maximum is 56, when $x_1 = 9$ and $x_2 = 4$.

17. The minimum is 180, when $x_1 = 10$ and $x_2 = 12$.

19. The answer matches that in Example 7. The minimum is 20, when $x_1 = 8$ and $x_2 = 6$.

21. The maximum profit is $1180, achieved by making 20 widgets and 30 whammies each day.

Section 9.4, page 568

1. Minimize $36y_1 + 55y_2$ subject to
$$\begin{aligned} 2y_1 + 5y_2 &\geq 10 \\ 3y_1 + 4y_2 &\geq 12 \end{aligned}$$
and $y_1 \geq 0$, $y_2 \geq 0$.

3. Minimize $26y_1 + 30y_2 + 13y_3$ subject to
$$\begin{aligned} y_1 + 2y_2 + y_3 &\geq 4 \\ 2y_1 + 3y_2 + y_3 &\geq 5 \end{aligned}$$
and $y_1 \geq 0$, $y_2 \geq 0$, $y_3 \geq 0$.

5. The minimum is $M = 150$, attained when $y_1 = \frac{20}{7}$ and $y_2 = \frac{6}{7}$.

7. The minimum is $M = 56$, attained when $y_1 = 0$, $y_2 = 1$, and $y_3 = 2$.

9–15. See the *Study Guide*.

17. The minimum is 43, when $x_1 = \frac{7}{4}$, $x_2 = 0$, and $x_3 = \frac{3}{4}$.

19. The minimum cost is $670, using 11 bags of Pixie Power and 3 bags of Misty Might.

21. The marginal value is zero. This corresponds to labor in the fabricating department being underutilized. That is, at the optimal production schedule with $x_1 = 20$ and $x_2 = 30$, only 160 of the 200 available hours in fabricating are needed. The extra labor is wasted, and so it has value zero.

23. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{1}{3} \end{bmatrix}, \hat{\mathbf{y}} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, v = 1$

25. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{bmatrix}, \hat{\mathbf{y}} = \begin{bmatrix} \frac{3}{7} \\ \frac{3}{7} \\ \frac{1}{7} \end{bmatrix}, v = 1$

27. Change this "game" into a linear programming problem and use the simplex method to analyze the game. The expected value of the game is $\frac{38}{35}$, based on a payoff matrix for an investment of $100. With $35,000 to invest, Bob "plays" this game 350 times. Thus, he expects to gain $380, and the expected value of his portfolio at the end of the year is $35,380. Using the optimal game strategy, Bob should invest $11,000 in stocks, $9000 in bonds, and $15,000 in gold.

29. **a.** The coordinates of $\bar{\mathbf{x}}$ are all nonnegative. From the definition of $\mathbf{u}$, λ is equal to the sum of these coordinates. It follows that the coordinates of $\hat{\mathbf{x}}$ are nonnegative and sum to 1. Thus, $\hat{\mathbf{x}}$ is a mixed strategy for the row player R. A similar argument holds for $\hat{\mathbf{y}}$ and the column player C.

b. If $\mathbf{y}$ is any mixed strategy for C, then
$$\begin{aligned} E(\hat{\mathbf{x}}, \mathbf{y}) &= \hat{\mathbf{x}}^T A\mathbf{y} = \frac{1}{\lambda}(\bar{\mathbf{x}}^T A\mathbf{y}) = \frac{1}{\lambda}\left[(A^T\bar{\mathbf{x}})\cdot\mathbf{y}\right] \\ &\geq \frac{1}{\lambda}(\mathbf{v}\cdot\mathbf{y}) = \frac{1}{\lambda} \end{aligned}$$

c. If $\mathbf{x}$ is any mixed strategy for R, then
$$\begin{aligned} E(\mathbf{x}, \hat{\mathbf{y}}) &= \mathbf{x}^T A\hat{\mathbf{y}} = \frac{1}{\lambda}(\mathbf{x}^T A\bar{\mathbf{y}}) = \frac{1}{\lambda}[\mathbf{x}\cdot A\bar{\mathbf{y}}] \\ &\leq \frac{1}{\lambda}(\mathbf{x}\cdot\mathbf{u}) = \frac{1}{\lambda} \end{aligned}$$

d. Part (b) implies $v(\hat{\mathbf{x}}) \geq 1/\lambda$, so $v_R \geq 1/\lambda$. Part (c) implies $v(\hat{\mathbf{y}}) \leq 1/\lambda$, so $v_C \leq 1/\lambda$. It follows from the Minimax Theorem in Section 9.1 that $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are optimal mixed strategies for R and C, respectively, and that the value of the game is $1/\lambda$.

Chapter 9 Supplementary Exercises, page 570

1. T **2.** F **3.** F **4.** F **5.** T **6.** F

7. T **8.** T **9.** T **10.** F **11.** T **12.** F

13. F **14.** T **15.** F **16.** T **17.** F **18.** F

19. T **20.** F **21.** F **22.** T **23.** T **24.** F

25. **b.** The extreme points are $(0, 0)$, $(0, 1)$, and $(1, 2)$.

27. $f(x_1, x_2) = x_1 + x_2$, $f(0, 0) = 0$, $f(0, 1) = 1$, and $f(1, 2) = 3$

29. *Hint:* there are no feasible solutions.

31. $\hat{\mathbf{x}} = \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}, \hat{\mathbf{y}} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$, and $v = 1$.

Index

Page numbers denoted with "C" are found within the online chapter 10 as available at `bit.ly/2nj1HhO`.

E

F

M

Q

R

References to Applications

WEB indicates material on the Web site.

Numerical Linear Algebra

Physical Sciences

Statistics

Preface

Our Goal

Fundamentals of Differential Equations is designed to serve the needs of a one-semester course in basic theory as well as applications of differential equations. The flexibility of the text provides the instructor substantial latitude in designing a syllabus to match the emphasis of the course. Sample syllabi are provided in this preface that illustrate the inherent flexibility of this text to balance theory, methodology, applications, and numerical methods, as well as the incorporation of commercially available computer software for this course.

New to This Edition

- This text now features a MyMathLab course with approximately 750 algorithmic online homework exercises, tutorial videos, and the complete eText. Please see the "Technology and Supplements" section below for more details.
- In the Laplace Transforms chapter (7), the treatments of discontinuous and periodic functions are now divided into two sections that are more appropriate for 50 minute lectures: Section 7.6 "Transforms of Discontinuous Functions" (page 383) and Section 7.7 "Transforms of Periodic and Power Functions" (page 392).
- New examples have been added dealing with variation of parameters, Laplace transforms, the Gamma function, and eigenvectors (among others).
- New problems added to exercise sets deal with such topics as axon gating variables and oscillations of a helium-filled balloon on a cord. Additionally, novel problems accompany the new projects, focusing on economic models, disease control, synchronization, signal propagation, and phase plane analyses of neural responses. We have also added a set of Review Problems for Chapter 1 (page 29).
- Several pedagogical changes were made including amplification of the distinction between phase plane solutions and actual trajectories in Chapter 5 and incorporation of matrix and Jacobian formulations for autonomous systems.
- A new appendix lists commercial software and freeware for direction fields, phase portraits, and numerical methods for solving differential equations. (Appendix G, page A-17.)
- "The 2014–2015 Ebola Epidemic" is a new Project in Chapter 5 that describes a system of differential equations for modelling for the spread of the disease in West Africa. The model incorporates such features as contact tracing, number of contacts, likelihood of infection, and efficacy of isolation. See Project F, page 314.
- A new project in Chapter 1 called "Applications to Economics" deals with models for an agrarian economy as well as the growth of capital. See Project C, page 35.

- A new project in Chapter 4 called "Gravity Train" invites to reader to utilize differential equations in the design of an underground tunnel from Moscow to St. Petersburg, Russia, using gravity for propulsion. See Project H, page 240.
- Phase-locked loops constitute the theme of a new project in Chapter 5 that utilizes differential equations to analyze a technique for measuring or matching high frequency radio oscillations. See Project G, page 317.
- A new Project in Chapter 10 broadens the analysis of the wave and heat equations to explore the telegrapher's and cable equations. See Project E, page 637.

Prerequisites

While some universities make linear algebra a prerequisite for differential equations, many schools (especially engineering) only require calculus. With this in mind, we have designed the text so that only Chapter 6 (Theory of Higher-Order Linear Differential Equations) and Chapter 9 (Matrix Methods for Linear Systems) require more than high school level linear algebra. Moreover, Chapter 9 contains review sections on matrices and vectors as well as specific references for the deeper results used from the theory of linear algebra. We have also written Chapter 5 so as to give an introduction to systems of differential equations—including methods of solving, phase plane analysis, applications, numerical procedures, and Poincaré maps—that does not require a background in linear algebra.

Sample Syllabi

As a rough guide in designing a one-semester syllabus related to this text, we provide three samples that can be used for a 15-week course that meets three hours per week. The first emphasizes applications and computations including phase plane analysis; the second is designed for courses that place more emphasis on theory; and the third stresses methodology and partial differential equations. Chapters 1, 2, and 4 provide the core for any first course. The rest of the chapters are, for the most part, independent of each other. For students with a background in linear algebra, the instructor may prefer to replace Chapter 7 (Laplace Transforms) or Chapter 8 (Series Solutions of Differential Equations) with sections from Chapter 9 (Matrix Methods for Linear Systems).

	Methods, Computations, and Applications	Theory and Methods (linear algebra prerequisite)	Methods and Partial Differential Equations
Week	**Sections**	**Sections**	**Sections**
1	1.1, 1.2, 1.3	1.1, 1.2, 1.3	1.1, 1.2, 1.3
2	1.4, 2.2	1.4, 2.2, 2.3	1.4, 2.2
3	2.3, 2.4, 3.2	2.4, 3.2, 4.1	2.3, 2.4
4	3.4, 3.5, 3.6	4.2, 4.3, 4.4	3.2, 3.4
5	3.7, 4.1	4.5, 4.6	4.2, 4.3
6	4.2, 4.3, 4.4	4.7, 5.2, 5.3	4.4, 4.5, 4.6
7	4.5, 4.6, 4.7	5.4, 6.1	4.7, 5.1, 5.2
8	4.8, 4.9	6.2, 6.3, 7.2	7.1, 7.2, 7.3
9	4.10, 5.1, 5.2	7.3, 7.4, 7.5	7.4, 7.5
10	5.3, 5.4, 5.5	7.6, 7.7, 7.8	7.6, 7.7
11	5.6, 5.7, 7.2	8.2, 8.3	7.8, 8.2
12	7.3, 7.4, 7.5	8.4, 8.6, 9.1	8.3, 8.5, 8.6
13	7.6, 7.7, 7.8	9.2, 9.3	10.2, 10.3
14	8.1, 8.2, 8.3	9.4, 9.5, 9.6	10.4, 10.5
15	8.4, 8.6	9.7, 9.8	10.6, 10.7

Retained Features

Flexible Organization Most of the material is modular in nature to allow for various course configurations and emphasis (theory, applications and techniques, and concepts).

Optional Use of Computer Software The availability of computer packages such as Mathcad®, Mathematica®, MATLAB®, and Maple™ provides an opportunity for the student to conduct numerical experiments and tackle realistic applications that give additional insights into the subject. Consequently, we have inserted several exercises and projects throughout the text that are designed for the student to employ available software in phase plane analysis, eigenvalue computations, and the numerical solutions of various equations.

Review of Integration In response to the perception that many of today's students' skills in integration have gotten rusty by the time they enter a differential equations course, we have included an appendix offering a quick review of the basic methods for integrating functions analytically.

Choice of Applications Because of syllabus constraints, some courses will have little or no time for sections (such as those in Chapters 3 and 5) that exclusively deal with applications. Therefore, we have made the sections in these chapters independent of each other. To afford the instructor even greater flexibility, we have built in a variety of applications in the exercises for the theoretical sections. In addition, we have included many projects that deal with such applications.

Projects At the end of each chapter are projects relating to the material covered in the chapter. Several of them have been contributed by distinguished researchers. A project might involve a more challenging application, delve deeper into the theory, or introduce more advanced topics in differential equations. Although these projects can be tackled by an individual student, classroom testing has shown that working in groups lends a valuable added dimension to the learning experience. Indeed, it simulates the interactions that take place in the professional arena.

Technical Writing Exercises

Communication skills are, of course, an essential aspect of professional activities. Yet few texts provide opportunities for the reader to develop these skills. Thus, we have added at the end of most chapters a set of clearly marked technical writing exercises that invite students to make documented responses to questions dealing with the concepts in the chapter. In so doing, students are encouraged to make comparisons between various methods and to present examples that support their analysis.

Historical Footnotes

Throughout the text historical footnotes are set off by colored daggers (†). These footnotes typically provide the name of the person who developed the technique, the date, and the context of the original research.

Motivating Problem

Most chapters begin with a discussion of a problem from physics or engineering that motivates the topic presented and illustrates the methodology.

Chapter Summary and Review Problems

All of the main chapters contain a set of review problems along with a synopsis of the major concepts presented.

Computer Graphics

Most of the figures in the text were generated via computer. Computer graphics not only ensure greater accuracy in the illustrations, they demonstrate the use of numerical experimentation in studying the behavior of solutions.

Proofs

While more pragmatic students may balk at proofs, most instructors regard these justifications as an essential ingredient in a textbook on differential equations. As with any text at this level, certain details in the proofs must be omitted. When this occurs, we flag the instance and refer readers either to a problem in the exercises or to another text. For convenience, the end of a proof is marked by the symbol ◆.

Linear Theory

We have developed the theory of linear differential equations in a gradual manner. In Chapter 4 (Linear Second-Order Equations) we first present the basic theory for linear second-order equations with constant coefficients and discuss various techniques for solving these equations. Section 4.7 surveys the extension of these ideas to variable-coefficient second-order equations. A more general and detailed discussion of linear differential equations is given in Chapter 6 (Theory of Higher-Order Linear Differential Equations). For a beginning course emphasizing methods of solution, the presentation in Chapter 4 may be sufficient and Chapter 6 can be skipped.

Numerical Algorithms

Several numerical methods for approximating solutions to differential equations are presented along with program outlines that are easily implemented on a computer. These methods are introduced early in the text so that teachers and/or students can use them for numerical experimentation and for tackling complicated applications. Where appropriate we direct the student to software packages or web-based applets for implementation of these algorithms.

Exercises

An abundance of exercises is graduated in difficulty from straightforward, routine problems to more challenging ones. Deeper theoretical questions, along with applications, usually occur toward the end of the exercise sets. Throughout the text we have included problems and projects that require the use of a calculator or computer. These exercises are denoted by the symbol .

Laplace Transforms

We provide a detailed chapter on Laplace transforms (Chapter 7), since this is a recurring topic for engineers. Our treatment emphasizes discontinuous forcing terms and includes a section on the Dirac delta function.

Power Series Power series solutions is a topic that occasionally causes student anxiety. Possibly, this is due to inadequate preparation in calculus where the more subtle subject of convergent series is (frequently) covered at a rapid pace. Our solution has been to provide a graceful initiation into the theory of power series solutions with an exposition of Taylor polynomial approximants to solutions, deferring the sophisticated issues of convergence to later sections. Unlike many texts, ours provides an extensive section on the method of Frobenius (Section 8.6) as well as a section on finding a second linearly independent solution. While we have given considerable space to power series solutions, we have also taken great care to accommodate the instructor who only wishes to give a basic introduction to the topic. An introduction to solving differential equations using power series and the method of Frobenius can be accomplished by covering the materials in Sections 8.1, 8.2, 8.3, and 8.6.

Partial Differential Equations An introduction to this subject is provided in Chapter 10, which covers the method of separation of variables, Fourier series, the heat equation, the wave equation, and Laplace's equation. Examples in two and three dimensions are included.

Phase Plane Chapter 5 describes how qualitative information for two-dimensional systems can be gleaned about the solutions to intractable autonomous equations by observing their direction fields and critical points on the phase plane. With the assistance of suitable software, this approach provides a refreshing, almost recreational alternative to the traditional analytic methodology as we discuss applications in nonlinear mechanics, ecosystems, and epidemiology.

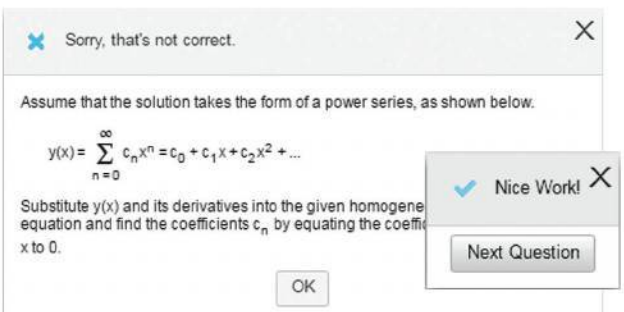
Vibrations Motivation for Chapter 4 on linear differential equations is provided in an introductory section describing the mass–spring oscillator. We exploit the reader's familiarity with common vibratory motions to anticipate the exposition of the theoretical and analytical aspects of linear equations. Not only does this model provide an anchor for the discourse on constant-coefficient equations, but a liberal interpretation of its features enables us to predict the qualitative behavior of variable-coefficient and nonlinear equations as well.

Review of Algebraic Equations and Matrices The chapter on matrix methods for linear systems (Chapter 9) begins with two (optional) introductory sections reviewing the theory of linear algebraic systems and matrix algebra.

Technology and Supplements

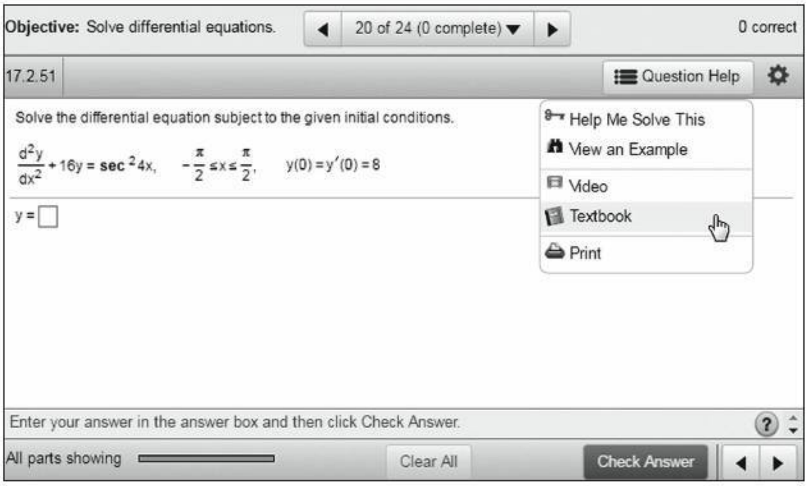
MyMathLab® Online Course (access code required) Built around Pearson's best-selling content, MyMathLab is an online homework, tutorial, and assessment program designed to work with this text to engage students and improve results. MyMathLab can be successfully implemented in any classroom environment—lab-based, hybrid, fully online, or traditional.

MyMathLab's online homework offers students immediate feedback and tutorial assistance that motivates them to do more, which means they retain more knowledge and improve their test scores. Used by more than 37 million students worldwide, MyMathLab delivers consistent, measurable gains in student learning outcomes, retention, and subsequent course success. Visit **www.mymathlab.com/results** to learn more.

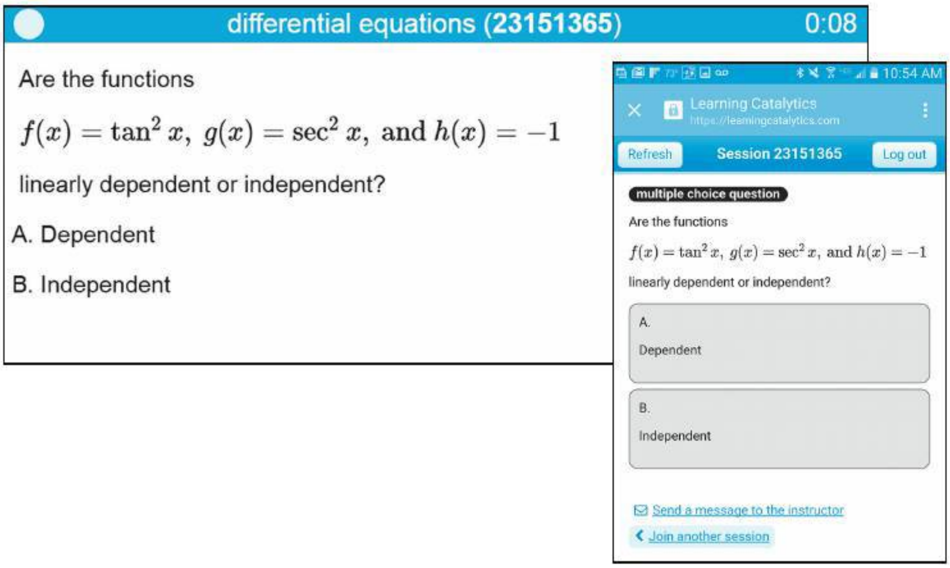


Learning and Teaching Tools

- **Exercises with immediate feedback**—Nearly 750 assignable exercises are based on the textbook exercises and regenerate algorithmically to give students unlimited opportunity for practice and mastery. MyMathLab provides helpful feedback when students enter incorrect answers and includes optional learning aids including Help Me Solve This, View an Example, videos, and an eText.

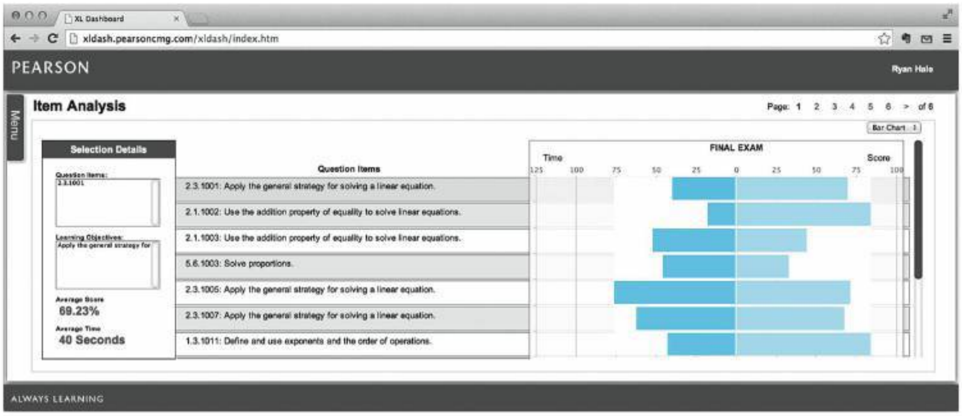


- **Learning Catalytics™** is a student response tool that uses students' smartphones, tablets, or laptops to engage them in more interactive tasks and thinking. Learning Catalytics fosters student engagement and peer-to-peer learning with real-time analytics.



- **Instructional videos** are available as learning aids within exercises and for self-study within the Multimedia Library. The Guide to Video-Based Assignments makes it easy to assign videos for homework by showing which MyMathLab exercises correspond to each video.
- **The complete eText** is available to students through their MyMathLab courses for the lifetime of the edition, giving students unlimited access to the eText within any course using that edition of the textbook.

- **Accessibility** and achievement go hand in hand. MyMathLab is compatible with the JAWS screen reader, and enables multiple-choice and free-response problem types to be read and interacted with via keyboard controls and math notation input. MyMathLab also works with screen enlargers, including ZoomText, MAGic, and SuperNova. And, all MyMathLab videos have closed-captioning. More information is available at **mymathlab.com/accessibility**.
- **A comprehensive gradebook** with enhanced reporting functionality allows you to efficiently manage your course.
  - **The Reporting Dashboard** provides insight to view, analyze, and report learning outcomes. Student performance data is presented at the class, section, and program levels in an accessible, visual manner so you'll have the information you need to keep your students on track.



  - **Item Analysis** tracks class-wide understanding of particular exercises so you can refine your class lectures or adjust the course/department syllabus. Just-in-time teaching has never been easier!

MyMathLab comes from an experienced partner with educational expertise and an eye on the future. Whether you are just getting started with MyMathLab, or have a question along the way, we're here to help you learn about our technologies and how to incorporate them into your course. To learn more about how MyMathLab helps students succeed, visit **mymathlab.com** or contact your Pearson rep.

MathXL® is the homework and assessment engine that runs MyMathLab. (MyMathLab is MathXL plus a learning management system.) MathXL access codes are also an option.

Student's Solutions Manual

ISBN-10: 0321977211 | ISBN-13: 9780321977212
Contains complete worked-out solutions to odd-numbered exercises, providing students with an excellent study tool. Available in print and for download within MyMathLab.

Instructor's Solutions Manual (downloadable)

ISBN-10: 0134659244 | ISBN-13: 9780134659244
Contains answers to all even-numbered exercises, detailed solutions to the even-numbered problems in several of the main chapters, and additional projects. Available for download in the Pearson Instructor Resource Center **www.pearsonhighered.com/irc** as well as within MyMathLab.

MATLAB, Maple, and Mathematica Manuals (downloadable)

By Thomas W. Polaski (Winthrop University), Bruno Welfert (Arizona State University), and Maurino Bautista (Rochester Institute of Technology), respectively. These manuals contain a collection of instructor tips, worksheets, and projects to aid instructors in integrating computer algebra systems into their courses. Complete manuals are available for instructor download within MyMathLab. Student worksheets and projects available for download within MyMathLab.

Acknowledgments

The staging of this text involved considerable behind-the-scenes activity. We thank, first of all, Philip Crooke, Glenn Webb (Vanderbilt University), and Joanna Wares (University of Richmond) who have continued to provide novel biomathematical projects, as well as Greg Huffman (Vanderbilt University) for his project on economics. We also want to thank Frank Glaser (California State Polytechnic University, Pomona) for many of the historical footnotes. We are indebted to Herbert E. Rauch (Lockheed Research Laboratory) for help with Section 3.3 on heating and cooling of buildings, Project B in Chapter 3 on aquaculture, and other application problems. Our appreciation goes to Richard H. Elderkin (Pomona College), Jerrold Marsden (California Institute of Technology), T. G. Proctor (Clemson University), and Philip W. Schaefer (University of Tennessee), who read and reread the manuscript for the original text, making numerous suggestions that greatly improved our work. Thanks also to the following reviewers of this and previous editions:

*Miklos Bona, University of Florida
Amin Boumenir, University of West Georgia
Mark Brittenham, University of Nebraska
*Jennifer Bryan, Oklahoma Christian University
Weiming Cao, University of Texas at San Antonio
Richard Carmichael, Wake Forest University
*Kwai-lee Chui, University of Florida
Karen Clark, The College of New Jersey
*Shaozhong Deng, University of North Carolina Charlotte
Patrick Dowling, Miami University
*Chris Fuller, Cumberland University
Sanford Geraci, Northern Virginia
David S. Gilliam, Texas Tech University at Lubbock
Scott Gordon, State University of West Georgia
*Irvin Hentzel, Iowa State University
*Mimi Rasky, Southwestern College
Richard Rubin, Florida International University
John Sylvester, University of Washington at Seattle
Steven Taliaferro, Texas A&M University at College Station
Michael M. Tom, Louisiana State University
Shu-Yi Tu, University of Michigan,
Flint Klaus Volpert, Villanova University
*Glenn F. Webb, Vanderbilt University

E. B. Saff, A. D. Snider

*Denotes reviewers of the current edition

CHAPTER 4

Linear Second-Order Equations

4.1 Introduction: The Mass–Spring Oscillator

A damped mass–spring oscillator consists of a mass m attached to a spring fixed at one end, as shown in Figure 4.1. Devise a differential equation that governs the motion of this oscillator, taking into account the forces acting on it due to the spring elasticity, damping friction, and possible external influences.

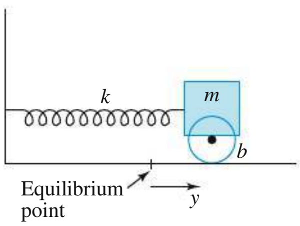


Figure 4.1 Damped mass–spring oscillator

Newton's second law—force equals mass times acceleration $(F = ma)$—is without a doubt the most commonly encountered differential equation in practice. It is an ordinary differential equation of the *second order* since acceleration is the second derivative of position (y) with respect to time $\left(a = d^2y/dt^2\right)$.

When the second law is applied to a mass–spring oscillator, the resulting motions are common experiences of everyday life, and we can exploit our familiarity with these vibrations to obtain a qualitative description of the solutions of more general second-order equations.

We begin by referring to Figure 4.1, which depicts the mass–spring oscillator. When the spring is unstretched and the inertial mass m is still, the system is at equilibrium; we measure the coordinate y of the mass by its displacement from the equilibrium position. When the mass m is displaced from equilibrium, the spring is stretched or compressed and it exerts a force that resists the displacement. For most springs this force is directly proportional to the displacement y and is thus given by

$$\textbf{(1)} \qquad F_{\text{spring}} = -ky\,,$$

where the positive constant k is known as the *stiffness* and the negative sign reflects the opposing nature of the force. **Hooke's law,** as equation (1) is commonly known, is only valid for sufficiently small displacements; if the spring is compressed so strongly that the coils press against each other, the opposing force obviously becomes much stronger.

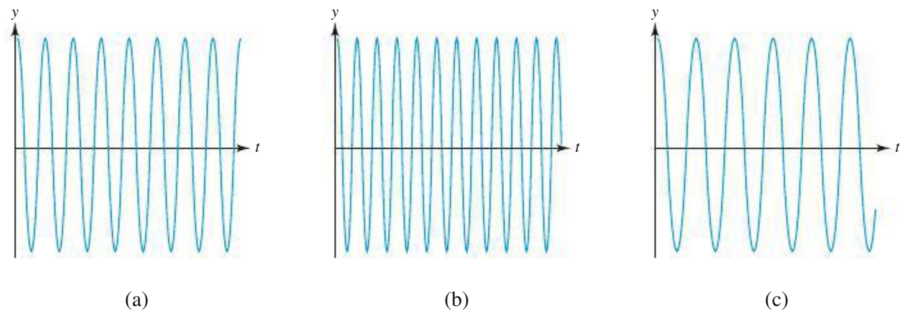


Figure 4.2 (a) Sinusoidal oscillation, (b) stiffer spring, and (c) heavier mass

Practically all mechanical systems also experience friction, and for vibrational motion this force is usually modeled accurately by a term proportional to velocity:

(2) $$F_{\text{friction}} = -b\frac{dy}{dt} = -by' ,$$

where $b\ (\geq 0)$ is the *damping coefficient* and the negative sign has the same significance as in equation (1).

The other forces on the oscillator are usually regarded as *external* to the system. Although they may be gravitational, electrical, or magnetic, commonly the most important external forces are transmitted to the mass by shaking the supports holding the system. For the moment we lump all the external forces into a single, *known* function $F_{\text{ext}}(t)$. Newton's law then provides the differential equation for the mass–spring oscillator:

$$my'' = -ky - by' + F_{\text{ext}}(t)$$

or

(3) $$\boldsymbol{my'' + by' + ky = F_{\text{ext}}(t)} .$$

What do mass–spring motions look like? From our everyday experience with weak auto suspensions, musical gongs, and bowls of jelly, we expect that when there is no friction $(b = 0)$ or external force, the (idealized) motions would be perpetual vibrations like the ones depicted in Figure 4.2. These vibrations resemble sinusoidal functions, with their amplitude depending on the initial displacement and velocity. The frequency of the oscillations increases for stiffer springs but decreases for heavier masses.

In Section 4.3 we will show how to find these solutions. Example 1 demonstrates a quick calculation that confirms our intuitive predictions.

Example 1 Verify that if $b = 0$ and $F_{\text{ext}}(t) = 0$, equation (3) has a solution of the form $y(t) = \cos \omega t$ and that the angular frequency ω increases with k and decreases with m.

Solution Under the conditions stated, equation (3) simplifies to

(4) $$my'' + ky = 0 .$$

The second derivative of $y(t)$ is $-\omega^2 \cos \omega t$, and if we insert it into (4), we find

$$my'' + ky = -m\omega^2 \cos \omega t + k \cos \omega t ,$$

which is indeed zero if $\omega = \sqrt{k/m}$. This ω increases with k and decreases with m, as predicted. ◆

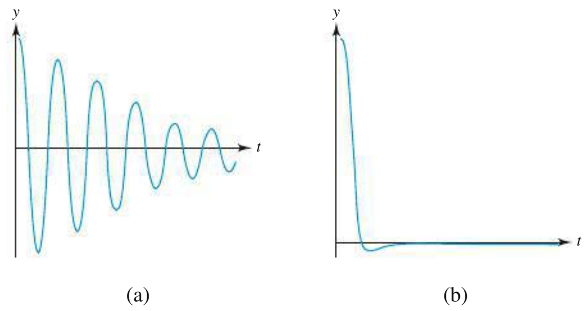


Figure 4.3 (a) Low damping and (b) high damping

When damping is present, the oscillations die out, and the motions resemble Figure 4.3. In Figure 4.3(a) the graph displays a damped oscillation; damping has slowed the frequency, and the amplitude appears to diminish exponentially with time. In Figure 4.3(b) the damping is so dominant that it has prevented the system from oscillating at all. Devices that are *supposed* to vibrate, like tuning forks or crystal oscillators, behave like Figure 4.3(a), and the damping effect is usually regarded as an undesirable loss mechanism. Good automotive suspension systems, on the other hand, behave like Figure 4.3(b); they exploit damping to *suppress* the oscillations.

The procedures for solving (unforced) mass–spring systems with damping are also described in Section 4.3, but as Examples 2 and 3 below show, the calculations are more complex. Example 2 has a relatively low damping coefficient $(b = 6)$ and illustrates the solutions for the "underdamped" case in Figure 4.3(a). In Example 3 the damping is more severe $(b = 10)$, and the solution is "overdamped" as in Figure 4.3(b).

Example 2 Verify that the exponentially damped sinusoid given by $y(t) = e^{-3t}\cos 4t$ is a solution to equation (3) if $F_{\text{ext}} = 0$, $m = 1$, $k = 25$, and $b = 6$.

Solution The derivatives of y are

$$\begin{aligned} y'(t) &= -3e^{-3t}\cos 4t - 4e^{-3t}\sin 4t\,, \\ y''(t) &= 9e^{-3t}\cos 4t + 12e^{-3t}\sin 4t + 12e^{-3t}\sin 4t - 16e^{-3t}\cos 4t \\ &= -7e^{-3t}\cos 4t + 24e^{-3t}\sin 4t\,, \end{aligned}$$

and insertion into (3) gives

$$\begin{aligned} my'' + by' + ky &= (1)y'' + 6y' + 25y \\ &= -7e^{-3t}\cos 4t + 24e^{-3t}\sin 4t + 6(-3e^{-3t}\cos 4t - 4e^{-3t}\sin t) \\ &\quad + 25e^{-3t}\cos 4t \\ &= 0\,. \end{aligned}$$ ◆

Example 3 Verify that the simple exponential function $y(t) = e^{-5t}$ is a solution to equation (3) if $F_{\text{ext}} = 0$, $m = 1$, $k = 25$, and $b = 10$.

Solution The derivatives of y are $y'(t) = -5e^{-5t}$, $y''(t) = 25e^{-5t}$ and insertion into (3) produces

$$my'' + by' + ky = (1)y'' + 10y' + 25y = 25e^{-5t} + 10(-5e^{-5t}) + 25e^{-5t} = 0\,.$$ ◆

Now if a mass–spring system is driven by an external force that is sinusoidal at the angular frequency ω, our experiences indicate that although the initial response of the system may be

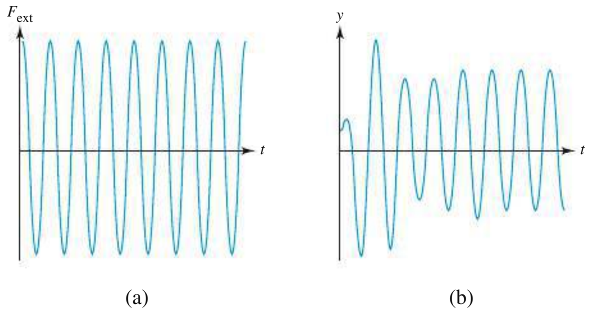


Figure 4.4 (a) Driving force and (b) response

somewhat erratic, eventually it will respond in "sync" with the driver and oscillate at the same frequency, as illustrated in Figure 4.4.

Common examples of systems vibrating in synchronization with their drivers are sound system speakers, cyclists bicycling over railroad tracks, electronic amplifier circuits, and ocean tides (driven by the periodic pull of the moon). However, there is more to the story than is revealed above. Systems can be enormously sensitive to the particular frequency ω at which they are driven. Thus, accurately tuned musical notes can shatter fine crystal, wind-induced vibrations at the right (wrong?) frequency can bring down a bridge, and a dripping faucet can cause inordinate headaches. These "resonance" responses (for which the responses have maximum amplitudes) may be quite destructive, and structural engineers have to be very careful to ensure that their products will not resonate with any of the vibrations likely to occur in the operating environment. Radio engineers, on the other hand, *do* want their receivers to resonate selectively to the desired broadcasting channel.

The calculation of these forced solutions is the subject of Sections 4.4 and 4.5. The next example illustrates some of the features of synchronous response and resonance.

Example 4 Find the synchronous response of the mass–spring oscillator with $m = 1, b = 1, k = 25$ to the force $\sin \Omega t$.

Solution We seek solutions of the differential equation

$$y'' + y' + 25y = \sin \Omega t \tag{5}$$

that are sinusoids in sync with $\sin \Omega t$; so let's try the form $y(t) = A \cos \Omega t + B \sin \Omega t$. Since

$$\begin{aligned} y' &= -\Omega A \sin \Omega t + \Omega B \cos \Omega t\,, \\ y'' &= -\Omega^2 A \cos \Omega t - \Omega^2 B \sin \Omega t\,, \end{aligned}$$

we can simply insert these forms into equation (5), collect terms, and match coefficients to obtain a solution:

$$\begin{aligned} \sin \Omega t &= y'' + y' + 25y \\ &= -\Omega^2 A \cos \Omega t - \Omega^2 B \sin \Omega t + [-\Omega A \sin \Omega t + \Omega B \cos \Omega t] \\ &\quad +25[A \cos \Omega t + B \sin \Omega t] \\ &= [-\Omega^2 B - \Omega A + 25B] \sin \Omega t + [-\Omega^2 A + \Omega B + 25A] \cos \Omega t\,, \end{aligned}$$

so

$$\begin{aligned} -\Omega A + (-\Omega^2 + 25)B &= 1 \\ (-\Omega^2 + 25)A + \Omega B &= 0\,. \end{aligned}$$

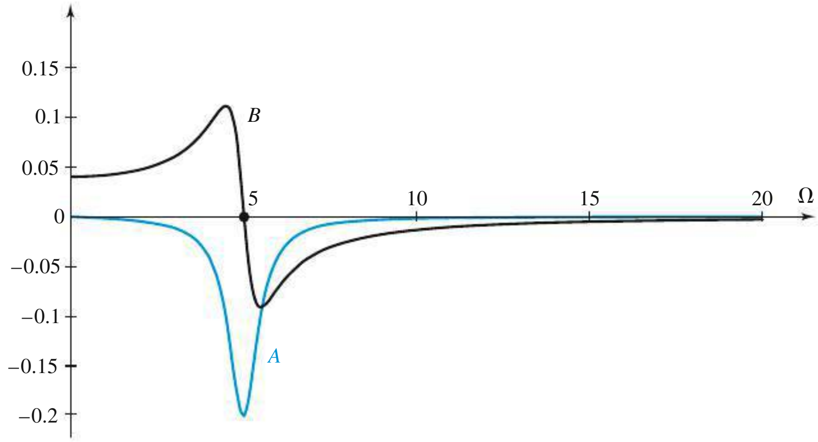


Figure 4.5 Vibration amplitudes around resonance

We find

$$A = \frac{-\Omega}{\Omega^2 + (\Omega^2 - 25)^2}, \qquad B = \frac{-\Omega^2 + 25}{\Omega^2 + (\Omega^2 - 25)^2}.$$

Figure 4.5 displays A and B as functions of the driving frequency Ω. A resonance clearly occurs around $\Omega \approx 5$. ◆

In most of this chapter, we are going to restrict our attention to differential equations of the form

$$ay'' + by' + cy = f(t)\,, \tag{6}$$

where $y(t)$ [or $y(x)$, or $x(t)$, etc.] is the unknown function that we seek; a, b, and c are constants; and $f(t)$ [or $f(x)$] is a *known* function. The proper nomenclature for (6) is the *linear, second-order ordinary differential equation with constant coefficients*. In Sections 4.7 and 4.8, we will generalize our focus to equations with nonconstant coefficients, as well as to nonlinear equations. However, (6) is an excellent starting point because we are able to obtain explicit solutions and observe, in concrete form, the theoretical properties that are predicted for more general equations. For motivation of the mathematical procedures and theory for solving (6), we will consistently compare it with the mass–spring paradigm:

$$[\textbf{inertia}] \times y'' + [\textbf{damping}] \times y' + [\textbf{stiffness}] \times y = F_{\text{ext}}\,.$$

4.1 EXERCISES

1. Verify that for $b = 0$ and $F_{\text{ext}}(t) = 0$, equation (3) has a solution of the form

$$y(t) = \cos \omega t, \text{ where } \omega = \sqrt{k/m}\,.$$

2. If $F_{\text{ext}}(t) = 0$, equation (3) becomes

$$my'' + by' + ky = 0\,.$$

For this equation, verify the following:

(a) If $y(t)$ is a solution, so is $cy(t)$, for any constant c.

(b) If $y_1(t)$ and $y_2(t)$ are solutions, so is their sum $y_1(t) + y_2(t)$.

3. Show that if $F_{\text{ext}}(t) = 0, m = 1, k = 9,$ and $b = 6$, then equation (3) has the "critically damped" solutions $y_1(t) = e^{-3t}$ and $y_2(t) = te^{-3t}$. What is the limit of these solutions as $t \to \infty$?

4. Verify that $y = \sin 3t + 2\cos 3t$ is a solution to the initial value problem

$$2y'' + 18y = 0\,; \qquad y(0) = 2\,, \qquad y'(0) = 3\,.$$

Find the maximum of $|y(t)|$ for $-\infty < t < \infty$.

5. Verify that the exponentially damped sinusoid $y(t) = e^{-3t}\sin(\sqrt{3}\,t)$ is a solution to equation (3) if $F_{\text{ext}}(t) = 0$, $m = 1$, $b = 6$, and $k = 12$. What is the limit of this solution as $t \to \infty$?

6. An external force $F(t) = 2\cos 2t$ is applied to a mass–spring system with $m = 1$, $b = 0$, and $k = 4$, which is initially at rest; i.e., $y(0) = 0$, $y'(0) = 0$. Verify that $y(t) = \frac{1}{2}t\sin 2t$ gives the motion of this spring. What will eventually (as t increases) happen to the spring?

In Problems 7–9, find a synchronous solution of the form $A\cos\Omega t + B\sin\Omega t$ to the given forced oscillator equation using the method of Example 4 to solve for A and B.

7. $y'' + 2y' + 4y = 5\sin 3t, \quad \Omega = 3$

8. $y'' + 2y' + 5y = -50\sin 5t, \quad \Omega = 5$

9. $y'' + 2y' + 4y = 6\cos 2t + 8\sin 2t, \quad \Omega = 2$

10. Undamped oscillators that are driven at resonance have unusual (and nonphysical) solutions.

(a) To investigate this, find the synchronous solution $A\cos\Omega t + B\sin\Omega t$ to the generic forced oscillator equation

$$(7) \qquad my'' + by' + ky = \cos\Omega t\,.$$

(b) Sketch graphs of the coefficients A and B, as functions of Ω, for $m = 1$, $b = 0.1$, and $k = 25$.

(c) Now set $b = 0$ in your formulas for A and B and resketch the graphs in part (b), with $m = 1$, and $k = 25$. What happens at $\Omega = 5$? Notice that the amplitudes of the synchronous solutions grow without bound as Ω approaches 5.

(d) Show directly, by substituting the form $A\cos\Omega t + B\sin\Omega t$ into equation (7), that when $b = 0$ there are *no* synchronous solutions if $\Omega = \sqrt{k/m}$.

(e) Verify that $(2m\Omega)^{-1}t\sin\Omega t$ solves equation (7) when $b = 0$ and $\Omega = \sqrt{k/m}$. Notice that this *non*synchronous solution grows in time, without bound.

Clearly one cannot neglect damping in analyzing an oscillator forced at resonance, because otherwise the solutions, as shown in part (e), are nonphysical. This behavior will be studied later in this chapter.

4.2 Homogeneous Linear Equations: The General Solution

We begin our study of the linear second-order constant-coefficient differential equation

$$(1) \qquad ay'' + by' + cy = f(t) \qquad (a \neq 0)$$

with the special case where the function $f(t)$ is zero:

$$(2) \qquad \boldsymbol{ay'' + by' + cy = 0\,.}$$

This case arises when we consider mass–spring oscillators vibrating freely—that is, without external forces applied. Equation (2) is called the *homogeneous form* of equation (1); $f(t)$ is the “nonhomogeneity” in (1). (This nomenclature is not related to the way we used the term for first-order equations in Section 2.6.)

A look at equation (2) tells us that a solution of (2) must have the property that its second derivative is expressible as a linear combination of its first and zeroth derivatives.† This suggests that we try to find a solution of the form $y = e^{rt}$, since derivatives of e^{rt} are just constants times e^{rt}. If we substitute $y = e^{rt}$ into (2), we obtain

$$\begin{aligned} ar^2e^{rt} + bre^{rt} + ce^{rt} &= 0\,, \\ e^{rt}(ar^2 + br + c) &= 0\,. \end{aligned}$$

†The zeroth derivative of a function is the function itself.

Because e^{rt} is never zero, we can divide by it to obtain

(3) $$ar^2 + br + c = 0 .$$

Consequently, $y = e^{rt}$ is a solution to (2) if and only if r satisfies equation (3). Equation (3) is called the **auxiliary equation** (also known as the **characteristic equation**) associated with the homogeneous equation (2).

Now the auxiliary equation is just a quadratic, and its roots are

$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} .$$

When the discriminant, $b^2 - 4ac$, is positive, the roots r_1 and r_2 are real and distinct. If $b^2 - 4ac = 0$, the roots are real and equal. And when $b^2 - 4ac < 0$, the roots are complex conjugate numbers. We consider the first two cases in this section; the complex case is deferred to Section 4.3.

Example 1 Find a pair of solutions to

(4) $$y'' + 5y' - 6y = 0 .$$

Solution The auxiliary equation associated with (4) is

$$r^2 + 5r - 6 = (r - 1)(r + 6) = 0 ,$$

which has the roots $r_1 = 1$, $r_2 = -6$. Thus, e^t and e^{-6t} are solutions. ◆

Notice that the identically zero function, $y(t) \equiv 0$, is always a solution to (2). Furthermore, when we have a pair of solutions $y_1(t)$ and $y_2(t)$ to this equation, as in Example 1, we can construct an infinite number of other solutions by forming linear combinations:

(5) $$y(t) = c_1 y_1(t) + c_2 y_2(t)$$

for any choice of the constants c_1 and c_2. The fact that (5) is a solution to (2) can be seen by direct substitution and rearrangement:

$$\begin{aligned} ay'' + by' + cy &= a(c_1y_1 + c_2y_2)'' + b(c_1y_1 + c_2y_2)' + c(c_1y_1 + c_2y_2) \\ &= a(c_1y_1'' + c_2y_2'') + b(c_1y_1' + c_2y_2') + c(c_1y_1 + c_2y_2) \\ &= c_1(ay_1'' + by_1' + cy_1) + c_2(ay_2'' + by_2' + cy_2) \\ &= 0 + 0 . \end{aligned}$$

The two "degrees of freedom" c_1 and c_2 in the combination (5) suggest that solutions to the differential equation (2) can be found meeting additional conditions, such as the initial conditions for the first-order equations in Chapter 1. But the presence of c_1 *and* c_2 leads one to anticipate that *two* such conditions, rather than just one, can be imposed. This is consistent with the mass–spring interpretation of equation (2), since predicting the motion of a mechanical system requires knowledge not only of the forces but also of the initial position $y(0)$ and velocity $y'(0)$ of the mass. A typical *initial value problem* for these second-order equations is given in the following example.

Example 2 Solve the initial value problem

(6) $$y'' + 2y' - y = 0 ; \qquad y(0) = 0 , \qquad y'(0) = -1 .$$

Solution We will first find a pair of solutions as in the previous example. Then we will adjust the constants c_1 and c_2 in (5) to obtain a solution that matches the initial conditions on $y(0)$ and $y'(0)$. The auxiliary equation is

$$r^2 + 2r - 1 = 0 .$$

Using the quadratic formula, we find that the roots of this equation are

$$r_1 = -1 + \sqrt{2} \qquad \text{and} \qquad r_2 = -1 - \sqrt{2} .$$

Consequently, the given differential equation has solutions of the form

$$y(t) = c_1 e^{(-1+\sqrt{2})t} + c_2 e^{(-1-\sqrt{2})t} . \tag{7}$$

To find the specific solution that satisfies the initial conditions given in (6), we first differentiate y as given in (7), then plug y and y' into the initial conditions of (6). This gives

$$y(0) = c_1 e^0 + c_2 e^0 ,$$
$$y'(0) = (-1 + \sqrt{2})c_1 e^0 + (-1 - \sqrt{2})c_2 e^0 ,$$

or

$$0 = c_1 + c_2 ,$$
$$-1 = (-1 + \sqrt{2})c_1 + (-1 - \sqrt{2})c_2 .$$

Solving this system yields $c_1 = -\sqrt{2}/4$ and $c_2 = \sqrt{2}/4$. Thus,

$$y(t) = -\frac{\sqrt{2}}{4} e^{(-1+\sqrt{2})t} + \frac{\sqrt{2}}{4} e^{(-1-\sqrt{2})t}$$

is the desired solution. ◆

To gain more insight into the significance of the two-parameter solution form (5), we need to look at some of the properties of the second-order equation (2). First of all, there is an existence-and-uniqueness theorem for solutions to (2); it is somewhat like the corresponding Theorem 1 in Section 1.2 for first-order equations but updated to reflect the fact that *two* initial conditions are appropriate for *second*-order equations. As motivation for the theorem, suppose the differential equation (2) were *really easy*, with $b = 0$ and $c = 0$. Then $y'' = 0$ would merely say that the graph of $y(t)$ is simply a *straight line*, so it is uniquely determined by specifying a point on the line,

$$y(t_0) = Y_0 , \tag{8}$$

and the slope of the line,

$$y'(t_0) = Y_1 . \tag{9}$$

Theorem 1 states that conditions (8) and (9) suffice to determine the solution uniquely for the more general equation (2).

Existence and Uniqueness: Homogeneous Case

Theorem 1. For any real numbers $a\ (\neq 0)$, b, c, t_0, Y_0, and Y_1, there exists a unique solution to the initial value problem

$$ay'' + by' + cy = 0 ; \qquad y(t_0) = Y_0 , \qquad y'(t_0) = Y_1 . \tag{10}$$

The solution is valid for all t in $(-\infty, +\infty)$.

Note in particular that if a solution $y(t)$ and its derivative vanish simultaneously at a point t_0 (i.e., $Y_0 = Y_1 = 0$), then $y(t)$ must be the identically zero solution.

In this section and the next, we will construct explicit solutions to (10), so the question of *existence* of a solution is not really an issue. It is extremely valuable to know, however, that the solution is *unique*. The proof of uniqueness is rather different from anything else in this chapter, so we defer it to Chapter 13.[†]

Now we want to use this theorem to show that, given two solutions $y_1(t)$ and $y_2(t)$ to equation (2), we can always find values of c_1 and c_2 so that $c_1y_1(t) + c_2y_2(t)$ meets specified initial conditions in (10) and therefore is the (unique) solution to the initial value problem. But we need to be a little more precise; if, for example, $y_2(t)$ is simply the identically zero solution, then $c_1y_1(t) + c_2y_2(t) = c_1y_1(t)$ actually has only *one* constant and cannot be expected to satisfy *two* conditions. Furthermore, if $y_2(t)$ is simply a constant multiple of $y_1(t)$—say, $y_2(t) = \kappa y_1(t)$—then again $c_1y_1(t) + c_2y_2(t) = (c_1 + \kappa c_2)y_1(t) = Cy_1(t)$ actually has only one constant. The condition we need is *linear independence*.

Linear Independence of Two Functions

Definition 1. A pair of functions $y_1(t)$ and $y_2(t)$ is said to be **linearly independent on the interval *I*** if and only if neither of them is a constant multiple of the other on all of I.[††] We say that y_1 and y_2 are **linearly dependent on *I*** if one of them *is* a constant multiple of the other on all of I.

Representation of Solutions to Initial Value Problem

Theorem 2. If $y_1(t)$ and $y_2(t)$ are any two solutions to the differential equation (2) that are linearly independent on $(-\infty, \infty)$, then unique constants c_1 and c_2 can always be found so that $c_1y_1(t) + c_2y_2(t)$ satisfies the initial value problem (10) on $(-\infty, \infty)$.

The proof of Theorem 2 will be easy once we establish the following technical lemma.

A Condition for Linear Dependence of Solutions

Lemma 1. For any real numbers $a\ (\neq 0)$, b, and c, if $y_1(t)$ and $y_2(t)$ are any two solutions to the differential equation (2) on $(-\infty, \infty)$ and if the equality

$$y_1(\tau)y_2'(\tau) - y_1'(\tau)y_2(\tau) = 0 \qquad \textbf{(11)}$$

holds at *any* point τ, then y_1 and y_2 are linearly dependent on $(-\infty, \infty)$. (The expression on the left-hand side of (11) is called the *Wronskian* of y_1 and y_2 at the point τ; see Problem 34 on page 164.)

[†]All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems,* 7th ed.

[††]This definition will be generalized to three or more functions in Problem 35 and Chapter 6.

Proof of Lemma 1. *Case 1*. If $y_1(\tau) \neq 0$, then let κ equal $y_2(\tau)/y_1(\tau)$ and consider the solution to (2) given by $y(t) = \kappa y_1(t)$. It satisfies the same "initial conditions" at $t = \tau$ as does $y_2(t)$:

$$y(\tau) = \frac{y_2(\tau)}{y_1(\tau)} y_1(\tau) = y_2(\tau) \, ; \qquad y'(\tau) = \frac{y_2(\tau)}{y_1(\tau)} y_1'(\tau) = y_2'(\tau) \, ,$$

where the last equality follows from (11). By uniqueness, $y_2(t)$ must be the same function as $\kappa y_1(t)$ on I.

Case 2. If $y_1(\tau) = 0$ but $y_1'(\tau) \neq 0$, then (11) implies $y_2(\tau) = 0$. Let $\kappa = y_2'(\tau)/y_1'(\tau)$. Then the solution to (2) given by $y(t) = \kappa y_1(t)$ (again) satisfies the same "initial conditions" at $t = \tau$ as does $y_2(t)$:

$$y(\tau) = \frac{y_2'(\tau)}{y_1'(\tau)} y_1(\tau) = 0 = y_2(\tau) \, ; \qquad y'(\tau) = \frac{y_2'(\tau)}{y_1'(\tau)} y_1'(\tau) = y_2'(\tau) \, .$$

By uniqueness, then, $y_2(t) = \kappa y_1(t)$ on I.

Case 3. If $y_1(\tau) = y_1'(\tau) = 0$, then $y_1(t)$ is a solution to the differential equation (2) satisfying the initial conditions $y_1(\tau) = y_1'(\tau) = 0$; but $y(t) \equiv 0$ is the *unique* solution to this initial value problem. Thus, $y_1(t) \equiv 0$ $[$and is a constant multiple of $y_2(t)]$. ◆

Proof of Theorem 2. We already know that $y(t) = c_1 y_1(t) + c_2 y_2(t)$ is a solution to (2); we must show that c_1 and c_2 can be chosen so that

$$y(t_0) = c_1 y_1(t_0) + c_2 y_2(t_0) = Y_0$$

and

$$y'(t_0) = c_1 y_1'(t_0) + c_2 y_2'(t_0) = Y_1 \, .$$

But simple algebra shows these equations have the solution†

$$c_1 = \frac{Y_0 y_2'(t_0) - Y_1 y_2(t_0)}{y_1(t_0) y_2'(t_0) - y_1'(t_0) y_2(t_0)} \quad \text{and} \quad c_2 = \frac{Y_1 y_1(t_0) - Y_0 y_1'(t_0)}{y_1(t_0) y_2'(t_0) - y_1'(t_0) y_2(t_0)}$$

as long as the denominator is nonzero, and the technical lemma assures us that this condition is met. ◆

Now we can honestly say that if y_1 and y_2 are linearly independent solutions to (2) on $(-\infty, +\infty)$, then (5) is a **general solution,** since *any* solution $y_g(t)$ of (2) can be expressed in this form; simply pick c_1 and c_2 so that $c_1 y_1 + c_2 y_2$ matches the value and the derivative of y_g *at any point*. By uniqueness, $c_1 y_1 + c_2 y_2$ and y_g have to be the same function. See Figure 4.6 on page 162.

How do we *find* a general solution for the differential equation (2)? We already know the answer if the roots of the auxiliary equation (3) are real and distinct because clearly $y_1(t) = e^{r_1 t}$ is not a constant multiple of $y_2(t) = e^{r_2 t}$ if $r_1 \neq r_2$.

Distinct Real Roots

If the auxiliary equation (3) has distinct real roots r_1 and r_2, then both $y_1(t) = e^{r_1 t}$ and $y_2(t) = e^{r_2 t}$ are solutions to (2) and $y(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t}$ is a general solution.

†To solve for c_1, for example, multiply the first equation by $y_2'(t_0)$ and the second by $y_2(t_0)$ and subtract.

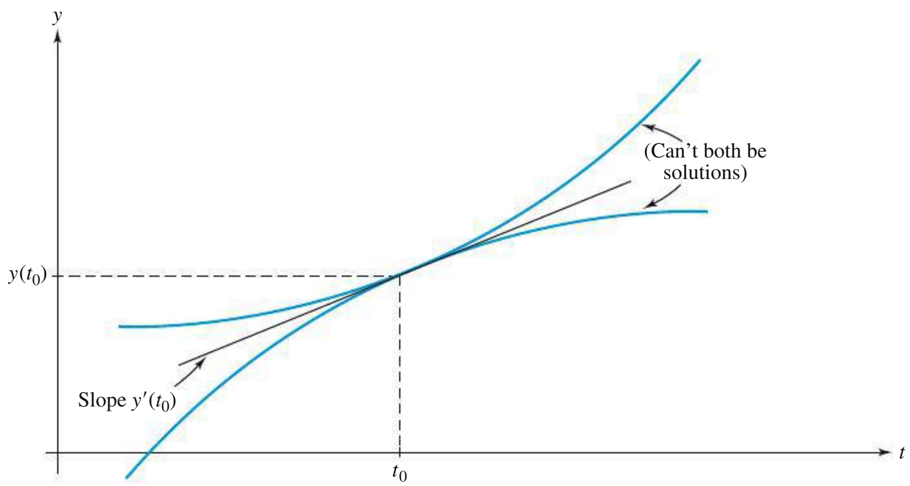


Figure 4.6 $y(t_0), y'(t_0)$ determine a *unique* solution.

When the roots of the auxiliary equation are equal, we only get one nontrivial solution, $y_1 = e^{rt}$. To satisfy *two* initial conditions, $y(t_0)$ and $y'(t_0)$, then we will need a second, linearly independent solution. The following rule is the key to finding a second solution.

Repeated Root

If the auxiliary equation (3) has a repeated root r, then both $y_1(t) = e^{rt}$ and $y_2(t) = te^{rt}$ are solutions to (2), and $y(t) = c_1 e^{rt} + c_2 te^{rt}$ is a general solution.

We illustrate this result before giving its proof.

Example 3 Find a solution to the initial value problem

$$\text{(12)} \qquad y'' + 4y' + 4y = 0 ; \qquad y(0) = 1 , \qquad y'(0) = 3 .$$

Solution The auxiliary equation for (12) is

$$r^2 + 4r + 4 = (r+2)^2 = 0 .$$

Because $r = -2$ is a double root, the rule says that (12) has solutions $y_1 = e^{-2t}$ and $y_2 = te^{-2t}$. Let's confirm that $y_2(t)$ is a solution:

$$\begin{aligned} y_2(t) &= te^{-2t} , \\ y_2'(t) &= e^{-2t} - 2te^{-2t} , \\ y_2''(t) &= -2e^{-2t} - 2e^{-2t} + 4te^{-2t} = -4e^{-2t} + 4te^{-2t} , \\ y_2'' + 4y_2' + 4y_2 &= -4e^{-2t} + 4te^{-2t} + 4(e^{-2t} - 2te^{-2t}) + 4te^{-2t} = 0 . \end{aligned}$$

Further observe that e^{-2t} and te^{-2t} are linearly independent since neither is a constant multiple of the other on $(-\infty, \infty)$. Finally, we insert the general solution $y(t) = c_1 e^{-2t} + c_2 te^{-2t}$ into the initial conditions,

$$\begin{aligned} y(0) &= c_1 e^0 + c_2(0)e^0 = 1 , \\ y'(0) &= -2c_1 e^0 + c_2 e^0 - 2c_2(0)e^0 = 3 , \end{aligned}$$

and solve to find $c_1 = 1$, $c_2 = 5$. Thus $y = e^{-2t} + 5te^{-2t}$ is the desired solution. ◆

Why is it that $y_2(t) = te^{rt}$ is a solution to the differential equation (2) when r is a double root (and not otherwise)? In later chapters we will see a theoretical justification of this rule in very general circumstances; for present purposes, though, simply note what happens if we substitute y_2 into the differential equation (2):

$$\begin{aligned} y_2(t) &= te^{rt} , \\ y_2'(t) &= e^{rt} + rte^{rt} , \\ y''_2(t) &= re^{rt} + re^{rt} + r^2te^{rt} = 2re^{rt} + r^2te^{rt} , \\ ay_2'' + by_2' + cy_2 &= [2ar + b]e^{rt} + [ar^2 + br + c]te^{rt} . \end{aligned}$$

Now if r is a root of the auxiliary equation (3), the expression in the second brackets is zero. However, *if r is a double root, the expression in the first brackets is zero also*:

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm (0)}{2a} ; \tag{13}$$

hence, $2ar + b = 0$ for a double root. In such a case, then, y_2 is a solution.

The method we have described for solving homogeneous linear second-order equations with constant coefficients applies to any order (even first-order) homogeneous linear equations with constant coefficients. We give a detailed treatment of such higher-order equations in Chapter 6. For now, we will be content to illustrate the method by means of an example. We remark briefly that a homogeneous linear nth-order equation has a general solution of the form

$$y(t) = c_1y_1(t) + c_2y_2(t) + \cdots + c_ny_n(t) ,$$

where the individual solutions $y_i(t)$ are "linearly independent." By this we mean that no y_i is expressible as a linear combination of the others; see Problem 35 on page 164.

Example 4 Find a general solution to

$$y''' + 3y'' - y' - 3y = 0 . \tag{14}$$

Solution If we try to find solutions of the form $y = e^{rt}$, then, as with second-order equations, we are led to finding roots of the auxiliary equation

$$r^3 + 3r^2 - r - 3 = 0 . \tag{15}$$

We observe that $r = 1$ is a root of the above equation, and dividing the polynomial on the left-hand side of (15) by $r - 1$ leads to the factorization

$$(r - 1)(r^2 + 4r + 3) = (r - 1)(r + 1)(r + 3) = 0 .$$

Hence, the roots of the auxiliary equation are 1, -1, and -3, and so three solutions of (14) are e^t, e^{-t}, and e^{-3t}. The linear independence of these three exponential functions is proved in Problem 36. A general solution to (14) is then

$$y(t) = c_1e^t + c_2e^{-t} + c_3e^{-3t} . \ ◆ \tag{16}$$

So far we have seen only exponential solutions to the linear second-order constant coefficient equation. You may wonder where the vibratory solutions that govern mass–spring oscillators are. In the next section, it will be seen that they arise when the solutions to the auxiliary equation are complex.

4.2 EXERCISES

In Problems 1–12, find a general solution to the given differential equation.

1. $2y'' + 7y' - 4y = 0$
2. $y'' + 6y' + 9y = 0$
3. $y'' + 5y' + 6y = 0$
4. $y'' - y' - 2y = 0$
5. $y'' + 8y' + 16y = 0$
6. $y'' - 5y' + 6y = 0$
7. $6y'' + y' - 2y = 0$
8. $z'' + z' - z = 0$
9. $4y'' - 4y' + y = 0$
10. $y'' - y' - 11y = 0$
11. $4w'' + 20w' + 25w = 0$
12. $3y'' + 11y' - 7y = 0$

In Problems 13–20, solve the given initial value problem.

13. $y'' + 2y' - 8y = 0; \quad y(0) = 3, \quad y'(0) = -12$
14. $y'' + y' = 0; \quad y(0) = 2, \quad y'(0) = 1$
15. $y'' - 4y' + 3y = 0; \quad y(0) = 1, \quad y'(0) = 1/3$
16. $y'' - 4y' - 5y = 0; \quad y(-1) = 3, \quad y'(-1) = 9$
17. $y'' - 6y' + 9y = 0; \quad y(0) = 2, \quad y'(0) = 25/3$
18. $z'' - 2z' - 2z = 0; \quad z(0) = 0, \quad z'(0) = 3$
19. $y'' + 2y' + y = 0; \quad y(0) = 1, \quad y'(0) = -3$
20. $y'' - 4y' + 4y = 0; \quad y(1) = 1, \quad y'(1) = 1$

21. First-Order Constant-Coefficient Equations.
(a) Substituting $y = e^{rt}$, find the auxiliary equation for the first-order linear equation
$$ay' + by = 0,$$
where a and b are constants with $a \neq 0$.
(b) Use the result of part (a) to find the general solution.

In Problems 22–25, use the method described in Problem 21 to find a general solution to the given equation.

22. $3y' - 7y = 0$
23. $5y' + 4y = 0$
24. $3z' + 11z = 0$
25. $6w' - 13w = 0$

26. Boundary Value Problems. When the values of a solution to a differential equation are specified at two different points, these conditions are called **boundary conditions.** (In contrast, initial conditions specify the values of a function and its derivative at the same point.) The purpose of this exercise is to show that for boundary value problems there is no existence–uniqueness theorem that is analogous to Theorem 1. Given that every solution to

$$y'' + y = 0 \tag{17}$$

is of the form
$$y(t) = c_1 \cos t + c_2 \sin t,$$
where c_1 and c_2 are arbitrary constants, show that
(a) There is a unique solution to (17) that satisfies the boundary conditions $y(0) = 2$ and $y(\pi/2) = 0$.
(b) There is no solution to (17) that satisfies $y(0) = 2$ and $y(\pi) = 0$.
(c) There are infinitely many solutions to (17) that satisfy $y(0) = 2$ and $y(\pi) = -2$.

In Problems 27–32, use Definition 1 to determine whether the functions y_1 and y_2 are linearly dependent on the interval $(0, 1)$.

27. $y_1(t) = \cos t \sin t, \quad y_2(t) = \sin 2t$
28. $y_1(t) = e^{3t}, \quad y_2(t) = e^{-4t}$
29. $y_1(t) = te^{2t}, \quad y_2(t) = e^{2t}$
30. $y_1(t) = t^2 \cos(\ln t), \quad y_2(t) = t^2 \sin(\ln t)$
31. $y_1(t) = \tan^2 t - \sec^2 t, \quad y_2(t) \equiv 3$
32. $y_1(t) \equiv 0, \quad y_2(t) = e^t$

33. Explain why two functions are linearly dependent on an interval I if and only if there exist constants c_1 and c_2, *not both zero*, such that
$$c_1 y_1(t) + c_2 y_2(t) = 0 \quad \text{for all } t \text{ in } I.$$

34. Wronskian. For any two differentiable functions y_1 and y_2, the function

$$W[y_1, y_2](t) = y_1(t)y_2'(t) - y_1'(t)y_2(t) \tag{18}$$

is called the *Wronskian*[†] of y_1 and y_2. This function plays a crucial role in the proof of Theorem 2.
(a) Show that $W[y_1, y_2]$ can be conveniently expressed as the 2×2 determinant
$$W[y_1, y_2](t) = \begin{vmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{vmatrix}.$$
(b) Let $y_1(t), y_2(t)$ be a pair of solutions to the homogeneous equation $ay'' + by' + cy = 0$ (with $a \neq 0$) on an open interval I. Prove that $y_1(t)$ and $y_2(t)$ are linearly independent on I if and only if their Wronskian is never zero on I. [*Hint:* This is just a reformulation of Lemma 1.]
(c) Show that if $y_1(t)$ and $y_2(t)$ are any two differentiable functions that are linearly dependent on I, then their Wronskian is identically zero on I.

35. Linear Dependence of Three Functions. Three functions $y_1(t)$, $y_2(t)$, and $y_3(t)$ are said to be linearly dependent on an interval I if, on I, at least one of these functions is a linear combination of the remaining two [e.g., if $y_1(t) = c_1 y_2(t) + c_2 y_3(t)$]. Equivalently (compare Problem 33), y_1, y_2, and y_3 are linearly dependent on I if there exist constants C_1, C_2, and C_3, *not all zero*, such that
$$C_1 y_1(t) + C_2 y_2(t) + C_3 y_3(t) = 0 \quad \text{for all } t \text{ in } I.$$
Otherwise, we say that these functions are linearly independent on I.

[†] *Historical Footnote:* The Wronskian was named after the Polish mathematician H. Wronski (1778–1863).

For each of the following, determine whether the given three functions are linearly dependent or linearly independent on $(-\infty, \infty)$:

(a) $y_1(t) = 1\,,\quad y_2(t) = t\,,\quad y_3(t) = t^2\,.$

(b) $y_1(t) = -3\,,\quad y_2(t) = 5\sin^2 t\,,\quad y_3(t) = \cos^2 t\,.$

(c) $y_1(t) = e^t\,,\quad y_2(t) = te^t\,,\quad y_3(t) = t^2e^t\,.$

(d) $y_1(t) = e^t\,,\quad y_2(t) = e^{-t}\,,\quad y_3(t) = \cosh t\,.$

36. Using the definition in Problem 35, prove that if r_1, r_2, and r_3 are distinct real numbers, then the functions e^{r_1t}, e^{r_2t}, and e^{r_3t} are linearly independent on $(-\infty, \infty)$. [*Hint*: Assume to the contrary that, say, $e^{r_1t} = c_1e^{r_2t} + c_2e^{r_3t}$ for all t. Divide by e^{r_2t} to get $e^{(r_1-r_2)t} = c_1 + c_2e^{(r_3-r_2)t}$ and then differentiate to deduce that $e^{(r_1-r_2)t}$ and $e^{(r_3-r_2)t}$ are linearly dependent, which is a contradiction. (Why?)]

In Problems 37–41, find three linearly independent solutions (see Problem 35) of the given third-order differential equation and write a general solution as an arbitrary linear combination of these.

37. $y''' + y'' - 6y' + 4y = 0$

38. $y''' - 6y'' - y' + 6y = 0$

39. $z''' + 2z'' - 4z' - 8z = 0$

40. $y''' - 7y'' + 7y' + 15y = 0$

41. $y''' + 3y'' - 4y' - 12y = 0$

42. (True or False): If f_1, f_2, f_3 are three functions defined on $(-\infty, \infty)$ that are *pairwise* linearly independent on $(-\infty, \infty)$, then f_1, f_2, f_3 form a linearly independent set on $(-\infty, \infty)$. Justify your answer.

43. Solve the initial value problem:

$$y''' - y' = 0\,;\qquad y(0) = 2\,,$$
$$y'(0) = 3\,,\qquad y''(0) = -1\,.$$

44. Solve the initial value problem:

$$y''' - 2y'' - y' + 2y = 0\,;$$
$$y(0) = 2\,,\quad y'(0) = 3\,,\quad y''(0) = 5\,.$$

45. By using Newton's method or some other numerical procedure to approximate the roots of the auxiliary equation, find general solutions to the following equations:

(a) $3y''' + 18y'' + 13y' - 19y = 0\,.$

(b) $y^{\text{iv}} - 5y'' + 5y = 0\,.$

(c) $y^{\text{v}} - 3y^{\text{iv}} - 5y''' + 15y'' + 4y' - 12y = 0\,.$

46. One way to define hyperbolic functions is by means of differential equations. Consider the equation $y'' - y = 0$. The *hyperbolic cosine*, $\cosh t$, is defined as the solution of this equation subject to the initial values: $y(0) = 1$ and $y'(0) = 0$. The *hyperbolic sine*, $\sinh t$, is defined as the solution of this equation subject to the initial values: $y(0) = 0$ and $y'(0) = 1$.

(a) Solve these initial value problems to derive explicit formulas for $\cosh t$ and $\sinh t$. Also show that $\frac{d}{dt}\cosh t = \sinh t$ and $\frac{d}{dt}\sinh t = \cosh t$.

(b) Prove that a general solution of the equation $y'' - y = 0$ is given by $y = c_1 \cosh t + c_2 \sinh t$.

(c) Suppose a, b, and c are given constants for which $ar^2 + br + c = 0$ has two distinct real roots. If the two roots are expressed in the form $\alpha - \beta$ and $\alpha + \beta$, show that a general solution of the equation $ay'' + by' + cy = 0$ is $y = c_1e^{\alpha t}\cosh(\beta t) + c_2e^{\alpha t}\sinh(\beta t)$.

(d) Use the result of part (c) to solve the initial value problem: $y'' + y' - 6y = 0$, $y(0) = 2$, $y'(0) = -17/2$.

4.3 Auxiliary Equations with Complex Roots

The *simple harmonic equation* $y'' + y = 0$, so called because of its relation to the fundamental vibration of a musical tone, has as solutions $y_1(t) = \cos t$ and $y_2(t) = \sin t$. Notice, however, that the auxiliary equation associated with the harmonic equation is $r^2 + 1 = 0$, which has imaginary roots $r = \pm i$, where i denotes $\sqrt{-1}$.† In the previous section, we expressed the solutions to a linear second-order equation with constant coefficients in terms of exponential functions. It would appear, then, that one might be able to attribute a meaning to the forms e^{it} and e^{-it} and that these "functions" should be related to $\cos t$ and $\sin t$. This matchup is accomplished by Euler's formula, which is discussed in this section.

When $b^2 - 4ac < 0$, the roots of the auxiliary equation

$$ar^2 + br + c = 0 \tag{1}$$

†Electrical engineers frequently use the symbol j to denote $\sqrt{-1}$.

associated with the homogeneous equation

$$ay'' + by' + cy = 0 \tag{2}$$

are the complex conjugate numbers

$$r_1 = \alpha + i\beta \quad \text{and} \quad r_2 = \alpha - i\beta \quad (i = \sqrt{-1}) ,$$

where α, β are the real numbers

$$\alpha = -\frac{b}{2a} \quad \text{and} \quad \beta = \frac{\sqrt{4ac - b^2}}{2a} . \tag{3}$$

As in the previous section, we would like to assert that the functions $e^{r_1 t}$ and $e^{r_2 t}$ are solutions to the equation (2). This is in fact the case, but before we can proceed, we need to address some fundamental questions. For example, if $r_1 = \alpha + i\beta$ is a complex number, what do we mean by the expression $e^{(\alpha+i\beta)t}$? If we assume that the law of exponents applies to complex numbers, then

$$e^{(\alpha+i\beta)t} = e^{\alpha t + i\beta t} = e^{\alpha t} e^{i\beta t} . \tag{4}$$

We now need only clarify the meaning of $e^{i\beta t}$.

For this purpose, let's assume that the Maclaurin series for e^z is the same for complex numbers z as it is for real numbers. Observing that $i^2 = -1$, then for θ real we have

$$\begin{aligned} e^{i\theta} &= 1 + (i\theta) + \frac{(i\theta)^2}{2!} + \cdots + \frac{(i\theta)^n}{n!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots\right) . \end{aligned}$$

Now recall the Maclaurin series for $\cos\theta$ and $\sin\theta$:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \cdots ,$$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots .$$

Recognizing these expansions in the proposed series for $e^{i\theta}$, we make the identification

$$\boldsymbol{e^{i\theta} = \cos\theta + i\sin\theta ,} \tag{5}$$

which is known as **Euler's formula.**†

When Euler's formula (with $\theta = \beta t$) is used in equation (4), we find

$$e^{(\alpha+i\beta)t} = e^{\alpha t}(\cos\beta t + i\sin\beta t) , \tag{6}$$

which expresses the complex function $e^{(\alpha+i\beta)t}$ in terms of familiar real functions. Having made sense out of $e^{(\alpha+i\beta)t}$, we can now show (see Problem 30 on page 172) that

$$\frac{d}{dt}e^{(\alpha+i\beta)t} = (\alpha + i\beta)e^{(\alpha+i\beta)t} , \tag{7}$$

†*Historical Footnote:* This formula first appeared in Leonhard Euler's monumental two-volume *Introductio in Analysin Infinitorum* (1748).

and, with the choices of α and β as given in (3), the complex function $e^{(\alpha+i\beta)t}$ is indeed a solution to equation (2), as is $e^{(\alpha-i\beta)t}$, and a general solution is given by

$$\begin{aligned} y(t) &= c_1 e^{(\alpha+i\beta)t} + c_2 e^{(\alpha-i\beta)t} \\ &= c_1 e^{\alpha t}(\cos\beta t + i\sin\beta t) + c_2 e^{\alpha t}(\cos\beta t - i\sin\beta t)\ . \end{aligned} \tag{8}$$

Example 1 shows that in general the constants c_1 and c_2 that go into (8), for a specific initial value problem, are complex.

Example 1 Use the general solution (8) to solve the initial value problem

$$y'' + 2y' + 2y = 0\,; \qquad y(0) = 0, \qquad y'(0) = 2\ .$$

Solution The auxiliary equation is $r^2 + 2r + 2 = 0$, which has roots

$$r = \frac{-2\sqrt{4-8}}{2} = -1 \pm i\,.$$

Hence, with $\alpha = -1$, $\beta = 1$, a general solution is given by

$$y(t) = c_1 e^{-t}(\cos t + i\sin t) + c_2 e^{-t}(\cos t - i\sin t)\ .$$

For initial conditions we have

$$\begin{aligned} y(0) &= c_1 e^0(\cos 0 + i\sin 0) + c_2 e^0(\cos 0 - i\sin 0) = c_1 + c_2 = 0\,, \\ y'(0) &= -c_1 e^0(\cos 0 + i\sin 0) + c_1 e^0(-\sin 0 + i\cos 0) \\ &\quad -c_2 e^0(\cos 0 - i\sin 0) + c_2 e^0(-\sin 0 - i\cos 0) \\ &= (-1+i)c_1 + (-1-i)c_2 \\ &= 2\,. \end{aligned}$$

As a result, $c_1 = -i$, $c_2 = i$, and $y(t) = -ie^{-t}(\cos t + i\sin t) + ie^{-t}(\cos t - i\sin t)$, or simply $2e^{-t}\sin t$. ◆

The final form of the answer to Example 1 suggests that we should seek an alternative pair of solutions to the differential equation (2) that don't require complex arithmetic, and we now turn to that task.

In general, if $z(t)$ is a complex-valued function of the real variable t, we can write $z(t) = u(t) + iv(t)$, where $u(t)$ and $v(t)$ are real-valued functions. The derivatives of $z(t)$ are then given by

$$\frac{dz}{dt} = \frac{du}{dt} + i\frac{dv}{dt}, \qquad \frac{d^2z}{dt^2} = \frac{d^2u}{dt^2} + i\frac{d^2v}{dt^2}\,.$$

With the following lemma, we show that the complex-valued solution $e^{(\alpha+i\beta)t}$ gives rise to two linearly independent *real-valued* solutions.

Real Solutions Derived from Complex Solutions

Lemma 2. Let $z(t) = u(t) + iv(t)$ be a solution to equation (2), where a, b, and c are real numbers. Then, the real part $u(t)$ and the imaginary part $v(t)$ are real-valued solutions of (2).[†]

[†]It will be clear from the proof that this property holds for any linear homogeneous differential equation having real-valued coefficients.

Proof. By assumption, $az'' + bz' + cz = 0$, and hence

$$a(u'' + iv'') + b(u' + iv') + c(u + iv) = 0\,,$$
$$(au'' + bu' + cu) + i(av'' + bv' + cv) = 0\,.$$

But a complex number is zero if and only if its real and imaginary parts are both zero. Thus, we must have

$$au'' + bu' + cu = 0 \quad \text{and} \quad av'' + bv' + cv = 0\,,$$

which means that both $u(t)$ and $v(t)$ are real-valued solutions of (2). ◆

When we apply Lemma 2 to the solution

$$e^{(\alpha+i\beta)t} = e^{\alpha t}\cos\beta t + ie^{\alpha t}\sin\beta t\,,$$

we obtain the following.

Complex Conjugate Roots

If the auxiliary equation has complex conjugate roots $\alpha \pm i\beta$, then two linearly independent solutions to (2) are

$$e^{\alpha t}\cos\beta t \quad \textbf{and} \quad e^{\alpha t}\sin\beta t\,,$$

and a general solution is

$$\textbf{(9)} \qquad \boldsymbol{y(t) = c_1 e^{\alpha t}\cos\beta t + c_2 e^{\alpha t}\sin\beta t}\,,$$

where c_1 and c_2 are arbitrary constants.

In the preceding discussion, we glossed over some important details concerning complex numbers and complex-valued functions. In particular, further analysis is required to justify the use of the law of exponents, Euler's formula, and even the fact that the derivative of e^{rt} is re^{rt} when r is a complex constant.† If you feel uneasy about our conclusions, we encourage you to substitute the expression in (9) into equation (2) to verify that it is, indeed, a solution.

You may also be wondering what would have happened if we had worked with the function $e^{(\alpha-i\beta)t}$ instead of $e^{(\alpha+i\beta)t}$. We leave it as an exercise to verify that $e^{(\alpha-i\beta)t}$ gives rise to the same general solution (9). Indeed, the sum of these two complex solutions, divided by two, gives the first real-valued solution, while their difference, divided by $2i$, gives the second.

Example 2 Find a general solution to

$$\textbf{(10)} \qquad y'' + 2y' + 4y = 0\,.$$

Solution The auxiliary equation is

$$r^2 + 2r + 4 = 0\,,$$

which has roots

$$r = \frac{-2 \pm \sqrt{4 - 16}}{2} = \frac{-2 \pm \sqrt{-12}}{2} = -1 \pm i\sqrt{3}\,.$$

†For a detailed treatment of these topics see, for example, *Fundamentals of Complex Analysis,* 3rd ed., by E. B. Saff and A. D. Snider (Prentice Hall, Upper Saddle River, New Jersey, 2003).

Hence, with $\alpha = -1$, $\beta = \sqrt{3}$, a general solution for (10) is

$$y(t) = c_1 e^{-t}\cos(\sqrt{3}\,t) + c_2 e^{-t}\sin(\sqrt{3}\,t)\,. \quad ◆$$

When the auxiliary equation has complex conjugate roots, the (real) solutions oscillate between positive and negative values. This type of behavior is observed in vibrating springs.

Example 3 In Section 4.1 we discussed the mechanics of the mass–spring oscillator (Figure 4.1, page 152), and we saw how Newton's second law implies that the position $y(t)$ of the mass m is governed by the second-order differential equation

$$my''(t) + by'(t) + ky(t) = 0\,, \tag{11}$$

where the terms are physically identified as

$$[\text{inertia}]y'' + [\text{damping}]y' + [\text{stiffness}]y = 0\,.$$

Determine the equation of motion for a spring system when $m = 36$ kg, $b = 12$ kg/sec (which is equivalent to 12 N-sec/m), $k = 37$ kg/sec^2, $y(0) = 0.7$ m, and $y'(0) = 0.1$ m/sec. After how many seconds will the mass first cross the equilibrium point?

Solution The equation of motion is given by $y(t)$, the solution of the initial value problem for the specified values of m, b, k, $y(0)$, and $y'(0)$. That is, we seek the solution to

$$36y'' + 12y' + 37y = 0\,; \qquad y(0) = 0.7\,,\ y'(0) = 0.1\,. \tag{12}$$

The auxiliary equation for (12) is

$$36r^2 + 12r + 37 = 0\,,$$

which has roots

$$r = \frac{-12 \pm \sqrt{144 - 4(36)(37)}}{72} = \frac{-12 \pm 12\sqrt{1-37}}{72} = -\frac{1}{6} \pm i\,.$$

Hence, with $\alpha = -1/6$, $\beta = 1$, the displacement $y(t)$ can be expressed in the form

$$y(t) = c_1 e^{-t/6}\cos t + c_2 e^{-t/6}\sin t\,. \tag{13}$$

We can find c_1 and c_2 by substituting $y(t)$ and $y'(t)$ into the initial conditions given in (12). Differentiating (13), we get a formula for $y'(t)$:

$$y'(t) = \left(-\frac{c_1}{6} + c_2\right)e^{-t/6}\cos t + \left(-c_1 - \frac{c_2}{6}\right)e^{-t/6}\sin t\,.$$

Substituting into the initial conditions now results in the system

$$\begin{aligned} c_1 &= 0.7\,, \\ -\frac{c_1}{6} + c_2 &= 0.1\,. \end{aligned}$$

Upon solving, we find $c_1 = 0.7$ and $c_2 = 1.3/6$. With these values, the equation of motion becomes

$$y(t) = 0.7e^{-t/6}\cos t + \frac{1.3}{6}e^{-t/6}\sin t\,.$$

To determine the times when the mass will cross the equilibrium point, we set $y(t) = 0$ and solve for t:

$$0 = y(t) = 0.7e^{-t/6}\cos t + \frac{1.3}{6}e^{-t/6}\sin t = (\cos t)\left(0.7e^{-t/6} + \frac{1.3}{6}e^{-t/6}\tan t\right).$$

But $y(t)$ is not zero for $\cos t = 0$, so only the zeros of the second factor are pertinent; that is,

$$\tan t = -\frac{4.2}{1.3}.$$

A quick glance at the graph of tan t reveals that the first positive t for which this is true lies between $\pi/2$ and π. Since the arctangent function takes only values between $-\pi/2$ and $\pi/2$, the appropriate adjustment is

$$t = \arctan\left(\frac{-4.2}{1.3}\right) + \pi \approx 1.87 \text{ seconds}. \quad ◆$$

From Example 3 we see that *any* second-order constant-coefficient differential equation $ay'' + by' + cy = 0$ can be interpreted as describing a mass–spring system with mass a, damping coefficient b, spring stiffness c, and displacement y, if these constants make sense physically; that is, if a is positive and b and c are nonnegative. From the discussion in Section 4.1, then, we expect on physical grounds to see damped oscillatory solutions in such a case. This is consistent with the display in equation (9). With $a = m$ and $c = k$, the exponential decay rate α equals $-b/(2m)$, and the angular frequency β equals $\sqrt{4mk - b^2}/(2m)$, by equation (3).

It is a little surprising, then, that the solutions to the equation $y'' + 4y' + 4y = 0$ do *not* oscillate; the general solution was shown in Example 3 of Section 4.2 (page 162) to be $c_1e^{-2t} + c_2te^{-2t}$. The physical significance of this is simply that when the damping coefficient b is too high, the resulting friction prevents the mass from oscillating. Rather than overshoot the spring's equilibrium point, it merely settles in lazily. This could happen if a light mass on a weak spring were submerged in a viscous fluid.

From the above formula for the oscillation frequency β, we can see that the oscillations will not occur for $b > \sqrt{4mk}$. This *overdamping* phenomenon is discussed in more detail in Section 4.9.

It is extremely enlightening to contemplate the predictions of the mass–spring analogy when the coefficients b and c in the equation $ay'' + by' + cy = 0$ are *negative*.

Example 4 Interpret the equation

$$36y'' - 12y' + 37y = 0 \tag{14}$$

in terms of the mass–spring system.

Solution Equation (14) is a minor alteration of equation (12) in Example 3; the auxiliary equation $36r^2 - 12r + 37$ has roots $r = (+)\frac{1}{6} \pm i$. Thus, its general solution becomes

$$y(t) = c_1e^{+t/6}\cos t + c_2e^{+t/6}\sin t. \tag{15}$$

Comparing equation (14) with the mass–spring model

$$[\textbf{inertia}]y'' + [\textbf{damping}]y' + [\textbf{stiffness}]y = 0, \tag{16}$$

we have to envision a *negative* damping coefficient $b = -12$, giving rise to a friction force $F_{\text{friction}} = -by'$ that *imparts* energy to the system instead of draining it. The increase in energy over time must then reveal itself in oscillations of ever-greater amplitude–precisely in accordance with formula (15), for which a typical graph is drawn in Figure 4.7. ◆

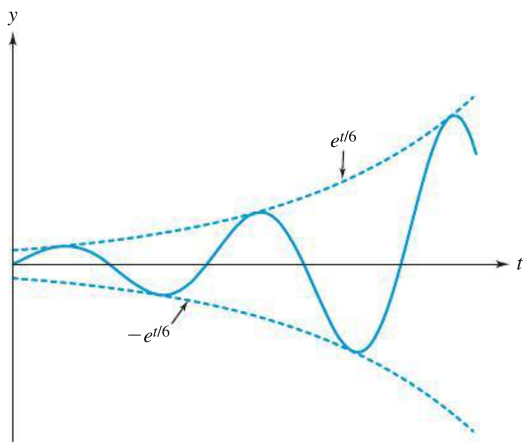


Figure 4.7 Solution graph for Example 4

Example 5 Interpret the equation

$$y'' + 5y' - 6y = 0 \tag{17}$$

in terms of the mass–spring system.

Solution Comparing the given equation with (16), we have to envision a spring with a *negative* stiffness $k = -6$. What does this mean? As the mass is moved away from the spring's equilibrium point, the spring *repels* the mass farther with a force $F_{\text{spring}} = -ky$ that intensifies as the displacement increases. Clearly the spring must "exile" the mass to (plus or minus) infinity, and we expect all solutions $y(t)$ to approach $\pm\infty$ as t increases (except for the equilibrium solution $y(t) \equiv 0$).

In fact, in Example 1 of Section 4.2, we showed the general solution to equation (17) to be

$$c_1 e^t + c_2 e^{-6t} . \tag{18}$$

Indeed, if we examine the solutions $y(t)$ that start with a unit displacement $y(0) = 1$ and velocity $y'(0) = v_0$, we find

$$y(t) = \frac{6 + v_0}{7} e^t + \frac{1 - v_0}{7} e^{-6t} , \tag{19}$$

and the plots in Figure 4.8 on page 172 confirm our prediction that all (nonequilibrium) solutions diverge—except for the one with $v_0 = -6$.

What is the physical significance of this isolated bounded solution? Evidently, if the mass is given an initial inwardly directed velocity of -6, it has barely enough energy to overcome the effect of the spring banishing it to $+\infty$ but not enough energy to cross the equilibrium point (and get pushed to $-\infty$). So it asymptotically approaches the (extremely delicate) equilibrium position $y = 0$. ◆

In Section 4.8, we will see that taking further liberties with the mass–spring interpretation enables us to predict qualitative features of more complicated equations.

Throughout this section we have assumed that the coefficients a, b, and c in the differential equation were real numbers. If we now allow them to be *complex* constants, then the roots

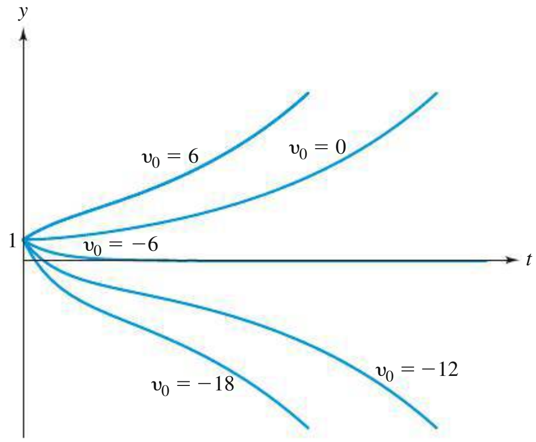


Figure 4.8 Solution graphs for Example 5

r_1, r_2 of the auxiliary equation (1) are, in general, also complex but not necessarily conjugates of each other. When $r_1 \neq r_2$, a general solution to equation (2) still has the form

$$y(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t},$$

but c_1 and c_2 are now arbitrary complex-valued constants, and we have to resort to the clumsy calculations of Example 1.

We also remark that a complex differential equation can be regarded as a system of two real differential equations since we can always work separately with its real and imaginary parts. Systems are discussed in Chapters 5 and 9.

4.3 EXERCISES

In Problems 1–8, the auxiliary equation for the given differential equation has complex roots. Find a general solution.

1. $y'' + 9y = 0$
2. $y'' + y = 0$
3. $z'' - 6z' + 10z = 0$
4. $y'' - 10y' + 26y = 0$
5. $w'' + 4w' + 6w = 0$
6. $y'' - 4y' + 7y = 0$
7. $4y'' + 4y' + 6y = 0$
8. $4y'' - 4y' + 26y = 0$

In Problems 9–20, find a general solution.

9. $y'' - 8y' + 7y = 0$
10. $y'' + 4y' + 8y = 0$
11. $z'' + 10z' + 25z = 0$
12. $u'' + 7u = 0$
13. $y'' - 2y' + 26y = 0$
14. $y'' + 2y' + 5y = 0$
15. $y'' - 3y' - 11y = 0$
16. $y'' + 10y' + 41y = 0$
17. $y'' - y' + 7y = 0$
18. $2y'' + 13y' - 7y = 0$
19. $y''' + y'' + 3y' - 5y = 0$
20. $y''' - y'' + 2y = 0$

In Problems 21–27, solve the given initial value problem.

21. $y'' + 2y' + 2y = 0; \quad y(0) = 2, \quad y'(0) = 1$
22. $y'' + 2y' + 17y = 0; \quad y(0) = 1, \quad y'(0) = -1$
23. $w'' - 4w' + 2w = 0; \quad w(0) = 0, \quad w'(0) = 1$
24. $y'' + 9y = 0; \quad y(0) = 1, \quad y'(0) = 1$
25. $y'' - 2y' + 2y = 0; \quad y(\pi) = e^{\pi}, \quad y'(\pi) = 0$
26. $y'' - 2y' + y = 0; \quad y(0) = 1, \quad y'(0) = -2$
27. $y''' - 4y'' + 7y' - 6y = 0; \quad y(0) = 1, \quad y'(0) = 0, \quad y''(0) = 0$

28. To see the effect of changing the parameter b in the initial value problem
$$y'' + by' + 4y = 0; \quad y(0) = 1, \quad y'(0) = 0,$$
solve the problem for $b = 5$, 4, and 2 and sketch the solutions.

29. Find a general solution to the following higher-order equations.
(a) $y''' - y'' + y' + 3y = 0$
(b) $y''' + 2y'' + 5y' - 26y = 0$
(c) $y^{\text{iv}} + 13y'' + 36y = 0$

30. Using the representation for $e^{(\alpha + i\beta)t}$ in (6), verify the differentiation formula (7).

31. Using the mass–spring analogy, predict the behavior as $t \to +\infty$ of the solution to the given initial value problem. Then confirm your prediction by actually solving the problem.

(a) $y'' + 16y = 0\,; \quad y(0) = 2\,, \quad y'(0) = 0$
(b) $y'' + 100y' + y = 0\,; \quad y(0) = 1\,, \quad y'(0) = 0$
(c) $y'' - 6y' + 8y = 0\,; \quad y(0) = 1\,, \quad y'(0) = 0$
(d) $y'' + 2y' - 3y = 0\,; \quad y(0) = -2\,, \quad y'(0) = 0$
(e) $y'' - y' - 6y = 0\,; \quad y(0) = 1\,, \quad y'(0) = 1$

32. Vibrating Spring without Damping. A vibrating spring without damping can be modeled by the initial value problem (11) in Example 3 by taking $b = 0$.

(a) If $m = 10$ kg, $k = 250$ kg/sec^2, $y(0) = 0.3$ m, and $y'(0) = -0.1$ m/sec, find the equation of motion for this undamped vibrating spring.
(b) After how many seconds will the mass in part (a) first cross the equilibrium point?
(c) When the equation of motion is of the form displayed in (9), the motion is said to be **oscillatory** with **frequency** $\beta/2\pi$. Find the frequency of oscillation for the spring system of part (a).

33. Vibrating Spring with Damping. Using the model for a vibrating spring with damping discussed in Example 3:

(a) Find the equation of motion for the vibrating spring with damping if $m = 10$ kg, $b = 60$ kg/sec, $k = 250$ kg/sec^2, $y(0) = 0.3$ m, and $y'(0) = -0.1$ m/sec.
(b) After how many seconds will the mass in part (a) first cross the equilibrium point?
(c) Find the frequency of oscillation for the spring system of part (a). [*Hint:* See the definition of frequency given in Problem 32(c).]
(d) Compare the results of Problems 32 and 33 and determine what effect the damping has on the frequency of oscillation. What other effects does it have on the solution?

34. *RLC* Series Circuit. In the study of an electrical circuit consisting of a resistor, capacitor, inductor, and an electromotive force (see Figure 4.9), we are led to an initial value problem of the form

$$\textbf{(20)} \qquad L\frac{dI}{dt} + RI + \frac{q}{C} = E(t)\,;$$
$$q(0) = q_0\,,$$
$$I(0) = I_0\,,$$

where L is the inductance in henrys, R is the resistance in ohms, C is the capacitance in farads, $E(t)$ is the electromotive force in volts, $q(t)$ is the charge in coulombs on the capacitor at time t, and $I = dq/dt$ is the current in amperes. Find the current at time t if the charge on the capacitor is initially zero, the initial current is zero, $L = 10$ H, $R = 20\,\Omega$, $C = (6260)^{-1}$ F, and $E(t) = 100$ V. [*Hint:* Differentiate both sides of the differential equation in (20) to obtain a homogeneous linear second-order equation for $I(t)$. Then use (20) to determine dI/dt at $t = 0$.]

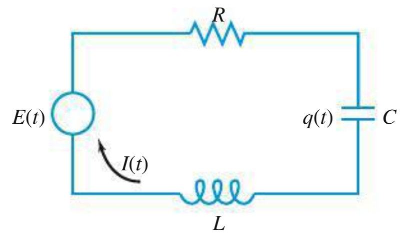


Figure 4.9 *RLC* series circuit

35. Swinging Door. The motion of a swinging door with an adjustment screw that controls the amount of friction on the hinges is governed by the initial value problem

$$I\theta'' + b\theta' + k\theta = 0\,; \quad \theta(0) = \theta_0\,, \quad \theta'(0) = v_0\,,$$

where θ is the angle that the door is open, I is the moment of inertia of the door about its hinges, $b > 0$ is a damping constant that varies with the amount of friction on the door, $k > 0$ is the spring constant associated with the swinging door, θ_0 is the initial angle that the door is opened, and v_0 is the initial angular velocity imparted to the door (see Figure 4.10). If I and k are fixed, determine for which values of b the door will *not* continually swing back and forth when closing.

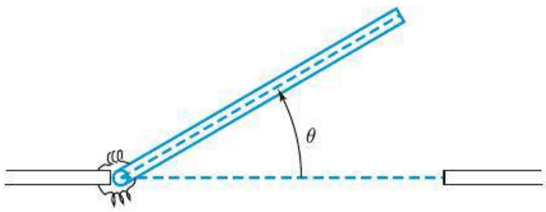


Figure 4.10 Top view of swinging door

36. Although the real general solution form (9) is convenient, it is also possible to use the form

$$\textbf{(21)} \qquad d_1 e^{(\alpha+i\beta)t} + d_2 e^{(\alpha-i\beta)t}$$

to solve initial value problems, as illustrated in Example 1. The coefficients d_1 and d_2 are complex constants.

(a) Use the form (21) to solve Problem 21. Verify that your form is equivalent to the one derived using (9).
(b) Show that, in general, d_1 and d_2 in (21) must be complex conjugates in order that the solution be real.

37. The auxiliary equations for the following differential equations have repeated complex roots. Adapt the "repeated root" procedure of Section 4.2 to find their general solutions:

(a) $y^{\text{iv}} + 2y'' + y = 0\,.$
(b) $y^{\text{iv}} + 4y''' + 12y'' + 16y' + 16y = 0\,.$ [*Hint:* The auxiliary equation is $(r^2 + 2r + 4)^2 = 0$.]

38. Prove the sum of angles formula for the sine function by following these steps. Fix x.

(a) Let $f(t) := \sin(x+t)$. Show that $f''(t)+f(t) = 0$, $f(0) = \sin x$, and $f'(0) = \cos x$.

(b) Use the auxiliary equation technique to solve the initial value problem $y''+y = 0$, $y(0) = \sin x$, and $y'(0) = \cos x$.

(c) By uniqueness, the solution in part (b) is the same as $f(t)$ from part (a). Write this equality; this should be the standard sum of angles formula for $\sin(x+t)$.

4.4 Nonhomogeneous Equations: the Method of Undetermined Coefficients

In this section we employ "judicious guessing" to derive a simple procedure for finding a solution to a *non*homogeneous linear equation with constant coefficients

$$ay''+by'+cy = f(t), \tag{1}$$

when the nonhomogeneity $f(t)$ is a single term of a special type. Our experience in Section 4.3 indicates that (1) will have an infinite number of solutions. For the moment we are content to find one particular solution. To motivate the procedure, let's first look at a few instructive examples.

Example 1 Find a particular solution to

$$y''+3y'+2y = 3t\,. \tag{2}$$

Solution We need to find a function $y(t)$ such that the combination $y''+3y'+2y$ is a linear function of t—namely, $3t$. Now what kind of function y "ends up" as a linear function after having its zeroth, first, and second derivatives combined? One immediate answer is: *another linear function*. So we might try $y_1(t) = At$ and attempt to match up $y_1''+3y_1'+2y_1$ with $3t$.

Perhaps you can see that this won't work: $y_1 = At$, $y_1' = A$ and $y_1'' = 0$ gives us

$$y_1''+3y_1'+2y_1 = 3A+2At\,,$$

and for this to equal $3t$, we require both that $A = 0$ and $A = 3/2$. We'll have better luck if we append a constant term to the trial function: $y_2(t) = At+B$. Then $y_2' = A$, $y_2'' = 0$, and

$$y_2''+3y_2'+2y_2 = 3A+2(At+B) = 2At+(3A+2B)\,,$$

which successfully matches up with $3t$ if $2A = 3$ and $3A+2B = 0$. Solving this system gives $A = 3/2$ and $B = -9/4$. Thus, the function

$$y_2(t) = \frac{3}{2}t-\frac{9}{4}$$

is a solution to (2). ◆

Example 1 suggests the following method for finding a particular solution to the equation

$$ay''+by'+cy = Ct^m, \quad m = 0, 1, 2, \ldots;$$

namely, we guess a solution of the form

$$y_p(t) = A_m t^m + \cdots + A_1 t + A_0\,,$$

with undetermined coefficients A_j, and match the corresponding powers of t in $ay''+by'+cy$ with Ct^m.[†] This procedure involves solving $m+1$ linear equations in the $m+1$ unknowns

[†]In this case the coefficient of t^k in $ay''+by'+cy$ will be zero for $k \neq m$ and C for $k = m$.

$A_0, A_1, \ldots, A_m$, and hopefully they have a solution. The technique is called the **method of undetermined coefficients.** Note that, as Example 1 demonstrates, we must retain *all* the powers $t^m, t^{m-1}, \ldots, t^1, t^0$ in the trial solution even though they are not present in the nonhomogeneity $f(t)$.

Example 2 Find a particular solution to

$$y'' + 3y' + 2y = 10e^{3t}\ . \tag{3}$$

Solution We guess $y_p(t) = Ae^{3t}$ because then y_p' and y_p'' will retain the same exponential form:

$$y_p'' + 3y_p' + 2y_p = 9Ae^{3t} + 3(3Ae^{3t}) + 2(Ae^{3t}) = 20Ae^{3t}\ .$$

Setting $20Ae^{3t} = 10e^{3t}$ and solving for A gives $A = 1/2$; hence,

$$y_p(t) = \frac{e^{3t}}{2}$$

is a solution to (3). ◆

Example 3 Find a particular solution to

$$y'' + 3y' + 2y = \sin t\ . \tag{4}$$

Solution Our initial action might be to guess $y_1(t) = A \sin t$, but this will fail because the derivatives introduce cosine terms:

$$y_1'' + 3y_1' + 2y_1 = -A\sin t + 3A\cos t + 2A\sin t = A\sin t + 3A\cos t\ ,$$

and matching this with $\sin t$ would require that A equal both 1 and 0. So we include the cosine term in the trial solution:

$$\begin{aligned} y_p(t) &= A\sin t + B\cos t\ , \\ y_p'(t) &= A\cos t - B\sin t\ , \\ y_p''(t) &= -A\sin t - B\cos t\ , \end{aligned}$$

and (4) becomes

$$\begin{aligned} y_p''(t) + 3y_p'(t) + 2y_p(t) &= -A\sin t - B\cos t + 3A\cos t - 3B\sin t \\ &\quad + 2A\sin t + 2B\cos t \\ &= (A - 3B)\sin t + (B + 3A)\cos t \\ &= \sin t\ . \end{aligned}$$

The equations $A - 3B = 1$, $B + 3A = 0$ have the solution $A = 0.1$, $B = -0.3$. Thus, the function

$$y_p(t) = 0.1\sin t - 0.3\cos t$$

is a particular solution to (4). ◆

More generally, for an equation of the form

$$ay'' + by' + cy = C\sin\beta t \quad (\text{or} \quad C\cos\beta t)\ , \tag{5}$$

the method of undetermined coefficients suggests that we guess

$$y_p(t) = A\cos\beta t + B\sin\beta t \tag{6}$$

and solve (5) for the unknowns A and B.

If we compare equation (5) with the mass–spring system equation

(7) $$[\text{inertia}] \times y'' + [\text{damping}] \times y' + [\text{stiffness}] \times y = F_{\text{ext}} ,$$

we can interpret (5) as describing a damped oscillator, shaken with a sinusoidal force. According to our discussion in Section 4.1, then, we would expect the mass ultimately to respond by moving in synchronization with the forcing sinusoid. In other words, the form (6) is suggested by physical, as well as mathematical, experience. A complete description of forced oscillators will be given in Section 4.10.

Example 4 Find a particular solution to

(8) $$y'' + 4y = 5t^2 e^t .$$

Solution Our experience with Example 1 suggests that we take a trial solution of the form $y_p(t) = (At^2 + Bt + C)e^t$, to match the nonhomogeneity in (8). We find

$$\begin{aligned} y_p &= (At^2 + Bt + C)e^t , \\ y_p' &= (2At + B)e^t + (At^2 + Bt + C)e^t , \\ y_p'' &= 2Ae^t + 2(2At + B)e^t + (At^2 + Bt + C)e^t , \\ y_p'' + 4y_p &= e^t(2A + 2B + C + 4C) + te^t(4A + B + 4B) + t^2e^t(A + 4A) \\ &= 5t^2e^t . \end{aligned}$$

Matching like terms yields $A = 1$, $B = -4/5$, and $C = -2/25$. A solution is given by

$$y_p(t) = \left(t^2 - \frac{4t}{5} - \frac{2}{25}\right)e^t . \ \blacklozenge$$

As our examples illustrate, when the nonhomogeneous term $f(t)$ is an exponential, a sine, a cosine function, or a nonnegative integer power of t times any of these, the function $f(t)$ itself suggests the form of a particular solution. However, certain situations thwart the straightforward application of the method of undetermined coefficients. Consider, for example, the equation

(9) $$y'' + y' = 5 .$$

Example 1 suggests that we guess $y_1(t) = A$, a zero-degree polynomial. But substitution into (9) proves futile:

$$(A)'' + (A)' = 0 \neq 5 .$$

The problem arises because any constant function, such as $y_1(t) = A$, is a solution to the corresponding homogeneous equation $y'' + y' = 0$, and the undetermined coefficient A gets lost upon substitution into the equation. We would encounter the same situation if we tried to find a solution to

(10) $$y'' - 6y' + 9y = e^{3t}$$

of the form $y_1 = Ae^{3t}$, because e^{3t} solves the associated homogeneous equation and

$$[Ae^{3t}]'' - 6[Ae^{3t}]' + 9[Ae^{3t}] = 0 \neq e^{3t} .$$

The "trick" for refining the method of undetermined coefficients in these situations smacks of the same logic as in Section 4.2, when a method was prescribed for finding second solutions to homogeneous equations with double roots. Basically, we append an extra factor of t to the trial solution suggested by the basic procedure. In other words, to solve (9) we try $y_p(t) = At$ instead of A:

(9′) $$y_p = At, \quad y_p' = A, \quad y_p'' = 0,$$
$$y_p'' + y_p' = 0 + A = 5\,,$$
$$A = 5, \qquad y_p(t) = 5t\,.$$

Similarly, to solve (10) we try $y_p = Ate^{3t}$ instead of Ae^{3t}. The trick won't work this time, because the characteristic equation of (10) has a double root and, consequently, Ate^{3t} also solves the homogeneous equation:

$$\left[Ate^{3t}\right]'' - 6\left[Ate^{3t}\right]' + 9\left[Ate^{3t}\right] = 0 \neq e^{3t}.$$

But if we append *another* factor of t, $y_p = At^2e^{3t}$, we succeed in finding a particular solution:†

$$y_p = At^2e^{3t},\ y_p' = 2Ate^{3t} + 3At^2e^{3t},\ y_p'' = 2Ae^{3t} + 12Ate^{3t} + 9At^2e^{3t}\,,$$
$$\begin{aligned} y_p'' - 6y_p' + 9y_p &= (2Ae^{3t} + 12Ate^{3t} + 9At^2e^{3t}) - 6(2Ate^{3t} + 3At^2e^{3t}) + 9(At^2e^{3t}) \\ &= 2Ae^{3t} = e^{3t}, \end{aligned}$$

so $A = 1/2$ and $y_p(t) = t^2e^{3t}/2$.

To see why this strategy resolves the problem and to generalize it, recall the form of the original differential equation (1), $ay'' + by' + cy = f(t)$. Its associated auxiliary equation is

(11) $$ar^2 + br + c = 0\,,$$

and if r is a double root, then

(12) $$2ar + b = 0$$

holds also [equation (13), Section 4.2, page 163].

Now suppose the nonhomogeneity $f(t)$ has the form Ct^me^{rt}, and we seek to match this $f(t)$ by substituting $y_p(t) = (A_nt^n + A_{n-1}t^{n-1} + \cdots + A_1t + A_0)e^{rt}$ into (1), *with the power n to be determined*. For simplicity we merely list the leading terms in y_p, y_p', and y_p'':

$$\begin{aligned} y_p &= A_nt^ne^{rt} + A_{n-1}t^{n-1}e^{rt} + A_{n-2}t^{n-2}e^{rt} + \ (\textit{lower-order terms}) \\ y_p' &= A_nrt^ne^{rt} + A_nnt^{n-1}e^{rt} + A_{n-1}rt^{n-1}e^{rt} + A_{n-1}(n-1)t^{n-2}e^{rt} \\ &\quad + A_{n-2}rt^{n-2}e^{rt} + (l.o.t.)\,, \\ y_p'' &= A_nr^2t^ne^{rt} + 2A_nnrt^{n-1}e^{rt} + A_nn(n-1)t^{n-2}e^{rt} \\ &\quad + A_{n-1}r^2t^{n-1}e^{rt} + 2A_{n-1}r(n-1)t^{n-2}e^{rt} + A_{n-2}r^2t^{n-2}e^{rt} + (l.o.t)\,. \end{aligned}$$

Then the left-hand member of (1) becomes

(13) $$\begin{aligned} &ay_p'' + by_p' + cy_p \\ &\quad = A_n(ar^2 + br + c)t^ne^{rt} + [A_nn(2ar + b) + A_{n-1}(ar^2 + br + c)]t^{n-1}e^{rt} \\ &\qquad + [A_nn(n-1)a + A_{n-1}(n-1)(2ar + b) + A_{n-2}(ar^2 + br + c)]t^{n-2}e^{rt} \\ &\qquad + (l.o.t.)\,, \end{aligned}$$

†Indeed, the solution t^2 to the equation $y'' = 2$, computed by simple integration, can also be derived by appending two factors of t to the solution $y \equiv 1$ of the associated homogeneous equation.

and we observe the following:

Case 1. If r is *not* a root of the auxiliary equation, the leading term in (13) is $A_n(ar^2 + br + c)t^n e^{rt}$, and to match $f(t) = Ct^m e^{rt}$ we must take $n = m$:

$$y_p(t) = (A_m t^m + \cdots + A_1 t + A_0)e^{rt} .$$

Case 2. If r is a *simple* root of the auxiliary equation, (11) holds and the leading term in (13) is $A_n n(2ar + b)t^{n-1}e^{rt}$, and to match $f(t) = Ct^m e^{rt}$ we must take $n = m + 1$:

$$y_p(t) = (A_{m+1}t^{m+1} + A_m t^m + \cdots + A_1 t + A_0)e^{rt} .$$

However, now the final term $A_0 e^{rt}$ can be dropped, since it solves the associated homogeneous equation, so we can factor out t and for simplicity renumber the coefficients to write

$$y_p(t) = t(A_m t^m + \cdots + A_1 t + A_0)e^{rt} .$$

Case 3. If r is a *double* root of the auxiliary equation, (11) and (12) hold and the leading term in (13) is $A_n n(n-1)at^{n-2}e^{rt}$, and to match $f(t) = Ct^m e^{rt}$ we must take $n = m + 2$:

$$y_p(t) = (A_{m+2}t^{m+2} + A_{m+1}t^{m+1} + \cdots + A_2 t^2 + A_1 t + A_0)e^{rt},$$

but again we drop the solutions to the associated homogeneous equation and renumber to write

$$y_p(t) = t^2(A_m t^m + \cdots + A_1 t + A_0)e^{rt} .$$

We summarize with the following rule.

Method of Undetermined Coefficients

To find a particular solution to the differential equation

$$ay'' + by' + cy = Ct^m e^{rt},$$

where m is a nonnegative integer, use the form

$$y_p(t) = t^s(A_m t^m + \cdots + A_1 t + A_0)e^{rt}, \tag{14}$$

with

(i) $s = 0$ if r is not a root of the associated auxiliary equation;
(ii) $s = 1$ if r is a simple root of the associated auxiliary equation; and
(iii) $s = 2$ if r is a double root of the associated auxiliary equation.

To find a particular solution to the differential equation

$$ay'' + by' + cy = \begin{cases} Ct^m e^{\alpha t}\cos\beta t \\ \text{or} \\ Ct^m e^{\alpha t}\sin\beta t \end{cases}$$

for $\beta \neq 0$, use the form

$$\begin{aligned} y_p(t) &= t^s(A_m t^m + \cdots + A_1 t + A_0)e^{\alpha t}\cos\beta t \\ &\quad + t^s(B_m t^m + \cdots + B_1 t + B_0)e^{\alpha t}\sin\beta t , \end{aligned} \tag{15}$$

with

(iv) $s = 0$ if $\alpha + i\beta$ is not a root of the associated auxiliary equation; and
(v) $s = 1$ if $\alpha + i\beta$ is a root of the associated auxiliary equation.

[The (cos, sin) formulation (15) is easily derived from the exponential formulation (14) by putting $r = \alpha + i\beta$ and employing Euler's formula, as in Section 4.3.]

Remark 1. The nonhomogeneity Ct^m corresponds to the case when $r = 0$.

Remark 2. The rigorous justification of the method of undetermined coefficients [including the analysis of the terms we dropped in (13)] will be presented in a more general context in Chapter 6.

Example 5 Find the form for a particular solution to

$$y'' + 2y' - 3y = f(t) \,, \tag{16}$$

where $f(t)$ equals

(a) $7\cos 3t$ **(b)** $2te^t \sin t$ **(c)** $t^2 \cos \pi t$ **(d)** $5e^{-3t}$ **(e)** $3te^t$ **(f)** t^2e^t

Solution The auxiliary equation for the homogeneous equation corresponding to (16), $r^2 + 2r - 3 = 0$, has roots $r_1 = 1$ and $r_2 = -3$. Notice that the functions in (a), (b), and (c) are associated with *complex* roots (because of the trigonometric factors). These are clearly different from r_1 and r_2, so the solution forms correspond to (15) with $s = 0$:

(a) $y_p(t) = A\cos 3t + B\sin 3t$

(b) $y_p(t) = (A_1 t + A_0)e^t \cos t + (B_1 t + B_0)e^t \sin t$

(c) $y_p(t) = (A_2 t^2 + A_1 t + A_0)\cos \pi t + (B_2 t^2 + B_1 t + B_0)\sin \pi t$

For the nonhomogeneity in (d) we appeal to (ii) and take $y_p(t) = Ate^{-3t}$ since -3 is a simple root of the auxiliary equation. Similarly, for (e) we take $y_p(t) = t(A_1 t + A_0)e^t$ and for (f) we take $y_p(t) = t(A_2 t^2 + A_1 t + A_0)e^t$. ◆

Example 6 Find the form of a particular solution to

$$y'' - 2y' + y = f(t),$$

for the same set of nonhomogeneities $f(t)$ as in Example 5.

Solution Now the auxiliary equation for the corresponding homogeneous equation is $r^2 - 2r + 1 = (r-1)^2 = 0$, with the double root $r = 1$. This root is not linked with any of the nonhomogeneities (a) through (d), so the same trial forms should be used for (a), (b), and (c) as in the previous example, and $y(t) = Ae^{-3t}$ will work for (d).

Since $r = 1$ is a double root, we have $s = 2$ in (14) and the trial forms for (e) and (f) have to be changed to

(e) $y_p(t) = t^2(A_1 t + A_0)e^t$

(f) $y_p(t) = t^2(A_2 t^2 + A_1 t + A_0)e^t$

respectively, in accordance with (iii). ◆

Example 7 Find the form of a particular solution to

$$y'' - 2y' + 2y = 5te^t \cos t \,.$$

Solution Now the auxiliary equation for the corresponding homogeneous equation is $r^2 - 2r + 2 = 0$, and it has complex roots $r_1 = 1 + i$, $r_2 = 1 - i$. Since the nonhomogeneity involves $e^{\alpha t}\cos \beta t$ with $\alpha = \beta = 1$; that is, $\alpha + i\beta = 1 + i = r_1$, the solution takes the form

$$y_p(t) = t(A_1 t + A_0)e^t \cos t + t(B_1 t + B_0)e^t \sin t \,.$$ ◆

The nonhomogeneity $\tan t$ in an equation like $y'' + y' + y = \tan t$ is not one of the forms for which the method of undetermined coefficients can be used; the derivatives of the "trial solution" $y(t) = A \tan t$, for example, get complicated, and it is not clear what additional terms need to be added to obtain a true solution. In Section 4.6 we discuss a different procedure that can handle such nonhomogeneous terms. **Keep in mind that the method of undetermined coefficients applies only to nonhomogeneities that are polynomials, exponentials, sines or cosines, or products of these functions.** The superposition principle in Section 4.5 shows how the method can be extended to the **sums** of such nonhomogeneities. Also, it provides the key to assembling a general solution to (1) that can accommodate initial value problems, which we have avoided so far in our examples.

4.4 EXERCISES

In Problems 1–8, decide whether or not the method of undetermined coefficients can be applied to find a particular solution of the given equation.

1. $y'' + 2y' - y = t^{-1}e^t$

2. $5y'' - 3y' + 2y = t^3 \cos 4t$

3. $2y''(x) - 6y'(x) + y(x) = (\sin x)/e^{4x}$

4. $x'' + 5x' - 3x = 3^t$

5. $y''(\theta) + 3y'(\theta) - y(\theta) = \sec\theta$

6. $2\omega''(x) - 3\omega(x) = 4x \sin^2 x + 4x \cos^2 x$

7. $8z'(x) - 2z(x) = 3x^{100}e^{4x} \cos 25x$

8. $ty'' - y' + 2y = \sin 3t$

In Problems 9–26, find a particular solution to the differential equation.

9. $y'' + 3y = -9$

10. $y'' + 2y' - y = 10$

11. $y''(x) + y(x) = 2^x$

12. $2x' + x = 3t^2$

13. $y'' - y' + 9y = 3 \sin 3t$

14. $2z'' + z = 9e^{2t}$

15. $\dfrac{d^2y}{dx^2} - 5\dfrac{dy}{dx} + 6y = xe^x$

16. $\theta''(t) - \theta(t) = t \sin t$

17. $y'' + 4y = 8 \sin 2t$

18. $y'' - 2y' + y = 8e^t$

19. $4y'' + 11y' - 3y = -2te^{-3t}$

20. $y'' + 4y = 16t \sin 2t$

21. $x''(t) - 4x'(t) + 4x(t) = te^{2t}$

22. $x''(t) - 2x'(t) + x(t) = 24t^2e^t$

23. $y''(\theta) - 7y'(\theta) = \theta^2$

24. $y''(x) + y(x) = 4x \cos x$

25. $y'' + 2y' + 4y = 111e^{2t} \cos 3t$

26. $y'' + 2y' + 2y = 4te^{-t} \cos t$

In Problems 27–32, determine the form of a particular solution for the differential equation. (Do not evaluate coefficients.)

27. $y'' + 9y = 4t^3 \sin 3t$

28. $y'' - 6y' + 9y = 5t^6e^{3t}$

29. $y'' + 3y' - 7y = t^4e^t$

30. $y'' - 2y' + y = 7e^t \cos t$

31. $y'' + 2y' + 2y = 8t^3e^{-t} \sin t$

32. $y'' - y' - 12y = 2t^6e^{-3t}$

In Problems 33–36, use the method of undetermined coefficients to find a particular solution to the given higher-order equation.

33. $y''' - y'' + y = \sin t$

34. $2y''' + 3y'' + y' - 4y = e^{-t}$

35. $y''' + y'' - 2y = te^t$

36. $y^{(4)} - 3y'' - 8y = \sin t$

4.5 The Superposition Principle and Undetermined Coefficients Revisited

The next theorem describes the superposition principle, a very simple observation which nonetheless endows the solution set for our equations with a powerful structure. It extends the applicability of the method of undetermined coefficients and enables us to solve initial value problems for nonhomogeneous differential equations.

Superposition Principle

Theorem 3. If y_1 is a solution to the differential equation

$$ay'' + by' + cy = f_1(t)\ ,$$

and y_2 is a solution to

$$ay'' + by' + cy = f_2(t)\ ,$$

then for any constants k_1 and k_2, the function $k_1y_1 + k_2y_2$ is a solution to the differential equation

$$ay'' + by' + cy = k_1f_1(t) + k_2f_2(t)\ .$$

Proof. This is straightforward; by substituting and rearranging we find

$$\begin{aligned} &a(k_1y_1 + k_2y_2)'' + b(k_1y_1 + k_2y_2)' + c(k_1y_1 + k_2y_2) \\ &\quad = k_1(ay_1'' + by_1' + cy_1) + k_2(ay_2'' + by_2' + cy_2) \\ &\quad = k_1f_1(t) + k_2f_2(t)\ .\ \blacklozenge \end{aligned}$$

Example 1 Find a particular solution to

(1) $$y'' + 3y' + 2y = 3t + 10e^{3t} \quad \text{and}$$

(2) $$y'' + 3y' + 2y = -9t + 20e^{3t}\ .$$

Solution In Example 1, Section 4.4, we found that $y_1(t) = 3t/2 - 9/4$ was a solution to $y'' + 3y' + 2y = 3t$, and in Example 2 we found that $y_2(t) = e^{3t}/2$ solved $y'' + 3y' + 2y = 10e^{3t}$. By superposition, then, $y_1 + y_2 = 3t/2 - 9/4 + e^{3t}/2$ solves equation (1).

The right-hand member of (2) equals minus three times $(3t)$ plus two times $(10e^{3t})$. Therefore, this same combination of y_1 and y_2 will solve (2):

$$y(t) = -3y_1 + 2y_2 = -3(3t/2 - 9/4) + 2(e^{3t}/2) = -9t/2 + 27/4 + e^{3t}\ .\ \blacklozenge$$

If we take a particular solution y_p to a nonhomogeneous equation like

(3) $$ay'' + by' + cy = f(t)$$

and add it to a general solution $c_1y_1 + c_2y_2$ of the homogeneous equation associated with (3),

(4) $$ay'' + by' + cy = 0\ ,$$

the sum

(5) $$y(t) = y_p(t) + c_1y_1(t) + c_2y_2(t)$$

is again, according to the superposition principle, a solution to (3):

$$\begin{aligned} &a(y_p + c_1y_1 + c_2y_2)'' + b(y_p + c_1y_1 + c_2y_2)' + c(y_p + c_1y_1 + c_2y_2) \\ &\quad = f(t) + 0 + 0 = f(t)\ . \end{aligned}$$

Since (5) contains two parameters, one would suspect that c_1 and c_2 can be chosen to make it satisfy arbitrary initial conditions. It is easy to verify that this is indeed the case.

Existence and Uniqueness: Nonhomogeneous Case

Theorem 4. For any real numbers $a(\neq 0)$, b, c, t_0, Y_0, and Y_1, suppose $y_p(t)$ is a particular solution to (3) in an interval I containing t_0 and that $y_1(t)$ and $y_2(t)$ are linearly independent solutions to the associated homogeneous equation (4) in I. Then there exists a unique solution in I to the initial value problem

$$ay'' + by' + cy = f(t), \qquad y(t_0) = Y_0, \quad y'(t_0) = Y_1, \tag{6}$$

and it is given by (5), for the appropriate choice of the constants c_1, c_2.

Proof. We have already seen that the superposition principle implies that (5) solves the differential equation. To satisfy the initial conditions in (6) we need to choose the constants so that

$$\begin{cases} y_p(t_0) + c_1 y_1(t_0) + c_2 y_2(t_0) = Y_0, \\ y_p'(t_0) + c_1 y_1'(t_0) + c_2 y_2'(t_0) = Y_1. \end{cases} \tag{7}$$

But as in the proof of Theorem 2 in Section 4.2, simple algebra shows that the choice

$$c_1 = \frac{[Y_0 - y_p(t_0)]y_2'(t_0) - [Y_1 - y_p'(t_0)]y_2(t_0)}{y_1(t_0)y_2'(t_0) - y_1'(t_0)y_2(t_0)}$$

and

$$c_2 = \frac{[Y_1 - y_p'(t_0)]y_1(t_0) - [Y_0 - y_p(t_0)]y_1'(t_0)}{y_1(t_0)y_2'(t_0) - y_1'(t_0)y_2(t_0)}$$

solves (7) unless the denominator is zero; Lemma 1, Section 4.2, assures us that it is not.

Why is the solution unique? If $y_{\mathrm{I}}(t)$ were another solution to (6), then the difference $y_{\mathrm{II}}(t) := y_p(t) + c_1 y_1(t) + c_2 y_2(t) - y_{\mathrm{I}}(t)$ would satisfy

$$\begin{cases} ay_{\mathrm{II}}'' + by_{\mathrm{II}}' + cy_{\mathrm{II}} = f(t) - f(t) = 0, \\ y_{\mathrm{II}}(t_0) = Y_0 - Y_0 = 0, \quad y_{\mathrm{II}}'(t_0) = Y_1 - Y_1 = 0. \end{cases} \tag{8}$$

But the initial value problem (8) admits the identically zero solution, and Theorem 1 in Section 4.2 applies since the differential equation in (8) is homogeneous. Consequently, (8) has *only* the identically zero solution. Thus, $y_{\mathrm{II}} \equiv 0$ and $y_{\mathrm{I}} = y_p + c_1 y_1 + c_2 y_2$. ◆

These deliberations entitle us to say that $y = y_p + c_1 y_1 + c_2 y_2$ is a **general solution** to the nonhomogeneous equation (3), since *any* solution $y_g(t)$ can be expressed in this form. (**Proof:** As in Section 4.2, we simply pick c_1 and c_2 so that $y_p + c_1 y_1 + c_2 y_2$ matches the value and the derivative of y_g *at any single point*; by uniqueness, $y_p + c_1 y_1 + c_2 y_2$ and y_g have to be the same function.)

Example 2 Given that $y_p(t) = t^2$ is a particular solution to

$$y'' - y = 2 - t^2,$$

find a general solution and a solution satisfying $y(0) = 1$, $y'(0) = 0$.

Solution The corresponding homogeneous equation,

$$y'' - y = 0\,,$$

has the associated auxiliary equation $r^2 - 1 = 0$. Because $r = \pm 1$ are the roots of this equation, a general solution to the homogeneous equation is $c_1e^t + c_2e^{-t}$. Combining this with the particular solution $y_p(t) = t^2$ of the nonhomogeneous equation, we find that a general solution is

$$y(t) = t^2 + c_1e^t + c_2e^{-t}\,.$$

To meet the initial conditions, set

$$\begin{aligned} y(0) &= 0^2 + c_1e^0 + c_2e^{-0} = 1\,, \\ y'(0) &= 2 \times 0 + c_1e^0 - c_2e^{-0} = 0\,, \end{aligned}$$

which yields $c_1 = c_2 = \frac{1}{2}$. The answer is

$$y(t) = t^2 + \frac{1}{2}(e^t + e^{-t}) = t^2 + \cosh t\,. \ ◆$$

Example 3 A mass–spring system is driven by a sinusoidal external force $(5\sin t + 5\cos t)$. The mass equals 1, the spring constant equals 2, and the damping coefficient equals 2 (in appropriate units), so the deliberations of Section 4.1 imply that the motion is governed by the differential equation

(9) $$y'' + 2y' + 2y = 5\sin t + 5\cos t\,.$$

If the mass is initially located at $y(0) = 1$, with a velocity $y'(0) = 2$, find its equation of motion.

Solution The associated homogeneous equation $y'' + 2y' + 2y = 0$ was studied in Example 1, Section 4.3; the roots of the auxiliary equation were found to be $-1 \pm i$, leading to a general solution $c_1e^{-t}\cos t + c_2e^{-t}\sin t$.

The method of undetermined coefficients dictates that we try to find a particular solution of the form $A\sin t + B\cos t$ for the first nonhomogeneity $5\sin t$:

(10) $$\begin{aligned} &y_p = A\sin t + B\cos t\,, \quad y_p' = A\cos t - B\sin t\,, \quad y_p'' = -A\sin t - B\cos t\,; \\ &y_p'' + 2y_p' + 2y_p = (-A - 2B + 2A)\sin t + (-B + 2A + 2B)\cos t = 5\sin t\,. \end{aligned}$$

Matching coefficients requires $A = 1$, $B = -2$ and so $y_p = \sin t - 2\cos t$.

The second nonhomogeneity $5\cos t$ calls for the identical form for y_p and leads to $(-A - 2B + 2A)\sin t + (-B + 2A + 2B)\cos t = 5\cos t$, or $A = 2, B = 1$. Hence $y_p = 2\sin t + \cos t$.

By the superposition principle, a general solution to (9) is given by the sum

$$\begin{aligned} y &= c_1e^{-t}\cos t + c_2e^{-t}\sin t + \sin t - 2\cos t + 2\sin t + \cos t \\ &= c_1e^{-t}\cos t + c_2e^{-t}\sin t + 3\sin t - \cos t\,. \end{aligned}$$

The initial conditions are

$$\begin{aligned} y(0) &= 1 = c_1e^{-0}\cos 0 + c_2e^{-0}\sin 0 + 3\sin 0 - \cos 0 = c_1 - 1\,, \\ y'(0) &= 2 = c_1\left[-e^{-t}\cos t - e^{-t}\sin t\right]_{t=0} + c_2\left[-e^{-t}\sin t + e^{-t}\cos t\right]_{t=0} \\ &\qquad\quad + 3\cos 0 + \sin 0 \\ &\quad = -c_1 + c_2 + 3\,, \end{aligned}$$

requiring $c_1 = 2$, $c_2 = 1$, and thus

(11) $$y(t) = 2e^{-t}\cos t + e^{-t}\sin t + 3\sin t - \cos t\,. \ ◆$$

The solution (11) exemplifies the features of forced, damped oscillations that we anticipated in Section 4.1. There is a sinusoidal component $(3\sin t - \cos t)$ that is synchronous with the driving force $(5\sin t + 5\cos t)$, and a component $(2e^{-t}\cos t + e^{-t}\sin t)$ that dies out. When the system is "pumped" sinusoidally, the response is a synchronous sinusoidal oscillation, after an initial transient that depends on the initial conditions; the synchronous response is the particular solution supplied by the method of undetermined coefficients, and the transient is the solution to the associated homogeneous equation. This interpretation will be discussed in detail in Sections 4.9 and 4.10.

You may have observed that, since the two undetermined-coefficient forms in the last example were identical and were destined to be added together, we could have used the form (10) to match both nonhomogeneities at the same time, deriving the condition

$$y_p'' + 2y_p' + 2y_p = (-A - 2B + 2A)\sin t + (-B + 2A + 2B)\cos t = 5\sin t + 5\cos t\,,$$

with solution $y_p = 3\sin t - \cos t$. The next example illustrates this "streamlined" procedure.

Example 4 Find a particular solution to

$$y'' - y = 8te^t + 2e^t\,. \tag{12}$$

Solution A general solution to the associated homogeneous equation is easily seen to be $c_1e^t + c_2e^{-t}$. Thus, a particular solution for matching the nonhomogeneity $8te^t$ has the form $t(A_1t + A_0)e^t$, whereas matching $2e^t$ requires the form A_0te^t. Therefore, we can match both with the first form:

$$\begin{aligned}
y_p &= t(A_1t + A_0)e^t = (A_1t^2 + A_0t)e^t\,,\\
y_p' &= (A_1t^2 + A_0t)e^t + (2A_1t + A_0)e^t = [A_1t^2 + (2A_1 + A_0)t + A_0]e^t\,,\\
y_p'' &= [2A_1t + (2A_1 + A_0)]e^t + [A_1t^2 + (2A_1 + A_0)t + A_0]e^t\\
&= [A_1t^2 + (4A_1 + A_0)t + (2A_1 + 2A_0)]e^t\,.
\end{aligned}$$

Thus

$$\begin{aligned}
y_p'' - y_p &= [4A_1t + (2A_1 + 2A_0)]e^t\\
&= 8te^t + 2e^t\,,
\end{aligned}$$

which yields $A_1 = 2$, $A_0 = -1$, and so $y_p = (2t^2 - t)e^t$. ◆

We generalize this procedure by modifying the method of undetermined coefficients as follows.

Method of Undetermined Coefficients (Revisited)

To find a particular solution to the differential equation

$$\boldsymbol{ay'' + by' + cy = P_m(t)e^{rt}},$$

where $P_m(t)$ is a polynomial of degree m, use the form

$$y_p(t) = \boldsymbol{t^s}(A_mt^m + \cdots + A_1t + A_0)e^{rt}\,; \tag{13}$$

if r is not a root of the associated auxiliary equation, take $s = 0$; if r is a simple root of the associated auxiliary equation, take $s = 1$; and if r is a double root of the associated auxiliary equation, take $s = 2$.

To find a particular solution to the differential equation

$$ay'' + by' + cy = P_m(t)e^{\alpha t}\cos\beta t + Q_n(t)e^{\alpha t}\sin\beta t\,, \quad \beta \neq 0\,,$$

where $P_m(t)$ is a polynomial of degree m and $Q_n(t)$ is a polynomial of degree n, use the form

$$y_p(t) = t^s(A_k t^k + \cdots + A_1 t + A_0)e^{\alpha t}\cos\beta t + t^s(B_k t^k + \cdots + B_1 t + B_0)e^{\alpha t}\sin\beta t\,, \tag{14}$$

where k is the larger of m and n. If $\alpha + i\beta$ is not a root of the associated auxiliary equation, take $s = 0$; if $\alpha + i\beta$ is a root of the associated auxiliary equation, take $s = 1$.

Example 5 Write down the form of a particular solution to the equation

$$y'' + 2y' + 2y = 5e^{-t}\sin t + 5t^3e^{-t}\cos t\,.$$

Solution The roots of the associated homogeneous equation $y'' + 2y' + 2y = 0$ were identified in Example 3 as $-1 \pm i$. Application of (14) dictates the form

$$y_p(t) = t(A_3t^3 + A_2t^2 + A_1t + A_0)e^{-t}\cos t + t(B_3t^3 + B_2t^2 + B_1t + B_0)e^{-t}\sin t\,. \quad ◆$$

The method of undetermined coefficients applies to higher-order linear differential equations with constant coefficients. Details will be provided in Chapter 6, but the following example should be clear.

Example 6 Write down the form of a particular solution to the equation

$$y''' + 2y'' + y' = 5e^{-t}\sin t + 3 + 7te^{-t}\,.$$

Solution The auxiliary equation for the associated homogeneous is $r^3 + 2r^2 + r = r(r+1)^2 = 0$, with a double root $r = -1$ and a single root $r = 0$. Term by term, the nonhomogeneities call for the forms

$$A_0 e^{-t}\cos t + B_0 e^{-t}\sin t \qquad (\text{for } 5e^{-t}\sin t)\,,$$
$$t A_0 \qquad (\text{for } 3)\,,$$
$$t^2(A_1t + A_0)e^{-t} \qquad (\text{for } 7te^{-t})\,.$$

(If -1 were a *triple* root, we would need $t^3(A_1t + A_0)e^{-t}$ for $7te^{-t}$.) Of course, we have to rename the coefficients, so the general form is

$$y_p(t) = Ae^{-t}\cos t + Be^{-t}\sin t + tC + t^2(Dt + E)e^{-t}\,. \quad ◆$$

4.5 EXERCISES

1. Given that $y_1(t) = \cos t$ is a solution to

$$y'' - y' + y = \sin t$$

and $y_2(t) = e^{2t}/3$ is a solution to

$$y'' - y' + y = e^{2t},$$

use the superposition principle to find solutions to the following differential equations:

(a) $y'' - y' + y = 5\sin t\,.$

(b) $y'' - y' + y = \sin t - 3e^{2t}\,.$

(c) $y'' - y' + y = 4\sin t + 18e^{2t}\,.$

2. Given that $y_1(t) = (1/4)\sin 2t$ is a solution to $y'' + 2y' + 4y = \cos 2t$ and that $y_2(t) = t/4 - 1/8$ is a solution to $y'' + 2y' + 4y = t$, use the superposition principle to find solutions to the following:

(a) $y'' + 2y' + 4y = t + \cos 2t$.

(b) $y'' + 2y' + 4y = 2t - 3\cos 2t$.

(c) $y'' + 2y' + 4y = 11t - 12\cos 2t$.

In Problems 3–8, a nonhomogeneous equation and a particular solution are given. Find a general solution for the equation.

3. $y'' - y = t, \quad y_p(t) = -t$

4. $y'' + y' = 1, \quad y_p(t) = t$

5. $\theta'' - \theta' - 2\theta = 1 - 2t, \quad \theta_p(t) = t - 1$

6. $y'' + 5y' + 6y = 6x^2 + 10x + 2 + 12e^x$, $y_p(x) = e^x + x^2$

7. $y'' = 2y + 2\tan^3 x, \quad y_p(x) = \tan x$

8. $y'' = 2y' - y + 2e^x, \quad y_p(x) = x^2e^x$

In Problems 9–16 decide whether the method of undetermined coefficients together with superposition can be applied to find a particular solution of the given equation. Do not solve the equation.

9. $3y'' + 2y' + 8y = t^2 + 4t - t^2e^t \sin t$

10. $y'' - y' + y = (e^t + t)^2$

11. $y'' - 6y' - 4y = 4\sin 3t - e^{3t}t^2 + 1/t$

12. $y'' + y' + ty = e^t + 7$

13. $y'' - 2y' + 3y = \cosh t + \sin^3 t$

14. $2y'' + 3y' - 4y = 2t + \sin^2 t + 3$

15. $y'' + e^t y' + y = 7 + 3t$

16. $2y'' - y' + 6y = t^2e^{-t}\sin t - 8t\cos 3t + 10^t$

In Problems 17–22, find a general solution to the differential equation.

17. $y'' - 2y' - 3y = 3t^2 - 5$

18. $y'' - y = -11t + 1$

19. $y''(x) - 3y'(x) + 2y(x) = e^x \sin x$

20. $y''(\theta) + 4y(\theta) = \sin\theta - \cos\theta$

21. $y''(\theta) + 2y'(\theta) + 2y(\theta) = e^{-\theta}\cos\theta$

22. $y''(x) + 6y'(x) + 10y(x) = 10x^4 + 24x^3 + 2x^2 - 12x + 18$

In Problems 23–30, find the solution to the initial value problem.

23. $y' - y = 1, \quad y(0) = 0$

24. $y'' = 6t; \quad y(0) = 3, \quad y'(0) = -1$

25. $z''(x) + z(x) = 2e^{-x}; \quad z(0) = 0, \quad z'(0) = 0$

26. $y'' + 9y = 27; \quad y(0) = 4, \quad y'(0) = 6$

27. $y''(x) - y'(x) - 2y(x) = \cos x - \sin 2x$; $y(0) = -7/20, \quad y'(0) = 1/5$

28. $y'' + y' - 12y = e^t + e^{2t} - 1; \; y(0) = 1, \quad y'(0) = 3$

29. $y''(\theta) - y(\theta) = \sin\theta - e^{2\theta}$; $y(0) = 1, \quad y'(0) = -1$

30. $y'' + 2y' + y = t^2 + 1 - e^t; \quad y(0) = 0, \quad y'(0) = 2$

In Problems 31–36, determine the form of a particular solution for the differential equation. Do not solve.

31. $y'' + y = \sin t + t\cos t + 10^t$

32. $y'' - y = e^{2t} + te^{2t} + t^2e^{2t}$

33. $x'' - x' - 2x = e^t\cos t - t^2 + \cos^3 t$

34. $y'' + 5y' + 6y = \sin t - \cos 2t$

35. $y'' - 4y' + 5y = e^{5t} + t\sin 3t - \cos 3t$

36. $y'' - 4y' + 4y = t^2e^{2t} - e^{2t}$

In Problems 37–40, find a particular solution to the given higher-order equation.

37. $y''' - 2y'' - y' + 2y = 2t^2 + 4t - 9$

38. $y^{(4)} - 5y'' + 4y = 10\cos t - 20\sin t$

39. $y''' + y'' - 2y = te^t + 1$

40. $y^{(4)} - 3y''' + 3y'' - y' = 6t - 20$

41. Discontinuous Forcing Term. In certain physical models, the nonhomogeneous term, or **forcing term,** $g(t)$ in the equation

$$ay'' + by' + cy = g(t)$$

may not be continuous but may have a jump discontinuity. If this occurs, we can still obtain a reasonable solution using the following procedure. Consider the initial value problem

$$y'' + 2y' + 5y = g(t); \quad y(0) = 0, \quad y'(0) = 0,$$

where

$$g(t) = \begin{cases} 10 & \text{if } 0 \le t \le 3\pi/2 \\ 0 & \text{if } t > 3\pi/2 \end{cases}.$$

(a) Find a solution to the initial value problem for $0 \le t \le 3\pi/2$.

(b) Find a general solution for $t > 3\pi/2$.

(c) Now choose the constants in the general solution from part (b) so that the solution from part (a) and the solution from part (b) agree, together with their first derivatives, at $t = 3\pi/2$. This gives us a continuously differentiable function that satisfies the differential equation except at $t = 3\pi/2$.

42. Forced Vibrations. As discussed in Section 4.1, a vibrating spring with damping that is under external force can be modeled by

$$my'' + by' + ky = g(t), \tag{15}$$

where $m > 0$ is the mass of the spring system, $b > 0$ is the damping constant, $k > 0$ is the spring constant, $g(t)$ is the force on the system at time t, and $y(t)$ is the displacement from the equilibrium of the spring system at time t. Assume $b^2 < 4mk$.

(a) Determine the form of the equation of motion for the spring system when $g(t) = \sin\beta t$ by finding a general solution to equation (15).

(b) Discuss the long-term behavior of this system. [*Hint:* Consider what happens to the general solution obtained in part (a) as $t \to +\infty$.]

43. A mass–spring system is driven by a sinusoidal external force $g(t) = 5\sin t$. The mass equals 1, the spring constant equals 3, and the damping coefficient equals 4. If the mass is initially located at $y(0) = 1/2$ and at rest, i.e., $y'(0) = 0$, find its equation of motion.

44. A mass–spring system is driven by the external force $g(t) = 2\sin 3t + 10\cos 3t$. The mass equals 1, the spring constant equals 5, and the damping coefficient equals 2. If the mass is initially located at $y(0) = -1$, with initial velocity $y'(0) = 5$, find its equation of motion.

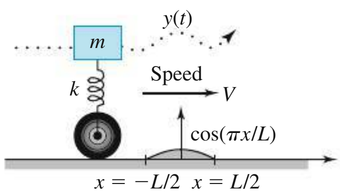


Figure 4.11 Speed bump

45. Speed Bumps. Often bumps like the one depicted in Figure 4.11 are built into roads to discourage speeding. The figure suggests that a crude model of the vertical motion $y(t)$ of a car encountering the speed bump with the speed V is given by

$$y(t) = 0 \qquad \text{for } t \le -L/(2V)\,,$$

$$my'' + ky = \begin{cases} F_0\cos(\pi V t/L) & \text{for } |t| < L/(2V) \\ 0 & \text{for } t \ge L/(2V)\,. \end{cases}$$

(The absence of a damping term indicates that the car's shock absorbers are not functioning.)

(a) Taking $m = k = 1$, $L = \pi$, and $F_0 = 1$ in appropriate units, solve this initial value problem. Thereby show that the formula for the oscillatory motion *after the car has traversed the speed bump* is $y(t) = A\sin t$, where the constant A depends on the speed V.

(b) Plot the amplitude $|A|$ of the solution $y(t)$ found in part (a) versus the car's speed V. From the graph, estimate the speed that produces the most violent shaking of the vehicle.

46. Show that the *boundary value problem*

$$y'' + \lambda^2 y = \sin t\,; \quad y(0) = 0\,, \quad y(\pi) = 1\,,$$

has a solution if and only if $\lambda \ne \pm 1, \pm 2, \pm 3, \ldots$.

47. Find the solution(s) to

$$y'' + 9y = 27\cos 6t$$

(if it exists) satisfying the *boundary conditions*

(a) $y(0) = -1\,, \quad y(\pi/6) = 3\,.$

(b) $y(0) = -1\,, \quad y(\pi/3) = 5\,.$

(c) $y(0) = -1\,, \quad y(\pi/3) = -1\,.$

48. All that is known concerning a mysterious second-order constant-coefficient differential equation $y'' + py' + qy = g(t)$ is that $t^2 + 1 + e^t\cos t$, $t^2 + 1 + e^t\sin t$, and $t^2 + 1 + e^t\cos t + e^t\sin t$ are solutions.

(a) Determine two linearly independent solutions to the corresponding homogeneous equation.

(b) Find a suitable choice of p, q, and $g(t)$ that enables these solutions.

4.6 Variation of Parameters

We have seen that the method of undetermined coefficients is a simple procedure for determining a particular solution when the equation has constant coefficients and the nonhomogeneous term is of a special type. Here we present a more general method, called **variation of parameters,**[†] for finding a particular solution.

Consider the nonhomogeneous linear second-order equation

$$ay'' + by' + cy = f(t) \tag{1}$$

and let $\{y_1(t), y_2(t)\}$ be two linearly independent solutions for the corresponding homogeneous equation

$$ay'' + by' + cy = 0\,.$$

[†] *Historical Footnote:* The method of variation of parameters was invented by Joseph Lagrange in 1774.

Then we know that a general solution to this homogeneous equation is given by

(2) $$y_h(t) = c_1y_1(t) + c_2y_2(t) ,$$

where c_1 and c_2 are constants. To find a particular solution to the nonhomogeneous equation, the strategy of variation of parameters is to replace the constants in (2) by functions of t. That is, we seek a solution of (1) of the form[†]

(3) $$\boldsymbol{y_p(t) = v_1(t)y_1(t) + v_2(t)y_2(t) .}$$

Because we have introduced two unknown functions, $v_1(t)$ and $v_2(t)$, it is reasonable to expect that we can impose two equations (requirements) on these functions. Naturally, one of these equations should come from (1). Let's therefore plug $y_p(t)$ given by (3) into (1). To accomplish this, we must first compute $y_p'(t)$ and $y_p''(t)$. From (3) we obtain

$$y_p' = (v_1'y_1 + v_2'y_2) + (v_1y_1' + v_2y_2') .$$

To simplify the computation and to avoid second-order derivatives for the unknowns v_1, v_2 in the expression for y_p'', we impose the requirement

(4) $$v_1'y_1 + v_2'y_2 = 0 .$$

Thus, the formula for y_p' becomes

(5) $$y_p' = v_1y_1' + v_2y_2' ,$$

and so

(6) $$y_p'' = v_1'y_1' + v_1y''_1 + v_2'y_2' + v_2y_2'' .$$

Now, substituting y_p, y_p', and y_p'', as given in (3), (5), and (6), into (1), we find

(7) $$\begin{aligned} f &= ay_p'' + by_p' + cy_p \\ &= a(v_1'y_1' + v_1y_1'' + v_2'y_2' + v_2y_2'') + b(v_1y_1' + v_2y_2') + c(v_1y_1 + v_2y_2) \\ &= a(v_1'y_1' + v_2'y_2') + v_1(ay_1'' + by_1' + cy_1) + v_2(ay_2'' + by_2' + cy_2) \\ &= a(v_1'y_1' + v_2'y_2') + 0 + 0 \end{aligned}$$

since y_1 and y_2 are solutions to the homogeneous equation. Thus, (7) reduces to

(8) $$v_1'y_1' + v_2'y_2' = \frac{f}{a} .$$

To summarize, if we can find v_1 and v_2 that satisfy both (4) and (8), that is,

(9) $$\begin{aligned} &\boldsymbol{y_1v_1' + y_2v_2' = 0 ,} \\ &\boldsymbol{y_1'v_1' + y_2'v_2' = \frac{f}{a} ,} \end{aligned}$$

then y_p given by (3) will be a particular solution to (1). To determine v_1 and v_2, we first solve the linear system (9) for v_1' and v_2'. Algebraic manipulation or Cramer's rule (see Appendix D) immediately gives

$$v_1'(t) = \frac{-f(t)y_2(t)}{a\left[y_1(t)y_2'(t) - y_1'(t)y_2(t)\right]} \quad \text{and} \quad v_2'(t) = \frac{f(t)y_1(t)}{a\left[y_1(t)y_2'(t) - y_1'(t)y_2(t)\right]} ,$$

[†]In Exercises 2.3, Problem 36, we developed this approach for first-order linear equations. Because of the similarity of equations (2) and (3), this technique is sometimes known as "variation of constants."

where the bracketed expression in the denominator (the Wronskian) is never zero because of Lemma 1, Section 4.2. Upon integrating these equations, we finally obtain

(10) $$v_1(t) = \int \frac{-f(t)y_2(t)}{a\left[y_1(t)y_2'(t) - y_1'(t)y_2(t)\right]}\,dt \quad \text{and} \quad v_2(t) = \int \frac{f(t)y_1(t)}{a\left[y_1(t)y_2'(t) - y_1'(t)y_2(t)\right]}\,dt\,.$$

Let's review this procedure.

Method of Variation of Parameters

To determine a particular solution to $ay'' + by' + cy = f$:

(a) Find two linearly independent solutions $\{y_1(t), y_2(t)\}$ to the corresponding homogeneous equation and take

$$y_p(t) = v_1(t)y_1(t) + v_2(t)y_2(t)\,.$$

(b) Determine $v_1(t)$ and $v_2(t)$ by solving the system in (9) for $v_1'(t)$ and $v_2'(t)$ and integrating.

(c) Substitute $v_1(t)$ and $v_2(t)$ into the expression for $y_p(t)$ to obtain a particular solution.

Of course, in step (b) one could use the formulas in (10), but $v_1(t)$ and $v_2(t)$ are so easy to derive that you are advised not to memorize them.

Example 1 Find a general solution on $(-\pi/2, \pi/2)$ to

(11) $$\frac{d^2y}{dt^2} + y = \tan t\,.$$

Solution Observe that two independent solutions to the homogeneous equation $y'' + y = 0$ are $\cos t$ and $\sin t$. We now set

(12) $$y_p(t) = v_1(t)\cos t + v_2(t)\sin t$$

and, referring to (9), solve the system

$$\begin{aligned}(\cos t)v_1'(t) + (\sin t)v_2'(t) &= 0\,,\\ (-\sin t)v_1'(t) + (\cos t)v_2'(t) &= \tan t\,,\end{aligned}$$

for $v_1'(t)$ and $v_2'(t)$. This gives

$$\begin{aligned}v_1'(t) &= -\tan t\ \sin t\,,\\ v_2'(t) &= \tan t\ \cos t = \sin t\,.\end{aligned}$$

Integrating, we obtain

(13) $$\begin{aligned}v_1(t) &= -\int \tan t \sin t\,dt = -\int \frac{\sin^2 t}{\cos t}\,dt\\ &= -\int \frac{1-\cos^2 t}{\cos t}\,dt = \int (\cos t - \sec t)\,dt\\ &= \sin t - \ln\lvert\sec t + \tan t\rvert + C_1\,,\end{aligned}$$

(14) $$v_2(t) = \int \sin t\,dt = -\cos t + C_2\,.$$

We need only one particular solution, so we take both C_1 and C_2 to be zero for simplicity. Then, substituting $v_1(t)$ and $v_2(t)$ in (12), we obtain

$$y_p(t) = (\sin t - \ln|\sec t + \tan t|)\cos t - \cos t \sin t\,,$$

which simplifies to

$$y_p(t) = -(\cos t)\ln|\sec t + \tan t|\,.$$

We may drop the absolute value symbols because $\sec t + \tan t = (1 + \sin t)/\cos t > 0$ for $-\pi/2 < t < \pi/2$.

Recall that a general solution to a nonhomogeneous equation is given by the sum of a general solution to the homogeneous equation and a particular solution. Consequently, a general solution to equation (11) on the interval $(-\pi/2, \pi/2)$ is

$$y(t) = c_1\cos t + c_2\sin t - (\cos t)\ln(\sec t + \tan t)\,. \tag{15}$$ ◆

Note that in the above example the constants C_1 and C_2 appearing in (13) and (14) were chosen to be zero. If we had retained these arbitrary constants, the ultimate effect would be just to add $C_1\cos t + C_2\sin t$ to (15), which is clearly redundant.

Example 2 Find a particular solution on $(-\pi/2, \pi/2)$ to

$$\frac{d^2y}{dt^2} + y = \tan t + 3t - 1\,. \tag{16}$$

Solution With $f(t) = \tan t + 3t - 1$, the variation of parameters procedure will lead to a solution. But it is simpler in this case to consider separately the equations

$$\frac{d^2y}{dt^2} + y = \tan t\,, \tag{17}$$

$$\frac{d^2y}{dt^2} + y = 3t - 1 \tag{18}$$

and then use the superposition principle (Theorem 3, page 181).

In Example 1 we found that

$$y_q(t) = -(\cos t)\ln(\sec t + \tan t)$$

is a particular solution for equation (17). For equation (18) the method of undetermined coefficients can be applied. On seeking a solution to (18) of the form $y_r(t) = At + B$, we quickly obtain

$$y_r(t) = 3t - 1\,.$$

Finally, we apply the superposition principle to get

$$\begin{aligned} y_p(t) &= y_q(t) + y_r(t) \\ &= -(\cos t)\ln(\sec t + \tan t) + 3t - 1 \end{aligned}$$

as a particular solution for equation (16). ◆

Note that we could not have solved Example 1 by the method of undetermined coefficients; the nonhomogeneity $\tan t$ is unsuitable. Another important advantage of the method of variation of parameters is its applicability to linear equations whose coefficients a, b, c are functions of t. Indeed, on reviewing the derivation of the system (9) and the formulas (10), one can check that we did not make any use of the constant coefficient property; i.e., the method works provided we know a pair of linearly independent solutions to the corresponding homogeneous equation. We illustrate the method in the next example.

Example 3 Find a particular solution of the variable coefficient linear equation

$$t^2 y'' - 4ty' + 6y = 4t^3 , \quad t > 0 , \tag{19}$$

given that $y_1(t) = t^2$ and $y_2(t) = t^3$ are solutions to the corresponding homogeneous equation.

Solution The functions t^2 and t^3 are linearly independent solutions to the corresponding homogeneous equation on $(0, \infty)$ (verify this!) and so (19) has a particular solution of the form

$$y_p(t) = v_1(t)t^2 + v_2(t)t^3 .$$

To determine the unknown functions v_1 and v_2, we solve the system (9) with $f(t) = 4t^3$ and $a = a(t) = t^2$:

$$t^2 v_1'(t) + t^3 v_2'(t) = 0$$

$$2t v_1'(t) + 3t^2 v_2'(t) = f/a = 4t .$$

The solutions are readily found to be $v_1'(t) = -4$ and $v_2'(t) = 4/t$, which gives $v_1(t) = -4t$ and $v_2(t) = 4 \ln t$. Consequently,

$$y_p(t) = (-4t)t^2 + (4 \ln t)t^3 = 4t^3(-1 + \ln t)$$

is a solution to (19). ◆

Variable coefficient linear equations will be discussed in more detail in the next section.

4.6 EXERCISES

In Problems 1–8, find a general solution to the differential equation using the method of variation of parameters.

1. $y'' + 4y = \tan 2t$
2. $y'' + y = \sec t$
3. $y'' - 2y' + y = t^{-1}e^t$
4. $y'' + 2y' + y = e^{-t}$
5. $y''(\theta) + 16y(\theta) = \sec 4\theta$
6. $y'' + 9y = \sec^2(3t)$
7. $y'' + 4y' + 4y = e^{-2t} \ln t$
8. $y'' + 4y = \csc^2(2t)$

In Problems 9 and 10, find a particular solution first by undetermined coefficients, and then by variation of parameters. Which method was quicker?

9. $y'' - y = 2t + 4$
10. $2x''(t) - 2x'(t) - 4x(t) = 2e^{2t}$

In Problems 11–18, find a general solution to the differential equation.

11. $y'' + y = \tan t + e^{3t} - 1$
12. $y'' + y = \tan^2 t$
13. $v'' + 4v = \sec^4(2t)$
14. $y''(\theta) + y(\theta) = \sec^3 \theta$
15. $y'' + y = 3 \sec t - t^2 + 1$
16. $y'' + 5y' + 6y = 18t^2$
17. $\frac{1}{2}y'' + 2y = \tan 2t - \frac{1}{2}e^t$
18. $y'' - 6y' + 9y = t^{-3}e^{3t}$
19. Express the solution to the initial value problem
$$y'' - y = \frac{1}{t}, \quad y(1) = 0 , \quad y'(1) = -2 ,$$
using definite integrals. Using numerical integration (Appendix C) to approximate the integrals, find an approximation for $y(2)$ to two decimal places.
20. Use the method of variation of parameters to show that
$$y(t) = c_1 \cos t + c_2 \sin t + \int_0^t f(s) \sin(t - s)\, ds$$
is a general solution to the differential equation
$$y'' + y = f(t) ,$$
where $f(t)$ is a continuous function on $(-\infty, \infty)$. [*Hint:* Use the trigonometric identity $\sin(t - s) = \sin t \cos s - \sin s \cos t$.]

21. Suppose y satisfies the equation $y'' + 10y' + 25y = e^{t^3}$ subject to $y(0) = 1$ and $y'(0) = -5$. Estimate $y(0.2)$ to within ± 0.0001 by numerically approximating the integrals in the variation of parameters formula.

In Problems 22 through 25, use variation of parameters to find a general solution to the differential equation given that the functions y_1 and y_2 are linearly independent solutions to the corresponding homogeneous equation for $t > 0$.

22. $t^2 y'' - 4ty' + 6y = t^3 + 1$;
$y_1 = t^2$, $y_2 = t^3$

23. $ty'' - (t+1)y' + y = t^2$;
$y_1 = e^t$, $y_2 = t + 1$

24. $ty'' + (1 - 2t)y' + (t - 1)y = te^t$;
$y_1 = e^t$, $y_2 = e^t \ln t$

25. $ty'' + (5t - 1)y' - 5y = t^2 e^{-5t}$;
$y_1 = 5t - 1$, $y_2 = e^{-5t}$

4.7 Variable-Coefficient Equations

The techniques of Sections 4.2 and 4.3 have explicitly demonstrated that solutions to a linear homogeneous constant-coefficient differential equation,

$$ay'' + by' + cy = 0\,, \tag{1}$$

are defined and satisfy the equation over the whole interval $(-\infty, +\infty)$. After all, such solutions are combinations of exponentials, sinusoids, and polynomials.

The variation of parameters formula of Section 4.6 extended this to nonhomogeneous constant-coefficient problems,

$$ay'' + by' + cy = f(t)\,, \tag{2}$$

yielding solutions valid *over all intervals where $f(t)$ is continuous* (ensuring that the integrals in (10) of Section 4.6 containing $f(t)$ exist and are differentiable). We could hardly hope for more; indeed, it is debatable what *meaning* the differential equation (2) would have at a point where $f(t)$ is undefined, or discontinuous.

Therefore, when we move to the realm of equations with *variable* coefficients of the form

$$a_2(t)y'' + a_1(t)y' + a_0(t)y = f(t)\,, \tag{3}$$

the most we can expect is that there are solutions that are valid over intervals where all four "governing" functions—$a_2(t)$, $a_1(t)$, $a_0(t)$, and $f(t)$—are continuous. Fortunately, this expectation is fulfilled except for an important technical requirement—namely, that the coefficient function $a_2(t)$ must be nonzero over the interval.†

Typically, one divides by the nonzero coefficient $a_2(t)$ and expresses the theorem for the equation in **standard form** [see (4), below] as follows.

Existence and Uniqueness of Solutions

Theorem 5. If $p(t)$, $q(t)$, and $g(t)$ are continuous on an interval (a, b) that contains the point t_0, then for any choice of the initial values Y_0 and Y_1, there exists a unique solution $y(t)$ on the same interval (a, b) to the initial value problem

$$y''(t) + p(t)y'(t) + q(t)y(t) = g(t)\,; \quad y(t_0) = Y_0\,, \quad y'(t_0) = Y_1\,. \tag{4}$$

†Indeed, the whole nature of the equation—reduction from *second*-order to *first*-order—changes at points where $a_2(t)$ is zero.

Example 1 Determine the largest interval for which Theorem 5 ensures the existence and uniqueness of a solution to the initial value problem

$$(5) \qquad (t-3)\frac{d^2y}{dt^2} + \frac{dy}{dt} + \sqrt{t}\,y = \ln t\,; \quad y(1) = 3\,, \quad y'(1) = -5\,.$$

Solution The data $p(t)$, $q(t)$, and $g(t)$ in the standard form of the equation,

$$y'' + py' + qy = \frac{d^2y}{dt^2} + \frac{1}{(t-3)}\frac{dy}{dt} + \frac{\sqrt{t}}{(t-3)}y = \frac{\ln t}{(t-3)} = g\,,$$

are simultaneously continuous in the intervals $0 < t < 3$ and $3 < t < \infty$. The former contains the point $t_0 = 1$, where the initial conditions are specified, so Theorem 5 guarantees (5) has a unique solution in $0 < t < 3$. ◆

Theorem 5, embracing existence and uniqueness for the variable-coefficient case, is difficult to prove because we can't construct explicit solutions in the general case. So the proof is deferred to Chapter 13.† However, it is instructive to examine a special case that we can solve explicitly.

Cauchy–Euler, or Equidimensional, Equations

Definition 2. A linear second-order equation that can be expressed in the form

$$(6) \qquad at^2y''(t) + bty'(t) + cy = f(t)\,,$$

where a, b, and c are constants, is called a **Cauchy–Euler,** or **equidimensional, equation.**

For example, the differential equation

$$3t^2y'' + 11ty' - 3y = \sin t$$

is a Cauchy–Euler equation, whereas

$$2y'' - 3ty' + 11y = 3t - 1$$

is *not* because the coefficient of y'' is 2, which is not a constant times t^2.

The nomenclature *equidimensional* comes about because if y has the dimensions of, say, meters and t has dimensions of time, then each term t^2y'', ty', and y has the same dimensions (meters). The coefficient of $y''(t)$ in (6) is at^2, and it is zero at $t = 0$; equivalently, the standard form

$$y'' + \frac{b}{at}y' + \frac{c}{at^2}y = \frac{f(t)}{at^2}$$

has discontinuous coefficients at $t = 0$. Therefore, we can expect the solutions to be valid only for $t > 0$ or $t < 0$. Discontinuities in f, of course, will impose further restrictions.

†All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems*, 7th ed.

To solve a *homogeneous* Cauchy–Euler equation, for $t > 0$, we exploit the equidimensional feature by looking for solutions of the form $y = t^r$, because then t^2y'', ty', and y each have the form *(constant)* $\times\, t^r$:

$$y = t^r\,, \quad ty' = trt^{r-1} = rt^r\,, \quad t^2y'' = t^2r(r-1)t^{r-2} = r(r-1)t^r\,,$$

and substitution into the homogeneous form of (6) (that is, with $g = 0$) yields a simple quadratic equation for r:

$$ar(r-1)t^r + brt^r + ct^r = [ar^2 + (b-a)r + c]t^r = 0\,, \quad \text{or}$$

(7) $$\boldsymbol{ar^2 + (b-a)r + c = 0\,,}$$

which we call the associated *characteristic equation.*

Example 2 Find two linearly independent solutions to the equation

$$3t^2y'' + 11ty' - 3y = 0\,, \quad t > 0\,.$$

Solution Inserting $y = t^r$ yields, according to (7),

$$3r^2 + (11-3)r - 3 = 3r^2 + 8r - 3 = 0\,,$$

whose roots $r = 1/3$ and $r = -3$ produce the independent solutions

$$y_1(t) = t^{1/3}\,, \quad y_2(t) = t^{-3} \quad (\text{for } t > 0)\,. \ ◆$$

Clearly, the substitution $y = t^r$ into a homogeneous *equidimensional* equation has the same simplifying effect as the insertion of $y = e^{rt}$ into the homogeneous *constant-coefficient* equation in Section 4.2. That means we will have to deal with the same encumbrances:

1. What to do when the roots of (7) are complex
2. What to do when the roots of (7) are equal

If r is complex, $r = \alpha + i\beta$, we can interpret $t^{\alpha+i\beta}$ by using the identity $t = e^{\ln t}$ and invoking Euler's formula [equation (5), Section 4.3]:

$$t^{\alpha+i\beta} = t^{\alpha}t^{i\beta} = t^{\alpha}e^{i\beta\ln t} = t^{\alpha}[\cos(\beta\ln t) + i\sin(\beta\ln t)]\,.$$

Then we simplify as in Section 4.3 by taking the real and imaginary parts to form independent solutions:

(8) $$\boldsymbol{y_1 = t^{\alpha}\cos(\beta\ln t)\,, \quad y_2 = t^{\alpha}\sin(\beta\ln t)\,.}$$

If r is a double root of the characteristic equation (7), then independent solutions of the Cauchy–Euler equation on $(0, \infty)$ are given by

(9) $$\boldsymbol{y_1 = t^r\,, \quad y_2 = t^r\ln t\,.}$$

This can be verified by direct substitution into the differential equation. Alternatively, the second, linearly independent, solution can be obtained by *reduction of order*, a procedure to be discussed shortly in Theorem 8. Furthermore, Problem 23 demonstrates that the substitution $t = e^x$ changes the homogeneous Cauchy–Euler equation into a homogeneous constant-coefficient equation, and the formats (8) and (9) then follow from our earlier deliberations.

We remark that if a homogeneous Cauchy–Euler equation is to be solved for $t < 0$, then one simply introduces the change of variable $t = -\tau$, where $\tau > 0$. The reader should verify via the chain rule that the identical characteristic equation (7) arises when $\tau^r = (-t)^r$ is substituted in the equation. Thus the solutions take the same form as (8), (9), but with t replaced by $-t$; for example, if r is a double root of (7), we get $(-t)^r$ and $(-t)^r \ln(-t)$ as two linearly independent solutions on $(-\infty, 0)$.

Example 3 Find a pair of linearly independent solutions to the following Cauchy–Euler equations for $t > 0$.

(a) $t^2y'' + 5ty' + 5y = 0$ **(b)** $t^2y'' + ty' = 0$

Solution For part (a), the characteristic equation becomes $r^2 + 4r + 5 = 0$, with the roots $r = -2 \pm i$, and (8) produces the real solutions $t^{-2}\cos(\ln t)$ and $t^{-2}\sin(\ln t)$.

For part (b), the characteristic equation becomes simply $r^2 = 0$ with the double root $r = 0$, and (9) yields the solutions $t^0 = 1$ and $\ln t$. ◆

In Chapter 8 we will see how one can obtain power series expansions for solutions to variable-coefficient equations when the coefficients are *analytic* functions. But, as we said, there is no procedure for explicitly solving the general case. Nonetheless, thanks to the existence/uniqueness result of Theorem 5, most of the other theorems and concepts of the preceding sections are easily extended to the variable-coefficient case, with the proviso that they apply only over intervals in which the governing functions $p(t)$, $q(t)$, $g(t)$ are continuous. Thus we have the following analog of Lemma 1, page 160.

A Condition for Linear Dependence of Solutions

Lemma 3. If $y_1(t)$ and $y_2(t)$ are any two solutions to the homogeneous differential equation

$$y''(t) + p(t)y'(t) + q(t)y(t) = 0 \tag{10}$$

on an interval *I* where the functions $p(t)$ and $q(t)$ are continuous and if the Wronskian†

$$W[y_1, y_2](t) := y_1(t)y_2'(t) - y_1'(t)y_2(t) = \begin{vmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{vmatrix}$$

is zero at *any* point t **of *I***, then y_1 and y_2 are linearly dependent **on *I***.

As in the constant-coefficient case, the Wronskian of two solutions is either identically zero or never zero on I, with the latter implying linear independence on I.

Precisely as in the proof for the constant-coefficient case, it can be verified that any linear combination $c_1y_1 + c_2y_2$ of solutions y_1 and y_2 to (10) is also a solution. In fact, these are the only solutions to (10) as stated in the following result.

Representation of Solutions to Initial Value Problems

Theorem 6. If $y_1(t)$ and $y_2(t)$ are any two solutions to the homogeneous differential equation (10) that are linearly independent on an interval I, then every solution to (10) on I is expressible as a linear combination of y_1 and y_2. Moreover, the initial value problem consisting of equation (10) and the initial conditions $y(t_0) = Y_0$, $y'(t_0) = Y_1$ has a unique solution on I for any point t_0 in I and any constants Y_0, Y_1.

† The determinant representation of the Wronskian was introduced in Problem 34, Section 4.2.

As in the constant-coefficient case, the linear combination $y_h = c_1y_1 + c_2y_2$ is called a **general solution** to (10) on I if y_1, y_2 are linearly independent solutions on I.

For the nonhomogeneous equation

$$y''(t) + p(t)y'(t) + q(t)y(t) = g(t)\ , \tag{11}$$

a general solution on I is given by $y = y_p + y_h$, where $y_h = c_1y_1 + c_2y_2$ is a general solution to the corresponding homogeneous equation (10) **on I and y_p is a particular solution to** (11) **on I.** In other words, the solution to the initial value problem stated in Theorem 5 must be of this form for a suitable choice of the constants c_1, c_2. This follows, just as before, from a straightforward extension of the superposition principle for variable-coefficient equations described in Problem 30.

As illustrated at the end of the Section 4.6, if linearly independent solutions to the homogeneous equation (10) are known, then y_p can be determined for (11) by the variation of parameters method.

Variation of Parameters

Theorem 7. If y_1 and y_2 are two linearly independent solutions to the homogeneous equation (10) on an interval I where $p(t)$, $q(t)$, and $g(t)$ are continuous, then a particular solution to (11) is given by $y_p = v_1y_1 + v_2y_2$, where v_1 and v_2 are determined up to a constant by the pair of equations

$$\begin{aligned} y_1v_1' + y_2v_2' &= 0\ ,\\ y_1'v_1' + y_2'v_2' &= g\ , \end{aligned}$$

which have the solution

$$v_1(t) = \int \frac{-g(t)\,y_2(t)}{W[y_1, y_2](t)}\,dt\ , \quad v_2(t) = \int \frac{g(t)y_1(t)}{W[y_1, y_2](t)}\,dt\ . \tag{12}$$

Note the formulation (12) presumes that the differential equation has been put into standard form [that is, divided by $a_2(t)$].

The proofs of the constant-coefficient versions of these theorems in Sections 4.2 and 4.5 did not make use of the constant-coefficient property, so one can prove them in the general case by literally copying those proofs but interpreting the coefficients as variables. Unfortunately, however, there is no construction analogous to the method of undetermined coefficients for the variable-coefficient case.

What does all this mean? The only stumbling block for our completely solving nonhomogeneous initial value problems for equations with variable coefficients,

$$y'' + p(t)y' + q(t)y = g(t)\ ; \quad y(t_0) = Y_0\ , \quad y'(t_0) = Y_1\ ,$$

is the lack of an explicit procedure for constructing independent solutions to the associated homogeneous equation (10). If we *had* y_1 and y_2 as described in the variation of parameters formula, we could implement (12) to find y_p, formulate the general solution of (11) as $y_p + c_1y_1 + c_2y_2$, and (with the assurance that the Wronskian is nonzero) fit the constants to the initial conditions. But with the exception of the Cauchy–Euler equation and the ponderous power series machinery of Chapter 8, we are stymied at the outset; there is no general procedure for finding y_1 and y_2.

Ironically, we only need *one* nontrivial solution to the associated homogeneous equation, thanks to a procedure known as *reduction of order* that constructs a second, linearly independent solution y_2 from a known one y_1. So one might well feel that the following theorem rubs salt into the wound.

Reduction of Order

Theorem 8. If $y_1(t)$ is a solution, not identically zero, to the homogeneous differential equation (10) in an interval I (see page 195), then

$$y_2(t) = y_1(t) \int \frac{e^{-\int p(t)dt}}{y_1(t)^2}\,dt \tag{13}$$

is a second, linearly independent solution.

This remarkable formula can be confirmed directly, but the following derivation shows how the procedure got its name.

Proof of Theorem 8. Our strategy is similar to that used in the derivation of the variation of parameters formula, Section 4.6. Bearing in mind that cy_1 is a solution of (10) for any constant c, we replace c by a *function* $v(t)$ and propose the trial solution $y_2(t) = v(t)y_1(t)$, spawning the formulas

$$y_2' = vy_1' + v'y_1\,, \quad y_2'' = vy_1'' + 2v'y_1' + v''y_1\,.$$

Substituting these expressions into the differential equation (10) yields

$$(vy_1'' + 2v'y_1' + v''y_1) + p(vy_1' + v'y_1) + qvy_1 = 0\,,$$

or, on regrouping,

$$(y_1'' + py_1' + qy_1)v + y_1v'' + (2y_1' + py_1)v' = 0\,. \tag{14}$$

The group in front of the undifferentiated $v(t)$ is simply a copy of the left-hand member of the original differential equation (10), so it is zero.† Thus (14) reduces to

$$y_1v'' + (2y_1' + py_1)v' = 0\,, \tag{15}$$

which is actually a *first*-order equation in the variable $w \equiv v'$:

$$y_1w' + (2y_1' + py_1)w = 0\,. \tag{16}$$

Indeed, (16) is separable and can be solved immediately using the procedure of Section 2.2. Problem 48 on page 201 requests the reader to carry out the details of this procedure to complete the derivation of (13). ◆

Example 4 Given that $y_1(t) = t$ is a solution to

$$y'' - \frac{1}{t}y' + \frac{1}{t^2}y = 0\,, \tag{17}$$

use the reduction of order procedure to determine a second linearly independent solution for $t > 0$.

---

†This is hardly a surprise; if v were constant, vy would be a solution with $v' = v'' = 0$ in (14).

Solution Rather than implementing the formula (13), let's apply the strategy used to derive it. We set $y_2(t) = v(t)y_1(t) = v(t)t$ and substitute $y_2' = v't + v$, $y_2'' = v''t + 2v'$ into (17) to find

$$v''t + 2v' - \frac{1}{t}(v't + v) + \frac{1}{t^2}vt = v''t + (2v' - v') = v''t + v' = 0\,. \tag{18}$$

As promised, (18) is a *separable first*-order equation in v', simplifying to $(v')'/(v') = -1/t$ with a solution $v' = 1/t$, or $v = \ln t$ (taking integration constants to be zero). Therefore, a second solution to (17) is $y_2 = vt = t\ln t$.

Of course (17) is a Cauchy–Euler equation for which (7) has equal roots:

$$ar^2 + (b - a)r + c = r^2 - 2r + 1 = (r - 1)^2 = 0\,,$$

and y_2 is precisely the form for the independent solution predicted by (9). ◆

Example 5 The following equation arises in the mathematical modeling of reverse osmosis.[†]

$$(\sin t)y'' - 2(\cos t)y' - (\sin t)y = 0\,, \quad 0 < t < \pi\,. \tag{19}$$

Find a general solution.

Solution As we indicated above, the tricky part is to find a single nontrivial solution. Inspection of (19) suggests that $y = \sin t$ or $y = \cos t$, combined with a little luck with trigonometric identities, might be solutions. In fact, trial and error shows that the cosine function works:

$$y_1 = \cos t\,, \quad y_1' = -\sin t\,, \quad y_1'' = -\cos t\,,$$
$$(\sin t)y_1'' - 2(\cos t)y_1' - (\sin t)y_1 = (\sin t)(-\cos t) - 2(\cos t)(-\sin t) - (\sin t)(\cos t) = 0\,.$$

Unfortunately, the sine function fails (try it).

So we use reduction of order to construct a second, independent solution. Setting $y_2(t) = v(t)y_1(t) = v(t)\cos t$ and computing $y_2' = v'\cos t - v\sin t$, $y_2'' = v''\cos t - 2v'\sin t - v\cos t$, we substitute into (19) to derive

$$\begin{aligned}(\sin t)[v''\cos t - 2v'\sin t - v\cos t] - 2(\cos t)[v'\cos t - v\sin t] - (\sin t)[v\cos t]\\ = v''(\sin t)(\cos t) - 2v'(\sin^2 t + \cos^2 t) = 0,\end{aligned}$$

which is equivalent to the separated first-order equation

$$\frac{(v')'}{(v')} = \frac{2}{(\sin t)(\cos t)} = 2\frac{\sec^2 t}{\tan t}\,.$$

Taking integration constants to be zero yields $\ln v' = 2\ln(\tan t)$ or $v' = \tan^2 t$, and $v = \tan t - t$. Therefore, a second solution to (19) is $y_2 = (\tan t - t)\cos t = \sin t - t\cos t$. We conclude that a general solution is $c_1\cos t + c_2(\sin t - t\cos t)$. ◆

In this section we have seen that the *theory* for variable-coefficient equations differs only slightly from the constant-coefficient case (in that solution domains are restricted to intervals), but explicit solutions can be hard to come by. In the next section, we will supplement our exposition by describing some nonrigorous procedures that sometimes can be used to predict qualitative features of the solutions.

[†]Reverse osmosis is a process used to fortify the alcoholic content of wine, among other applications.

4.7 EXERCISES

In Problems 1 through 4, use Theorem 5 to discuss the existence and uniqueness of a solution to the differential equation that satisfies the initial conditions $y(1) = Y_0$, $y'(1) = Y_1$, where Y_0 and Y_1 are real constants.

1. $(1 + t^2)y'' + ty' - y = \tan t$

2. $t(t - 3)y'' + 2ty' - y = t^2$

3. $t^2y'' + y = \cos t$

4. $e^t y'' - \dfrac{y'}{t - 3} + y = \ln t$

In Problems 5 through 8, determine whether Theorem 5 applies. If it does, then discuss what conclusions can be drawn. If it does not, explain why.

5. $t^2z'' + tz' + z = \cos t\,;\quad z(0) = 1\,,\quad z'(0) = 0$

6. $y'' + yy' = t^2 - 1\,;\quad y(0) = 1\,,\quad y'(0) = -1$

7. $y'' + ty' - t^2y = 0\,;\quad y(0) = 0\,,\quad y(1) = 0$

8. $(1 - t)y'' + ty' - 2y = \sin t\,;\quad y(0) = 1\,,\quad y'(0) = 1$

In Problems 9 through 14, find a general solution to the given Cauchy–Euler equation for $t > 0$.

9. $t^2y''(t) + 7ty'(t) - 7y(t) = 0$

10. $t^2\dfrac{d^2y}{dt^2} + 2t\dfrac{dy}{dt} - 6y = 0$

11. $t^2\dfrac{d^2z}{dt^2} + 5t\dfrac{dz}{dt} + 4z = 0$

12. $\dfrac{d^2w}{dt^2} + \dfrac{6}{t}\dfrac{dw}{dt} + \dfrac{4}{t^2}w = 0$

13. $9t^2y''(t) + 15ty'(t) + y(t) = 0$

14. $t^2y''(t) - 3ty'(t) + 4y(t) = 0$

In Problems 15 through 18, find a general solution for $t < 0$.

15. $y''(t) - \dfrac{1}{t}y'(t) + \dfrac{5}{t^2}y(t) = 0$

16. $t^2y''(t) - 3ty'(t) + 6y(t) = 0$

17. $t^2y''(t) + 9ty'(t) + 17y(t) = 0$

18. $t^2y''(t) + 3ty'(t) + 5y(t) = 0$

In Problems 19 and 20, solve the given initial value problem for the Cauchy–Euler equation.

19. $t^2y''(t) - 4ty'(t) + 4y(t) = 0\,;$
$y(1) = -2\,,\quad y'(1) = -11$

20. $t^2y''(t) + 7ty'(t) + 5y(t) = 0\,;$
$y(1) = -1\,,\quad y'(1) = 13$

In Problems 21 and 22, devise a modification of the method for Cauchy–Euler equations to find a general solution to the given equation.

21. $(t - 2)^2y''(t) - 7(t - 2)y'(t) + 7y(t) = 0\,,\quad t > 2$

22. $(t + 1)^2y''(t) + 10(t + 1)y'(t) + 14y(t) = 0\,,$
$t > -1$

23. To justify the solution formulas (8) and (9), perform the following analysis.

(a) Show that if the substitution $t = e^x$ is made in the function $y(t)$ and x is regarded as the new independent variable in $Y(x) := y(e^x)$, the chain rule implies the following relationships:

$$t\frac{dy}{dt} = \frac{dY}{dx}\,,\qquad t^2\frac{d^2y}{dt^2} = \frac{d^2Y}{dx^2} - \frac{dY}{dx}\,.$$

(b) Using part (a), show that if the substitution $t = e^x$ is made in the Cauchy–Euler differential equation (6), the result is a constant-coefficient equation for $Y(x) = y(e^x)$, namely,

$$a\frac{d^2Y}{dx^2} + (b - a)\frac{dY}{dx} + cY = f(e^x)\,. \qquad \textbf{(20)}$$

(c) Observe that the auxiliary equation (recall Section 4.2) for the homogeneous form of (20) is the same as (7) in this section. If the roots of the former are complex, linearly independent solutions of (20) have the form $e^{\alpha x}\cos\beta x$ and $e^{\alpha x}\sin\beta x$; if they are equal, linearly independent solutions of (20) have the form e^{rx} and xe^{rx}. Express x in terms of t to derive the corresponding solution forms (8) and (9).

24. Solve the following Cauchy–Euler equations by using the substitution described in Problem 23 to change them to constant coefficient equations, finding their general solutions by the methods of the preceding sections, and restoring the original independent variable t.

(a) $t^2y'' + ty' - 9y = 0$

(b) $t^2y'' + 3ty' + 10y = 0$

(c) $t^2y'' + 3ty' + y = t + t^{-1}$

(d) $t^2y'' + ty' + 9y = -\tan(3\ln t)$

25. Let y_1 and y_2 be two functions defined on $(-\infty, \infty)$.

(a) True or False: If y_1 and y_2 are linearly dependent on the interval $[a, b]$, then y_1 and y_2 are linearly dependent on the smaller interval $[c, d] \subset [a, b]$.

(b) True or False: If y_1 and y_2 are linearly dependent on the interval $[a, b]$, then y_1 and y_2 are linearly dependent on the larger interval $[C, D] \supset [a, b]$.

26. Let $y_1(t) = t^3$ and $y_2(t) = |t^3|$. Are y_1 and y_2 linearly independent on the following intervals?

(a) $[0, \infty)$ **(b)** $(-\infty, 0]$ **(c)** $(-\infty, \infty)$

(d) Compute the Wronskian $W[y_1, y_2](t)$ on the interval $(-\infty, \infty)$.

27. Consider the linear equation

$$t^2y'' - 3ty' + 3y = 0\,,\quad \text{for } -\infty < t < \infty\,. \qquad \textbf{(21)}$$

(a) Verify that $y_1(t) := t$ and $y_2(t) := t^3$ are two solutions to (21) on $(-\infty, \infty)$. Furthermore, show that $y_1(t_0)y_2'(t_0) - y_1'(t_0)y_2(t_0) \neq 0$ for $t_0 = 1$.

(b) Prove that $y_1(t)$ and $y_2(t)$ are linearly independent on $(-\infty, \infty)$.

(c) Verify that the function $y_3(t) := |t|^3$ is also a solution to (21) on $(-\infty, \infty)$.

(d) Prove that there is *no* choice of constants c_1, c_2 such that $y_3(t) = c_1y_1(t) + c_2y_2(t)$ for *all* t in $(-\infty, \infty)$. [*Hint:* Argue that the contrary assumption leads to a contradiction.]

(e) From parts (c) and (d), we see that there is at least one solution to (21) on $(-\infty, \infty)$ that is not expressible as a linear combination of the solutions $y_1(t)$, $y_2(t)$. Does this provide a counterexample to the theory in this section? Explain.

28. Let $y_1(t) = t^2$ and $y_2(t) = 2t|t|$. Are y_1 and y_2 linearly independent on the interval:

(a) $[0, \infty)$? **(b)** $(-\infty, 0]$? **(c)** $(-\infty, \infty)$?

(d) Compute the Wronskian $W[y_1, y_2](t)$ on the interval $(-\infty, \infty)$.

29. Prove that if y_1 and y_2 are linearly independent solutions of $y'' + py' + qy = 0$ on (a, b), then they cannot both be zero at the same point t_0 in (a, b).

30. Superposition Principle. Let y_1 be a solution to

$$y''(t) + p(t)y'(t) + q(t)y(t) = g_1(t)$$

on the interval I and let y_2 be a solution to

$$y''(t) + p(t)y'(t) + q(t)y(t) = g_2(t)$$

on the same interval. Show that for any constants k_1 and k_2, the function $k_1y_1 + k_2y_2$ is a solution on I to

$$y''(t) + p(t)y'(t) + q(t)y(t) = k_1g_1(t) + k_2g_2(t) .$$

31. Determine whether the following functions can be Wronskians on $-1 < t < 1$ for a pair of solutions to some equation $y'' + py' + qy = 0$ (with p and q continuous).

(a) $w(t) = 6e^{4t}$ **(b)** $w(t) = t^3$

(c) $w(t) = (t + 1)^{-1}$ **(d)** $w(t) \equiv 0$

32. By completing the following steps, prove that the Wronskian of any two solutions y_1, y_2 to the equation $y'' + py' + qy = 0$ on (a, b) is given by **Abel's formula**†

$$W[y_1, y_2](t) = C \exp\left\{-\int_{t_0}^{t} p(\tau)\, d\tau\right\} ,$$

t_0 and t in (a, b) ,

where the constant C depends on y_1 and y_2.

(a) Show that the Wronskian W satisfies the equation $W' + pW = 0$.

(b) Solve the separable equation in part (a).

(c) How does Abel's formula clarify the fact that the Wronskian is either identically zero or never zero on (a, b)?

33. Use Abel's formula (Problem 32) to determine (up to a constant multiple) the Wronskian of two solutions on $(0, \infty)$ to

$$ty'' + (t - 1)y' + 3y = 0 .$$

34. All that is known concerning a mysterious differential equation $y'' + p(t)y' + q(t)y = g(t)$ is that the functions t, t^2, and t^3 are solutions.

(a) Determine two linearly independent solutions to the corresponding homogeneous differential equation.

(b) Find the solution to the original equation satisfying the initial conditions $y(2) = 2$, $y'(2) = 5$.

(c) What is $p(t)$? [*Hint:* Use Abel's formula for the Wronskian, Problem 32.]

35. Given that $1 + t$, $1 + 2t$, and $1 + 3t^2$ are solutions to the differential equation $y'' + p(t)y' + q(t)y = g(t)$, find the solution to this equation that satisfies $y(1) = 2$, $y'(1) = 0$.

36. Verify that the given functions y_1 and y_2 are linearly independent solutions of the following differential equation and find the solution that satisfies the given initial conditions.

$$ty'' - (t + 2)y' + 2y = 0 ;$$
$$y_1(t) = e^t , \quad y_2(t) = t^2 + 2t + 2 ;$$
$$y(1) = 0 , \quad y'(1) = 1$$

In Problems 37 through 39, find general solutions to the nonhomogeneous Cauchy–Euler equations using variation of parameters.

37. $t^2z'' + tz' + 9z = -\tan(3 \ln t)$

38. $t^2y'' + 3ty' + y = t^{-1}$

39. $t^2z'' - tz' + z = t\left(1 + \dfrac{3}{\ln t}\right)$

40. The **Bessel equation** of order one-half

$$t^2y'' + ty' + \left(t^2 - \frac{1}{4}\right)y = 0 , \quad t > 0$$

has two linearly independent solutions,

$$y_1(t) = t^{-1/2} \cos t , \quad y_2(t) = t^{-1/2} \sin t .$$

Find a general solution to the nonhomogeneous equation

$$t^2y'' + ty' + \left(t^2 - \frac{1}{4}\right)y = t^{5/2} , \quad t > 0 .$$

In Problems 41 through 44, a differential equation and a nontrivial solution f are given. Find a second linearly independent solution using reduction of order.

41. $t^2y'' - 2ty' - 4y = 0 , \quad t > 0 ; \quad f(t) = t^{-1}$

42. $t^2y'' + 6ty' + 6y = 0 , \quad t > 0 ; \quad f(t) = t^{-2}$

43. $tx'' - (t + 1)x' + x = 0 , \quad t > 0 ; \quad f(t) = e^t$

44. $ty'' + (1 - 2t)y' + (t - 1)y = 0 , \quad t > 0 ; \quad f(t) = e^t$

† *Historical Footnote:* Niels Abel derived this identity in 1827.

45. Find a particular solution to the nonhomogeneous equation

$$ty'' - (t+1)y' + y = t^2e^{2t},$$

given that $f(t) = e^t$ is a solution to the corresponding homogeneous equation.

46. Find a particular solution to the nonhomogeneous equation

$$(1-t)y'' + ty' - y = (1-t)^2,$$

given that $f(t) = t$ is a solution to the corresponding homogeneous equation.

47. In quantum mechanics, the study of the Schrödinger equation for the case of a harmonic oscillator leads to a consideration of **Hermite's equation,**

$$y'' - 2ty' + \lambda y = 0,$$

where λ is a parameter. Use the reduction of order formula to obtain an integral representation of a second linearly independent solution to Hermite's equation for the given value of λ and corresponding solution $f(t)$.

(a) $\lambda = 4, \quad f(t) = 1 - 2t^2$

(b) $\lambda = 6, \quad f(t) = 3t - 2t^3$

48. Complete the proof of Theorem 8 by solving equation (16).

49. The reduction of order procedure can be used more generally to reduce a homogeneous linear nth-order equation to a homogeneous linear $(n-1)$th-order equation. For the equation

$$ty''' - ty'' + y' - y = 0,$$

which has $f(t) = e^t$ as a solution, use the substitution $y(t) = v(t)f(t)$ to reduce this third-order equation to a homogeneous linear second-order equation in the variable $w = v'$.

50. The equation

$$ty''' + (1-t)y'' + ty' - y = 0$$

has $f(t) = t$ as a solution. Use the substitution $y(t) = v(t)f(t)$ to reduce this third-order equation to a homogeneous linear second-order equation in the variable $w = v'$.

51. **Isolated Zeros.** Let $\phi(t)$ be a solution to $y'' + py' + qy = 0$ on (a, b), where p, q are continuous on (a, b). By completing the following steps, prove that if ϕ is not identically zero, then its zeros in (a, b) are *isolated*, i.e., if $\phi(t_0) = 0$, then there exists a $\delta > 0$ such that $\phi(t) \neq 0$ for $0 < |t - t_0| < \delta$.

(a) Suppose $\phi(t_0) = 0$ and assume to the contrary that for each $n = 1, 2, \ldots$, the function ϕ has a zero at t_n, where $0 < |t_0 - t_n| < 1/n$. Show that this implies $\phi'(t_0) = 0$. [*Hint:* Consider the difference quotient for ϕ at t_0.]

(b) With the assumptions of part (a), we have $\phi(t_0) = \phi'(t_0) = 0$. Conclude from this that ϕ must be identically zero, which is a contradiction. Hence, there is some integer n_0 such that $\phi(t)$ is not zero for $0 < |t - t_0| < 1/n_0$.

52. The reduction of order formula (13) can also be derived from Abels' identity (Problem 32). Let $f(t)$ be a nontrivial solution to (10) and $y(t)$ a second linearly independent solution. Show that

$$\left(\frac{y}{f}\right)' = \frac{W[f, y]}{f^2}$$

and then use Abel's identity for the Wronskian $W[f, y]$ to obtain the reduction of order formula.

4.8 Qualitative Considerations for Variable-Coefficient and Nonlinear Equations

There are no techniques for obtaining explicit, closed-form solutions to second-order linear differential equations with variable coefficients (with certain exceptions) or for nonlinear equations. In general, we will have to settle for numerical solutions or power series expansions. So it would be helpful to be able to derive, with simple calculations, some nonrigorous, qualitative conclusions about the behavior of the solutions before we launch the heavy computational machinery. In this section we first display a few examples that illustrate the profound differences that can occur when the equations have variable coefficients or are nonlinear. Then we show how the mass–spring analogy, discussed in Section 4.1, can be exploited to predict some of the attributes of solutions of these more complicated equations.

To begin our discussion we display a linear constant-coefficient, a linear variable-coefficient, and two nonlinear equations.

(a) The equation

$$3y'' + 2y' + 4y = 0 \tag{1}$$

is linear, homogeneous with constant coefficients. We know everything about such equations; the solutions are, at worst, polynomials times exponentials times sinusoids in t, and unique solutions can be found to match any prescribed data $y(a)$, $y'(a)$ at any instant $t = a$. It has the superposition property: If $y_1(t)$ and $y_2(t)$ are solutions, so is $y(t) = c_1y_1(t) + c_2y_2(t)$.

(b) The equation

$$\boldsymbol{(1 - t^2)y'' - 2ty' + 2y = 0} \tag{2}$$

also has the superposition property (Problem 30, Exercises 4.7). It is a linear variable-coefficient equation and is a special case of **Legendre's equation** $(1 - t^2)y'' - 2ty' + \lambda y = 0$, which arises in the analysis of wave and diffusion phenomena in spherical coordinates.

(c) The equations

$$y'' - 6y^2 = 0\,, \tag{3}$$

$$y'' - 24y^{1/3} = 0 \tag{4}$$

do not share the superposition property because of the square and the cube root of y terms (e.g., the quadratic term y^2 does not reduce to $y_1^2 + y_2^2$). They are *nonlinear*[†] equations.

The Legendre equation (2) has one simple solution, $y_1(t) = t$, as can easily be verified by mental calculation. A second, linearly independent, solution for $-1 < t < 1$ can be derived by the reduction of order procedure of Section 4.7. Traditionally, the second solution is taken to be

$$y_2(t) = \frac{t}{2}\ln\left(\frac{1+t}{1-t}\right) - 1\,. \tag{5}$$

Notice in particular the behavior near $t = \pm 1$; none of the solutions of our *constant*-coefficient equations ever diverged at a finite point!

We would have anticipated troublesome behavior for (2) at $t = \pm 1$ if we had rewritten it in standard form as

$$y'' - \frac{2t}{1 - t^2}y' + \frac{2}{1 - t^2}y = 0\,, \tag{6}$$

since Theorem 5, page 192, only promises existence and uniqueness *between* these points.

As we have noted, there are no general solution procedures for solving nonlinear equations. However, the following lemma is very useful in some situations such as equations (3), (4). It has an extremely significant physical interpretation, which we will explore in Project D of Chapter 5; for now we will merely tantalize the reader by giving it a suggestive name.

[†]Although the quadratic y^2 renders equation (3) nonlinear, the occurrence of t^2 in (2) does not spoil its linearity (in y).

The Energy Integral Lemma

Lemma 4. Let $y(t)$ be a solution to the differential equation

$$y'' = f(y) \, , \tag{7}$$

where $f(y)$ is a continuous function that does not depend on y' or the independent variable t. Let $F(y)$ be an indefinite integral of $f(y)$, that is,

$$f(y) = \frac{d}{dy}F(y) \, .$$

Then the quantity

$$E(t) := \frac{1}{2}y'(t)^2 - F\big(y(t)\big) \tag{8}$$

is constant; i.e.,

$$\frac{d}{dt}E(t) = 0 \, . \tag{9}$$

Proof. This is immediate; we insert (8), differentiate, and apply the differential equation (7):

$$\begin{aligned}\frac{d}{dt}E(t) &= \frac{d}{dt}\left[\frac{1}{2}y'(t)^2 - F\big(y(t)\big)\right] \\ &= \frac{1}{2}2y'y'' - \frac{dF}{dy}y' \\ &= y'[y'' - f(y)] \\ &= 0 \, . \; \blacklozenge\end{aligned}$$

As a result, an equation of the form (7) can be reduced to

$$\frac{1}{2}(y')^2 - F(y) = K \, , \tag{10}$$

for some constant K, which is equivalent to the *separable* first-order equation

$$\frac{dy}{dt} = \pm\sqrt{2[F(y) + K]}$$

having the implicit two-parameter solution (Section 2.2)

$$t = \pm\int\frac{dy}{\sqrt{2[F(y) + K]}} + c \, . \tag{11}$$

We will use formula (11) to illustrate some startling features of nonlinear equations.

Example 1 Apply the energy integral lemma to explore the solutions of the nonlinear equation $y'' = 6y^2$, given in (3).

Solution Since $6y^2 = \frac{d}{dy}(2y^3)$, the solution form (11) becomes

$$t = \pm \int \frac{dy}{\sqrt{2[y^3 + K]}} + c\,.$$

For simplicity we take the plus sign and focus attention on solutions with $K = 0$. Then we find $t = \int \frac{1}{2} y^{-3/2}\,dy = -y^{-1/2} + c$, or

$$y(t) = (c - t)^{-2}\,, \tag{12}$$

for any value of the constant c.

Clearly, this equation is an enigma; solutions can blow up at $t = 1, t = 2, t = \pi$, or anywhere—and there is no clue in equation (3) as to why this should happen! Moreover, we have found an *infinite* family of pairwise linearly independent solutions (rather than the expected *two*). Yet we still cannot assemble, out of these, a solution matching (say) $y(0) = 1, y'(0) = 3$; all our solutions (12) have $y'(0) = 2y(0)^{3/2}$, and the absence of a superposition principle voids the use of linear combinations $c_1y_1(t) + c_2y_2(t)$. ◆

Example 2 Apply the energy integral lemma to explore the solutions of the nonlinear equation $y'' = 24y^{1/3}$, given in (4).

Solution Since $24y^{1/3} = \frac{d}{dy}(18y^{4/3})$, formula (11) gives

$$t = \pm \int \frac{dy}{\sqrt{2[18y^{4/3} + K]}} + c\,.$$

Again we take the plus sign and focus attention on solutions with $K = 0$. Then we find $t = y^{1/3}/2 + c$. In particular, $y_1(t) = 8t^3$ is a solution, and it satisfies the initial conditions $y(0) = 0, y'(0) = 0$. But note that $y_2(t) \equiv 0$ is also a solution to (4), *and it satisfies the same initial conditions*! Hence, the *uniqueness* feature, guaranteed for linear equations by Theorem 5 on page 192 of Section 4.7, can fail in the nonlinear case. ◆

So these examples have demonstrated violations of the existence and uniqueness properties (as well as "finiteness") that we have come to expect from the constant-coefficient case. It should not be surprising, then, that the solution techniques for variable-coefficient and nonlinear second-order equations are more complicated—when, indeed, they exist. Recall, however, that in Section 4.3 we saw that our familiarity with the mass–spring oscillator equation

$$\begin{aligned} F_{\text{ext}} &= [\text{inertia}]y'' + [\text{damping}]y' + [\text{stiffness}]y \\ &= my'' + by' + ky \end{aligned} \tag{13}$$

was helpful in picturing the qualitative features of the solutions of other constant-coefficient equations. (See Figure 4.1, page 152.) By pushing these analogies further, we can also anticipate some of the features of the solutions in the variable-coefficient and nonlinear cases. One of the simplest linear second-order differential equations with variable coefficients is

$$y'' + ty = 0\,. \tag{14}$$

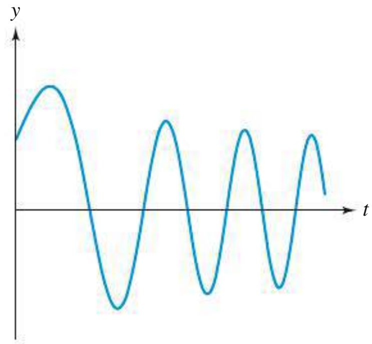


Figure 4.12 Solution to equation (14)

Example 3 Using the mass–spring analogy, predict the nature of the solutions to equation (14) for $t > 0$.

Solution Comparing (13) with (14), we see that the latter equation describes a mass–spring oscillator where the *spring stiffness* "k" varies in time—in fact, it stiffens as time passes ["k" $= t$ in equation (14)]. Physically, then, we would expect to see oscillations whose frequency increases with time, while the amplitude of the oscillations diminishes (because the spring gets harder to stretch). The numerically computed solution in Figure 4.12 displays precisely this behavior. ◆

Remark Schemes for numerically computing solutions to second-order equations will be discussed in Section 5.3. If such schemes are available, then why is there a need for the qualitative analysis discussed in this section? The answer is that numerical methods provide only *approximations* to solutions of initial value problems, and their accuracy is sometimes difficult to predict (especially for nonlinear equations). For example, numerical methods are often ineffective near points of discontinuity or over the long time intervals needed to study the asymptotic behavior. And this is precisely when qualitative arguments can lend insight into the reasonableness of the computed solution.

It is easy to verify (Problem 1) that if $y(t)$ is a solution of the **Airy equation**

(15) $$y'' - ty = 0\,,$$

then $y(-t)$ solves $y'' + ty = 0$, so "Airy functions" exhibit the behavior shown in Figure 4.12 for *negative time*. For positive t the Airy equation has a *negative* stiffness "k" $= -t$, with magnitude increasing in time. As we observed in Example 5 of Section 4.3, negative stiffness tends to reinforce, rather than oppose, displacements, and the solutions $y(t)$ grow rapidly with (positive) time. The solution known as the Airy function of the second kind $\mathrm{Bi}(t)$, depicted in Figure 4.13, behaves exactly as expected.

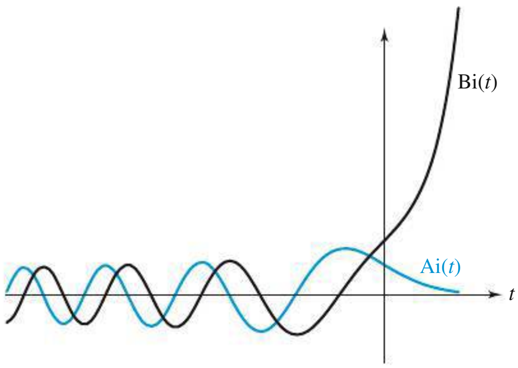


Figure 4.13 Airy functions

In Section 4.3 we also pointed out that mass–spring systems with negative spring stiffness can have isolated *bounded* solutions if the initial displacement $y(0)$ and velocity $y'(0)$ are precisely selected to counteract the repulsive spring force. The Airy function of the first kind $\mathrm{Ai}(t)$, also depicted in Figure 4.13, is such a solution for the "Airy spring"; the initial inwardly directed velocity is just adequate to overcome the outward push of the stiffening spring, and the mass approaches a delicate equilibrium state $y(t) \equiv 0$.

Now let's look at **Bessel's equation.** It arises in the analysis of wave or diffusion phenomena in cylindrical coordinates. The *Bessel equation of order n* is written

$$\mathbf{(16)} \qquad y'' + \frac{1}{t}y' + \left(1 - \frac{n^2}{t^2}\right)y = 0\,.$$

Clearly, there are irregularities at $t = 0$, analogous to those at $t = \pm 1$ for the Legendre equation (2); we will explore these in depth in Chapter 8.

Example 4 Apply the mass–spring analogy to predict qualitative features of solutions to Bessel's equation for $t > 0$.

Solution Comparing (16) with the paradigm (13), we observe that

- the inertia "m" $= 1$ is fixed at unity;
- there is positive damping ("b" $= 1/t$), although it weakens with time; and
- the stiffness ("k" $= 1 - n^2/t^2$) is positive when $t > n$ and tends to 1 as $t \to +\infty$.

Solutions, then, should be expected to oscillate with amplitudes that diminish slowly (due to the damping), and the frequency of the oscillations should settle at a constant value (given, according to the procedures of Section 4.3, by $\sqrt{k/m} = 1$ radian per unit time). The graphs of the Bessel functions $J_n(t)$ and $Y_n(t)$ of the first and second kind of order $n = \frac{1}{2}$ exemplify these qualitative predictions; see Figure 4.14. The effect of the singularities in the coefficients at $t = 0$ is manifested in the graph of $Y_{1/2}(t)$. ◆

Although most Bessel functions have to be computed by power series methods, if the *order n* is a half-integer, then $J_n(t)$ and $Y_n(t)$ have closed-form expressions. In fact, $J_{1/2}(t) = \sqrt{2/(\pi t)}\sin t$ and $Y_{1/2}(t) = -\sqrt{2/(\pi t)}\cos t$. You can verify directly that these functions solve equation (16).

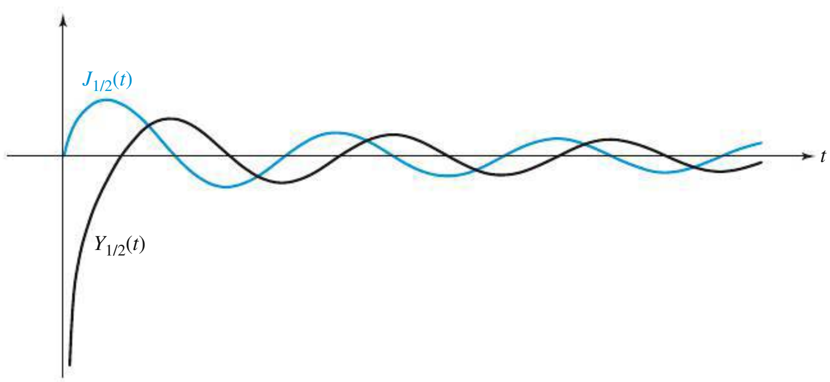


Figure 4.14 Bessel functions

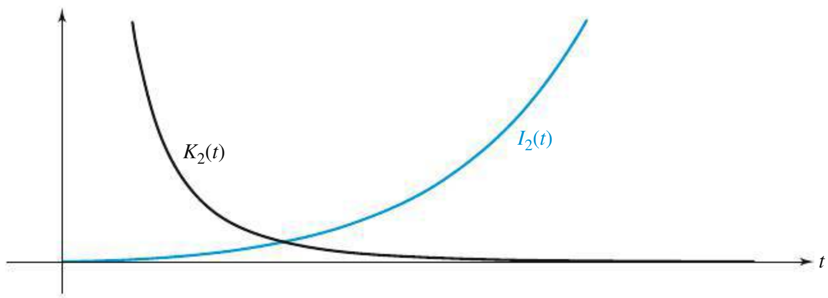


Figure 4.15 Modified Bessel functions

Example 5 Give a qualitative analysis of the **modified Bessel equation** of order n:

$$y'' + \frac{1}{t}y' - \left(1 - \frac{n^2}{t^2}\right)y = 0 . \tag{17}$$

Solution This equation also exhibits unit mass and positive, diminishing, damping. However, the stiffness now converges to *negative* 1. Accordingly, we expect typical solutions to diverge as $t \to +\infty$. The modified Bessel function of the first kind, $I_n(t)$ of order $n = 2$ in Figure 4.15 follows this prediction, whereas the modified Bessel function of the second kind, $K_n(t)$ of order $n = 2$ in Figure 4.15, exhibits the same sort of balance of initial position and velocity as we saw for the Airy function $\mathrm{Ai}(t)$. Again, the effect of the singularity at $t = 0$ is evident. ◆

Example 6 Use the mass–spring model to predict qualitative features of the solutions to the nonlinear **Duffing equation**

$$y'' + y + y^3 = y'' + (1 + y^2)y = 0 . \tag{18}$$

Solution Although equation (18) is nonlinear, it can be matched with the paradigm (13) if we envision unit mass, no damping, and a (positive) stiffness "k" $= 1 + y^2$, which increases as the *displacement* y gets larger. (This increasing-stiffness effect is built into some popular mattresses for therapeutic reasons.)† Such a spring grows stiffer as the mass moves farther away, but it restores to its original value when the mass returns. Thus, high-amplitude excursions should oscillate faster than low-amplitude ones, and the sinusoidal shapes in the graphs of $y(t)$ should be "pinched in" somewhat at their peaks. These qualitative predictions are demonstrated by the numerically computed solutions plotted in Figure 4.16. ◆

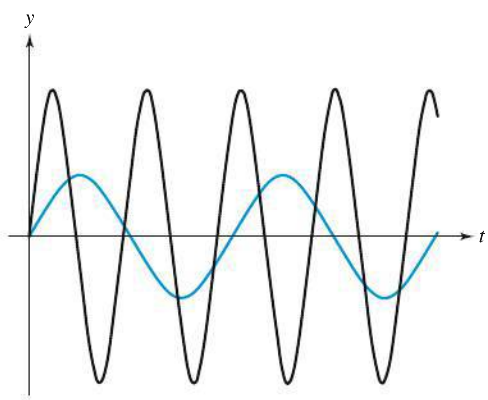


Figure 4.16 Solution graphs for the Duffing equation

†Graphic depictions of oscillations on a therapeutic mattress are best left to one's imagination, the editors say.

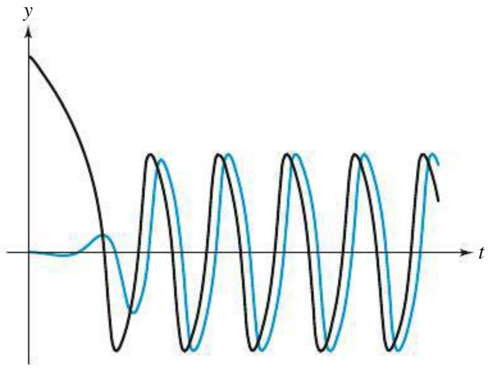


Figure 4.17 Solutions to the van der Pol equation

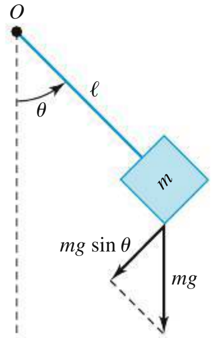


Figure 4.18 A pendulum

The fascinating **van der Pol equation**

$$y'' - (1 - y^2)y' + y = 0 \tag{19}$$

originated in the study of the electrical oscillations observed in vacuum tubes.

Example 7 Predict the behavior of the solutions to equation (19) using the mass–spring model.

Solution By comparison to the paradigm (13), we observe unit mass and stiffness, positive damping [“b” $= -(1-y^2)$] when $|y(t)| > 1$, and negative damping when $|y(t)| < 1$. *Friction thus dampens large-amplitude motions but energizes small oscillations*. The result, then, is that all (nonzero) solutions tend to a *limit cycle* whose friction penalty incurred while $|y(t)| > 1$ is balanced by the negative-friction boost received while $|y(t)| < 1$. The computer-generated Figure 4.17 illustrates the convergence to the limit cycle for some solutions to the van der Pol equation†. ◆

Finally, we consider the motion of the pendulum depicted in Figure 4.18. This motion is measured by the angle $\theta(t)$ that the pendulum makes with the vertical line through O at time t. As the diagram shows, the component of gravity, which exerts a torque on the pendulum and thus accelerates the angular velocity $d\theta/dt$, is given by $-mg \sin\theta$. Consequently, the rotational analog of Newton’s second law, *torque equals rate of change of angular momentum,* dictates (see Problem 7)

$$m\ell^2\theta'' = -\ell mg \sin\theta \,, \tag{20}$$

or

$$m\theta'' + m\frac{g}{\ell}\sin\theta = 0 \,. \tag{21}$$

Example 8 Give a qualitative analysis of the motion of the pendulum.

Solution If we rewrite (21) as

$$m\theta'' + \frac{mg}{\ell}\frac{\sin\theta}{\theta}\theta = 0$$

---

†The vdP equation also provides a model for the world's most accurate pendulum clock, a detail of which appears on the cover.

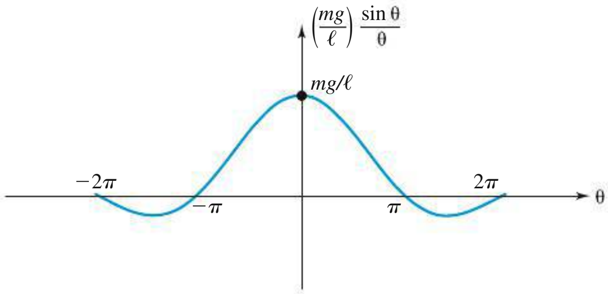


Figure 4.19 Pendulum "stiffness"

and compare with the paradigm (13), we see fixed mass m, no damping, and a stiffness given by

$$\text{“}k\text{”} = \frac{mg}{\ell}\frac{\sin\theta}{\theta}.$$

This stiffness is plotted in Figure 4.19, where we see that small-amplitude motions are driven by a nearly constant spring stiffness of value mg/ℓ, and the considerations of Section 4.3 dictate the familiar formula

$$\omega = \sqrt{\frac{k}{m}} = \sqrt{\frac{g}{\ell}}$$

for the angular frequency of the nearly sinusoidal oscillations. See Figure 4.20, which compares a computer-generated solution to equation (21) to the solution of the constant-stiffness equation with the same initial conditions.[†]

For larger motions, however, the diminishing stiffness distorts the sinusoidal nature of the graph of $\theta(t)$, and lowers the frequency. This is evident in Figure 4.21 on page 210.

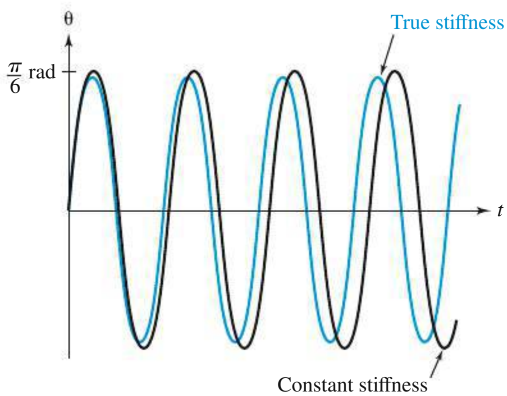


Figure 4.20 Small-amplitude pendulum motion

[†]The latter is identified as the *linearized equation* in Project D on page 236.

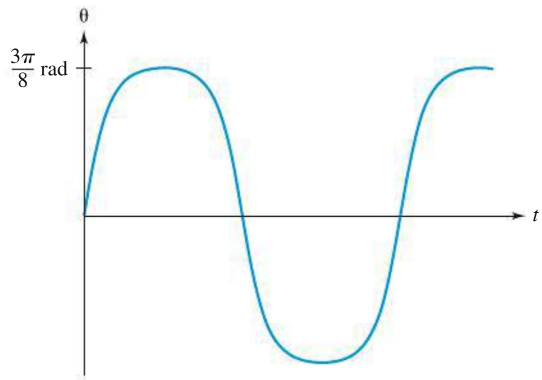


Figure 4.21 Large-amplitude pendulum motion

Finally, if the motion is so energetic that θ reaches the value π, the stiffness changes sign and abets the displacement; the pendulum passes the apex and gains speed as it falls, and this spinning motion repeats continuously. See Figure 4.22. ◆

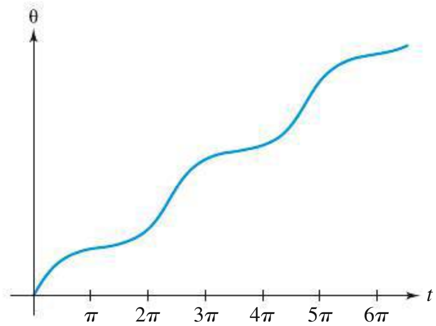


Figure 4.22 Very-large-amplitude pendulum motion

The computation of the solutions of the Legendre, Bessel, and Airy equations and the analysis of the nonlinear equations of Duffing, van der Pol, and the pendulum have challenged many of the great mathematicians of the past. It is gratifying, then, to note that so many of their salient features are susceptible to the qualitative reasoning we have used herein.

4.8 EXERCISES

1. Show that if $y(t)$ satisfies $y'' - ty = 0$, then $y(-t)$ satisfies $y'' + ty = 0$.
2. Using the paradigm (13), what are the inertia, damping, and stiffness for the equation $y'' - 6y^2 = 0$? If $y > 0$, what is the sign of the "stiffness constant"? Does your answer help explain the runaway behavior of the solutions $y(t) = 1/(c-t)^2$?
3. Try to predict the qualitative features of the solution to $y'' - 6y^2 = 0$ that satisfies the initial conditions $y(0) = -1, y'(0) = -1$. Compare with the computer-generated Figure 4.23. [*Hint:* Consider the sign of the spring stiffness.]

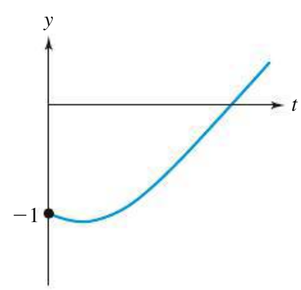


Figure 4.23 Solution for Problem 3

4. Show that the *three* solutions $1/(1-t)^2$, $1/(2-t)^2$, and $1/(3-t)^2$ to $y'' - 6y^2 = 0$ are linearly independent on $(-1, 1)$. (See Problem 35, Exercises 4.2, page 164.)

5. **(a)** Use the energy integral lemma to derive the family of solutions $y(t) = 1/(t-c)$ to the equation $y'' = 2y^3$.

(b) For $c \neq 0$ show that these solutions are pairwise linearly independent for different values of c in an appropriate interval around $t = 0$.

(c) Show that none of these solutions satisfies the initial conditions $y(0) = 1, y'(0) = 2$.

6. Use the energy integral lemma to show that motions of the free undamped mass–spring oscillator $my'' + ky = 0$ obey

$$m(y')^2 + ky^2 = \text{ constant }.$$

7. **Pendulum Equation**. To derive the pendulum equation (21), complete the following steps.

(a) The *angular momentum* of the pendulum mass m measured about the support O in Figure 4.18 on page 208 is given by the product of the "lever arm" length ℓ and the component of the vector momentum mv perpendicular to the lever arm. Show that this gives

$$\text{angular momentum } = m\ell^2\frac{d\theta}{dt}.$$

(b) The *torque* produced by gravity equals the product of the lever arm length ℓ and the component of gravitational (vector) force $m\mathbf{g}$ perpendicular to the lever arm. Show that this gives

$$\text{torque } = -\ell mg \sin\theta .$$

(c) Now use Newton's law of rotational motion to deduce the pendulum equation (20).

8. Use the energy integral lemma to show that pendulum motions obey

$$\frac{(\theta')^2}{2} - \frac{g}{\ell}\cos\theta = \text{ constant }.$$

9. Use the result of Problem 8 to find the value of $\theta'(0)$, the initial velocity, that must be imparted to a pendulum at rest to make it approach (but not cross over) the apex of its motion. Take $\ell = g$ for simplicity.

10. Use the result of Problem 8 to prove that if the pendulum in Figure 4.18 on page 208 is released *from rest* at the angle α, $0 < \alpha < \pi$, then $|\theta(t)| \le \alpha$ for all t. [*Hint:* The initial conditions are $\theta(0) = \alpha, \theta'(0) = 0$; argue that the *constant* in Problem 8 equals $-(g/\ell)\cos\alpha$.]

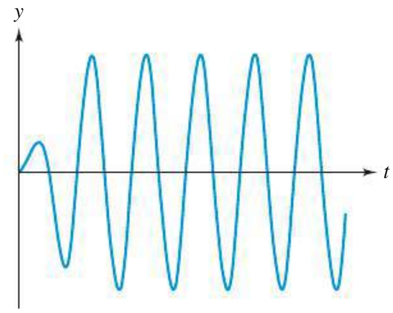


Figure 4.24 Solution to the Rayleigh equation

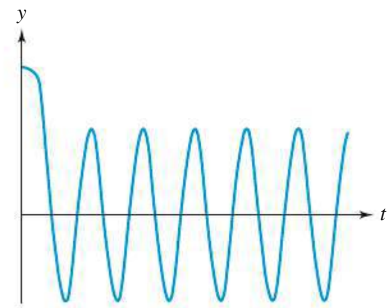


Figure 4.25 Solution to the Rayleigh equation

11. Use the mass–spring analogy to explain the qualitative nature of the solutions to the **Rayleigh equation**

$$y'' - [1 - (y')^2]y' + y = 0 \tag{22}$$

depicted in Figures 4.24 and 4.25.

12. Use reduction of order to derive the solution $y_2(t)$ in equation (5) for Legendre's equation.

13. Figure 4.26 contains graphs of solutions to the Duffing, Airy, and van der Pol equations. Try to match the solution to the equation.

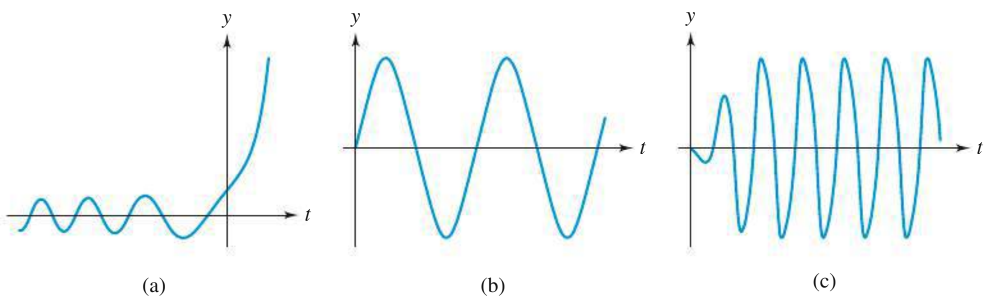


Figure 4.26 Solution graphs for Problem 13

14. Verify that the formulas for the Bessel functions $J_{1/2}(t)$, $Y_{1/2}(t)$ do indeed solve equation (16).

15. Use the mass–spring oscillator analogy to decide whether all solutions to each of the following differential equations are bounded as $t \to +\infty$.

(a) $y'' + t^2 y = 0$ **(b)** $y'' - t^2 y = 0$

(c) $y'' + y^5 = 0$ **(d)** $y'' + y^6 = 0$

(e) $y'' + (4 + 2\cos t)y = 0$ (**Mathieu's equation**)

(f) $y'' + ty' + y = 0$

(g) $y'' - ty' - y = 0$

16. Use the energy integral lemma to show that every solution to the Duffing equation (18) is bounded; that is, $|y(t)| \le M$ for some M. [*Hint:* First argue that $y^2/2 + y^4/4 \le K$ for some K.]

17. Armageddon. Earth revolves around the sun in an approximately circular orbit with radius $r = a$, completing a revolution in the time $T = 2\pi(a^3/GM)^{1/2}$, which is one Earth year; here M is the mass of the sun and G is the universal gravitational constant. The gravitational force of the sun on Earth is given by GMm/r^2, where m is the mass of Earth. Therefore, if Earth "stood still," losing its orbital velocity, it would fall on a straight line into the sun in accordance with Newton's second law:

$$m\frac{d^2r}{dt^2} = -\frac{GMm}{r^2}\,.$$

If this calamity occurred, what fraction of the normal year T would it take for Earth to splash into the sun (i.e., achieve $r = 0$)? [*Hint:* Use the energy integral lemma and the initial conditions $r(0) = a$, $r'(0) = 0$.]

4.9 A Closer Look at Free Mechanical Vibrations

In this section we return to the mass–spring system depicted in Figure 4.1 (page 152) and analyze its motion in more detail. The governing equation is

$$F_{\text{ext}} = [\text{inertia}]\frac{d^2y}{dt^2} + [\text{damping}]\frac{dy}{dt} + [\text{stiffness}]y = my'' + by' + ky\,. \tag{1}$$

Let's focus on the simple case in which $b = 0$ and $F_{\text{ext}} = 0$, the so-called **undamped, free** case. Then equation (1) reduces to

$$m\frac{d^2y}{dt^2} + ky = 0 \tag{2}$$

and, when divided by m, becomes

$$\frac{d^2y}{dt^2} + \omega^2 y = 0\,, \tag{3}$$

where $\omega = \sqrt{k/m}$. The auxiliary equation associated with (3) is $r^2 + \omega^2 = 0$, which has complex conjugate roots $\pm\omega i$. Hence, a general solution to (3) is

$$y(t) = C_1\cos\omega t + C_2\sin\omega t\,. \tag{4}$$

We can express $y(t)$ in the more convenient form

$$y(t) = A\sin(\omega t + \phi)\,, \tag{5}$$

with $A \ge 0$, by letting $C_1 = A\sin\phi$ and $C_2 = A\cos\phi$. That is,

$$\begin{aligned} A\sin(\omega t + \phi) &= A\cos\omega t\sin\phi + A\sin\omega t\cos\phi \\ &= C_1\cos\omega t + C_2\sin\omega t\,. \end{aligned}$$

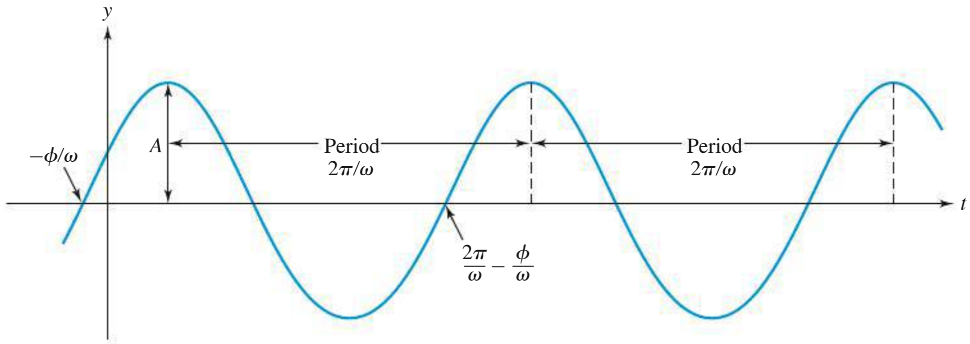


Figure 4.27 Simple harmonic motion of undamped, free vibrations

Solving for A and ϕ in terms of C_1 and C_2, we find

$$\textbf{(6)} \qquad A = \sqrt{C_1^2 + C_2^2} \quad \text{and} \quad \tan\phi = \frac{C_1}{C_2},$$

where the quadrant in which ϕ lies is determined by the signs of C_1 and C_2. This is because $\sin\phi$ has the same sign as C_1 $(\sin\phi = C_1/A)$ and $\cos\phi$ has the same sign as C_2 $(\cos\phi = C_2/A)$. For example, if $C_1 > 0$ and $C_2 < 0$, then ϕ is in Quadrant II. (Note, in particular, that ϕ is not simply the arctangent of C_1/C_2, which would lie in Quadrant IV.)

It is evident from (5) that, as we predicted in Section 4.1, the motion of a mass in an *undamped, free* system is simply a sine wave, or what is called **simple harmonic motion.** (See Figure 4.27.) The constant A is the amplitude of the motion and ϕ is the phase angle. The motion is periodic with **period** $2\pi/\omega$ and **natural frequency** $\omega/2\pi$, where $\omega = \sqrt{k/m}$. The period is measured in units of time, and the natural frequency has the dimensions of periods (or cycles) per unit time. The constant ω is the **angular frequency** for the sine function in (5) and has dimensions of radians per unit time. To summarize:

$$\text{angular frequency} = \omega = \sqrt{k/m} \quad (\text{rad/sec}),$$
$$\text{natural frequency} = \omega/2\pi \quad (\text{cycles/sec}),$$
$$\text{period} = 2\pi/\omega \quad (\text{sec}).$$

Observe that the amplitude and phase angle depend on the constants C_1 and C_2, which, in turn, are determined by the initial position and initial velocity of the mass. However, the period and frequency depend only on k and m and not on the initial conditions.

Example 1 A $1/8$-kg mass is attached to a spring with stiffness $k = 16$ N/m, as depicted in Figure 4.1. The mass is displaced $1/2$ m to the right of the equilibrium point and given an outward velocity (to the right) of $\sqrt{2}$ m/sec. Neglecting any damping or external forces that may be present, determine the equation of motion of the mass along with its amplitude, period, and natural frequency. How long after release does the mass pass through the equilibrium position?

Solution Because we have a case of undamped, free vibration, the equation governing the motion is (3). Thus, we find the angular frequency to be

$$\omega = \sqrt{\frac{k}{m}} = \sqrt{\frac{16}{1/8}} = 8\sqrt{2} \text{ rad/sec}.$$

Substituting this value for ω into (4) gives

$$\textbf{(7)} \qquad y(t) = C_1 \cos(8\sqrt{2}\,t) + C_2 \sin(8\sqrt{2}\,t).$$

Now we use the initial conditions, $y(0) = 1/2$ m and $y'(0) = \sqrt{2}$ m/sec, to solve for C_1 and C_2 in (7). That is,

$$1/2 = y(0) = C_1 ,$$
$$\sqrt{2} = y'(0) = 8\sqrt{2}C_2 ,$$

and so $C_1 = 1/2$ and $C_2 = 1/8$. Hence, the equation of motion of the mass is

(8) $$y(t) = \frac{1}{2}\cos(8\sqrt{2}t) + \frac{1}{8}\sin(8\sqrt{2}t) .$$

To express $y(t)$ in the alternative form (5), we set

$$A = \sqrt{C_1^2 + C_2^2} = \sqrt{(1/2)^2 + (1/8)^2} = \frac{\sqrt{17}}{8} ,$$
$$\tan\phi = \frac{C_1}{C_2} = \frac{1/2}{1/8} = 4 .$$

Since both C_1 and C_2 are positive, ϕ is in Quadrant I, so $\phi = \arctan 4 \approx 1.326$. Hence,

(9) $$y(t) = \frac{\sqrt{17}}{8}\sin(8\sqrt{2}t + \phi) .$$

Thus, the amplitude A is $\sqrt{17}/8$ m, and the phase angle ϕ is approximately 1.326 rad. The period is $P = 2\pi/\omega = 2\pi/(8\sqrt{2}) = \sqrt{2}\pi/8$ sec, and the natural frequency is $1/P = 8/(\sqrt{2}\pi)$ cycles per sec.

Finally, to determine when the mass will pass through the equilibrium position, $y = 0$, we must solve the trigonometric equation

(10) $$y(t) = \frac{\sqrt{17}}{8}\sin(8\sqrt{2}t + \phi) = 0$$

for t. Equation (10) will be satisfied whenever

(11) $$8\sqrt{2}t + \phi = n\pi \qquad \text{or} \qquad t = \frac{n\pi - \phi}{8\sqrt{2}} \approx \frac{n\pi - 1.326}{8\sqrt{2}} ,$$

n an integer. Putting $n = 1$ in (11) determines the first time t when the mass crosses its equilibrium position:

$$t = \frac{\pi - \phi}{8\sqrt{2}} \approx 0.16 \text{ sec} . \ ◆$$

In most applications of vibrational analysis, of course, there is some type of frictional or damping force affecting the vibrations. This force may be due to a component in the system, such as a shock absorber in a car, or to the medium that surrounds the system, such as air or some liquid. So we turn to a study of the effects of damping on free vibrations, and equation (2) generalizes to

(12) $$m\frac{d^2y}{dt^2} + b\frac{dy}{dt} + ky = 0 .$$

The auxiliary equation associated with (12) is

(13) $$mr^2 + br + k = 0 ,$$

and its roots are

(14) $$\frac{-b \pm \sqrt{b^2 - 4mk}}{2m} = -\frac{b}{2m} \pm \frac{1}{2m}\sqrt{b^2 - 4mk}\,.$$

As we found in Sections 4.2 and 4.3, the form of the solution to (12) depends on the nature of these roots and, in particular, on the discriminant $b^2 - 4mk$.

Underdamped or Oscillatory Motion ($b^2 < 4mk$)

When $b^2 < 4mk$, the discriminant $b^2 - 4mk$ is negative, and there are two complex conjugate roots to the auxiliary equation (13). These roots are $\alpha \pm i\beta$, where

(15) $$\alpha := -\frac{b}{2m}\,, \quad \beta := \frac{1}{2m}\sqrt{4mk - b^2}\,.$$

Hence, a general solution to (12) is

(16) $$y(t) = e^{\alpha t}(C_1 \cos \beta t + C_2 \sin \beta t)\ .$$

As we did with simple harmonic motion, we can express $y(t)$ in the alternate form

(17) $$y(t) = Ae^{\alpha t}\sin(\beta t + \phi)\ ,$$

where $A = \sqrt{C_1^2 + C_2^2}$ and $\tan \phi = C_1/C_2$. It is now evident that $y(t)$ is the product of an exponential **damping factor,**

$$Ae^{\alpha t} = Ae^{-(b/2m)t}\,,$$

and a sine factor, $\sin(\beta t + \phi)$, which accounts for the oscillatory motion. Because the sine factor varies between –1 and 1 with period $2\pi/\beta$, the solution $y(t)$ varies between $-Ae^{\alpha t}$ and $Ae^{\alpha t}$ with **quasiperiod** $P = 2\pi/\beta = 4m\pi/\sqrt{4mk - b^2}$ and **quasifrequency** $1/P$. Moreover, since b and m are positive, $\alpha = -b/2m$ is negative, and thus the exponential factor tends to zero as $t \to +\infty$. A graph of a typical solution $y(t)$ is given in Figure 4.28. The system is called **underdamped** because there is not enough damping present (b is too small) to prevent the system from oscillating.

It is easily seen that as $b \to 0$ the damping factor approaches the constant A and the quasi-frequency approaches the natural frequency of the corresponding undamped harmonic motion. Figure 4.28 demonstrates that the values of t where the graph of $y(t)$ touches the exponential curves $\pm Ae^{\alpha t}$ are close to (but not exactly) the same values of t at which $y(t)$ attains its relative maximum and minimum values (see Problem 13 on page 221).

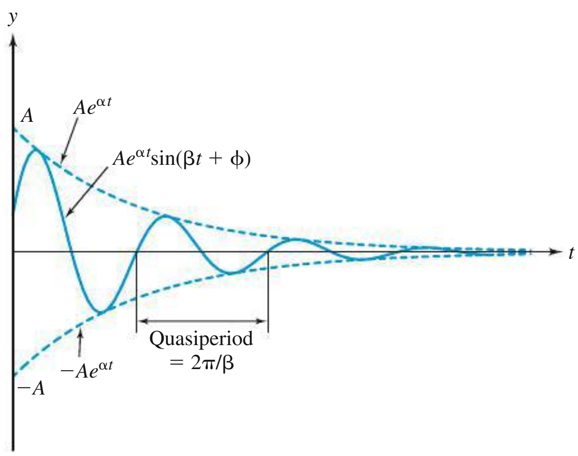


Figure 4.28 Damped oscillatory motion

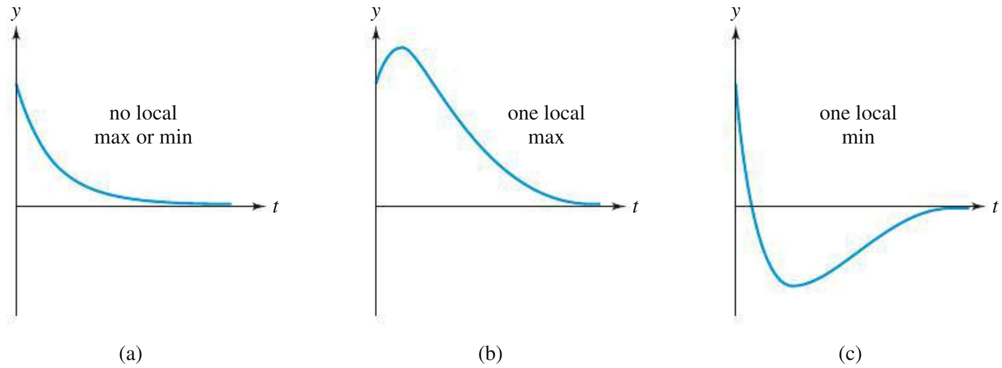


Figure 4.29 Overdamped vibrations

Overdamped Motion ($b^2 > 4mk$)

When $b^2 > 4mk$, the discriminant $b^2 - 4mk$ is positive, and there are two distinct real roots to the auxiliary equation (13):

(18) $$r_1 = -\frac{b}{2m} + \frac{1}{2m}\sqrt{b^2 - 4mk}\,, \quad r_2 = -\frac{b}{2m} - \frac{1}{2m}\sqrt{b^2 - 4mk}\,.$$

Hence, a general solution to (12) in this case is

(19) $$y(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t}\,.$$

Obviously, r_2 is negative. And since $b^2 > b^2 - 4mk$ (that is, $b > \sqrt{b^2 - 4mk}$), it follows that r_1 is also negative. Therefore, as $t \to +\infty$, both of the exponentials in (19) decay and $y(t) \to 0$. Moreover, since

$$y'(t) = C_1 r_1 e^{r_1 t} + C_2 r_2 e^{r_2 t} = e^{r_1 t}\left(C_1 r_1 + C_2 r_2 e^{(r_2 - r_1)t}\right)\,,$$

we see that the derivative is either identically zero (when $C_1 = C_2 = 0$) or vanishes for at most one value of t (when the factor in parentheses is zero). If the trivial solution $y(t) \equiv 0$ is ignored, it follows that $y(t)$ has at most one local maximum or minimum for $t > 0$. Therefore, $y(t)$ *does not oscillate*. This leaves, qualitatively, only three possibilities for the motion of $y(t)$, depending on the initial conditions. These are illustrated in Figure 4.29. This case where $b^2 > 4mk$ is called **overdamped** motion.

Critically Damped Motion ($b^2 = 4mk$)

When $b^2 = 4mk$, the discriminant $b^2 - 4mk$ is zero, and the auxiliary equation has the repeated root $-b/2m$. Hence, a general solution to (12) is now

(20) $$y(t) = (C_1 + C_2 t)e^{-(b/2m)t}\,.$$

To understand the motion described by $y(t)$ in (20), we first consider the behavior of $y(t)$ as $t \to +\infty$. By L'Hôpital's rule,

(21) $$\lim_{t \to +\infty} y(t) = \lim_{t \to +\infty} \frac{C_1 + C_2 t}{e^{(b/2m)t}} = \lim_{t \to +\infty} \frac{C_2}{(b/2m)e^{(b/2m)t}} = 0$$

(recall that $b/2m > 0$). Hence, $y(t)$ dies off to zero as $t \to +\infty$. Next, since

$$y'(t) = \left(C_2 - \frac{b}{2m}C_1 - \frac{b}{2m}C_2 t\right)e^{-(b/2m)t},$$

we see again that a nontrivial solution can have at most one local maximum or minimum for $t > 0$, so motion is *non*oscillatory. If b were any smaller, oscillation would occur. Thus, the special case where $b^2 = 4mk$ is called **critically damped** motion. Qualitatively, critically damped motions are similar to overdamped motions (see Figure 4.29 again).

Example 2 Assume that the motion of a mass–spring system with damping is governed by

$$\frac{d^2y}{dt^2} + b\frac{dy}{dt} + 25y = 0\,; \quad y(0) = 1\,, \quad y'(0) = 0\,. \tag{22}$$

Find the equation of motion and sketch its graph for the three cases where $b = 6$, 10, and 12.

Solution The auxiliary equation for (22) is

$$r^2 + br + 25 = 0\,, \tag{23}$$

whose roots are

$$r = -\frac{b}{2} \pm \frac{1}{2}\sqrt{b^2 - 100}\,. \tag{24}$$

Case 1. When $b = 6$, the roots (24) are $-3 \pm 4i$. This is thus a case of underdamping, and the equation of motion has the form

$$y(t) = C_1 e^{-3t}\cos 4t + C_2 e^{-3t}\sin 4t\,. \tag{25}$$

Setting $y(0) = 1$ and $y'(0) = 0$ gives the system

$$C_1 = 1\,, \quad -3C_1 + 4C_2 = 0\,,$$

whose solution is $C_1 = 1$, $C_2 = 3/4$. To express $y(t)$ as the product of a damping factor and a sine factor [recall equation (17)], we set

$$A = \sqrt{C_1^2 + C_2^2} = \frac{5}{4}\,, \quad \tan\phi = \frac{C_1}{C_2} = \frac{4}{3}\,,$$

where ϕ is a Quadrant I angle, since C_1 and C_2 are both positive. Then

$$y(t) = \frac{5}{4}e^{-3t}\sin(4t + \phi)\,, \tag{26}$$

where $\phi = \arctan(4/3) \approx 0.9273$. The underdamped spring motion is shown in Figure 4.30(a) on page 218.

Case 2. When $b = 10$, there is only one (repeated) root to the auxiliary equation (23), namely, $r = -5$. This is a case of critical damping, and the equation of motion has the form

$$y(t) = (C_1 + C_2 t)e^{-5t}\,. \tag{27}$$

Setting $y(0) = 1$ and $y'(0) = 0$ now gives

$$C_1 = 1\,, \quad C_2 - 5C_1 = 0\,,$$

and so $C_1 = 1$, $C_2 = 5$. Thus,

$$y(t) = (1 + 5t)e^{-5t}\,. \tag{28}$$

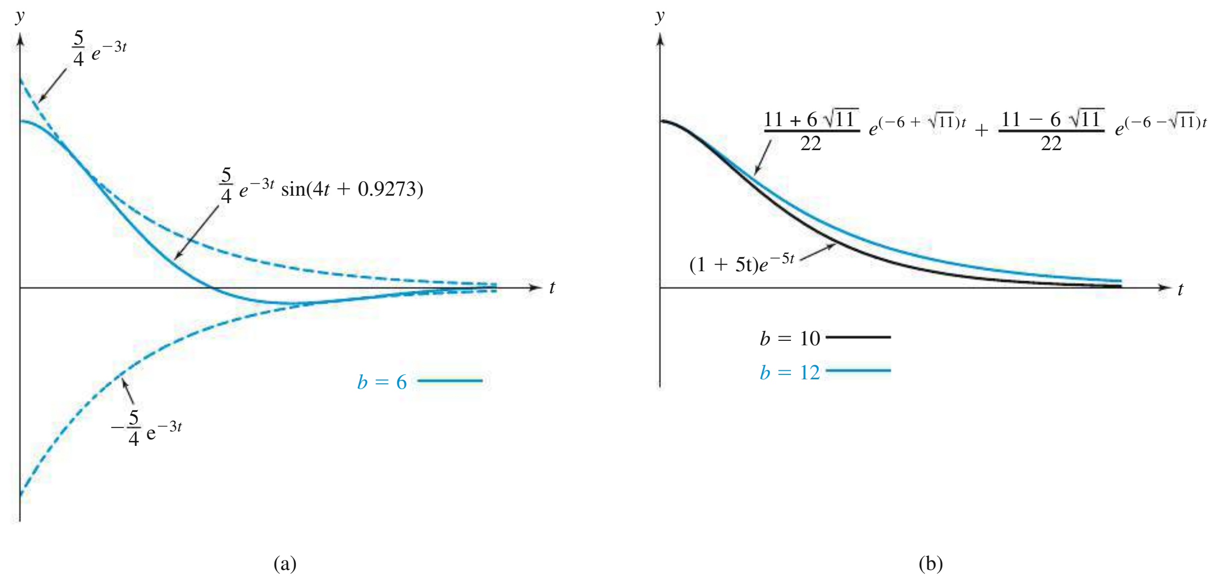


Figure 4.30 Solutions for various values of b

The graph of $y(t)$ given in (28) is represented by the lower curve in Figure 4.30(b). Notice that $y(t)$ is zero only for $t = -1/5$ and hence does not cross the t-axis for $t > 0$.

Case 3. When $b = 12$, the roots to the auxiliary equation are $-6 \pm \sqrt{11}$. This is a case of overdamping, and the equation of motion has the form

$$y(t) = C_1 e^{(-6+\sqrt{11})t} + C_2 e^{(-6-\sqrt{11})t} . \tag{29}$$

Setting $y(0) = 1$ and $y'(0) = 0$ gives

$$C_1 + C_2 = 1 , \quad (-6+\sqrt{11})C_1 + (-6-\sqrt{11})C_2 = 0 ,$$

from which we find $C_1 = (11 + 6\sqrt{11})/22$ and $C_2 = (11 - 6\sqrt{11})/22$. Hence,

$$\begin{aligned} y(t) &= \frac{11+6\sqrt{11}}{22}e^{(-6+\sqrt{11})t} + \frac{11-6\sqrt{11}}{22}e^{(-6-\sqrt{11})t} \\ &= \frac{e^{(-6+\sqrt{11})t}}{22}\left\{11 + 6\sqrt{11} + (11 - 6\sqrt{11})e^{-2\sqrt{11}t}\right\} . \end{aligned} \tag{30}$$

The graph of this overdamped motion is represented by the upper curve in Figure 4.30(b). ◆

It is interesting to observe in Example 2 that when the system is underdamped $(b = 6)$, the solution goes to zero like e^{-3t}; when the system is critically damped $(b = 10)$, the solution tends to zero roughly like e^{-5t}; and when the system is overdamped $(b = 12)$, the solution goes to zero like $e^{(-6+\sqrt{11})t} \approx e^{-2.68t}$. This means that if the system is underdamped, it not only oscillates but also dies off slower than if it were critically damped. Moreover, if the system is overdamped, it again dies off more slowly than if it were critically damped (in agreement with our physical intuition that the damping forces hinder the return to equilibrium).

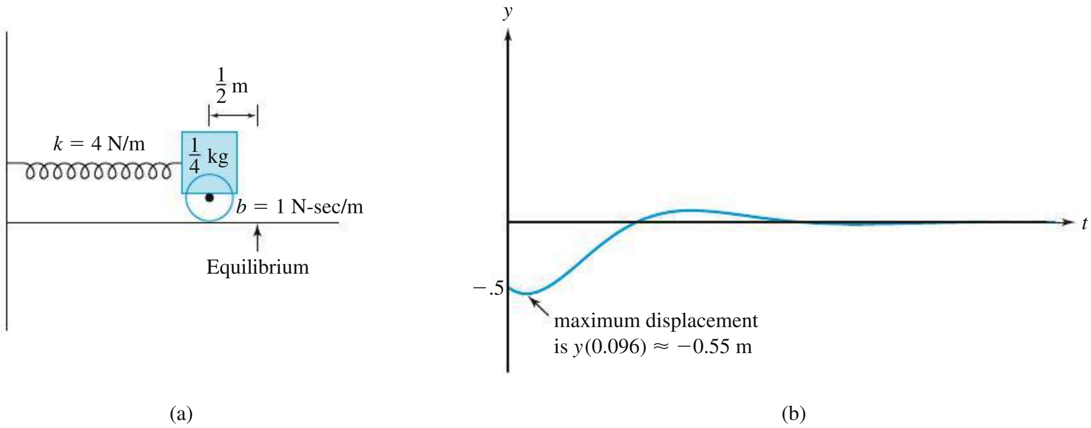


Figure 4.31 Mass–spring system and graph of motion for Example 3

Example 3 A $1/4$-kg mass is attached to a spring with a stiffness 4 N/m as shown in Figure 4.31(a). The damping constant b for the system is 1 N-sec/m. If the mass is displaced $1/2$ m to the left and given an initial velocity of 1 m/sec to the left, find the equation of motion. What is the maximum displacement that the mass will attain?

Solution Substituting the values for m, b, and k into equation (12) and enforcing the initial conditions, we obtain the initial value problem

$$\frac{1}{4}\frac{d^2y}{dt^2} + \frac{dy}{dt} + 4y = 0\,; \quad y(0) = -\frac{1}{2}, \quad y'(0) = -1\,. \tag{31}$$

The negative signs for the initial conditions reflect the facts that the initial displacement and push are to the left.

It can readily be verified that the solution to (31) is

$$y(t) = -\frac{1}{2}e^{-2t}\cos\left(2\sqrt{3}\,t\right) - \frac{1}{\sqrt{3}}e^{-2t}\sin\left(2\sqrt{3}\,t\right)\,, \tag{32}$$

or

$$y(t) = \sqrt{\frac{7}{12}}e^{-2t}\sin\left(2\sqrt{3}\,t + \phi\right)\,, \tag{33}$$

where $\tan\phi = \sqrt{3}/2$ and ϕ lies in Quadrant III because $C_1 = -1/2$ and $C_2 = -1/\sqrt{3}$ are both negative. [See Figure 4.31(b) for a sketch of $y(t)$.]

To determine the maximum displacement from equilibrium, we must determine the maximum value of $|y(t)|$ on the graph in Figure 4.31(b). Because $y(t)$ dies off exponentially, this will occur at the first critical point of $y(t)$. Computing $y'(t)$ from (32), setting it equal to zero, and solving gives

$$y'(t) = e^{-2t}\left\{\frac{5}{\sqrt{3}}\sin\left(2\sqrt{3}\,t\right) - \cos\left(2\sqrt{3}\,t\right)\right\} = 0\,,$$

$$\frac{5}{\sqrt{3}}\sin(2\sqrt{3}\,t) = \cos(2\sqrt{3}\,t)\,,$$

$$\tan(2\sqrt{3}\,t) = \frac{\sqrt{3}}{5}\,.$$

Thus, the first positive root is

$$t = \frac{1}{2\sqrt{3}} \arctan \frac{\sqrt{3}}{5} \approx 0.096\,.$$

Substituting this value for t back into equation (32) or (33) gives $y(0.096) \approx -0.55$. Hence, the maximum displacement, which occurs to the left of equilibrium, is approximately 0.55 m. ◆

4.9 EXERCISES

All problems refer to the mass–spring configuration depicted in Figure 4.1, page 152.

1. A 2-kg mass is attached to a spring with stiffness $k = 50$ N/m. The mass is displaced $1/4$ m to the left of the equilibrium point and given a velocity of 1 m/sec to the left. Neglecting damping, find the equation of motion of the mass along with the amplitude, period, and frequency. How long after release does the mass pass through the equilibrium position?

2. A 3-kg mass is attached to a spring with stiffness $k = 48$ N/m. The mass is displaced $1/2$ m to the left of the equilibrium point and given a velocity of 2 m/sec to the right. The damping force is negligible. Find the equation of motion of the mass along with the amplitude, period, and frequency. How long after release does the mass pass through the equilibrium position?

3. The motion of a mass–spring system with damping is governed by

$$y''(t) + by'(t) + 16y(t) = 0\,;$$
$$y(0) = 1\,, \qquad y'(0) = 0\,.$$

Find the equation of motion and sketch its graph for $b = 0$, 6, 8, and 10.

4. The motion of a mass–spring system with damping is governed by

$$y''(t) + by'(t) + 64y(t) = 0\,;$$
$$y(0) = 1\,, \qquad y'(0) = 0\,.$$

Find the equation of motion and sketch its graph for $b = 0$, 10, 16, and 20.

5. The motion of a mass–spring system with damping is governed by

$$y''(t) + 10y'(t) + ky(t) = 0\,;$$
$$y(0) = 1\,, \qquad y'(0) = 0\,.$$

Find the equation of motion and sketch its graph for $k = 20$, 25, and 30.

6. The motion of a mass–spring system with damping is governed by

$$y''(t) + 4y'(t) + ky(t) = 0\,;$$
$$y(0) = 1\,, \qquad y'(0) = 0\,.$$

Find the equation of motion and sketch its graph for $k = 2$, 4, and 6.

7. A $1/8$-kg mass is attached to a spring with stiffness 16 N/m. The damping constant for the system is 2 N-sec/m. If the mass is moved $3/4$ m to the left of equilibrium and given an initial leftward velocity of 2 m/sec, determine the equation of motion of the mass and give its damping factor, quasiperiod, and quasifrequency.

8. A 20-kg mass is attached to a spring with stiffness 200 N/m. The damping constant for the system is 140 N-sec/m. If the mass is pulled 25 cm to the right of equilibrium and given an initial leftward velocity of 1 m/sec, when will it first return to its equilibrium position?

9. A 2-kg mass is attached to a spring with stiffness 40 N/m. The damping constant for the system is $8\sqrt{5}$ N-sec/m. If the mass is pulled 10 cm to the right of equilibrium and given an initial rightward velocity of 2 m/sec, what is the maximum displacement from equilibrium that it will attain?

10. A $1/4$-kg mass is attached to a spring with stiffness 8 N/m. The damping constant for the system is $1/4$ N-sec/m. If the mass is moved 1 m to the left of equilibrium and released, what is the maximum displacement to the right that it will attain?

11. A 1-kg mass is attached to a spring with stiffness 100 N/m. The damping constant for the system is 0.2 N-sec/m. If the mass is pushed rightward from the equilibrium position with a velocity of 1 m/sec, when will it attain its maximum displacement to the right?

12. A $1/4$-kg mass is attached to a spring with stiffness 8 N/m. The damping constant for the system is 2 N-sec/m. If the mass is pushed 50 cm to the left of equilibrium and given a leftward velocity of 2 m/sec, when will the mass attain its maximum displacement to the left?

13. Show that for the underdamped system of Example 3, the times when the solution curve $y(t)$ in (33) touches the exponential curves $\pm\sqrt{7/12}e^{-2t}$ are not the same values of t for which the function $y(t)$ attains its relative extrema.

14. For an underdamped system, verify that as $b \to 0$ the damping factor approaches the constant A and the quasifrequency approaches the natural frequency $\sqrt{k/m}/(2\pi)$.

15. How can one deduce the value of the damping constant b by observing the motion of an underdamped system? Assume that the mass m is known.

16. A mass attached to a spring oscillates with a period of 3 sec. After 2 kg are added, the period becomes 4 sec. Assuming that we can neglect any damping or external forces, determine how much mass was originally attached to the spring.

17. Consider the equation for free mechanical vibration, $my'' + by' + ky = 0$, and assume the motion is critically damped. Let $y(0) = y_0, y'(0) = v_0$ and assume $y_0 \neq 0$.

(a) Prove that the mass will pass through its equilibrium at exactly one positive time if and only if
$$\frac{-2my_0}{2mv_0 + by_0} > 0.$$

(b) Use computer software to illustrate part (a) for a specific choice of m, b, k, y_0, and v_0. Be sure to include an appropriate graph in your illustration.

18. Consider the equation for free mechanical vibration, $my'' + by' + ky = 0$, and assume the motion is overdamped. Suppose $y(0) > 0$ and $y'(0) > 0$. Prove that the mass will never pass through its equilibrium at any positive time.

4.10 A Closer Look at Forced Mechanical Vibrations

We now consider the vibrations of a mass–spring system when an external force is applied. Of particular interest is the response of the system to a *sinusoidal* forcing term. As a paradigm, let's investigate the effect of a cosine forcing function on the system governed by the differential equation

$$m\frac{d^2y}{dt^2} + b\frac{dy}{dt} + ky = F_0 \cos \gamma t\,, \tag{1}$$

where F_0 and γ are nonnegative constants and $0 < b^2 < 4mk$ (so the system is *underdamped*).

A solution to (1) has the form $y = y_h + y_p$, where y_p is a particular solution and y_h is a general solution to the corresponding homogeneous equation. We found in equation (17) of Section 4.9 that

$$y_h(t) = Ae^{-(b/2m)t}\sin\left(\frac{\sqrt{4mk - b^2}}{2m}t + \phi\right), \tag{2}$$

where A and ϕ are constants.

To determine y_p, we can use the method of undetermined coefficients (Section 4.4). From the form of the nonhomogeneous term, we know that

$$y_p(t) = A_1 \cos \gamma t + A_2 \sin \gamma t\,, \tag{3}$$

where A_1 and A_2 are constants to be determined. Substituting this expression into equation (1) and simplifying gives

$$\left[\left(k - m\gamma^2\right) A_1 + b\gamma A_2\right]\cos \gamma t + \left[\left(k - m\gamma^2\right) A_2 - b\gamma A_1\right]\sin \gamma t = F_0 \cos \gamma t\,. \tag{4}$$

Setting the corresponding coefficients on both sides equal, we have

$$(k - m\gamma^2)A_1 + b\gamma A_2 = F_0\,,$$
$$-b\gamma A_1 + (k - m\gamma^2)A_2 = 0\,.$$

Solving, we obtain

(5) $$A_1 = \frac{F_0(k - m\gamma^2)}{(k - m\gamma^2)^2 + b^2\gamma^2}, \quad A_2 = \frac{F_0 b\gamma}{(k - m\gamma^2)^2 + b^2\gamma^2}.$$

Hence, a particular solution to (1) is

(6) $$y_p(t) = \frac{F_0}{(k - m\gamma^2)^2 + b^2\gamma^2}\left[(k - m\gamma^2)\cos\gamma t + b\gamma \sin\gamma t \right].$$

The expression in brackets can also be written as

$$\sqrt{(k - m\gamma^2)^2 + b^2\gamma^2}\,\sin(\gamma t + \theta)\,,$$

so we can express y_p in the alternative form

(7) $$y_p(t) = \frac{F_0}{\sqrt{(k - m\gamma^2)^2 + b^2\gamma^2}}\sin(\gamma t + \theta)\,,$$

where $\tan\theta = A_1/A_2 = (k - m\gamma^2)/(b\gamma)$ and the quadrant in which θ lies is determined by the signs of A_1 and A_2.

Combining equations (2) and (7), we have the following representation of a general solution to (1) in the case $0 < b^2 < 4mk$:

(8) $$y(t) = Ae^{-(b/2m)t}\sin\left(\frac{\sqrt{4mk - b^2}}{2m}t + \phi\right) + \frac{F_0}{\sqrt{(k - m\gamma^2)^2 + b^2\gamma^2}}\sin(\gamma t + \theta)\,.$$

The solution (8) is the sum of two terms. The first term, y_h, represents damped oscillation and depends only on the parameters of the system and the initial conditions. Because of the damping factor $Ae^{-(b/2m)t}$, this term tends to zero as $t \to +\infty$. Consequently, it is referred to as the **transient** part of the solution. The second term, y_p, in (8) is the offspring of the external forcing function $f(t) = F_0\cos\gamma t$. Like the forcing function, y_p is a sinusoid with angular frequency γ. It is the synchronous solution that we anticipated in Section 4.1. However, y_p is out of phase with $f(t)$ (by the angle $\theta - \pi/2$), and its magnitude is different by the factor

(9) $$\frac{1}{\sqrt{(k - m\gamma^2)^2 + b^2\gamma^2}}\,.$$

As the transient term dies off, the motion of the mass–spring system becomes essentially that of the second term y_p (see Figure 4.32, page 223). Hence, this term is called the **steady-state** solution. The factor appearing in (9) is referred to as the **frequency gain,** or **gain factor,** since it represents the ratio of the magnitude of the sinusoidal response to that of the input force. Note that this factor depends on the frequency γ and has units of length/force.

Example 1 A 10-kg mass is attached to a spring with stiffness $k = 49$ N/m. At time $t = 0$, an external force $f(t) = 20\cos 4t$ N is applied to the system. The damping constant for the system is 3 N-sec/m. Determine the steady-state solution for the system.

Solution Substituting the given parameters into equation (1), we obtain

(10) $$10\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 49y = 20\cos 4t\,,$$

where $y(t)$ is the displacement (from equilibrium) of the mass at time t.

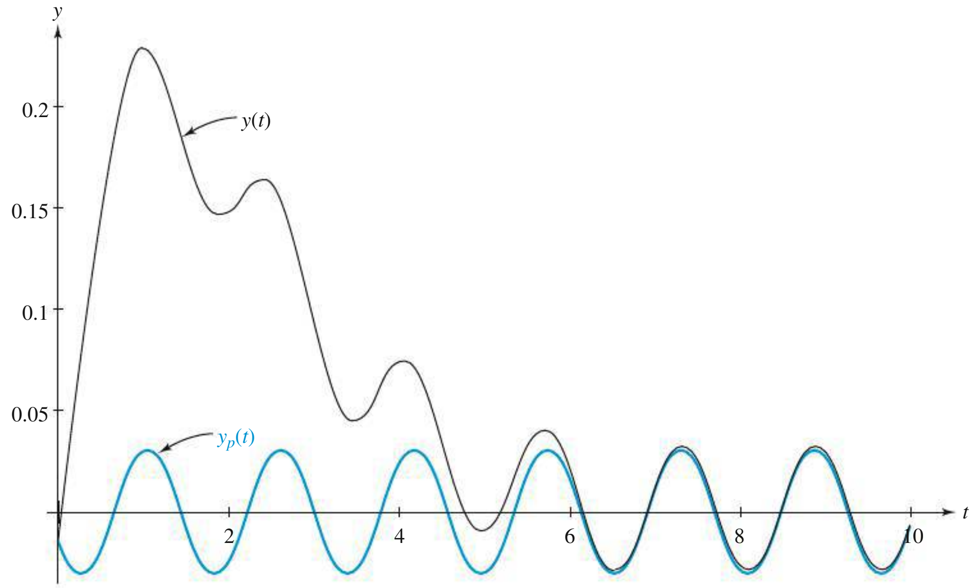


Figure 4.32 Convergence of $y(t)$ to the steady-state solution $y_p(t)$ when $m = 4$, $b = 6$, $k = 3$, $F_0 = 2$, $\gamma = 4$

To find the steady-state response, we must produce a particular solution to (10) that is a sinusoid. We can do this using the method of undetermined coefficients, guessing a solution of the form $A_1 \cos 4t + A_2 \sin 4t$. But this is precisely how we derived equation (7). Thus, we substitute directly into (7) and find

(11) $$y_p(t) = \frac{20}{\sqrt{(49-160)^2 + (9)(16)}} \sin(4t+\theta) \approx (0.18)\sin(4t+\theta),$$

where $\tan\theta = (49-160)/12 \approx -9.25$. Since the numerator, $(49-160)$, is negative and the denominator, 12, is positive, θ is a Quadrant IV angle. Thus,

$$\theta \approx \arctan(-9.25) \approx -1.46,$$

and the steady-state solution is given (approximately) by

(12) $$y_p(t) = (0.18)\sin(4t - 1.46). \ \blacklozenge$$

The above example illustrates an important point made earlier: The steady-state response (12) to the sinusoidal forcing function $20 \cos 4t$ is a sinusoid of the same frequency but different amplitude. The gain factor [see (9)] in this case is $(0.18)/20 = 0.009$ m/N.

In general, the amplitude of the steady-state solution to equation (1) depends on the angular frequency γ of the forcing function and is given by $A(\gamma) = F_0 M(\gamma)$, where

(13) $$M(\gamma) := \frac{1}{\sqrt{(k - m\gamma^2)^2 + b^2\gamma^2}}$$

is the frequency gain [see (9)]. This formula is valid even when $b^2 \geq 4mk$. For a given system (m, b, and k fixed), it is often of interest to know how this system reacts to sinusoidal inputs of various frequencies (γ is a variable). For this purpose the graph of the gain $M(\gamma)$, called the **frequency response curve,** or **resonance curve,** for the system, is enlightening.

To sketch the frequency response curve, we first observe that for $\gamma = 0$ we find $M(0) = 1/k$. Of course, $\gamma = 0$ implies the force $F_0 \cos \gamma t$ is static; there is no motion in the steady state, so this value of $M(0)$ is appropriate. Also note that as $\gamma \to \infty$ the gain $M(\gamma) \to 0$; the inertia of the system limits the extent to which it can respond to extremely rapid vibrations. As a further aid in describing the graph, we compute from (13)

$$M'(\gamma) = -\frac{2m^2\gamma\left[\gamma^2 - \left(\dfrac{k}{m} - \dfrac{b^2}{2m^2}\right)\right]}{\left[(k - m\gamma^2)^2 + b^2\gamma^2\right]^{3/2}} . \tag{14}$$

It follows from (14) that $M'(\gamma) = 0$ if and only if

$$\gamma = 0 \quad \text{or} \quad \gamma = \gamma_r := \sqrt{\frac{k}{m} - \frac{b^2}{2m^2}} . \tag{15}$$

Now when the system is overdamped or critically damped, so $b^2 \geq 4mk > 2mk$, the term inside the radical in (15) is negative, and hence $M'(\gamma) = 0$ only when $\gamma = 0$. In this case, as γ increases from 0 to infinity, $M(\gamma)$ decreases from $M(0) = 1/k$ to a limiting value of zero.

When $b^2 < 2mk$ (which implies the system is underdamped), then γ_r is real and positive, and it is easy to verify that $M(\gamma)$ has a *maximum* at γ_r. Substituting γ_r into (13) gives

$$M(\gamma_r) = \frac{1/b}{\sqrt{\dfrac{k}{m} - \dfrac{b^2}{4m^2}}} . \tag{16}$$

The value $\gamma_r/2\pi$ is called the **resonance frequency** for the system. When the system is stimulated by an external force at this frequency, it is said to be **at resonance.**

To illustrate the effect of the damping constant b on the resonance curve, we consider a system in which $m = k = 1$. In this case the frequency response curves are given by

$$M(\gamma) = \frac{1}{\sqrt{(1 - \gamma^2)^2 + b^2\gamma^2}} , \tag{17}$$

and, for $b < \sqrt{2}$, the resonance frequency is $\gamma_r/2\pi = (1/2\pi)\sqrt{1 - b^2/2}$. Figure 4.33 displays the graphs of these frequency response curves for $b = 1/4$, $1/2$, 1, $3/2$, and 2. Observe that as $b \to 0$ the maximum magnitude of the frequency gain increases and the resonance frequency $\gamma_r/2\pi$ for the damped system approaches $\sqrt{k/m}/2\pi = 1/2\pi$, the natural frequency for the undamped system.

To understand what is occurring, consider the undamped system $(b = 0)$ with forcing term $F_0 \cos \gamma t$. This system is governed by

$$m\frac{d^2y}{dt^2} + ky = F_0 \cos \gamma t . \tag{18}$$

A general solution to (18) is the sum of a particular solution and a general solution to the homogeneous equation. In Section 4.9 we showed that the latter describes simple harmonic motion:

$$y_h(t) = A \sin(\omega t + \phi) , \qquad \omega := \sqrt{k/m} . \tag{19}$$

The formula for the particular solution given in (7) is valid for $b = 0$, provided $\gamma \neq \omega = \sqrt{k/m}$. However, when $b = 0$ and $\gamma = \omega$, then the form we used with undetermined coefficients

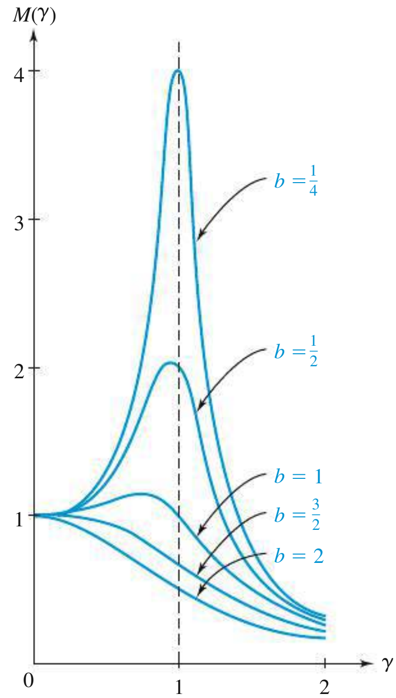


Figure 4.33 Frequency response curves for various values of b

to derive (7) does not work because $\cos \omega t$ and $\sin \omega t$ are solutions to the corresponding homogeneous equation. The correct form is

(20) $$y_p(t) = A_1 t \cos \omega t + A_2 t \sin \omega t ,$$

which leads to the solution

(21) $$y_p(t) = \frac{F_0}{2m\omega} t \sin \omega t .$$

[The verification of (21) is straightforward.] Hence, in the *undamped resonant case* (when $\gamma = \omega$), a general solution to (18) is

(22) $$y(t) = A \sin(\omega t + \phi) + \frac{F_0}{2m\omega} t \sin \omega t .$$

Returning to the question of resonance, observe that the particular solution in (21) oscillates between $\pm (F_0 t)/(2m\omega)$. Hence, as $t \to +\infty$ the maximum magnitude of (21) approaches ∞ (see Figure 4.34 on page 226).

It is obvious from the above discussion that **if the damping constant b is very small, the system is subject to large oscillations when the forcing function has a frequency near the resonance frequency for the system.** It is these large vibrations at resonance that concern engineers. Indeed, resonance vibrations have been known to cause airplane wings to snap, bridges to collapse,† and (less catastrophically) wine glasses to shatter.

†An interesting discussion of one such disaster, involving the Tacoma Narrows bridge in Washington State, can be found in *Differential Equations and Their Applications,* 4th ed., by M. Braun (Springer-Verlag, New York, 1993). See also the articles, "Large-Amplitude Periodic Oscillations in Suspension Bridges: Some New Connections with Nonlinear Analysis," by A. C. Lazer and P. J. McKenna, *SIAM Review,* Vol. 32 (1990): 537–578; or "Still Twisting," by Henry Petroski, *American Scientist,* Vol. 19 (1991): 398–401.

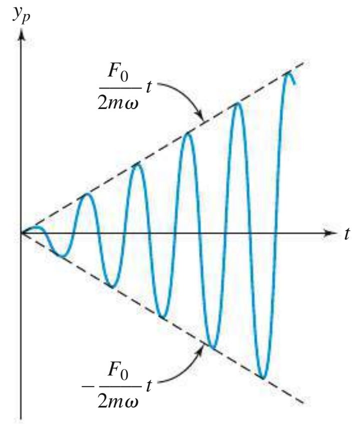


Figure 4.34 Undamped oscillation of the particular solution in (21)

When the mass–spring system is hung vertically as in Figure 4.35, the force of gravity must be taken into account. This is accomplished very easily. With y measured downward from the unstretched spring position, the governing equation is

$$my'' + by' + ky = mg\,,$$

and if the right-hand side is recognized as a sinusoidal forcing term with frequency zero $(mg\cos 0t)$, then the synchronous steady-state response is a constant, which is easily seen to be $y_p(t) = mg/k$. Now if we redefine $y(t)$ to be measured *from this (true) equilibrium level,* as indicated in Figure 4.35(c),

$$y_{\text{new}}(t) := y(t) - mg/k\,,$$

then the governing equation

$$my'' + by' + ky = mg + F_{\text{ext}}(t)$$

simplifies; we find

$$\begin{aligned} my''_{\text{new}} + by'_{\text{new}} + ky_{\text{new}} &= m(y - mg/k)'' + b(y - mg/k)' + k(y - mg/k) \\ &= my'' + by' + ky - mg \\ &= mg + F_{\text{ext}}(t) - mg\,, \end{aligned}$$

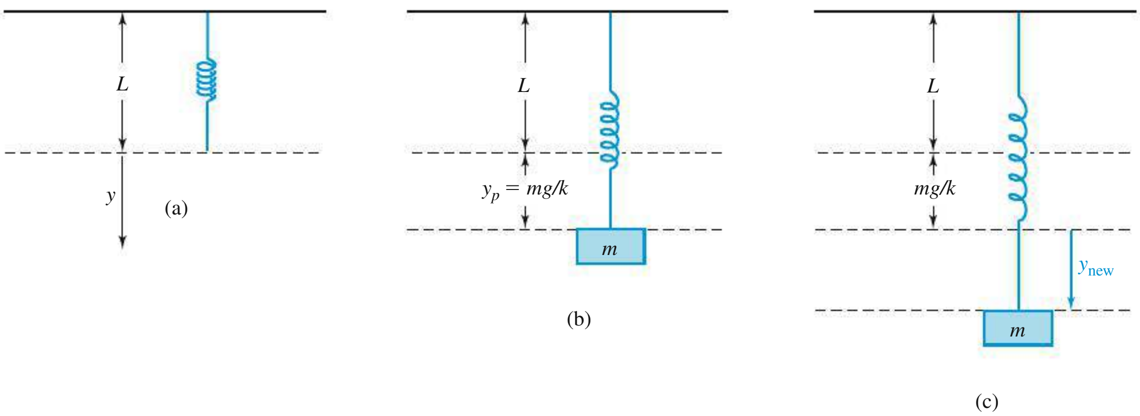


Figure 4.35 Spring (a) in natural position, (b) in equilibrium, and (c) in motion

or

(23) $$my''_{\text{new}} + by'_{\text{new}} + ky_{\text{new}} = F_{\text{ext}}(t)\ .$$

Thus, the gravitational force can be ignored if $y(t)$ is measured from the equilibrium position. Adopting this convention, we drop the "new" subscript hereafter.

Example 2 Suppose the mass–spring system in Example 1 is hung vertically. Find the steady-state solution.

Solution This is trivial; the steady-state solution is identical to what we derived before,

$$y_p(t) = \frac{20}{\sqrt{(49-160)^2 + (9)(16)}}\sin(4t+\theta)$$

[equation (11)], but now y_p is measured from the equilibrium position, which is $mg/k \approx 10 \times 9.8/49 = 2$ m below the unstretched spring position. ◆

Example 3 A 64-lb weight is attached to a vertical spring, causing it to stretch 3 in. upon coming to rest at equilibrium. The damping constant for the system is 3 lb-sec/ft. An external force $F(t) = 3\cos 12t$ lb is applied to the weight. Find the steady-state solution for the system.

Solution If a weight of 64 lb stretches a spring by 3 in. (0.25 ft), then the spring stiffness must be $64/0.25 = 256$ lb/ft. Thus, if we measured the displacement from the (true) equilibrium level, equation (23) becomes

(24) $$my'' + by' + ky = 3\cos 12t\ ,$$

with $b = 3$ and $k = 256$. But recall that the unit of mass in the U.S. Customary System is the *slug*, which equals the weight divided by the gravitational acceleration constant $g \approx 32$ ft/sec^2 (Table 3.2, page 110). Therefore, m in equation (24) is $64/32 = 2$ slugs, and we have

$$2y'' + 3y' + 256y = 3\cos 12t\ .$$

The steady-state solution is given by equation (6) with $F_0 = 3$ and $\gamma = 12$:

$$y_p(t) = \frac{3}{(256 - 2\cdot 12^2)^2 + 3^2\cdot 12^2}\left[(256 - 2\cdot 12^2)\cos 12t + 3\cdot 12\sin 12t\right]$$

$$= \frac{3}{580}(-8\cos 12t + 9\sin 12t)\ .$$ ◆

4.10 EXERCISES

In the following problems, take $g = 32$ ft/sec^2 for the U.S. Customary System and $g = 9.8$ m/sec^2 for the MKS system.

1. Sketch the frequency response curve (13) for the system in which $m = 4$, $k = 1$, $b = 2$.

2. Sketch the frequency response curve (13) for the system in which $m = 2$, $k = 3$, $b = 3$.

3. Determine the equation of motion for an undamped system at resonance governed by

$$\frac{d^2y}{dt^2} + 9y = 2\cos 3t\ \ ;$$

$$y(0) = 1\ , \quad y'(0) = 0\ .$$

Sketch the solution.

4. Determine the equation of motion for an undamped system at resonance governed by

$$\frac{d^2y}{dt^2} + y = 5\cos t \ ;$$
$$y(0) = 0\,, \quad y'(0) = 1\,.$$

Sketch the solution.

5. An undamped system is governed by

$$m\frac{d^2y}{dt^2} + ky = F_0 \cos\gamma t \ ;$$
$$y(0) = y'(0) = 0\,,$$

where $\gamma \neq \omega := \sqrt{k/m}$.

(a) Find the equation of motion of the system.

(b) Use trigonometric identities to show that the solution can be written in the form

$$y(t) = \frac{2F_0}{m(\omega^2 - \gamma^2)} \sin\left(\frac{\omega+\gamma}{2}t\right) \sin\left(\frac{\omega-\gamma}{2}t\right).$$

(c) When γ is near ω, then $\omega - \gamma$ is small, while $\omega + \gamma$ is relatively large compared with $\omega - \gamma$. Hence, $y(t)$ can be viewed as the product of a slowly varying sine function, $\sin[(\omega-\gamma)t/2]$, and a rapidly varying sine function, $\sin[(\omega+\gamma)t/2]$. The net effect is a sine function $y(t)$ with frequency $(\omega-\gamma)/4\pi$, which serves as the time-varying amplitude of a sine function with frequency $(\omega+\gamma)/4\pi$. This vibration phenomenon is referred to as **beats** and is used in tuning stringed instruments. This same phenomenon in electronics is called **amplitude modulation.** To illustrate this phenomenon, sketch the curve $y(t)$ for $F_0 = 32$, $m = 2$, $\omega = 9$, and $\gamma = 7$.

6. Derive the formula for $y_p(t)$ given in (21).

7. Shock absorbers in automobiles and aircraft can be described as forced *overdamped* mass–spring systems. Derive an expression analogous to equation (8) for the general solution to the differential equation (1) when $b^2 > 4mk$.

8. The response of an overdamped system to a constant force is governed by equation (1) with $m = 2$, $b = 8$, $k = 6$, $F_0 = 18$, and $\gamma = 0$. If the system starts from rest $[y(0) = y'(0) = 0]$, compute and sketch the displacement $y(t)$. What is the limiting value of $y(t)$ as $t \to +\infty$? Interpret this physically.

9. An 8-kg mass is attached to a spring hanging from the ceiling, thereby causing the spring to stretch 1.96 m upon coming to rest at equilibrium. At time $t = 0$, an external force $F(t) = \cos 2t\ N$ is applied to the system. The damping constant for the system is 3 N-sec/m. Determine the steady-state solution for the system.

10. Show that the period of the simple harmonic motion of a mass hanging from a spring is $2\pi\sqrt{l/g}$, where l denotes the amount (beyond its natural length) that the spring is stretched when the mass is at equilibrium.

11. A mass weighing 8 lb is attached to a spring hanging from the ceiling and comes to rest at its equilibrium position. At $t = 0$, an external force $F(t) = 2\cos 2t$ lb is applied to the system. If the spring constant is 10 lb/ft and the damping constant is 1 lb-sec/ft, find the equation of motion of the mass. What is the resonance frequency for the system?

12. A 2-kg mass is attached to a spring hanging from the ceiling, thereby causing the spring to stretch 20 cm upon coming to rest at equilibrium. At time $t = 0$, the mass is displaced 5 cm below the equilibrium position and released. At this same instant, an external force $F(t) = 0.3\cos t$ N is applied to the system. If the damping constant for the system is 5 N-sec/m, determine the equation of motion for the mass. What is the resonance frequency for the system?

13. A mass weighing 32 lb is attached to a spring hanging from the ceiling and comes to rest at its equilibrium position. At time $t = 0$, an external force $F(t) = 3\cos 4t$ lb is applied to the system. If the spring constant is 5 lb/ft and the damping constant is 2 lb-sec/ft, find the steady-state solution for the system.

14. An 8-kg mass is attached to a spring hanging from the ceiling and allowed to come to rest. Assume that the spring constant is 40 N/m and the damping constant is 3 N-sec/m. At time $t = 0$, an external force of $2\sin(2t + \pi/4)$ N is applied to the system. Determine the amplitude and frequency of the steady-state solution.

15. An 8-kg mass is attached to a spring hanging from the ceiling and allowed to come to rest. Assume that the spring constant is 40 N/m and the damping constant is 3 N-sec/m. At time $t = 0$, an external force of $2\sin 2t\cos 2t$ N is applied to the system. Determine the amplitude and frequency of the steady-state solution.

16. A helium-filled balloon on a cord, hanging y km above a level surface, is subjected to three forces:

(i) the (constant) buoyant force B exerted by the external air pressure;

(ii) the weight of the balloon mg, where m is the mass of the balloon;

(iii) the weight of the part of the cord that has been lifted off the surface, dgy, where d is the linear density (kg/m) of the cord. (We assume that the cord is longer than the elevation of the balloon.)

Express Newton's second law for the balloon and show that it is exactly analogous to the governing equation for a mass vertically hung from a spring (page 226). Find the equation of motion. What is the frequency of the balloon's oscillations?

Chapter 4 Summary

In this chapter we discussed the theory of second-order linear differential equations and presented explicit solution methods for equations with constant coefficients. Much of the methodology can also be applied to the more general case of variable coefficients. We also studied the mathematical description of vibrating mechanical systems, and we saw how the mass–spring analogy could be used to predict qualitative features of solutions to some variable-coefficient and nonlinear equations.

The important features and solution techniques for the constant-coefficient case are listed below.

Homogeneous Linear Equations (Constant Coefficients)

$$ay'' + by' + cy = 0, \quad a(\neq 0), b, c \text{ constants.}$$

Linearly Independent Solutions: y_1, y_2. Two solutions y_1 and y_2 to the homogeneous equation on the interval I are said to be linearly independent on I if neither function is a constant times the other on I. This will be true provided their Wronskian,

$$W[y_1, y_2](t) := y_1(t)y_2'(t) - y_1'(t)y_2(t)\,,$$

is different from zero for some (and hence all) t in I.

General Solution to Homogeneous Equation: $c_1y_1 + c_2y_2$. If y_1 and y_2 are linearly independent solutions to the homogeneous equation, then a general solution is

$$y(t) = c_1y_1(t) + c_2y_2(t)\,,$$

where c_1 and c_2 are arbitrary constants.

Form of General Solution. The form of a general solution for a homogeneous equation with constant coefficients depends on the roots

$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}\,, \qquad r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

of the auxiliary equation

$$ar^2 + br + c = 0\,, \qquad a \neq 0\,.$$

(a) When $b^2 - 4ac > 0$, the auxiliary equation has two distinct real roots r_1 and r_2 and a general solution is

$$y(t) = c_1e^{r_1t} + c_2e^{r_2t}\,.$$

(b) When $b^2 - 4ac = 0$, the auxiliary equation has a repeated real root $r = r_1 = r_2$ and a general solution is

$$y(t) = c_1e^{rt} + c_2te^{rt}\,.$$

(c) When $b^2 - 4ac < 0$, the auxiliary equation has complex conjugate roots $r = \alpha \pm i\beta$ and a general solution is

$$y(t) = c_1e^{\alpha t}\cos\beta t + c_2e^{\alpha t}\sin\beta t\,.$$

Nonhomogeneous Linear Equations (Constant Coefficients)

$$ay'' + by' + cy = f(t)$$

General Solution to Nonhomogeneous Equation: $y_p + c_1y_1 + c_2y_2$. If y_p is any particular solution to the nonhomogeneous equation and y_1 and y_2 are linearly independent solutions to the corresponding homogeneous equation, then a general solution is

$$y(t) = y_p(t) + c_1y_1(t) + c_2y_2(t) ,$$

where c_1 and c_2 are arbitrary constants.

Two methods for finding a particular solution y_p are those of undetermined coefficients and variation of parameters.

Undetermined Coefficients: $f(t) = p_n(t)e^{\alpha t}\begin{Bmatrix}\cos\beta t\\ \sin\beta t\end{Bmatrix}$. If the right-hand side $f(t)$ of a nonhomogeneous equation with constant coefficients is a polynomial $p_n(t)$, an exponential of the form $e^{\alpha t}$, a trigonometric function of the form $\cos\beta t$ or $\sin\beta t$, or any product of these special types of functions, then a particular solution of an appropriate form can be found. The form of the particular solution involves unknown coefficients and depends on whether $\alpha + i\beta$ is a root of the corresponding auxiliary equation. See the summary box on page 184. The unknown coefficients are found by substituting the form into the differential equation and equating coefficients of like terms.

Variation of Parameters: $y(t) = v_1(t)y_1(t) + v_2(t)y_2(t)$. If y_1 and y_2 are two linearly independent solutions to the corresponding homogeneous equation, then a particular solution to the nonhomogeneous equation is

$$y(t) = v_1(t)y_1(t) + v_2(t)y_2(t) ,$$

where v_1' and v_2' are determined by the equations

$$v_1'y_1 + v_2'y_2 = 0$$
$$v_1'y_1' + v_2'y_2' = f(t)/a \ .$$

Superposition Principle. If y_1 and y_2 are solutions to the equations

$$ay'' + by' + cy = f_1 \quad \text{and} \quad ay'' + by' + cy = f_2 ,$$

respectively, then $k_1y_1 + k_2y_2$ is a solution to the equation

$$ay'' + by' + cy = k_1f_1 + k_2f_2 .$$

The superposition principle facilitates finding a particular solution when the nonhomogeneous term is the sum of nonhomogeneities for which particular solutions can be determined.

Cauchy–Euler (Equidimensional) Equations

$$at^2y'' + bty' + cy = f(t)$$

Substituting $y = t^r$ yields the associated characteristic equation

$$ar^2 + (b - a)r + c = 0$$

for the corresponding *homogeneous* Cauchy–Euler equation. A general solution to the homogeneous equation for $t > 0$ is given by

(i) $c_1t^{r_1} + c_2t^{r_2}$, if r_1 and r_2 are distinct real roots;
(ii) $c_1t^r + c_2t^r \ln t$, if r is a repeated root;
(iii) $c_1t^\alpha\cos(\beta\ln t) + c_2t^\alpha\sin(\beta\ln t)$, if $\alpha + i\beta$ is a complex root.

A general solution to the nonhomogeneous equation is $y = y_p + y_h$, where y_p is a particular solution and y_h is a general solution to the corresponding homogeneous equation. The method of variation of parameters (but not the method of undetermined coefficients) can be used to find a particular solution.

REVIEW PROBLEMS FOR CHAPTER 4

In Problems 1–28, find a general solution to the given differential equation.

1. $y'' + 8y' - 9y = 0$

2. $49y'' + 14y' + y = 0$

3. $4y'' - 4y' + 10y = 0$

4. $9y'' - 30y' + 25y = 0$

5. $6y'' - 11y' + 3y = 0$

6. $y'' + 8y' - 14y = 0$

7. $36y'' + 24y' + 5y = 0$

8. $25y'' + 20y' + 4y = 0$

9. $16z'' - 56z' + 49z = 0$

10. $u'' + 11u = 0$

11. $t^2x''(t) + 5x(t) = 0, \quad t > 0$

12. $2y''' - 3y'' - 12y' + 20y = 0$

13. $y'' + 16y = te^t$

14. $v'' - 4v' + 7v = 0$

15. $3y''' + 10y'' + 9y' + 2y = 0$

16. $y''' + 3y'' + 5y' + 3y = 0$

17. $y''' + 10y' - 11y = 0$

18. $y^{(4)} = 120t$

19. $4y''' + 8y'' - 11y' + 3y = 0$

20. $2y'' - y = t \sin t$

21. $y'' - 3y' + 7y = 7t^2 - e^t$

22. $y'' - 8y' - 33y = 546 \sin t$

23. $y''(\theta) + 16y(\theta) = \tan 4\theta$

24. $10y'' + y' - 3y = t - e^{t/2}$

25. $4y'' - 12y' + 9y = e^{5t} + e^{3t}$

26. $y'' + 6y' + 15y = e^{2t} + 75$

27. $x^2y'' + 2xy' - 2y = 6x^{-2} + 3x, \quad x > 0$

28. $y'' = 5x^{-1}y' - 13x^{-2}y, \quad x > 0$

In Problems 29–36, find the solution to the given initial value problem.

29. $y'' + 4y' + 7y = 0$;
$y(0) = 1, \quad y'(0) = -2$

30. $y''(\theta) + 2y'(\theta) + y(\theta) = 2\cos\theta$;
$y(0) = 3, \quad y'(0) = 0$

31. $y'' - 2y' + 10y = 6\cos 3t - \sin 3t$;
$y(0) = 2, \quad y'(0) = -8$

32. $4y'' - 4y' + 5y = 0$;
$y(0) = 1, \quad y'(0) = -11/2$

33. $y''' - 12y'' + 27y' + 40y = 0$;
$y(0) = -3, \quad y'(0) = -6, \quad y''(0) = -12$

34. $y'' + 5y' - 14y = 0$;
$y(0) = 5, \quad y'(0) = 1$

35. $y''(\theta) + y(\theta) = \sec\theta$; $\quad y(0) = 1, \quad y'(0) = 2$

36. $9y'' + 12y' + 4y = 0$;
$y(0) = -3, \quad y'(0) = 3$

37. Use the mass–spring oscillator analogy to decide whether all solutions to each of the following differential equations are bounded as $t \to +\infty$.

(a) $y'' + t^4y = 0$

(b) $y'' - t^4y = 0$

(c) $y'' + y^7 = 0$

(d) $y'' + y^8 = 0$

(e) $y'' + (3 + \sin t)y = 0$

(f) $y'' + t^2y' + y = 0$

(g) $y'' - t^2y' - y = 0$

38. A 3-kg mass is attached to a spring with stiffness $k = 75$ N/m, as in Figure 4.1, page 152. The mass is displaced $1/4$ m to the left and given a velocity of 1 m/sec to the right. The damping force is negligible. Find the equation of motion of the mass along with the amplitude, period, and frequency. How long after release does the mass pass through the equilibrium position?

39. A 32-lb weight is attached to a vertical spring, causing it to stretch 6 in. upon coming to rest at equilibrium. The damping constant for the system is 2 lb-sec/ft. An external force $F(t) = 4\cos 8t$ lb is applied to the weight. Find the steady-state solution for the system. What is its resonant frequency?

TECHNICAL WRITING EXERCISES FOR CHAPTER 4

1. Compare the two methods—undetermined coefficients and variation of parameters—for determining a particular solution to a nonhomogeneous equation. What are the advantages and disadvantages of each?

2. Consider the differential equation

$$\frac{d^2y}{dx^2} + 2b\frac{dy}{dx} + y = 0\,,$$

where b is a constant. Describe how the behavior of solutions to this equation changes as b varies.

3. Consider the differential equation

$$\frac{d^2y}{dx^2} + cy = 0\,,$$

where c is a constant. Describe how the behavior of solutions to this equation changes as c varies.

4. For students with a background in linear algebra: Compare the theory for linear second-order equations with that for systems of n linear equations in n unknowns whose coefficient matrix has rank $n - 2$. Use the terminology from linear algebra; for example, subspace, basis, dimension, linear transformation, and kernel. Discuss both homogeneous and nonhomogeneous equations.

Projects for Chapter 4

A Nonlinear Equations Solvable by First-Order Techniques

Certain *nonlinear* second-order equations—namely, those with dependent or independent variables missing—can be solved by reducing them to a pair of *first-order equations.* This is accomplished by making the substitution $w = dy/dx$, where x is the independent variable.

(a) To solve an equation of the form $y'' = F(x, y')$ in which the dependent variable y is missing, setting $w = y'$ (so that $w' = y''$) yields the pair of equations

$$w' = F(x, w)\,,$$
$$y' = w\,.$$

Because $w' = F(x, w)$ is a first-order equation, we have available the techniques of Chapter 2 to solve it for $w(x)$. Once $w(x)$ is determined, we integrate it to obtain $y(x)$.

Using this method, solve

$$2xy'' - y' + \frac{1}{y'} = 0\,, \qquad x > 0\,.$$

(b) To solve an equation of the form $y'' = F(y, y')$ in which the independent variable x is missing, setting $w = dy/dx$ yields, via the chain rule,

$$\frac{d^2y}{dx^2} = \frac{dw}{dx} = \frac{dw}{dy}\frac{dy}{dx} = w\frac{dw}{dy}\,.$$

Thus, $y'' = F(y, y')$ is equivalent to the pair of equations

$$\textbf{(1)} \qquad w\frac{dw}{dy} = F(y, w)\,,$$

$$\textbf{(2)} \qquad \frac{dy}{dx} = w\,.$$

In equation (1) notice that y plays the role of the *independent* variable; hence, solving it yields $w(y)$. Then substituting $w(y)$ into (2), we obtain a separable equation that determines $y(x)$.

Using this method, solve the following equations:

$$\textbf{(i)} \quad 2y\frac{d^2y}{dx^2} = 1 + \left(\frac{dy}{dx}\right)^2. \qquad \textbf{(ii)} \quad \frac{d^2y}{dx^2} + y\frac{dy}{dx} = 0\,.$$

(c) **Suspended Cable.** In the study of a cable suspended between two fixed points (see Figure 4.36 on page 234), one encounters the initial value problem

$$\frac{d^2y}{dx^2} = \frac{1}{a}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,; \qquad y(0) = a\,, \qquad y'(0) = 0\,,$$

where a $(\neq 0)$ is a constant. Solve this initial value problem for y. The resulting curve is called a **catenary.**

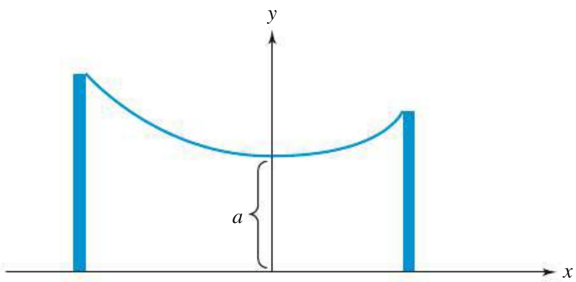


Figure 4.36 Suspended cable

B *Apollo* Reentry

Courtesy of Alar Toomre, Massachusetts Institute of Technology

Each time the Apollo astronauts returned from the moon circa 1970, they took great care to reenter Earth's atmosphere along a path that was only a small angle α from the horizontal. (See Figure 4.37.) This was necessary in order to avoid intolerably large "g" forces during their reentry.

To appreciate their grounds for concern, consider the idealized problem

$$\frac{d^2s}{dt^2} = -Ke^{s/H}\left(\frac{ds}{dt}\right)^2,$$

where K and H are constants and distance s is measured downrange from some reference point on the trajectory, as shown in the figure. This approximate equation pretends that the only force on the capsule during reentry is air drag. For a bluff body such as the *Apollo,* drag is proportional to the square of the speed and to the local atmospheric density, which falls off exponentially with height. Intuitively, one might expect that the deceleration predicted by this model would depend heavily on the constant K (which takes into account the vehicle's mass, area, etc.); but, remarkably, for capsules entering the atmosphere (at "$s = -\infty$") with a common speed V_0, the *maximum* deceleration turns out to be independent of K.

(a) Verify this last assertion by demonstrating that this maximum deceleration is just $V_0^2/(2eH)$. [*Hint:* The independent variable t does not appear in the differential equation, so it is helpful to make the substitution $v = ds/dt$; see Project A, part (b).]

(b) Also verify that any such spacecraft at the instant when it is decelerating most fiercely will be traveling exactly with speed $V_0/\sqrt{e}$, having by then lost almost 40% of its original velocity.

(c) Using the plausible data $V_0 = 11$ km/sec and $H = 10/(\sin\alpha)$ km, estimate how small α had to be chosen so as to inconvenience the returning travelers with no more than 10 g's.

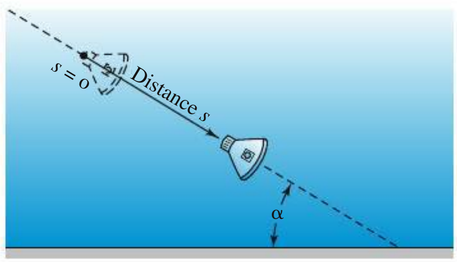


Figure 4.37 Reentry path

C Simple Pendulum

In Section 4.8, we discussed the simple pendulum consisting of a mass m suspended by a rod of length ℓ having negligible mass and derived the nonlinear initial value problem

$$\textbf{(3)} \qquad \frac{d^2\theta}{dt^2} + \frac{g}{\ell}\sin\theta = 0\,; \quad \theta(0) = \alpha\,, \quad \theta'(0) = 0\,,$$

where g is the acceleration due to gravity and $\theta(t)$ is the angle the rod makes with the vertical at time t (see Figure 4.18, page 208). Here it is assumed that the mass is released with zero velocity at an initial angle α, $0 < \alpha < \pi$. We would like to determine the equation of motion for the pendulum and its period of oscillation.

(a) Use equation (3) and the energy integral lemma discussed in Section 4.8 to show that

$$\left(\frac{d\theta}{dt}\right)^2 = \frac{2g}{\ell}(\cos\theta - \cos\alpha)$$

and hence

$$dt = -\sqrt{\frac{\ell}{2g}}\frac{d\theta}{\sqrt{\cos\theta - \cos\alpha}}\,.$$

(b) Use the trigonometric identity $\cos x = 1 - 2\sin^2(x/2)$ to express dt by

$$dt = -\frac{1}{2}\sqrt{\frac{\ell}{g}}\frac{d\theta}{\sqrt{\sin^2(\alpha/2) - \sin^2(\theta/2)}}\,.$$

(c) Make the change of variables $\sin(\theta/2) = \sin(\alpha/2)\sin\phi$ and show that the elapsed time, T, for the pendulum to fall from the angle $\theta = \alpha$ (corresponding to $\phi = \pi/2$) to the angle $\theta = \beta$ (corresponding to $\phi = \Phi$), when $\alpha \ge \beta \ge 0$, is given by

$$\textbf{(4)} \qquad T = \int_0^T dt = -\int_\alpha^\beta \frac{1}{2}\sqrt{\frac{\ell}{g}}\frac{d\theta}{\sqrt{\sin^2(\alpha/2) - \sin^2(\theta/2)}} = -\sqrt{\frac{\ell}{g}}\int_{\pi/2}^{\Phi}\frac{d\phi}{\sqrt{1 - k^2\sin^2\phi}}\,,$$

where $k := \sin(\alpha/2)$.

(d) The period P of the pendulum is defined to be the time required for it to swing from one extreme to the other and back—that is, from α to $-\alpha$ and back to α. Show that the period is given by

$$\textbf{(5)} \qquad P = 4\sqrt{\frac{\ell}{g}}\int_0^{\pi/2}\frac{d\phi}{\sqrt{1 - k^2\sin^2\phi}}\,.$$

The integral in (5) is called an **elliptic integral of the first kind** and is denoted by $F(k, \pi/2)$. As you might expect, the period of the simple pendulum depends on the length ℓ of the rod and the initial displacement α. In fact, a check of an elliptic integral table will show that the period nearly doubles as the initial displacement increases from $\pi/8$ to $15\pi/16$ (for fixed ℓ). What happens as α approaches π?

(e) From equation (5) show that

$$\textbf{(6)} \qquad -T + P/4 = \sqrt{\frac{\ell}{g}}F(k, \Phi), \text{ where } F(k, \Phi) := \int_{\phi=0}^{\Phi}\frac{d\phi}{\sqrt{1 - k^2\sin^2\phi}}\,.$$

For fixed k, $F(k, \Phi)$ has an "inverse," denoted by $\operatorname{sn}(k, u)$, that satisfies $u = F(k, \Phi)$ if and only if $\operatorname{sn}(k, u) = \sin \Phi$. The function $\operatorname{sn}(k, u)$ is called a **Jacobi elliptic function** and has many properties that resemble those of the sine function. Using the Jacobi elliptic function $\operatorname{sn}(k, u)$, express the equation of motion for the pendulum in the form

$$\beta = 2 \arcsin \left\{ k \operatorname{sn} \left[k, \sqrt{\frac{g}{\ell}} (-T + P/4) \right] \right\}, \quad 0 \le T \le P/4 . \tag{7}$$

(f) Take $\ell = 1$ m, $g = 9.8$ m/sec, $\alpha = \pi/4$ radians. Use Runge–Kutta algorithms or tabulated values of the Jacobi elliptic function to determine the period of the pendulum.

D Linearization of Nonlinear Problems

A useful approach to analyzing a nonlinear equation is to study its **linearized equation,** which is obtained by replacing the nonlinear terms by linear approximations. For example, the nonlinear equation

$$\frac{d^2\theta}{dt^2} + \sin \theta = 0 , \tag{8}$$

which governs the motion of a simple pendulum, has

$$\frac{d^2\theta}{dt^2} + \theta = 0 \tag{9}$$

as a linearization for small θ. (The nonlinear term $\sin \theta$ has been replaced by the linear approximation θ.)

A general solution to equation (8) involves Jacobi elliptic functions (see Project C), which are rather complicated, so let's try to approximate the solutions. For this purpose we consider two methods: Taylor series and linearization.

(a) Derive the first six terms of the Taylor series about $t = 0$ of the solution to equation (8) with initial conditions $\theta(0) = \pi/12, \theta'(0) = 0$. (The Taylor series method is discussed in Project D of Chapter 1 and Section 8.1.)
(b) Solve equation (9) subject to the same initial conditions $\theta(0) = \pi/12, \theta'(0) = 0$.
(c) On the same coordinate axes, graph the two approximations found in parts (a) and (b).
(d) Discuss the advantages and disadvantages of the Taylor series method and the linearization method.
(e) Give a linearization for the initial value problem.

$$x''(t) + 0.1[1 - x^2(t)]x'(t) + x(t) = 0 \quad x(0) = 0.4 , \quad x'(0) = 0 ,$$

for x small. Solve this linearized problem to obtain an approximation for the nonlinear problem.

E Convolution Method

The **convolution** of two functions g and f is the function $g * f$ defined by

$$(g * f)(t) \coloneqq \int_0^t g(t-v)f(v)\,dv\,.$$

The aim of this project is to show how convolutions can be used to obtain a particular solution to a nonhomogeneous equation of the form

(10) $ay'' + by' + cy = f(t)$, where a, b, and c are constants, $a \neq 0$.

(a) Use Leibniz's rule,

$$\frac{d}{dt}\int_a^t h(t,v)\,dv = \int_a^t \frac{\partial h}{\partial t}(t,v)\,dv + h(t,t)\,,$$

to show the following:

$$(y * f)'(t) = (y' * f)(t) + y(0)f(t)$$
$$(y * f)''(t) = (y'' * f)(t) + y'(0)f(t) + y(0)f'(t)\,,$$

assuming y and f are sufficiently differentiable.

(b) Let $y_s(t)$ be the solution to the homogeneous equation $ay'' + by' + cy = 0$ that satisfies $y_s(0) = 0$, $y_s'(0) = 1/a$. Show that $y_s * f$ is the particular solution to equation (10) satisfying $y(0) = y'(0) = 0$.

(c) Let $y_k(t)$ be the solution to the homogeneous equation $ay'' + by' + cy = 0$ that satisfies $y(0) = Y_0$, $y'(0) = Y_1$, and let y_s be as defined in part (b). Show that

$$(y_s * f)(t) + y_k(t)$$

is the unique solution to the initial value problem

(11) $ay'' + by' + cy = f(t)\,;\quad y(0) = Y_0\,,\quad y'(0) = Y_1\,.$

(d) Use the result of part (c) to determine the solution to each of the following initial value problems. Carry out all integrations and express your answers in terms of elementary functions.

(i) $y'' + y = \tan t\,;\quad y(0) = 0\,,\quad y'(0) = -1$

(ii) $2y'' + y' - y = e^{-t}\sin t\,;\quad y(0) = 1\,,\quad y'(0) = 1$

(iii) $y'' - 2y' + y = \sqrt{t}\,e^t\,;\quad y(0) = 2\,,\quad y'(0) = 0$

F Undetermined Coefficients Using Complex Arithmetic

The technique of undetermined coefficients described in Section 4.5 can be streamlined with the aid of complex arithmetic and the properties of the complex exponential function. The essential formulas are

$$e^{(\alpha+i\beta)t} = e^{\alpha t}(\cos\beta t + i\sin\beta t)\,,\qquad \frac{d}{dt}e^{(\alpha+i\beta)t} = (\alpha+i\beta)e^{(\alpha+i\beta)t}\,,$$
$$\operatorname{Re} e^{(\alpha+i\beta)t} = e^{\alpha t}\cos\beta t\,,\qquad \operatorname{Im} e^{(\alpha+i\beta)t} = e^{\alpha t}\sin\beta t\,.$$

(a) From the preceding formulas derive the equations

(12) $$\text{Re}[(a+ib)e^{(\alpha+i\beta)t}] = e^{\alpha t}(a\cos\beta t - b\sin\beta t),$$

(13) $$\text{Im}[(a+ib)e^{(\alpha+i\beta)t}] = e^{\alpha t}(b\cos\beta t + a\sin\beta t).$$

Now consider a second-order equation of the form

(14) $$L[y] := ay'' + by' + cy = g,$$

where a, b, and c are real numbers and g is of the special form

(15) $$\boldsymbol{g(t) = e^{\alpha t}[(a_n t^n + \cdots + a_1 t + a_0)\cos\beta t + (b_n t^n + \cdots + b_1 t + b_0)\sin\beta t],}$$

with the a_j's, b_j's, α, and β real numbers. Such a function can always be expressed as the real or imaginary part of a function involving the complex exponential. For example, using equation (12), one can quickly check that

(16) $$\boldsymbol{g(t) = \text{Re}[G(t)],}$$

where

(17) $$\boldsymbol{G(t) = e^{(\alpha - i\beta)t}[(a_n + ib_n)t^n + \cdots + (a_1 + ib_1)t + (a_0 + ib_0)].}$$

Now suppose for the moment that we can find a complex-valued solution Y to the equation

(18) $$L[Y] = aY'' + bY' + cY = G.$$

Then, since a, b, and c are real numbers, we get a real-valued solution y to (14) by simply taking the real part of Y; that is, $y = \text{Re}\,Y$ solves (14). (Recall that in Lemma 2, page 167, we proved this fact for homogeneous equations.) Thus, we need focus only on finding a solution to (18).

The method of undetermined coefficients implies that any differential equation of the form

(19) $$L[Y] = e^{(\alpha \pm i\beta)t}[(a_n + ib_n)t^n + \cdots + (a_1 + ib_1)t + (a_0 + ib_0)]$$

has a solution of the form

(20) $$Y_p(t) = t^s e^{(\alpha \pm i\beta)t}[A_n t^n + \cdots + A_1 t + A_0],$$

where $A_n, \ldots, A_0$ are complex constants and s is the multiplicity of $\alpha + i\beta$ as a root of the auxiliary equation for the corresponding homogeneous equation $L[Y] = 0$. We can solve for the unknown constants A_j by substituting (20) into (19) and equating coefficients of like terms. With these facts in mind, we can (for the small price of using complex arithmetic) dispense with the methods of Section 4.5 and avoid the unpleasant task of computing derivatives of a function like $e^{3t}(2 + 3t + t^2)\sin(2t)$, which involves both exponential and trigonometric factors.

Carry out this procedure to determine particular solutions to the following equations:

(b) $y'' - y' - 2y = \cos t - \sin 2t$.
(c) $y'' + y = e^{-t}(\cos 2t - 3\sin 2t)$.
(d) $y'' - 2y' + 10y = te^t \sin 3t$.

The use of complex arithmetic not only streamlines the computations but also proves very useful in analyzing the response of a linear system to a sinusoidal input. Electrical engineers make good use of this in their study of *RLC* circuits by introducing the concept of **impedance.**

G Asymptotic Behavior of Solutions

In the application of linear systems theory to mechanical problems, we have encountered the equation

(21) $$y'' + py' + qy = f(t)\ ,$$

where p and q are positive constants with $p^2 < 4q$ and $f(t)$ is a forcing function for the system. In many cases it is important for the design engineer to know that a bounded forcing function gives rise only to bounded solutions. More specifically, how does the behavior of $f(t)$ for large values of t affect the asymptotic behavior of the solution? To answer this question, do the following:

(a) Show that the homogeneous equation associated with equation (21) has two linearly independent solutions given by

$$e^{\alpha t}\cos\beta t\ , \qquad e^{\alpha t}\sin\beta t\ ,$$

where $\alpha = -p/2 < 0$ and $\beta = \frac{1}{2}\sqrt{4q - p^2}$.

(b) Let $f(t)$ be a continuous function defined on the interval $[0, \infty)$. Use the variation of parameters formula to show that any solution to (21) on $[0, \infty)$ can be expressed in the form

(22) $$\begin{aligned} y(t) &= c_1 e^{\alpha t}\cos\beta t + c_2 e^{\alpha t}\sin\beta t \\ &\quad -\frac{1}{\beta}e^{\alpha t}\cos\beta t\int_0^t f(v)e^{-\alpha v}\sin\beta v\,dv \\ &\quad +\frac{1}{\beta}e^{\alpha t}\sin\beta t\int_0^t f(v)e^{-\alpha v}\cos\beta v\,dv\ . \end{aligned}$$

(c) Assuming that f is bounded on $[0, \infty)$ (that is, there exists a constant K such that $|f(v)| \le K$ for all $v \ge 0$), use the triangle inequality and other properties of the absolute value to show that $y(t)$ given in (22) satisfies

$$|y(t)| \le (|c_1| + |c_2|)e^{\alpha t} + \frac{2K}{|\alpha|\beta}(1 - e^{\alpha t})$$

for all $t > 0$.

(d) In a similar fashion, show that if $f_1(t)$ and $f_2(t)$ are two bounded continuous functions on $[0, \infty)$ such that $|f_1(t) - f_2(t)| \le \varepsilon$ for all $t > t_0$, and if ϕ_1 is a solution to (21) with $f = f_1$ and ϕ_2 is a solution to (21) with $f = f_2$, then

$$|\phi_1(t) - \phi_2(t)| \le Me^{\alpha t} + \frac{2\varepsilon}{|\alpha|\beta}(1 - e^{\alpha(t-t_0)})$$

for all $t > t_0$, where M is a constant that depends on ϕ_1 and ϕ_2 but not on t.

(e) Now assume $f(t) \to F_0$ as $t \to +\infty$, where F_0 is a constant. Use the result of part (d) to prove that any solution ϕ to (21) must satisfy $\phi(t) \to F_0/q$ as $t \to +\infty$. [*Hint:* Choose $f_1 = f$, $f_2 \equiv F_0$, $\phi_1 = \phi$, $\phi_2 \equiv F_0/q$.]

H GRAVITY TRAIN†

Legend has it that Robert Hooke proposed to Isaac Newton that a straight tunnel be drilled from Moscow to St. Petersburg to accommodate a "gravity train" serving the two cities. Upon departure gravity would accelerate the train toward the center of the tunnel, then decelerate it as it continued toward the terminal. It is amusing to investigate the mechanics of such a conveyance, under idealized conditions.

As depicted in Figure 4.38, x denotes the distance of the train from the center of the tunnel, R denotes its distance from the center of the earth, and mg denotes the pull of gravity.

(a) Show that Newton's law yields the differential equation

$$mx'' = -mgx/R \tag{23}$$

if friction is neglected.

(b) Assume that the tunnel lies so close to the surface of the earth that it is reasonable to regard g and R as constants—in which case (23) is analogous to the undamped mass–spring equation (2) on page 212 of Section 4.9, with mg/R as the effective "spring constant" k. Take $g = 9.8$ km/s^2 and $R = 6400$ km. The overland (surface) distance from Moscow to St. Petersburg is about 650 km. Find the equation of motion for the train departing at rest from Moscow at time $t = 0$.

(c) How long will it take the train to get to St. Petersburg?

(d) The American mathematician/entertainer Tom Lehrer composed a humorous song‡ about Nicolai Ivanovich Lobachevsky, in which a plagiarized dissertation travels from Dnepropetrovsk to Petropavlovsk (16,000 km) to Iliysk (8,500 km) to Novorossiysk (17,000 km) to Alexandrovsk (15,000 km) to Akmolinsk (1,000 km) to Tomsk (1,300 km) to Omsk (900 km) to Pinsk (21,000 km) to Minsk (3,000 km) to Moscow (5,000 km). If gravity trains were to connect all these cities, what would be unusual about their timetables?

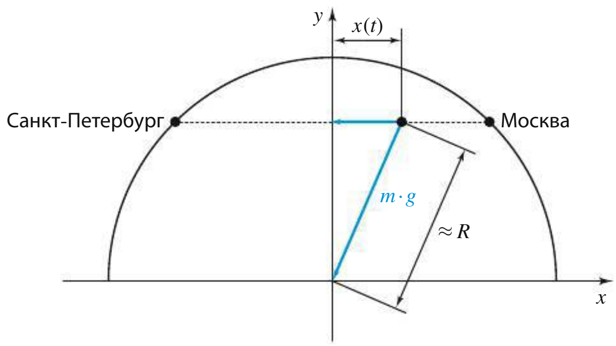


Figure 4.38 Force diagram of gravity train

†*Suggested by Valery Ochkov, National Research University Moscow; and Katarina Pisacic, Sveucilisni Centar Varazdin.*

‡As we go to press the song is archived at https://en.wikipedia.org/wiki/Lobachevsky_%28song%29, and Lehrer's performance can be viewed at https://www.youtube.com/watch?v=UQHaGhC7C2E.

CHAPTER

5

Introduction to Systems and Phase Plane Analysis

5.1 Interconnected Fluid Tanks

Two large tanks, each holding 24 liters of a brine solution, are interconnected by pipes as shown in Figure 5.1. Fresh water flows into tank A at a rate of 6 L/min, and fluid is drained out of tank B at the same rate; also 8 L/min of fluid are pumped from tank A to tank B, and 2 L/min from tank B to tank A. The liquids inside each tank are kept well stirred so that each mixture is homogeneous. If, initially, the brine solution in tank A contains x_0 kg of salt and that in tank B initially contains y_0 kg of salt, determine the mass of salt in each tank at time $t > 0$.†

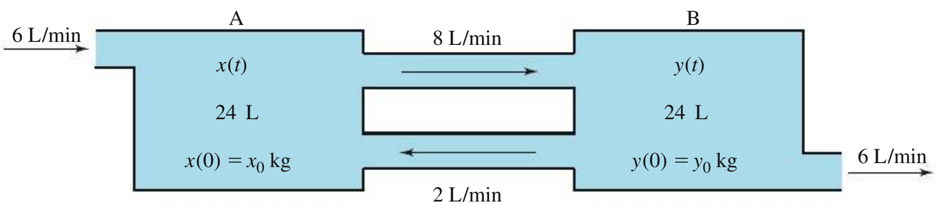


Figure 5.1 Interconnected fluid tanks

Note that the *volume* of liquid in each tank remains constant at 24 L because of the balance between the inflow and outflow volume rates. Hence, we have two unknown functions of t: the mass of salt $x(t)$ in tank A and the mass of salt $y(t)$ in tank B. By focusing attention on one tank at a time, we can derive two equations relating these unknowns. Since the system is being flushed with freshwater, we expect that the salt content of each tank will diminish to zero as $t \to +\infty$.

To formulate the equations for this system, we equate the rate of change of salt in each tank with the *net* rate at which salt is transferred to that tank. The salt *concentration* in tank A is $x(t)/24$ kg/L, so the upper interconnecting pipe carries salt out of tank A at a rate of $8x/24$ kg/min; similarly, the lower interconnecting pipe brings salt into tank A at the rate $2y/24$ kg/min (the concentration of salt in tank B is $y/24$ kg/L). The freshwater inlet, of course, transfers no salt (it simply maintains the volume in tank A at 24 L). From our premise,

$$\frac{dx}{dt} = \text{input rate} - \text{output rate} ,$$

†For this application we simplify the analysis by assuming the lengths and volumes of the pipes are sufficiently small that we can ignore the diffusive and advective dynamics taking place therein.

so the rate of change of the mass of salt in tank A is

$$\frac{dx}{dt} = \frac{2}{24}y - \frac{8}{24}x = \frac{1}{12}y - \frac{1}{3}x\,.$$

The rate of change of salt in tank B is determined by the same interconnecting pipes *and* by the drain pipe, carrying away $6y/24$ kg/min:

$$\frac{dy}{dt} = \frac{8}{24}x - \frac{2}{24}y - \frac{6}{24}y = \frac{1}{3}x - \frac{1}{3}y\,.$$

The interconnected tanks are thus governed by a *system* of differential equations:

$$\begin{aligned} x' &= -\frac{1}{3}x + \frac{1}{12}y\,, \\ y' &= \frac{1}{3}x - \frac{1}{3}y\,. \end{aligned} \tag{1}$$

Although both unknowns $x(t)$ and $y(t)$ appear in each of equations (1) (they are "coupled"), the structure is so transparent that we can obtain an equation for y alone by solving the second equation for x,

$$x = 3y' + y\,, \tag{2}$$

and substituting (2) in the first equation to eliminate x:

$$\begin{aligned} (3y' + y)' &= -\frac{1}{3}(3y' + y) + \frac{1}{12}y\,, \\ 3y'' + y' &= -y' - \frac{1}{3}y + \frac{1}{12}y\,, \end{aligned}$$

or

$$3y'' + 2y' + \frac{1}{4}y = 0\,.$$

This last equation, which is linear with constant coefficients, is readily solved by the methods of Section 4.2. Since the auxiliary equation

$$3r^2 + 2r + \frac{1}{4} = 0$$

has roots $-1/2$, $-1/6$, a general solution is given by

$$y(t) = c_1e^{-t/2} + c_2e^{-t/6}\,. \tag{3}$$

Having determined y, we use equation (2) to deduce a formula for x:

$$x(t) = 3\left(-\frac{c_1}{2}e^{-t/2} - \frac{c_2}{6}e^{-t/6}\right) + c_1e^{-t/2} + c_2e^{-t/6} = -\frac{1}{2}c_1e^{-t/2} + \frac{1}{2}c_2e^{-t/6}\,. \tag{4}$$

Formulas (3) and (4) contain two undetermined parameters, c_1 and c_2, which can be adjusted to meet the specified initial conditions:

$$x(0) = -\frac{1}{2}c_1 + \frac{1}{2}c_2 = x_0\,, \qquad y(0) = c_1 + c_2 = y_0\,,$$

or

$$c_1 = \frac{y_0 - 2x_0}{2}, \quad c_2 = \frac{y_0 + 2x_0}{2}.$$

Thus, the mass of salt in tanks A and B at time t are, respectively,

$$\text{(5)} \quad \begin{aligned} x(t) &= -\left(\frac{y_0 - 2x_0}{4}\right)e^{-t/2} + \left(\frac{y_0 + 2x_0}{4}\right)e^{-t/6}, \\ y(t) &= \left(\frac{y_0 - 2x_0}{2}\right)e^{-t/2} + \left(\frac{y_0 + 2x_0}{2}\right)e^{-t/6}. \end{aligned}$$

The ad hoc elimination procedure that we used to solve this example will be generalized and formalized in the next section, to find solutions of all *linear systems with constant coefficients*. Furthermore, in later sections we will show how to extend our numerical algorithms for first-order equations to *general* systems and will consider applications to coupled oscillators and electrical systems.

It is interesting to note from (5) that all solutions of the interconnected-tanks problem tend to the constant solution $x(t) \equiv 0, y(t) \equiv 0$ as $t \to +\infty$. (This is of course consistent with our physical expectations.) This constant solution will be identified as a *stable equilibrium solution* in Section 5.4, in which we introduce phase plane analysis. It turns out that, for a general class of systems, equilibria can be identified and classified so as to give qualitative information about the other solutions even when we cannot solve the system explicitly.

5.2 Differential Operators and the Elimination Method* for Systems

The notation $y'(t) = \dfrac{dy}{dt} = \dfrac{d}{dt}y$ was devised to suggest that the derivative of a function y is the result of *operating* on the function y with the differentiation operator $\dfrac{d}{dt}$. Indeed, second derivatives are formed by iterating the operation: $y''(t) = \dfrac{d^2y}{dt^2} = \dfrac{d}{dt}\dfrac{d}{dt}y$. Commonly, the symbol D is used instead of $\dfrac{d}{dt}$, and the second-order differential equation

$$y'' + 4y' + 3y = 0$$

is represented† by

$$D^2y + 4Dy + 3y = (D^2 + 4D + 3)[y] = 0.$$

So, we have implicitly adopted the convention that the operator "product," D times D, is interpreted as the *composition* of D with itself when it operates on functions: D^2y means $D(D[y])$; i.e., the second derivative. Similarly, the product $(D+3)(D+1)$ operates on a function via

$$\begin{aligned} (D+3)(D+1)[y] &= (D+3)[(D+1)[y]] = (D+3)[y' + y] \\ &= D[y' + y] + 3[y' + y] \\ &= (y'' + y') + (3y' + 3y) = y'' + 4y' + 3y = (D^2 + 4D + 3)[y]. \end{aligned}$$

---

*An alternative procedure to the methodology of this section will be described in Chapter 9. Although it involves the machinery of matrix analysis, it is preferable for large systems.

†Some authors utilize the identity operator I, defined by $I[y] = y$, and write more formally $D^2 + 4D + 3I$ instead of $D^2 + 4D + 3$.

Thus, $(D+3)(D+1)$ is the same operator as D^2+4D+3; when they are applied to twice-differentiable functions, the results are identical.

Example 1 Show that the operator $(D+1)(D+3)$ is also the same as D^2+4D+3.

Solution For any twice-differentiable function $y(t)$, we have

$$\begin{aligned}(D+1)(D+3)[y] &= (D+1)[(D+3)[y]] = (D+1)[y'+3y]\\ &= D[y'+3y]+1[y'+3y] = (y''+3y')+(y'+3y)\\ &= y''+4y'+3y = (D^2+4D+3)[y]\,.\end{aligned}$$

Hence, $(D+1)(D+3) = D^2+4D+3$. ◆

Since $(D+1)(D+3) = (D+3)(D+1) = D^2+4D+3$, it is tempting to generalize and propose that one can treat expressions like aD^2+bD+c as if they were ordinary polynomials in D. This is true, as long as we restrict the coefficients a, b, c to be *constants*. The following example, which has *variable* coefficients, is instructive.

Example 2 Show that $(D+3t)D$ is *not* the same as $D(D+3t)$.

Solution With $y(t)$ as before,

$$\begin{aligned}(D+3t)D[y] &= (D+3t)[y'] = y''+3ty'\,;\\ D(D+3t)[y] &= D[y'+3ty] = y''+3y+3ty'\,.\end{aligned}$$

They are not the same! ◆

Because the coefficient $3t$ is not a constant, it "interrupts" the interaction of the differentiation operator D with the function $y(t)$. As long as we only deal with expressions like aD^2+bD+c with *constant* coefficients $a, b,$ and $c,$ the "algebra" of differential operators follows the same rules as the algebra of polynomials. (See Problem 39 for elaboration on this point.)

This means that the familiar elimination method, used for solving *algebraic* systems like

$$\begin{aligned}3x-2y+z &= 4\,,\\ x+y-z &= 0\,,\\ 2x-y+3z &= 6\,,\end{aligned}$$

can be adapted to solve any system of *linear differential equations with constant coefficients.* In fact, we used this approach in solving the system that arose in the interconnected tanks problem of Section 5.1. Our goal in this section is to formalize this **elimination method** so that we can tackle more general linear constant coefficient systems.

We first demonstrate how the method applies to a linear system of two first-order differential equations of the form

$$\begin{aligned}a_1x'(t)+a_2x(t)+a_3y'(t)+a_4y(t) &= f_1(t)\,,\\ a_5x'(t)+a_6x(t)+a_7y'(t)+a_8y(t) &= f_2(t)\,,\end{aligned}$$

where $a_1, a_2, \ldots, a_8$ are constants and $x(t), y(t)$ is the function pair to be determined. In operator notation this becomes

$$\begin{aligned}(a_1D+a_2)[x]+(a_3D+a_4)[y] &= f_1\,,\\ (a_5D+a_6)[x]+(a_7D+a_8)[y] &= f_2\,.\end{aligned}$$

Example 3 Solve the system

$$\text{(1)} \qquad \begin{aligned} x'(t) &= 3x(t) - 4y(t) + 1\,, \\ y'(t) &= 4x(t) - 7y(t) + 10t\,. \end{aligned}$$

Solution The alert reader may observe that since y' is absent from the first equation, we could use the latter to express y in terms of x and x' and substitute into the second equation to derive an "uncoupled" equation containing only x and its derivatives. However, this simple trick will not work on more general systems (Problem 18 is an example).

To utilize the elimination method, we first write the system using the operator notation:

$$\text{(2)} \qquad \begin{aligned} (D-3)[x] + 4y &= 1\,, \\ -4x + (D+7)[y] &= 10t\,. \end{aligned}$$

Imitating the elimination procedure for algebraic systems, we can eliminate x from this system by adding 4 times the first equation to $(D-3)$ applied to the second equation. This gives

$$\big(16 + (D-3)(D+7)\big)[y] = 4 \cdot 1 + (D-3)[10t] = 4 + 10 - 30t\,,$$

which simplifies to

$$\text{(3)} \qquad (D^2 + 4D - 5)[y] = 14 - 30t\,.$$

Now equation (3) is just a second-order linear equation in y with constant coefficients that has the general solution

$$\text{(4)} \qquad y(t) = C_1 e^{-5t} + C_2 e^{t} + 6t + 2\,,$$

which can be found using undetermined coefficients.

To find $x(t)$, we have two options.

Method 1. We return to system (2) and eliminate y. This is accomplished by "multiplying" the first equation in (2) by $(D+7)$ and the second equation by -4 and then adding to obtain

$$(D^2 + 4D - 5)[x] = 7 - 40t\,.$$

This equation can likewise be solved using undetermined coefficients to yield

$$\text{(5)} \qquad x(t) = K_1 e^{-5t} + K_2 e^{t} + 8t + 5\,,$$

where we have taken K_1 and K_2 to be the arbitrary constants, which are not necessarily the same as C_1 and C_2 used in formula (4).

It is reasonable to expect that system (1) will involve only *two* arbitrary constants, since it consists of two first-order equations. Thus, the four constants C_1, C_2, K_1, and K_2 are not independent. To determine the relationships, we substitute the expressions for $x(t)$ and $y(t)$ given in (4) and (5) into one of the equations in (1)—say, the first one. This yields

$$\begin{aligned} -5K_1 e^{-5t} + K_2 e^{t} + 8 = \\ &3K_1 e^{-5t} + 3K_2 e^{t} + 24t + 15 - 4C_1 e^{-5t} - 4C_2 e^{t} - 24t - 8 + 1\,, \end{aligned}$$

which simplifies to

$$(4C_1 - 8K_1)e^{-5t} + (4C_2 - 2K_2)e^{t} = 0\,.$$

Because e^t and e^{-5t} are linearly independent functions on any interval, this last equation holds for all t only if

$$4C_1 - 8K_1 = 0 \qquad \text{and} \qquad 4C_2 - 2K_2 = 0\,.$$

Therefore, $K_1 = C_1/2$ and $K_2 = 2C_2$.

A solution to system (1) is then given by the pair

$$x(t) = \frac{1}{2}C_1e^{-5t} + 2C_2e^t + 8t + 5\,, \quad y(t) = C_1e^{-5t} + C_2e^t + 6t + 2\,. \tag{6}$$

As you might expect, this pair is a **general solution** to (1) in the sense that *any* solution to (1) can be expressed in this fashion.

Method 2. A simpler method for determining $x(t)$ once $y(t)$ is known is to use the system to obtain an equation for $x(t)$ in terms of $y(t)$ and $y'(t)$. In this example we can directly solve the second equation in (1) for $x(t)$:

$$x(t) = \frac{1}{4}y'(t) + \frac{7}{4}y(t) - \frac{5}{2}t\,.$$

Substituting $y(t)$ as given in (4) yields

$$\begin{aligned} x(t) &= \frac{1}{4}[-5C_1e^{-5t} + C_2e^t + 6] + \frac{7}{4}[C_1e^{-5t} + C_2e^t + 6t + 2] - \frac{5}{2}t \\ &= \frac{1}{2}C_1e^{-5t} + 2C_2e^t + 8t + 5\,, \end{aligned}$$

which agrees with (6). ◆

The above procedure works, more generally, for any linear system of two equations and two unknowns with *constant coefficients* regardless of the order of the equations. For example, if we let L_1, L_2, L_3, and L_4 denote linear differential operators with constant coefficients (i.e., polynomials in D), then the method can be applied to the linear system

$$\begin{aligned} L_1[x] + L_2[y] &= f_1\,, \\ L_3[x] + L_4[y] &= f_2\,. \end{aligned}$$

Because the system has constant coefficients, the operators commute (e.g., $L_2L_4 = L_4L_2$) and we can eliminate variables in the usual algebraic fashion. Eliminating the variable y gives

$$(L_1L_4 - L_2L_3)[x] = g_1\,, \tag{7}$$

where $g_1 := L_4[f_1] - L_2[f_2]$. Similarly, eliminating the variable x yields

$$(L_1L_4 - L_2L_3)[y] = g_2\,, \tag{8}$$

where $g_2 := L_1[f_2] - L_3[f_1]$. Now if $L_1L_4 - L_2L_3$ is a differential operator of order n, then a general solution for (7) contains n arbitrary constants, and a general solution for (8) also contains n arbitrary constants. Thus, a total of $2n$ constants arise. However, as we saw in Example 3, there are only n of these that are independent for the system; the remaining constants can be expressed in terms of these.† The pair of general solutions to (7) and (8) written in terms of the n independent constants is called a **general solution for the system.**

†For a proof of this fact, see *Ordinary Differential Equations*, by M. Tenenbaum and H. Pollard (Dover, New York, 1985), Chapter 7.

If it turns out that $L_1L_4 - L_2L_3$ is the zero operator, the system is said to be **degenerate.** As with the anomalous problem of solving for the points of intersection of two parallel or coincident lines, a degenerate system may have no solutions, or if it does possess solutions, they may involve any number of arbitrary constants (see Problems 23 and 24).

Elimination Procedure for 2 × 2 Systems

To find a general solution for the system

$$\begin{aligned} L_1[x] + L_2[y] &= f_1 , \\ L_3[x] + L_4[y] &= f_2 , \end{aligned}$$

where L_1, L_2, L_3, and L_4 are polynomials in $D = d/dt$:

(a) Make sure that the system is written in operator form.
(b) Eliminate one of the variables, say, y, and solve the resulting equation for $x(t)$. If the system is degenerate, stop! A separate analysis is required to determine whether or not there are solutions.
(c) (*Shortcut*) If possible, use the system to derive an equation that involves $y(t)$ but not its derivatives. [Otherwise, go to step (d).] Substitute the found expression for $x(t)$ into this equation to get a formula for $y(t)$. The expressions for $x(t)$, $y(t)$ give the desired general solution. ◆
(d) Eliminate x from the system and solve for $y(t)$. [Solving for $y(t)$ gives more constants—in fact, twice as many as needed.]
(e) Remove the extra constants by substituting the expressions for $x(t)$ and $y(t)$ into one or both of the equations in the system. Write the expressions for $x(t)$ and $y(t)$ in terms of the remaining constants. ◆

Example 4 Find a general solution for

$$\begin{aligned} x''(t) + y'(t) - x(t) + y(t) &= -1 , \\ x'(t) + y'(t) - x(t) &= t^2 . \end{aligned} \tag{9}$$

Solution We begin by expressing the system in operator notation:

$$\begin{aligned} (D^2 - 1)[x] + (D + 1)[y] &= -1 , \\ (D - 1)[x] + D[y] &= t^2 . \end{aligned} \tag{10}$$

Here $L_1 := D^2 - 1$, $L_2 := D + 1$, $L_3 := D - 1$, and $L_4 := D$.

Eliminating y gives [see (7)]:

$$\big((D^2 - 1)D - (D + 1)(D - 1)\big)[x] = D[-1] - (D + 1)[t^2] ,$$

which reduces to

$$(D^2 - 1)(D - 1)[x] = -2t - t^2 , \quad \text{or}$$
$$(D - 1)^2(D + 1)[x] = -2t - t^2 . \tag{11}$$

Since $(D - 1)^2(D + 1)$ is third order, we should expect three arbitrary constants in a general solution to system (9).

Although the methods of Chapter 4 focused on solving second-order equations, we have seen several examples of how they extend in a natural way to higher-order equations.[†] Applying this strategy to the third-order equation (11), we observe that the corresponding homogeneous equation has the auxiliary equation $(r-1)^2(r+1) = 0$ with roots $r = 1, 1, -1$. Hence, a general solution for the homogeneous equation is

$$x_h(t) = C_1e^t + C_2te^t + C_3e^{-t} .$$

To find a particular solution to (11), we use the method of undetermined coefficients with $x_p(t) = At^2 + Bt + C$. Substituting into (11) and solving for A, B, and C yields (after a little algebra)

$$x_p(t) = -t^2 - 4t - 6 .$$

Thus, a general solution to equation (11) is

$$x(t) = x_h(t) + x_p(t) = C_1e^t + C_2te^t + C_3e^{-t} - t^2 - 4t - 6 . \tag{12}$$

To find $y(t)$, we take the shortcut described in step (c) of the elimination procedure box. Subtracting the second equation in (10) from the first, we find

$$(D^2 - D)[x] + y = -1 - t^2 ,$$

so that

$$y = (D - D^2)[x] - 1 - t^2 .$$

Inserting the expression for $x(t)$, given in (12), we obtain

$$\begin{aligned} y(t) &= C_1e^t + C_2(te^t + e^t) - C_3e^{-t} - 2t - 4 \\ &\quad -[C_1e^t + C_2(te^t + 2e^t) + C_3e^{-t} - 2] - 1 - t^2 , \quad \text{or} \end{aligned}$$

$$y(t) = -C_2e^t - 2C_3e^{-t} - t^2 - 2t - 3 . \tag{13}$$

The formulas for $x(t)$ in (12) and $y(t)$ in (13) give the desired general solution to (9). ◆

The elimination method also applies to linear systems with three or more equations and unknowns; however, the process becomes more cumbersome as the number of equations and unknowns increases. The matrix methods presented in Chapter 9 are better suited for handling larger systems. Here we illustrate the elimination technique for a 3×3 system.

Example 5 Find a general solution to

$$\begin{aligned} x'(t) &= x(t) + 2y(t) - z(t) , \\ y'(t) &= x(t) + z(t) , \\ z'(t) &= 4x(t) - 4y(t) + 5z(t) . \end{aligned} \tag{14}$$

Solution We begin by expressing the system in operator notation:

$$\begin{aligned} (D-1)[x] - 2y + z &= 0 , \\ -x + D[y] - z &= 0 , \\ -4x + 4y + (D-5)[z] &= 0 . \end{aligned} \tag{15}$$

Eliminating z from the first two equations (by adding them) and then from the last two equations yields (after some algebra, which we omit)

$$\begin{aligned} (D-2)[x] + (D-2)[y] &= 0 , \\ -(D-1)[x] + (D-1)(D-4)[y] &= 0 . \end{aligned} \tag{16}$$

[†]More detailed treatment of higher-order equations is given in Chapter 6.

On eliminating x from this 2×2 system, we eventually obtain

$$(D-1)(D-2)(D-3)[y] = 0 ,$$

which has the general solution

$$y(t) = C_1e^t + C_2e^{2t} + C_3e^{3t} . \tag{17}$$

Taking the shortcut approach, we add the two equations in (16) to get an expression for x in terms of y and its derivatives, which simplifies to

$$x = (D^2 - 4D + 2)[y] = y'' - 4y' + 2y .$$

When we substitute the expression (17) for $y(t)$ into this equation, we find

$$x(t) = -C_1e^t - 2C_2e^{2t} - C_3e^{3t} . \tag{18}$$

Finally, using the second equation in (14) to solve for $z(t)$, we get

$$z(t) = y'(t) - x(t) ,$$

and substituting in for $y(t)$ and $x(t)$ yields

$$z(t) = 2C_1e^t + 4C_2e^{2t} + 4C_3e^{3t} . \tag{19}$$

The expressions for $x(t)$ in (18), $y(t)$ in (17), and $z(t)$ in (19) give a general solution with C_1, C_2, and C_3 as arbitrary constants. ◆

5.2 EXERCISES

1. Let $A = D - 1$, $B = D + 2$, $C = D^2 + D - 2$, where $D = d/dt$. For $y = t^3 - 8$, compute
   (a) $A[y]$ **(b)** $B[A[y]]$ **(c)** $B[y]$
   (d) $A[B[y]]$ **(e)** $C[y]$
2. Show that the operator $(D-1)(D+2)$ is the same as the operator $D^2 + D - 2$.

In Problems 3–18, use the elimination method to find a general solution for the given linear system, where differentiation is with respect to t.

3. $x' + 2y = 0 ,$
   $x' - y' = 0$
4. $x' = x - y ,$
   $y' = y - 4x$
5. $x' + y' - x = 5 ,$
   $x' + y' + y = 1$
6. $x' = 3x - 2y + \sin t ,$
   $y' = 4x - y - \cos t$
7. $(D+1)[u] - (D+1)[v] = e^t ,$
   $(D-1)[u] + (2D+1)[v] = 5$
8. $(D-3)[x] + (D-1)[y] = t ,$
   $(D+1)[x] + (D+4)[y] = 1$
9. $x' + y' + 2x = 0 ,$
   $x' + y' - x - y = \sin t$
10. $2x' + y' - x - y = e^{-t} ,$
    $x' + y' + 2x + y = e^t$
11. $(D^2 - 1)[u] + 5v = e^t ,$
    $2u + (D^2 + 2)[v] = 0$
12. $D^2[u] + D[v] = 2 ,$
    $4u + D[v] = 6$
13. $\dfrac{dx}{dt} = x - 4y ,$
    $\dfrac{dy}{dt} = x + y$
14. $\dfrac{dx}{dt} + y = t^2 ,$
    $-x + \dfrac{dy}{dt} = 1$
15. $\dfrac{dw}{dt} = 5w + 2z + 5t ,$
    $\dfrac{dz}{dt} = 3w + 4z + 17t$
16. $\dfrac{dx}{dt} + x + \dfrac{dy}{dt} = e^{4t} ,$
    $2x + \dfrac{d^2y}{dt^2} = 0$
17. $x'' + 5x - 4y = 0 ,$
    $-x + y'' + 2y = 0$
18. $x'' + y'' - x' = 2t ,$
    $x'' + y' - x + y = -1$

In Problems 19–21, solve the given initial value problem.

19. $\dfrac{dx}{dt} = 4x + y ; \quad x(0) = 1 ,$
    $\dfrac{dy}{dt} = -2x + y ; \quad y(0) = 0$
20. $\dfrac{dx}{dt} = 2x + y - e^{2t} ; \quad x(0) = 1 ,$
    $\dfrac{dy}{dt} = x + 2y ; \quad y(0) = -1$

21. $\dfrac{d^2x}{dt^2} = y\,;\quad x(0) = 3\,,\qquad x'(0) = 1\,,$

$\dfrac{d^2y}{dt^2} = x\,;\quad y(0) = 1\,,\qquad y'(0) = -1$

22. Verify that the solution to the initial value problem

$$x' = 5x - 3y - 2\,;\quad x(0) = 2\,,$$
$$y' = 4x - 3y - 1\,;\quad y(0) = 0$$

satisfies $|x(t)| + |y(t)| \to +\infty$ as $t \to +\infty$.

In Problems 23 and 24, show that the given linear system is degenerate. In attempting to solve the system, determine whether it has no solutions or infinitely many solutions.

23. $(D-1)[x] + (D-1)[y] = -3e^{-2t}\,,$

$(D+2)[x] + (D+2)[y] = 3e^{t}$

24. $D[x] + (D+1)[y] = e^{t}\,,$

$D^2[x] + (D^2 + D)[y] = 0$

In Problems 25–28, use the elimination method to find a general solution for the given system of three equations in the three unknown functions $x(t)$, $y(t)$, $z(t)$.

25. $x' = x + 2y - z\,,$

$y' = x + z\,,$

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

26. $x' = 3x + y - z\,,$

$y' = x + 2y - z\,,$

$z' = 3x + 3y - z$

27. $x' = 4x - 4z\,,$

$y' = 4y - 2z\,,$

$z' = -2x - 4y + 4z$

28. $x' = x + 2y + z\,,$

$y' = 6x - y\,,$

$z' = -x - 2y - z$

*In Problems 29 and 30, determine the range of values (if any) of the parameter λ that will ensure **all** solutions $x(t)$, $y(t)$ of the given system remain bounded as $t \to +\infty$.*

29. $\dfrac{dx}{dt} = \lambda x - y\,,$

$\dfrac{dy}{dt} = 3x + y$

30. $\dfrac{dx}{dt} = -x + \lambda y\,,$

$\dfrac{dy}{dt} = x - y$

31. Two large tanks, each holding 100 L of liquid, are interconnected by pipes, with the liquid flowing from tank A into tank B at a rate of 3 L/min and from B into A at a rate of 1 L/min (see Figure 5.2). The liquid inside each tank is kept well stirred. A brine solution with a concentration of 0.2 kg/L of salt flows into tank A at a rate of 6 L/min. The (diluted) solution flows out of the system from tank A at 4 L/min and from tank B at 2 L/min. If, initially, tank A contains pure water and tank B contains 20 kg of salt, determine the mass of salt in each tank at time $t \geq 0$.

32. In Problem 31, 3 L/min of liquid flowed from tank A into tank B and 1 L/min from B into A. Determine the mass of salt in each tank at time $t \geq 0$ if, instead, 5 L/min flows from A into B and 3 L/min flows from B into A, with all other data the same.

33. In Problem 31, assume that no solution flows out of the system from tank B, only 1 L/min flows from A into B, and only 4 L/min of brine flows into the system at tank A, other data being the same. Determine the mass of salt in each tank at time $t \geq 0$.

34. **Feedback System with Pooling Delay.** Many physical and biological systems involve time delays. A pure time delay has its output the same as its input but shifted in time. A more common type of delay is *pooling delay*. An example of such a feedback system is shown in Figure 5.3 on page 251. Here the level of fluid in tank B determines the rate at which fluid enters tank A. Suppose this rate is given by $R_1(t) = \alpha[V - V_2(t)]$, where α and V are positive constants and $V_2(t)$ is the volume of fluid in tank B at time t.

(a) If the outflow rate R_3 from tank B is constant and the flow rate R_2 from tank A into B is $R_2(t) = KV_1(t)$, where K is a positive constant and $V_1(t)$ is the volume of fluid in tank A at time t, then show that this feedback system is governed by the system

$$\frac{dV_1}{dt} = \alpha\big(V - V_2(t)\big) - KV_1(t)\,,$$
$$\frac{dV_2}{dt} = KV_1(t) - R_3\,.$$

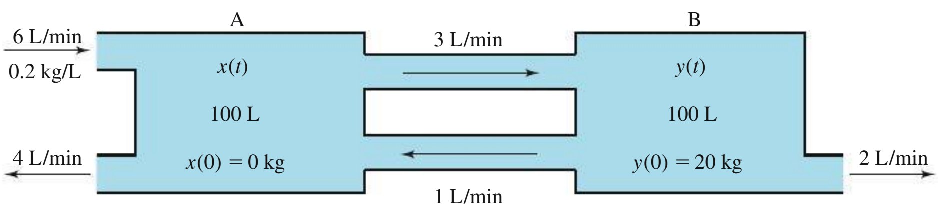


Figure 5.2 Mixing problem for interconnected tanks

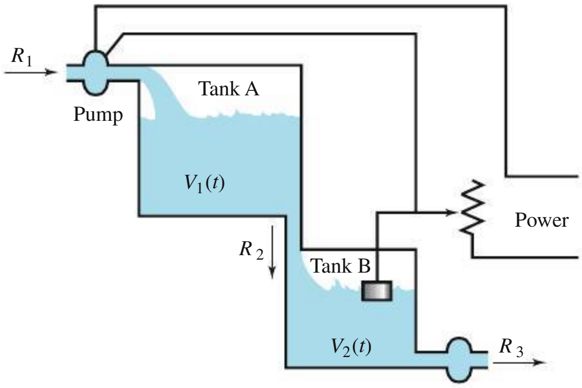


Figure 5.3 Feedback system with pooling delay

(b) Find a general solution for the system in part (a) when $\alpha = 5\ (\text{min})^{-1}$, $V = 20$ L, $K = 2\ (\text{min})^{-1}$, and $R_3 = 10$ L/min.

(c) Using the general solution obtained in part (b), what can be said about the volume of fluid in each of the tanks as $t \to +\infty$?

35. A house, for cooling purposes, consists of two zones: the attic area zone A and the living area zone B (see Figure 5.4). The living area is cooled by a 2-ton air conditioning unit that removes 24,000 Btu/hr. The heat capacity of zone B is $1/2$°F per thousand Btu. The time constant for heat transfer between zone A and the outside is 2 hr, between zone B and the outside is 4 hr, and between the two zones is 4 hr. If the outside temperature stays at 100°F, how warm does it eventually get in the attic zone A? (Heating and cooling of buildings was treated in Section 3.3 on page 102.)

36. A building consists of two zones A and B (see Figure 5.5). Only zone A is heated by a furnace, which generates 80,000 Btu/hr. The heat capacity of zone A is $1/4$°F per thousand Btu. The time constant for heat transfer between zone A and the outside is 4 hr, between the unheated zone B and the outside is 5 hr, and between the two zones is 2 hr. If the outside temperature stays at 0°F, how cold does it eventually get in the unheated zone B?

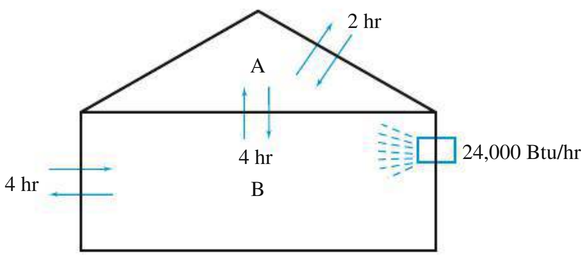


Figure 5.4 Air-conditioned house with attic

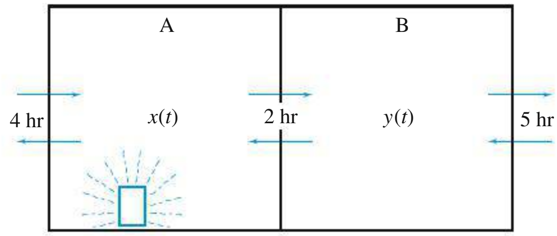


Figure 5.5 Two-zone building with one zone heated

37. In Problem 36, if a small furnace that generates 1000 Btu/hr is placed in zone B, determine the coldest it would eventually get in zone B if zone B has a heat capacity of 2°F per thousand Btu.

38. Arms Race. A simplified mathematical model for an arms race between two countries whose expenditures for defense are expressed by the variables $x(t)$ and $y(t)$ is given by the linear system

$$\frac{dx}{dt} = 2y - x + a\,; \qquad x(0) = 1\,,$$
$$\frac{dy}{dt} = 4x - 3y + b\,; \qquad y(0) = 4\,,$$

where a and b are constants that measure the trust (or distrust) each country has for the other. Determine whether there is going to be disarmament (x and y approach 0 as t increases), a stabilized arms race (x and y approach a constant as $t \to +\infty$), or a runaway arms race (x and y approach $+\infty$ as $t \to +\infty$).

39. Let A, B, and C represent three linear differential operators with constant coefficients; for example,

$$A := a_2D^2 + a_1D + a_0,\ B := b_2D^2 + b_1D + b_0\,,$$
$$C := c_2D^2 + c_1D + c_0\,,$$

where the a's, b's, and c's are constants. Verify the following properties:†

(a) Commutative laws:

$$A + B = B + A\,,$$
$$AB = BA\,.$$

(b) Associative laws:

$$(A + B) + C = A + (B + C)\,,$$
$$(AB)C = A(BC)\,.$$

(c) Distributive law: $A(B + C) = AB + AC$.

†We say that two operators A and B are *equal* if $A[y] = B[y]$ for all functions y with the necessary derivatives.

5.3 Solving Systems and Higher-Order Equations Numerically

Although we studied a half-dozen analytic methods for obtaining solutions to first-order ordinary differential equations in Chapter 2, the techniques for higher-order equations, or systems of equations, are much more limited. Chapter 4 focused on solving the linear constant-coefficient second-order equation. The elimination method of the previous section is also restricted to constant-coefficient systems. And, indeed, higher-order linear constant-coefficient equations and systems can be solved analytically by extensions of these methods, as we will see in Chapters 6, 7, and 9.

However, if the equations—even a single second-order linear equation—have variable coefficients, the solution process is much less satisfactory. As will be seen in Chapter 8, the solutions are expressed as infinite series, and their computation can be very laborious (with the notable exception of the Cauchy–Euler, or equidimensional, equation). And we know virtually nothing about how to obtain exact solutions to nonlinear second-order equations.

Fortunately, all the cases that arise (constant or variable coefficients, nonlinear, higher-order equations or systems) can be addressed by a single formulation that lends itself to a multitude of *numerical* approaches. In this section we'll see how to express differential equations as a *system in normal form* and then show how the basic Euler method for computer solution can be easily "vectorized" to apply to such systems. Although subsequent chapters will return to analytic solution methods, the vectorized version of the Euler technique or the more efficient Runge–Kutta technique will hereafter be available as fallback methods for numerical exploration of intractable problems.

Normal Form

A system of m differential equations in the m unknown functions $x_1(t), x_2(t), \ldots, x_m(t)$ expressed as

$$\begin{aligned} x_1'(t) &= f_1(t, x_1, x_2, \ldots, x_m) \ , \\ x_2'(t) &= f_2(t, x_1, x_2, \ldots, x_m) \ , \\ &\vdots \\ x_m'(t) &= f_m(t, x_1, x_2, \ldots, x_m) \end{aligned} \tag{1}$$

is said to be in **normal form.** Notice that (1) consists of m first-order equations that collectively look like a *vectorized* version of the single generic first-order equation

$$x' = f(t, x) \ , \tag{2}$$

and that the system expressed in equation (1) of Section 5.1 takes this form, as do equations (1) and (14) in Section 5.2. An initial value problem for (1) entails finding a solution to this system that satisfies the initial conditions

$$x_1(t_0) = a_1, \quad x_2(t_0) = a_2, \quad \ldots, \quad x_m(t_0) = a_m$$

for prescribed values $t_0, a_1, a_2, \ldots, a_m$.

The importance of the normal form is underscored by the fact that most professional codes for initial value problems presume that the system is written in this form. Furthermore, for a *linear* system in normal form, the powerful machinery of linear algebra can be readily applied. [Indeed, in Chapter 9 we will show how the solutions $x(t) = ce^{at}$ of the simple equation $x' = ax$ can be generalized to constant-coefficient systems in normal form.]

For these reasons it is gratifying to note that a (single) higher-order equation can always be converted to an equivalent system of first-order equations.

To convert an mth-order differential equation

$$y^{(m)}(t) = f\left(t, y, y', \ldots, y^{(m-1)}\right) \tag{3}$$

into a first-order system, we introduce, as additional unknowns, the sequence of derivatives of y:

$$x_1(t) := y(t), \quad x_2(t) := y'(t), \quad \ldots, \quad x_m(t) := y^{(m-1)}(t)\,.$$

With this scheme, we obtain $m - 1$ first-order equations quite trivially:

$$\begin{aligned} x_1'(t) &= y'(t) = x_2(t)\,, \\ x_2'(t) &= y''(t) = x_3(t)\,, \\ &\vdots \\ x_{m-1}'(t) &= y^{(m-1)}(t) = x_m(t)\,. \end{aligned} \tag{4}$$

The mth and final equation then constitutes a restatement of the original equation (3) in terms of the new unknowns:

$$x_m'(t) = y^{(m)}(t) = f(t, x_1, x_2, \ldots, x_m)\,. \tag{5}$$

If equation (3) has initial conditions $y(t_0) = a_1, y'(t_0) = a_2, \ldots, y^{(m-1)}(t_0) = a_m$, then the system (4)–(5) has initial conditions $x_1(t_0) = a_1, x_2(t_0) = a_2, \ldots, x_m(t_0) = a_m$.

Example 1 Convert the initial value problem

$$y''(t) + 3ty'(t) + y(t)^2 = \sin t\,; \quad y(0) = 1\,, \quad y'(0) = 5 \tag{6}$$

into an initial value problem for a system in normal form.

Solution We first express the differential equation in (6) as

$$y''(t) = -3ty'(t) - y(t)^2 + \sin t\,.$$

Setting $x_1(t) := y(t)$ and $x_2(t) := y'(t)$, we obtain

$$\begin{aligned} x_1'(t) &= x_2(t)\,, \\ x_2'(t) &= -3tx_2(t) - x_1(t)^2 + \sin t\,. \end{aligned}$$

The initial conditions transform to $x_1(0) = 1, x_2(0) = 5$. ◆

Euler's Method for Systems in Normal Form

Recall from Section 1.4 that Euler's method for solving a single first-order equation (2) is based on estimating the solution x at time $(t_0 + h)$ using the approximation

$$x(t_0 + h) \approx x(t_0) + hx'(t_0) = x(t_0) + hf\left(t_0, x(t_0)\right)\,, \tag{7}$$

and that as a consequence the algorithm can be summarized by the recursive formulas

$$t_{n+1} = t_n + h\,, \tag{8}$$

$$x_{n+1} = x_n + hf(t_n, x_n), \qquad n = 0, 1, 2, \ldots \tag{9}$$

[compare equations (2) and (3), Section 1.4]. Now we can apply the approximation (7) to each of the equations in the system (1):

$$x_k(t_0 + h) \approx x_k(t_0) + hx_k'(t_0) = x_k(t_0) + hf_k\left(t_0, x_1(t_0), x_2(t_0), \ldots, x_m(t_0)\right)\,, \tag{10}$$

and for $k = 1, 2, \ldots m$, we are led to the recursive formulas

$$t_{n+1} = t_n + h\,, \tag{11}$$

$$\begin{aligned} x_{1;n+1} &= x_{1;n} + hf_1(t_n, x_{1;n}, x_{2;n}, \ldots, x_{m;n})\,, \\ x_{2;n+1} &= x_{2;n} + hf_2(t_n, x_{1;n}, x_{2;n}, \ldots, x_{m;n})\,, \\ &\vdots \\ x_{m;n+1} &= x_{m;n} + hf_m(t_n, x_{1;n}, x_{2;n}, \ldots, x_{m;n}) \qquad (n = 0, 1, 2, \ldots)\,. \end{aligned} \tag{12}$$

Here we are burdened with the ungainly notation $x_{p;n}$ for the approximation to the value of the pth-function x_p at time $t = t_0 + nh$; i.e., $x_{p;n} \approx x_p(t_0 + nh)$. However, if we treat the unknowns and right-hand members of (1) as components of vectors

$$\begin{aligned} \mathbf{x}(t) &\coloneqq [x_1(t), x_2(t), \ldots, x_m(t)]\,, \\ \mathbf{f}(t, \mathbf{x}) &\coloneqq [f_1(t, x_1, x_2, \ldots, x_m), f_2(t, x_1, x_2, \ldots, x_m), \ldots, f_m(t, x_1, x_2, \ldots, x_m)]\,, \end{aligned}$$

then (12) can be expressed in the much neater form

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h\mathbf{f}(t_n, \mathbf{x}_n)\,. \tag{13}$$

Example 2 Use the vectorized Euler method with step size $h = 0.1$ to find an approximation for the solution to the initial value problem

$$y''(t) + 4y'(t) + 3y(t) = 0; \quad y(0) = 1.5\,, \quad y'(0) = -2.5\,, \tag{14}$$

on the interval $[0, 1]$.

Solution For the given step size, the method will yield approximations for $y(0.1), y(0.2), \ldots, y(1.0)$. To apply the vectorized Euler method to (14), we first convert it to normal form. Setting $x_1 = y$ and $x_2 = y'$, we obtain the system

$$\begin{aligned} x_1' &= x_2; & x_1(0) &= 1.5\,, \\ x_2' &= -4x_2 - 3x_1; & x_2(0) &= -2.5\,. \end{aligned} \tag{15}$$

Comparing (15) with (1) we see that $f_1(t, x_1, x_2) = x_2$ and $f_2(t, x_1, x_2) = -4x_2 - 3x_1$. With the starting values of $t_0 = 0$, $x_{1;0} = 1.5$, and $x_{2;0} = -2.5$, we compute

$$\begin{cases} x_1(0.1) \approx x_{1;1} = x_{1;0} + hx_{2;0} = 1.5 + 0.1(-2.5) = 1.25\,, \\ x_2(0.1) \approx x_{2;1} = x_{2;0} + h(-4x_{2;0} - 3x_{1;0}) = -2.5 + 0.1[-4(-2.5) - 3 \cdot 1.5] = -1.95\,; \end{cases}$$

$$\begin{cases} x_1(0.2) \approx x_{1;2} = x_{1;1} + hx_{2;1} = 1.25 + 0.1(-1.95) = 1.055\,, \\ x_2(0.2) \approx x_{2;2} = x_{2;1} + h(-4x_{2;1} - 3x_{1;1}) = -1.95 + 0.1[-4(-1.95) - 3 \cdot 1.25] = -1.545\,. \end{cases}$$

Continuing the algorithm we compute the remaining values. These are listed in Table 5.1 on page 255, along with the exact values calculated via the methods of Chapter 4. Note that the $x_{2;n}$ column gives approximations to $y'(t)$, since $x_2(t) \equiv y'(t)$. ◆

TABLE 5.1 Approximations of the Solution to (14) in Example 2

$t = n(0.1)$	$x_{1;n}$	y Exact	$x_{2;n}$	y' Exact
0	1.5	1.5	−2.5	−2.5
0.1	1.25	1.275246528	−1.95	−2.016064749
0.2	1.055	1.093136571	−1.545	−1.641948207
0.3	0.9005	0.944103051	−1.2435	−1.35067271
0.4	0.77615	0.820917152	−1.01625	−1.122111364
0.5	0.674525	0.71809574	−0.842595	−0.9412259
0.6	0.5902655	0.63146108	−0.7079145	−0.796759968
0.7	0.51947405	0.557813518	−0.60182835	−0.680269946
0.8	0.459291215	0.494687941	−0.516939225	−0.585405894
0.9	0.407597293	0.440172416	−0.4479509	−0.507377929
1	0.362802203	0.392772975	−0.391049727	−0.442560044

Euler's method is modestly accurate for this problem with a step size of $h = 0.1$. The next example demonstrates the effects of using a sequence of smaller values of h to improve the accuracy.

Example 3 For the initial value problem of Example 2, use Euler's method to estimate $y(1)$ for successively halved step sizes $h = 0.1, 0.05, 0.025, 0.0125, 0.00625$.

Solution Using the same scheme as in Example 2, we find the following approximations, denoted by $y(1;h)$ (obtained with step size h):

h	0.1	0.05	0.025	0.0125	0.00625
$y(1;h)$	0.36280	0.37787	0.38535	0.38907	0.39092

[Recall that the exact value, rounded to 5 decimal places, is $y(1) = 0.39277$.] ◆

The Runge–Kutta scheme described in Section 3.7 is easy to vectorize also; details are given on the following page. As would be expected, its performance is considerably more accurate, yielding five-decimal agreement with the exact solution for a step size of 0.05:

h	0.1	0.05	0.025	0.0125	0.00625
$y(1;h)$	0.39278	0.39277	0.39277	0.39277	0.39277

As in Section 3.7, both algorithms can be coded so as to repeat the calculation of $y(1)$ with a sequence of smaller step sizes until two consecutive estimates agree to within some prespecified tolerance ε. Here one should interpret "two estimates agree to within ε" to mean that *each component* of the successive vector approximants [i.e., approximants to $y(1)$ and $y'(1)$] should agree to within ε.

An Application to Population Dynamics

A mathematical model for the population dynamics of competing species, one a predator with population $x_2(t)$ and the other its prey with population $x_1(t)$, was developed independently in the early 1900s by A. J. Lotka and V. Volterra. It assumes that there is plenty of food available for the prey to eat, so the birthrate of the prey should follow the Malthusian or exponential law (see Section 3.2); that is, the birthrate of the prey is Ax_1, where A is a positive constant. The death rate of the prey depends on the number of interactions between the predators and the prey. This is modeled by the expression Bx_1x_2, where B is a positive constant. Therefore, the rate of change in the population of the prey per unit time is $dx_1/dt = Ax_1 - Bx_1x_2$. Assuming that the predators depend entirely on the prey for their food, it is argued that the birthrate of the predators depends on the number of interactions with the prey; that is, the birthrate of predators is Dx_1x_2, where D is a positive constant. The death rate of the predators is assumed to be Cx_2 because without food the population would die off at a rate proportional to the population present. Hence, the rate of change in the population of predators per unit time is $dx_2/dt = -Cx_2 + Dx_1x_2$. Combining these two equations, we obtain the Volterra–Lotka system for the population dynamics of two competing species:

$$
\begin{aligned}
x_1' &= Ax_1 - Bx_1x_2 , \\
x_2' &= -Cx_2 + Dx_1x_2 .
\end{aligned}
\tag{16}
$$

Such systems are in general not explicitly solvable. In the following example, we obtain an approximate solution for such a system by utilizing the vectorized form of the Runge–Kutta algorithm.

For the system of two equations

$$
\begin{aligned}
x_1' &= f_1(t, x_1, x_2) , \\
x_2' &= f_2(t, x_1, x_2) ,
\end{aligned}
$$

with initial conditions $x_1(t_0) = x_{1;0}$, $x_2(t_0) = x_{2;0}$, the vectorized form of the Runge–Kutta recursive equations (cf. (14), page 254) becomes

$$
\begin{cases}
t_{n+1} := t_n + h \qquad (n = 0, 1, 2, \ldots) , \\
x_{1;n+1} := x_{1;n} + \frac{1}{6}(k_{1,1} + 2k_{1,2} + 2k_{1,3} + k_{1,4}) , \\
x_{2;n+1} := x_{2;n} + \frac{1}{6}(k_{2,1} + 2k_{2,2} + 2k_{2,3} + k_{2,4}) ,
\end{cases}
\tag{17}
$$

where h is the step size and, for $i = 1$ and 2,

$$
\begin{cases}
k_{i,1} := hf_i(t_n, x_{1;n}, x_{2;n}) , \\
k_{i,2} := hf_i\left(t_n + \frac{h}{2},\ x_{1;n} + \frac{1}{2}k_{1,1},\ x_{2;n} + \frac{1}{2}k_{2,1}\right) , \\
k_{i,3} := hf_i\left(t_n + \frac{h}{2},\ x_{1;n} + \frac{1}{2}k_{1,2},\ x_{2;n} + \frac{1}{2}k_{2,2}\right) , \\
k_{i,4} := hf_i(t_n + h,\ x_{1;n} + k_{1,3},\ x_{2;n} + k_{2,3}) .
\end{cases}
\tag{18}
$$

It is important to note that both $k_{1,1}$ and $k_{2,1}$ must be computed before either $k_{1,2}$ or $k_{2,2}$. Similarly, both $k_{1,2}$ and $k_{2,2}$ are needed to compute $k_{1,3}$ and $k_{2,3}$, etc. In Appendix F, program outlines are given for applying the method to graph approximate solutions over a specified interval $[t_0, t_1]$ or to obtain approximations of the solutions at a specified point to within a desired tolerance.

Example 4 Use the classical fourth-order Runge–Kutta algorithm for systems to approximate the solution of the initial value problem

$$\textbf{(19)} \qquad \begin{aligned} x_1' &= 2x_1 - 2x_1x_2 ; \qquad & x_1(0) &= 1 , \\ x_2' &= x_1x_2 - x_2 ; \qquad & x_2(0) &= 3 \end{aligned}$$

at $t = 1$. Starting with $h = 1$, continue halving the step size until two successive approximations of $x_1(1)$ and of $x_2(1)$ differ by at most 0.0001.

Solution Here $f_1(t, x_1, x_2) = 2x_1 - 2x_1x_2$ and $f_2(t, x_1, x_2) = x_1x_2 - x_2$. With the inputs $t_0 = 0$, $x_{1;0} = 1$, $x_{2;0} = 3$, we proceed with the algorithm to compute $x_1(1; 1)$ and $x_2(1; 1)$, the approximations to $x_1(1)$, $x_2(1)$ using $h = 1$. We find from the formulas in (18) that

$$\begin{aligned}
k_{1,1} &= h(2x_{1;0} - 2x_{1;0}x_{2;0}) = 2(1) - 2(1)(3) = -4 , \\
k_{2,1} &= h(x_{1;0}x_{2;0} - x_{2;0}) = (1)(3) - 3 = 0 , \\
k_{1,2} &= h\left[2\left(x_{1;0} + \tfrac{1}{2}k_{1,1}\right) - 2\left(x_{1;0} + \tfrac{1}{2}k_{1,1}\right)\left(x_{2;0} + \tfrac{1}{2}k_{2,1}\right)\right] \\
&= 2\left[1 + \tfrac{1}{2}(-4)\right] - 2\left[1 + \tfrac{1}{2}(-4)\right]\left[3 + \tfrac{1}{2}(0)\right] \\
&= -2 + 2(3) = 4 , \\
k_{2,2} &= h\left[\left(x_{1;0} + \tfrac{1}{2}k_{1,1}\right)\left(x_{2;0} + \tfrac{1}{2}k_{2,1}\right) - \left(x_{2;0} + \tfrac{1}{2}k_{2,1}\right)\right] \\
&= \left[1 + \tfrac{1}{2}(-4)\right]\left[3 + \tfrac{1}{2}(0)\right] - \left[3 + \tfrac{1}{2}(0)\right] \\
&= (-1)(3) - 3 = -6 ,
\end{aligned}$$

and similarly we compute

$$\begin{aligned}
k_{1,3} &= h\left[2\left(x_{1;0} + \tfrac{1}{2}k_{1,2}\right) - 2\left(x_{1;0} + \tfrac{1}{2}k_{1,2}\right)\left(x_{2;0} + \tfrac{1}{2}k_{2,2}\right)\right] = 6 , \\
k_{2,3} &= h\left[\left(x_{1;0} + \tfrac{1}{2}k_{1,2}\right)\left(x_{2;0} + \tfrac{1}{2}k_{2,2}\right) - \left(x_{2;0} + \tfrac{1}{2}k_{2,2}\right)\right] = 0 , \\
k_{1,4} &= h[2(x_{1;0} + k_{1,3}) - 2(x_{1;0} + k_{1,3})(x_{2;0} + k_{2,3})] = -28 , \\
k_{2,4} &= h[(x_{1;0} + k_{1,3})(x_{2;0} + k_{2,3}) - (x_{2;0} + k_{2,3})] = -18 .
\end{aligned}$$

Inserting these values into formula (17), we get

$$\begin{aligned}
x_{1;1} &= x_{1;0} + \frac{1}{6}(k_{1,1} + 2k_{1,2} + 2k_{1,3} + k_{1,4}) \\
&= 1 + \frac{1}{6}(-4 + 8 + 12 - 28) = -1 , \\
x_{2;1} &= x_{2;0} + \frac{1}{6}(k_{2,1} + 2k_{2,2} + 2k_{2,3} + k_{2,4}) , \\
&= 3 + \frac{1}{6}(0 - 12 + 0 + 18) = 4 ,
\end{aligned}$$

as the respective approximations to $x_1(1)$ and $x_2(1)$.

Repeating the algorithm with $h = 1/2$ ($N = 2$) we obtain the approximations $x_1(1; 2^{-1})$ and $x_2(1; 2^{-1})$ for $x_1(1)$ and $x_2(1)$. In Table 5.2, we list the approximations $x_1(1; 2^{-m})$ and $x_2(1; 2^{-m})$ for $x_1(1)$ and $x_2(1)$ using step size $h = 2^{-m}$ for $m = 0, 1, 2, 3$, and 4. We stopped at $m = 4$, since both

$$|x_1(1; 2^{-3}) - x_1(1; 2^{-4})| = 0.00006 < 0.0001$$

and

$$|x_2(1; 2^{-3}) - x_2(1; 2^{-4})| = 0.00001 < 0.0001 .$$

Hence, $x_1(1) \approx 0.07735$ and $x_2(1) \approx 1.46445$, with tolerance 0.0001. ◆

TABLE 5.2 Approximations of the Solution to System (19) in Example 4

m	h	$x_1(1; h)$	$x_2(1; h)$
0	1.0	−1.0	4.0
1	0.5	0.14662	1.47356
2	0.25	0.07885	1.46469
3	0.125	0.07741	1.46446
4	0.0625	0.07735	1.46445

To get a better feel for the solution to system (19), we have graphed in Figure 5.6 an approximation of the solution for $0 \le t \le 12$, using linear interpolation to connect the vectorized Runge–Kutta approximants for the points $t = 0, 0.125, 0.25, \ldots, 12.0$ (i.e., with $h = 0.125$). From the graph it appears that the components x_1 and x_2 are periodic in the variable t. Phase plane analysis is used in Section 5.5 to show that, indeed, Volterra–Lotka equations have periodic solutions.

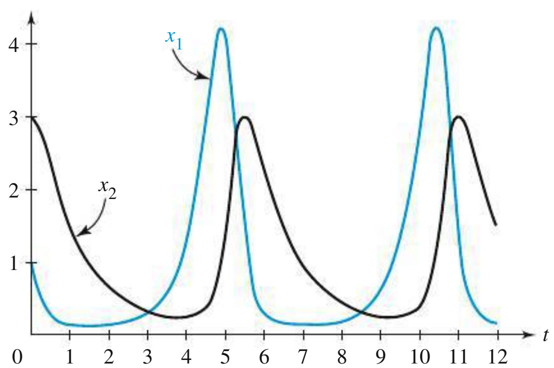


Figure 5.6 Graphs of the components of an approximate solution to the Volterra–Lotka system (17)

5.3 EXERCISES

In Problems 1–7, convert the given initial value problem into an initial value problem for a system in normal form.

1. $y''(t) + ty'(t) - 3y(t) = t^2\,;$
$y(0) = 3\,, \quad y'(0) = -6$

2. $y''(t) = \cos(t - y) + y^2(t)\,;$
$y(0) = 1\,, \quad y'(0) = 0$

3. $y^{(4)}(t) - y^{(3)}(t) + 7y(t) = \cos t\,;$
$y(0) = y'(0) = 1\,, \quad y''(0) = 0\,, \quad y^{(3)}(0) = 2$

4. $y^{(6)}(t) = [y'(t)]^3 - \sin\big(y(t)\big) + e^{2t}\,;$
$y(0) = y'(0) = \cdots = y^{(5)}(0) = 0$

5. $x'' + y - x' = 2t\,; \quad x(3) = 5\,, \quad x'(3) = 2\,,$
$y'' - x + y = -1\,; \quad y(3) = 1\,, \quad y'(3) = -1$
[*Hint:* Set $x_1 = x\,, \quad x_2 = x'\,, \quad x_3 = y\,, \quad x_4 = y'$.]

6. $3x'' + 5x - 2y = 0\,; \quad x(0) = -1\,, \quad x'(0) = 0\,,$
$4y'' + 2y - 6x = 0\,; \quad y(0) = 1\,, \quad y'(0) = 2$

7. $x''' - y = t\,; \quad x(0) = x'(0) = x''(0) = 4\,,$
$2x'' + 5y'' - 2y = 1\,; \quad y(0) = y'(0) = 1$

8. Sturm–Liouville Form. A second-order equation is said to be in **Sturm–Liouville form** if it is expressed as

$$[p(t)y'(t)]' + q(t)y(t) = 0\,.$$

Show that the substitutions $x_1 = y, x_2 = py'$ result in the normal form

$$x_1' = x_2/p\,,$$
$$x_2' = -qx_1\,.$$

If $y(0) = a$ and $y'(0) = b$ are the initial values for the Sturm–Liouville problem, what are $x_1(0)$ and $x_2(0)$?

9. In Section 3.6, we discussed the improved Euler's method for approximating the solution to a first-order equation. Extend this method to normal systems and give the recursive formulas for solving the initial value problem.

In Problems 10–13, use the vectorized Euler method with $h = 0.25$ to find an approximation for the solution to the given initial value problem on the specified interval.

10. $y'' + ty' + y = 0\,;$
$y(0) = 1\,, \quad y'(0) = 0 \quad \text{on } [0, 1]$

11. $(1 + t^2)y'' + y' - y = 0\,;$
$y(0) = 1\,, \quad y'(0) = -1 \quad \text{on } [0, 1]$

12. $t^2y'' + y = t + 2\,;$
$y(1) = 1\,, \quad y'(1) = -1 \quad \text{on } [1, 2]$

13. $y'' = t^2 - y^2\,;$
$y(0) = 0\,, \quad y'(0) = 1 \quad \text{on } [0, 1]$
(Can you guess the solution?)

In Problems 14–24, you will need a computer and a programmed version of the vectorized classical fourth-order Runge–Kutta algorithm. (At the instructor's discretion, other algorithms may be used.)†

14. Using the vectorized Runge–Kutta algorithm with $h = 0.5$, approximate the solution to the initial value problem

$$3t^2y'' - 5ty' + 5y = 0\,;$$
$$y(1) = 0\,, \quad y'(1) = \frac{2}{3}$$

at $t = 8$. Compare this approximation to the actual solution $y(t) = t^{5/3} - t$.

15. Using the vectorized Runge–Kutta algorithm, approximate the solution to the initial value problem

$$y'' = t^2 + y^2\,; \quad y(0) = 1\,, \quad y'(0) = 0$$

at $t = 1$. Starting with $h = 1$, continue halving the step size until two successive approximations [of both $y(1)$ and $y'(1)$] differ by at most 0.01.

16. Using the vectorized Runge–Kutta algorithm for systems with $h = 0.125$, approximate the solution to the initial value problem

$$x' = 2x - y\,; \quad x(0) = 0\,,$$
$$y' = 3x + 6y\,; \quad y(0) = -2$$

at $t = 1$. Compare this approximation to the actual solution

$$x(t) = e^{5t} - e^{3t}\,, \quad y(t) = e^{3t} - 3e^{5t}\,.$$

17. Using the vectorized Runge–Kutta algorithm, approximate the solution to the initial value problem

$$\frac{du}{dx} = 3u - 4v\,; \quad u(0) = 1\,,$$
$$\frac{dv}{dx} = 2u - 3v\,; \quad v(0) = 1$$

at $x = 1$. Starting with $h = 1$, continue halving the step size until two successive approximations of $u(1)$ and $v(1)$ differ by at most 0.001.

†Appendix G describes various websites and commercial software that sketch direction fields and automate most of the differential equation algorithms discussed in this book.

18. Combat Model. A simplified mathematical model for conventional versus guerrilla combat is given by the system

$$x_1' = -(0.1)x_1x_2\,; \qquad x_1(0) = 10\,,$$
$$x_2' = -x_1\,; \qquad x_2(0) = 15\,,$$

where x_1 and x_2 are the strengths of guerrilla and conventional troops, respectively, and 0.1 and 1 are the *combat effectiveness coefficients*. Who will win the conflict: the conventional troops or the guerrillas? [*Hint:* Use the vectorized Runge–Kutta algorithm for systems with $h = 0.1$ to approximate the solutions.]

19. Predator–Prey Model. The Volterra–Lotka predator–prey model predicts some rather interesting behavior that is evident in certain biological systems. For example, suppose you fix the initial population of prey but increase the initial population of predators. Then the population cycle for the prey becomes more severe in the sense that there is a long period of time with a reduced population of prey followed by a short period when the population of prey is very large. To demonstrate this behavior, use the vectorized Runge–Kutta algorithm for systems with $h = 0.5$ to approximate the populations of prey x and of predators y over the period $[0, 5]$ that satisfy the Volterra–Lotka system

$$x' = x(3 - y)\,,$$
$$y' = y(x - 3)$$

under each of the following initial conditions:

(a) $x(0) = 2\,, \quad y(0) = 4\,.$
(b) $x(0) = 2\,, \quad y(0) = 5\,.$
(c) $x(0) = 2\,, \quad y(0) = 7\,.$

20. In Project C of Chapter 4, it was shown that the simple pendulum equation

$$\theta''(t) + \sin\theta(t) = 0$$

has periodic solutions when the initial displacement and velocity are small. Show that the period of the solution may depend on the initial conditions by using the vectorized Runge–Kutta algorithm with $h = 0.02$ to approximate the solutions to the simple pendulum problem on $[0, 4]$ for the initial conditions:

(a) $\theta(0) = 0.1\,, \quad \theta'(0) = 0\,.$
(b) $\theta(0) = 0.5\,, \quad \theta'(0) = 0\,.$
(c) $\theta(0) = 1.0\,, \quad \theta'(0) = 0\,.$

[*Hint:* Approximate the length of time it takes to reach $-\theta(0)$.]

21. Fluid Ejection. In the design of a sewage treatment plant, the following equation arises:†

$$60 - H = (77.7)H'' + (19.42)(H')^2\,;$$
$$H(0) = H'(0) = 0\,,$$

where H is the level of the fluid in an ejection chamber and t is the time in seconds. Use the vectorized Runge–Kutta algorithm with $h = 0.5$ to approximate $H(t)$ over the interval $[0, 5]$.

22. Oscillations and Nonlinear Equations. For the initial value problem

$$x'' + (0.1)(1 - x^2)x' + x = 0\,;$$
$$x(0) = x_0\,, \qquad x'(0) = 0\,,$$

use the vectorized Runge–Kutta algorithm with $h = 0.02$ to illustrate that as t increases from 0 to 20, the solution x exhibits damped oscillations when $x_0 = 1$, whereas x exhibits expanding oscillations when $x_0 = 2.1$.

23. Nonlinear Spring. The **Duffing equation**

$$y'' + y + ry^3 = 0\,,$$

where r is a constant, is a model for the vibrations of a mass attached to a *nonlinear* spring. For this model, does the period of vibration vary as the parameter r is varied? Does the period vary as the initial conditions are varied? [*Hint:* Use the vectorized Runge–Kutta algorithm with $h = 0.1$ to approximate the solutions for $r = 1$ and 2, with initial conditions $y(0) = a$, $y'(0) = 0$ for $a = 1$, 2, and 3.]

24. Pendulum with Varying Length. A pendulum is formed by a mass m attached to the end of a wire that is attached to the ceiling. Assume that the length $l(t)$ of the wire varies with time in some predetermined fashion. If $\theta(t)$ is the angle in radians between the pendulum and the vertical, then the motion of the pendulum is governed for small angles by the initial value problem

$$l^2(t)\theta''(t) + 2l(t)l'(t)\theta'(t) + gl(t)\sin\big(\theta(t)\big) = 0\,;$$
$$\theta(0) = \theta_0\,, \qquad \theta'(0) = \theta_1\,,$$

where g is the acceleration due to gravity. Assume that

$$l(t) = l_0 + l_1\cos(\omega t - \phi)\,,$$

where l_1 is much smaller than l_0. (This might be a model for a person on a swing, where the *pumping* action changes the distance from the center of mass of the swing to the point where the swing is attached.) To simplify the computations, take $g = 1$. Using the Runge–Kutta algorithm with $h = 0.1$, study the motion of the pendulum when $\theta_0 = 0.05$, $\theta_1 = 0$, $l_0 = 1$, $l_1 = 0.1$, $\omega = 1$, and $\phi = 0.02$. In particular, does the pendulum ever attain an angle greater in absolute value than the initial angle θ_0?

†See *Numerical Solution of Differential Equations*, by William Milne (Dover, New York, 1970), p. 82.

In Problems 25–30, use a software package or the SUBROUTINE in Appendix F.

25. Using the Runge–Kutta algorithm for systems with $h = 0.05$, approximate the solution to the initial value problem

$$y''' + y'' + y^2 = t\,;$$
$$y(0) = 1\,, \qquad y'(0) = 0\,, \qquad y''(0) = 1$$

at $t = 1$.

26. Use the Runge–Kutta algorithm for systems with $h = 0.1$ to approximate the solution to the initial value problem

$$x' = yz\,; \qquad x(0) = 0\,,$$
$$y' = -xz\,; \qquad y(0) = 1\,,$$
$$z' = -xy/2\,; \qquad z(0) = 1\,,$$

at $t = 1$.

27. Generalized Blasius Equation. H. Blasius, in his study of laminar flow of a fluid, encountered an equation of the form

$$y''' + yy'' = (y')^2 - 1\,.$$

Use the Runge–Kutta algorithm for systems with $h = 0.1$ to approximate the solution that satisfies the initial conditions $y(0) = 0, y'(0) = 0$, and $y''(0) = 1.32824$. Sketch this solution on the interval $[0, 2]$.

28. Lunar Orbit. The motion of a moon moving in a planar orbit about a planet is governed by the equations

$$\frac{d^2x}{dt^2} = -G\frac{mx}{r^3}\,, \qquad \frac{d^2y}{dt^2} = -G\frac{my}{r^3}\,,$$

where $r := (x^2 + y^2)^{1/2}$, G is the gravitational constant, and m is the mass of the planet. Assume $Gm = 1$. When $x(0) = 1, x'(0) = y(0) = 0$, and $y'(0) = 1$, the motion is a circular orbit of radius 1 and period 2π.

(a) Setting $x_1 = x$, $x_2 = x'$, $x_3 = y$, $x_4 = y'$, express the governing equations as a first-order system in normal form.

(b) Using $h = 2\pi/100 \approx 0.0628318$, compute one orbit of this moon (i.e., do $N = 100$ steps.). Do your approximations agree with the fact that the orbit is a circle of radius 1?

29. Competing Species. Let $p_i(t)$ denote, respectively, the populations of three competing species S_i, $i = 1, 2, 3$. Suppose these species have the same growth rates, and the maximum population that the habitat can support is the same for each species. (We assume it to be one unit.) Also suppose the competitive advantage that S_1 has over S_2 is the same as that of S_2 over S_3 and S_3 over S_1. This situation is modeled by the system

$$p_1' = p_1(1 - p_1 - ap_2 - bp_3)\,,$$
$$p_2' = p_2(1 - bp_1 - p_2 - ap_3)\,,$$
$$p_3' = p_3(1 - ap_1 - bp_2 - p_3)\,,$$

where a and b are positive constants. To demonstrate the population dynamics of this system when $a = b = 0.5$, use the Runge–Kutta algorithm for systems with $h = 0.1$ to approximate the populations p_i over the time interval $[0, 10]$ under each of the following initial conditions:

(a) $p_1(0) = 1.0\,, \quad p_2(0) = 0.1\,, \quad p_3(0) = 0.1\,.$
(b) $p_1(0) = 0.1\,, \quad p_2(0) = 1.0\,, \quad p_3(0) = 0.1\,.$
(c) $p_1(0) = 0.1\,, \quad p_2(0) = 0.1\,, \quad p_3(0) = 1.0\,.$

On the basis of the results of parts (a)–(c), decide what you think will happen to these populations as $t \to +\infty$.

30. Spring Pendulum. Let a mass be attached to one end of a spring with spring constant k and the other end attached to the ceiling. Let l_0 be the natural length of the spring and let $l(t)$ be its length at time t. If $\theta(t)$ is the angle between the pendulum and the vertical, then the motion of the spring pendulum is governed by the system

$$l''(t) - l(t)\theta'(t) - g\cos\theta(t) + \frac{k}{m}(l - l_0) = 0\,,$$
$$l^2(t)\theta''(t) + 2l(t)l'(t)\theta'(t) + gl(t)\sin\theta(t) = 0\,.$$

Assume $g = 1$, $k = m = 1$, and $l_0 = 4$. When the system is at rest, $l = l_0 + mg/k = 5$.

(a) Describe the motion of the pendulum when $l(0) = 5.5, l'(0) = 0, \theta(0) = 0$, and $\theta'(0) = 0$.

(b) When the pendulum is both stretched and given an angular displacement, the motion of the pendulum is more complicated. Using the Runge–Kutta algorithm for systems with $h = 0.1$ to approximate the solution, sketch the graphs of the length l and the angular displacement θ on the interval $[0, 10]$ if $l(0) = 5.5$, $l'(0) = 0, \theta(0) = 0.5$, and $\theta'(0) = 0$.

5.4 Introduction to the Phase Plane

In this section, we study systems of two first-order equations of the form

$$\textbf{(1)}\qquad \begin{aligned} \frac{dx}{dt} &= f(x,y)\,, \\ \frac{dy}{dt} &= g(x,y)\,. \end{aligned}$$

Note that the independent variable t does not appear in the right-hand terms $f(x, y)$ and $g(x, y)$; such systems are called **autonomous.** For example, the system that modeled the interconnected tanks problem in Section 5.1,

$$\begin{aligned} x' &= -\frac{1}{3}x + \frac{1}{12}y\,, \\ y' &= \frac{1}{3}x - \frac{1}{3}y\,, \end{aligned}$$

is autonomous. So is the Volterra–Lotka system,

$$\begin{aligned} x' &= Ax - Bxy\,, \\ y' &= -Cy + Dxy\,, \end{aligned}$$

(with A, B, C, D constants), which was discussed in Example 4 of Section 5.3 as a model for population dynamics.

For future reference, we note that the solutions to autonomous systems have a "time-shift immunity," in the sense that if the pair $x(t), y(t)$ solves (1), so does the **time-shifted** pair $x(t+c), y(t+c)$ for any constant c. Specifically, if we let $X(t) \coloneqq x(t+c)$ and $Y(t) \coloneqq y(t+c)$, then by the chain rule

$$\begin{aligned} \frac{dX}{dt}(t) &= \frac{dx}{dt}(t+c) = f\big(x(t+c), y(t+c)\big) = f\big(X(t), Y(t)\big)\,, \\ \frac{dY}{dt}(t) &= \frac{dy}{dt}(t+c) = g\big(x(t+c), y(t+c)\big) = g\big(X(t), Y(t)\big)\,, \end{aligned}$$

proving that $X(t), Y(t)$ is also a solution to (1).

Since a solution to the autonomous system is a pair of functions $x(t), y(t)$ that satisfies (1) for all t in some interval I, it can be visualized as a pair of graphs, as shown in Figure 5.7. For purposes of comparing one of the solution pair with its mate, we can compress these two graphs into one by charting the path in the plane traced out by the ordered pair $\big(x(t), y(t)\big)$ as t increases.

Trajectories and the Phase Plane

Definition 1. If $x(t), y(t)$ is a solution pair to (1) for t in the interval I, then a plot in the xy-plane of the parametrized curve $x = x(t), y = y(t)$ for t in I, together with arrows indicating its direction with increasing t, is said to be a **trajectory** of the system (1). In such a context we call the xy-plane the **phase plane** and refer to a representative set of trajectories in this plane as a **phase portrait** of the system.

A trajectory provides a visualization of the motion of a particle in the plane that starts at some point (x_0, y_0) and follows a path as time increases that is determined by the solution to the autonomous system with initial conditions $x(0) = x_0, y(0) = y_0$. The trajectory for the

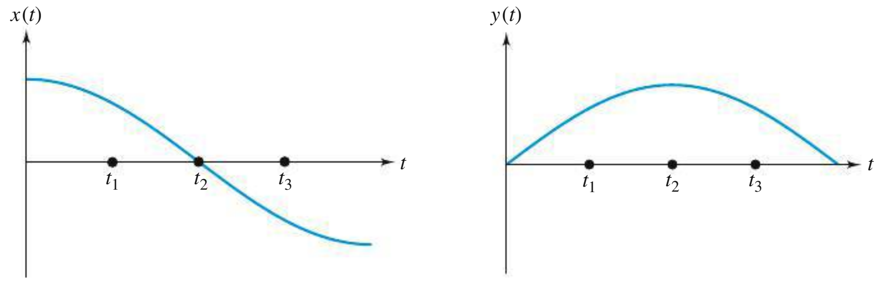


Figure 5.7 Solution pair for system (1)

solution pair of Figure 5.7 is shown in Figure 5.8. Note, however, that the trajectory contains less information than the individual graphs for $x(t)$ and $y(t)$, because the t-dependence has been suppressed. (We can't tell if the particle's speed is fast or slow.) In principle we can construct, point by point, the trajectory from the solution graphs, but we cannot reconstruct the solution graphs from the phase plane trajectory alone (because we would not know what value of t to assign to each point). Nonetheless, trajectories provide valuable information about the qualitative behavior of solutions to the autonomous system.

Can we find trajectories without knowing a solution pair to the system? Yes; the key is to observe that since t does not appear explicitly on the right-hand side of system (1), if we divide the two equations and invoke the chain rule

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt},$$

we get the "related phase plane differential equation" (*without t*)

$$\textbf{(2)} \qquad \frac{dy}{dx} = \frac{g(x, y)}{f(x, y)}.$$

We mastered equations like (2) in Chapter 2; their solution curves provide highways along which the trajectories travel. What is missing from these highways is the orientation of a trajectory, which can be gleaned from the autonomous system (1) by determining the sign of f or g at a point and recalling that $(dx/dt, dy/dt) = (f, g)$ is the velocity of the moving particle. Furthermore, as illustrated in Example 1 below, a single curve arising from the related phase plane differential equation might well accommodate several different trajectories.

As we shall see in the examples to follow as well as in Figure 5.12, trajectories are typically of the following types: (i) a single point, indicating a stationary particle; (ii) a non-closed arc, without self-intersections; or (iii) a closed curve (cycle), indicating periodic motion. Moreover, thanks to the Existence-Uniqueness Theorem for first-order equations in Chapter 1 (page 11), as long as the functions f and g in (1) are continuously differentiable in the plane, the trajectories of a phase portrait will not intersect each other (unless they coincide).

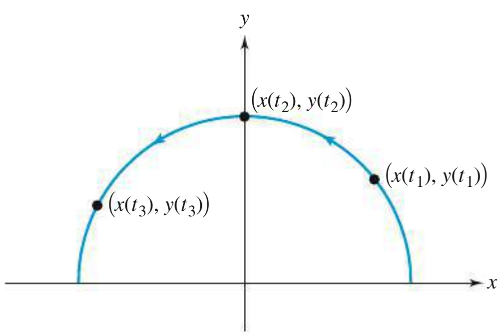


Figure 5.8 Phase plane trajectory of the solution pair for system (1)

Except for the very special case of linear systems with constant coefficients that was discussed in Section 5.2, finding all solutions to the system (1) is generally an impossible task. But it is relatively easy to find *constant solutions;* if $f(x_0, y_0) = 0$ and $g(x_0, y_0) = 0$, then the constant functions $x(t) \equiv x_0$, $y(t) \equiv y_0$ solve (1). For such solutions the following terminology is used.

Critical Points and Equilibrium Solutions

Definition 2. A point (x_0, y_0) where $f(x_0, y_0) = 0$ and $g(x_0, y_0) = 0$ is called a **critical point,** or **equilibrium point,** of the system $dx/dt = f(x, y)$, $dy/dt = g(x, y)$, and the corresponding constant solution $x(t) \equiv x_0$, $y(t) \equiv y_0$ is called an **equilibrium solution.**

Notice that trajectories of equilibrium solutions consist of just single points (the equili-brium points). But what can be said about the other trajectories? Can we predict any of their features from closer examination of the equilibrium points? To explore this we focus on the related phase plane differential equation (2) and exploit its direction field (recall Section 1.3, page 15). However, we'll augment the direction field plot by attaching arrowheads to the line segments, indicating the direction of the "flow" of solutions as t increases. This is easy: When dx/dt is positive, $x(t)$ increases so the trajectory flows to the right. Therefore, according to (1), all direction field segments drawn in a region where $f(x, y)$ is positive should point to the right [and, of course, they point to the left if $f(x, y)$ is negative]. If $f(x, y)$ is zero, we can use $g(x, y)$ to decide if the flow is upward [$y(t)$ increases] or downward [$y(t)$ decreases]. [*What if both* $f(x, y)$ *and* $g(x, y)$ *are zero?*]

In the examples that follow, one can use computers or calculators for generating these direction fields.

Example 1 Sketch the direction field in the phase plane for the system

$$\text{(3)} \qquad \frac{dx}{dt} = -x\,, \qquad \frac{dy}{dt} = -2y$$

and identify its critical point.

Solution Here $f(x, y) = -x$ and $g(x, y) = -2y$ are both zero when $x = y = 0$, so $(0, 0)$ is the critical point. The direction field for the related phase plane differential equation

$$\text{(4)} \qquad \frac{dy}{dx} = \frac{-2y}{-x} = \frac{2y}{x}$$

is given in Figure 5.9 on page 265. Since $dx/dt = -x$ in (3), trajectories in the right half-plane (where $x > 0$) flow to the left, and vice versa. From the figure we can see that all solutions "flow into" the critical point $(0, 0)$. Such a critical point is called **asymptotically stable.**† ◆

Remark. For this simple example, we can actually solve the system (3) explicitly; indeed, (3) constitutes an uncoupled pair of linear equations whose solutions are $x(t) = c_1e^{-t}$ and $y(t) = c_2e^{-2t}$. By elimination of t, we obtain the equation $y = c_2e^{-2t} = c_2[x(t)/c_1]^2 = cx^2$. So the trajectories lie along the parabolas $y = cx^2$. [Alternatively, we could have solved (4) by separating variables and obtained these same parabolas as the "highways" on which the trajectories travel.] Notice that each such parabola is made up of three trajectories: an incoming trajectory approaching the origin in the right half-plane; its mirror-image trajectory approaching

†See Section 12.3 for a rigorous exposition of stability and critical points. All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems,* 7th ed.

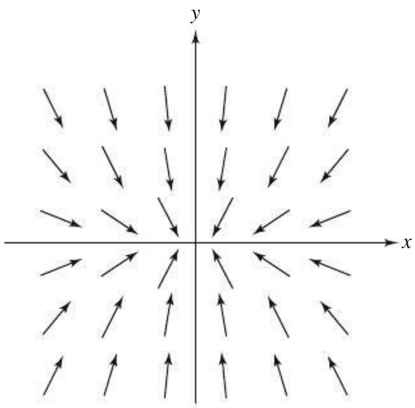


Figure 5.9 Direction field for Example 1

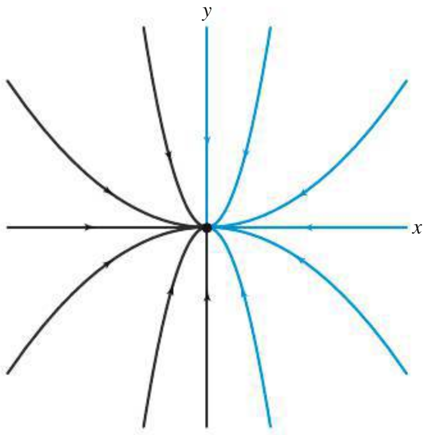


Figure 5.10 Trajectories for Example 1

the origin in the left half-plane; and the origin itself, an equilibrium point. Sample trajectories are indicated in Figure 5.10.

Example 2 Sketch the direction field in the phase plane for the system

(5)
$$\frac{dx}{dt} = x\,, \qquad \frac{dy}{dt} = 2y$$

and describe the behavior of solutions near the critical point $(0, 0)$.

Solution This example is almost identical to the previous one; in fact, one could say we have merely "reversed time" in (3). The direction field segments for

(6)
$$\frac{dy}{dx} = \frac{2y}{x}$$

are the same as those of (4), but the direction arrows are reversed. Now all solutions flow *away* from the critical point $(0, 0)$; the equilibrium is **unstable.** ◆

Example 3 For the system (7) below, find the critical points, sketch the direction field in the phase plane, and predict the asymptotic nature (i.e., behavior as $t \to +\infty$) of the solution starting at $x = 2$, $y = 0$ when $t = 0$.

(7)
$$\frac{dx}{dt} = 5x - 3y - 2\,, \qquad \frac{dy}{dt} = 4x - 3y - 1\,.$$

Solution The only critical point is the solution of the simultaneous equations $f(x, y) = g(x, y) = 0$:

(8)
$$5x_0 - 3y_0 - 2 = 0\,,$$
$$4x_0 - 3y_0 - 1 = 0\,,$$

from which we find $x_0 = y_0 = 1$. The direction field for the related phase plane differential equation

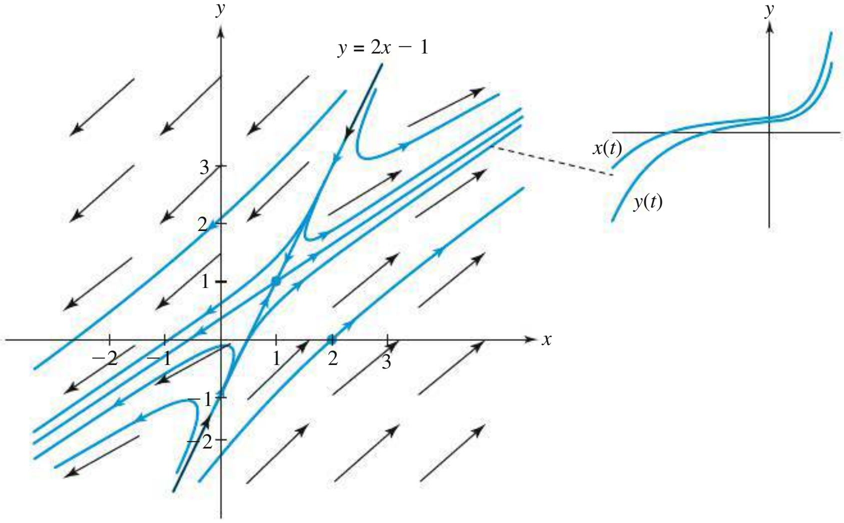


Figure 5.11 Direction field, trajectories, and a typical solution pair for a trajectory (Example 3)

$$\frac{dy}{dx} = \frac{4x - 3y - 1}{5x - 3y - 2} \tag{9}$$

is shown in Figure 5.11, along with some trajectories and a typical solution pair.† Note that solutions flow to the right for $5x - 3y - 2 > 0$, i.e., for all points *below* the line $5x - 3y - 2 = 0$, and that they slow down near the critical point.

The solution curve to (9) passing through $(2, 0)$ in Figure 5.11 apparently is a "highway" extending to infinity. Does this imply the corresponding *system* solution $x(t)$, $y(t)$ also approaches infinity in the sense that $|x(t)| + |y(t)| \to +\infty$ as $t \to +\infty$, or could its trajectory "stall" at some point along the highway, or possibly even "backtrack?" *It cannot backtrack,* because the direction arrows along the trajectory point, unambiguously, to the right. And if $\big(x(t), y(t)\big)$ stalls at some point (x_1, y_1), then intuitively we would conclude that (x_1, y_1) was an equilibrium point (since the "speeds" dx/dt and dy/dt would approach zero there). But we have already found the only critical point. So we conclude, with a high degree of confidence,‡ that the system solution does indeed go to infinity.

The critical point $(1, 1)$ is **unstable** because, although many solutions get arbitrarily close to $(1, 1)$, most of them eventually flow away. Solutions that lie on the line $y = 2x - 1$, however, do converge to $(1, 1)$. Such an equilibrium is an example of a **saddle point.** ◆

In the preceding example, we informally argued that if a trajectory "stalls"—that is, if it has an endpoint—then this endpoint would have to be a critical point. This is more carefully stated in the following theorem, whose proof is outlined in Problem 30.

Endpoints Are Critical Points

Theorem 1. Let the pair $x(t)$, $y(t)$ be a solution on $[0, +\infty)$ to the autonomous system $dx/dt = f(x, y)$, $dy/dt = g(x, y)$, where f and g are continuous in the plane. If the limits

$$x^* := \lim_{t \to +\infty} x(t) \qquad \text{and} \qquad y^* := \lim_{t \to +\infty} y(t)$$

exist and are finite, then the point (x^*, y^*) is a critical point for the system.

†Solutions to equation (9) can be obtained analytically using the methods of Section 2.6.

‡These informal arguments are made more rigorous in Chapter 12. All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems,* 7th edition.

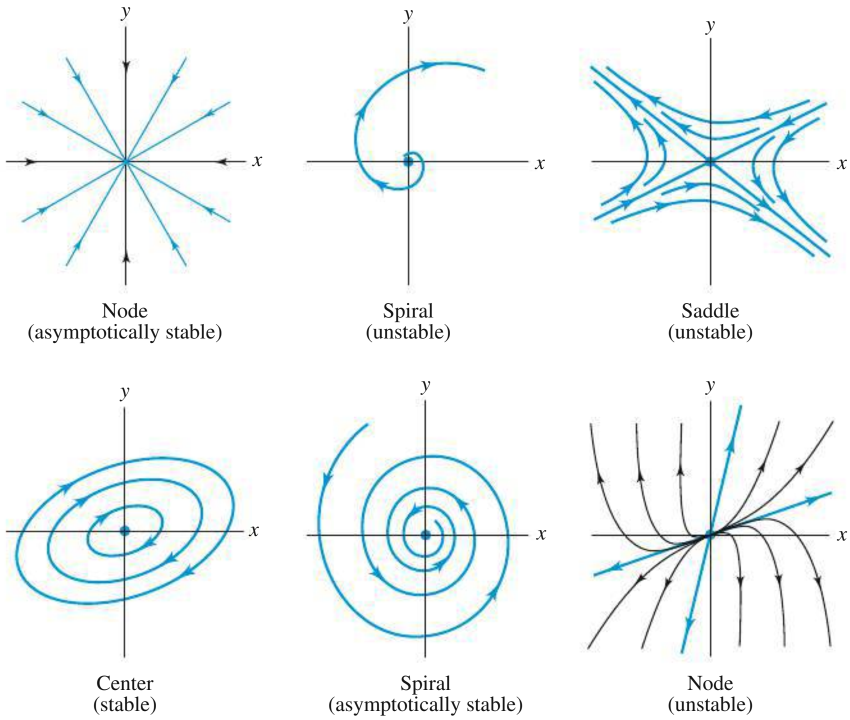


Figure 5.12 Examples of different trajectory behaviors near critical point at origin

Some typical trajectory configurations near critical points are displayed and classified in Figure 5.12. These phase plane portraits were generated by software packages having trajectory-sketching options†. A more complete discussion of the nature of various types of equilibrium solutions and their stability is deferred to Chapter 12.‡ For the moment, however, notice that unstable critical points are distinguished by "runaway" trajectories emanating from arbitrarily nearby points, while stable equilibria "trap" all neighboring trajectories. The asymptotically stable critical points *attract* their neighboring trajectories as $t \to +\infty$.

Historically, the phase plane was introduced to facilitate the analysis of mechanical systems governed by Newton's second law, force equals mass times acceleration. An autonomous mechanical system arises when this force is independent of time and can be modeled by a second-order equation of the form

$$y'' = f(y, y') \,. \tag{10}$$

As we have seen in Section 5.3, this equation can be converted to a normal first-order system by introducing the velocity $v = dy/dt$ and writing

$$\begin{aligned} \frac{dy}{dt} &= v \,, \\ \frac{dv}{dt} &= f(y, v) \,. \end{aligned} \tag{11}$$

---

†Appendix G describes various websites and commercial software that sketch direction fields and automate most of the differential equation algorithms discussed in this book.

‡All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems,* 7th edition.

Thus, we can analyze the behavior of an autonomous mechanical system by studying its phase portrait in the yv-plane. Notice that with v as the vertical axis, trajectories $(y(t), v(t))$ flow to the right in the upper half-plane (where $v > 0$), and to the left in the lower half-plane.

Example 4 Sketch the direction field in the phase plane for the first-order system corresponding to the unforced, undamped mass–spring oscillator described in Section 4.1 (Figure 4.1, page 152). Sketch several trajectories and interpret them physically.

Solution The equation derived in Section 4.1 for this oscillator is $my'' + ky = 0$ or, equivalently, $y'' = -ky/m$. Hence, the system (11) takes the form

$$\text{(12)} \qquad \begin{aligned} y' &= v\,, \\ v' &= -\frac{ky}{m}\,. \end{aligned}$$

The critical point is at the origin $y = v = 0$. The direction field in Figure 5.13 indicates that the trajectories appear to be either closed curves (ellipses?) or spirals that encircle the critical point.

We saw in Section 4.9 that the undamped oscillator motions are periodic; they cycle repeatedly through the same sets of points, with the same velocities. Their trajectories in the phase plane, then, must be closed curves.† Let's confirm this mathematically by solving the related phase plane differential equation

$$\text{(13)} \qquad \frac{dv}{dy} = -\frac{ky}{mv}\,.$$

Equation (13) is separable, and we find

$$v\,dv = -\frac{ky}{m}dy \qquad \text{or} \qquad d\left(\frac{v^2}{2}\right) = -\frac{k}{m}d\left(\frac{y^2}{2}\right),$$

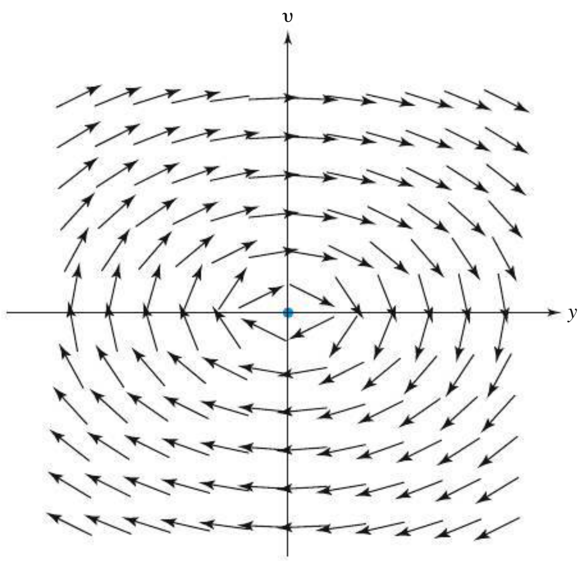


Figure 5.13 Direction field for Example 4

†By the same reasoning, *underdamped* oscillations would correspond to spiral trajectories asymptotically approaching the origin as $t \to +\infty$.

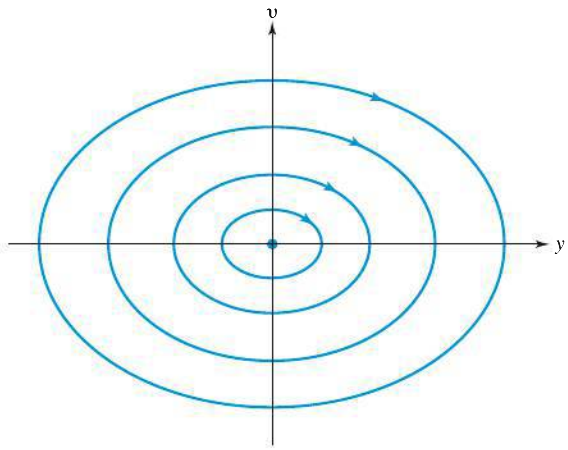


Figure 5.14 Trajectories for Example 4

so its solutions are the ellipses $v^2/2 + ky^2/2m = C$ as shown in Figure 5.14. The trajectories of (12) are confined to these ellipses, and hence neither come arbitrarily close to nor wander arbitrarily far from the equilibrium solution. The critical point is thus identified as a **center** in Figure 5.12 on page 267.

Furthermore, the system solutions must continually circulate around the ellipses, since there are no critical points to stop them. This confirms that all solutions are periodic. ◆

Remark. More generally, we argue that if a solution to an autonomous system like (1) passes through a point in the phase plane *twice* and if it is sufficiently well behaved to satisfy a *uniqueness theorem,* then the second "tour" satisfies the same initial conditions as the first tour and so must replicate it. In other words, *closed trajectories containing no critical points correspond to periodic solutions.*

Through these examples we have seen how, by studying the phase plane, one can often anticipate some of the features (boundedness, periodicity, etc.) of solutions of autonomous systems without solving them explicitly. Much of this information can be predicted simply from the critical points and the direction field (oriented by arrowheads), which are obtainable through standard software packages. The final example ties together several of these ideas.

Example 5 Find the critical points and solve the related phase plane equation for

$$\begin{aligned} \frac{dx}{dt} &= -y(y-2)\,, \\ \frac{dy}{dt} &= (x-2)(y-2)\,. \end{aligned} \tag{14}$$

What is the asymptotic behavior of the solutions starting from $(3, 0)$, $(5, 0)$, and $(2, 3)$?

Solution To find the critical points, we solve the system

$$-y(y-2) = 0, \qquad (x-2)(y-2) = 0\,.$$

One family of solutions to this system is given by $y = 2$ with x arbitrary; that is, the line $y = 2$. If $y \neq 2$, then the system simplifies to $-y = 0$, and $x - 2 = 0$, which has the solution $x = 2$, $y = 0$. Hence, the critical point set consists of the isolated point $(2, 0)$ and the horizontal line $y = 2$. The corresponding equilibrium solutions are $x(t) \equiv 2$, $y(t) \equiv 0$, and the family $x(t) \equiv c$, $y(t) \equiv 2$, where c is an arbitrary constant.

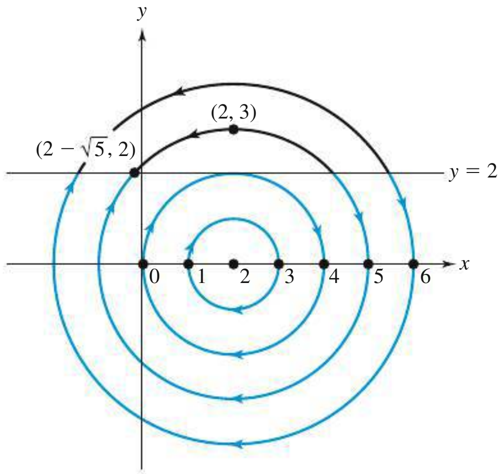


Figure 5.15 Phase portrait for Example 5

The related phase plane differential equation for the system (14) is

(15) $$\frac{dy}{dx} = \frac{(x-2)(y-2)}{-y(y-2)} = -\frac{x-2}{y} .$$

Solving this equation by separating variables,

$$y\,dy = -(x-2)\,dx \qquad \text{or} \qquad y^2 + (x-2)^2 = C ,$$

demonstrates that the trajectories lie on concentric circles centered at $(2, 0)$. See Figure 5.15.

Next we analyze the flow along each trajectory. From the equation $dx/dt = -y(y-2)$, we see that x is decreasing when $y > 2$. This means the flow is from right to left along the arc of a circle that lies above the line $y = 2$. For $0 < y < 2$, we have $dx/dt > 0$, so in this region the flow is from left to right. Furthermore, for $y < 0$, we have $dx/dt < 0$, and again the flow is from right to left.

We now observe in Figure 5.15 that there are four types of trajectories associated with system (14): (a) those that begin above the line $y = 2$ and follow the arc of a circle counterclockwise back to that line; (b) those that begin below the line $y = 2$ and follow the arc of a circle clockwise back to that line; (c) those that continually move clockwise around a circle centered at $(2, 0)$ with radius less than 2 (i.e., they do not intersect the line $y = 2$); and finally, (d) the critical points $(2, 0)$ and $y = 2$, x arbitrary.

The solution starting at $(3, 0)$ lies on a circle with no critical points; therefore, it is a *periodic* solution, and the critical point $(2, 0)$ is a center. But the circle containing the solutions starting at $(5, 0)$ and at $(2, 3)$ has critical points at $(2 - \sqrt{5}, 2)$ and $(2 + \sqrt{5}, 2)$. The direction arrows indicate that both solutions approach $(2 - \sqrt{5}, 2)$ asymptotically (as $t \to +\infty$). They lie on the same solution curve for the related phase plane differential equation (a circle), but they are quite different trajectories. ◆

Note that for the system (14) the critical points on the line $y = 2$ are not isolated, so they do not fit into any of the categories depicted in Figure 5.12 on page 267. Observe also that all solutions of this system are bounded, since they are confined to circles.

5.4 EXERCISES

In Problems 1 and 2, verify that the pair $x(t)$, $y(t)$ is a solution to the given system. Sketch the trajectory of the given solution in the phase plane.

1. $\dfrac{dx}{dt} = 3y^3\,, \quad \dfrac{dy}{dt} = y\,;$

$x(t) = e^{3t}\,, \quad y(t) = e^t$

2. $\dfrac{dx}{dt} = 1\,, \quad \dfrac{dy}{dt} = 3t^2\,;$

$x(t) = t+1\,, \quad y(t) = t^3 + 3t^2 + 3t$

In Problems 3–6, find the critical point set for the given system.

3. $\dfrac{dx}{dt} = x - y\,, \quad \dfrac{dy}{dt} = x^2 + y^2 - 1$

4. $\dfrac{dx}{dt} = y - 1\,, \quad \dfrac{dy}{dt} = x + y + 5$

5. $\dfrac{dx}{dt} = x^2 - 2xy\,, \quad \dfrac{dy}{dt} = 3xy - y^2$

6. $\dfrac{dx}{dt} = y^2 - 3y + 2\,, \quad \dfrac{dy}{dt} = (x-1)(y-2)$

In Problems 7–9, solve the related phase plane differential equation (2), page 263, for the given system.

7. $\dfrac{dx}{dt} = y - 1\,, \quad \dfrac{dy}{dt} = e^{x+y}$

8. $\dfrac{dx}{dt} = x^2 - 2y^{-3}\,, \quad \dfrac{dy}{dt} = 3x^2 - 2xy$

9. $\dfrac{dx}{dt} = 2y - x\,, \quad \dfrac{dy}{dt} = e^x + y$

10. Find all the critical points of the system

$$\frac{dx}{dt} = x^2 - 1\,,$$
$$\frac{dy}{dt} = xy\,,$$

and the solution curves for the related phase plane differential equation. Thereby prove that there are two trajectories that lie on semicircles. What are the endpoints of the semicircles?

In Problems 11–14, solve the related phase plane differential equation for the given system. Then sketch by hand several representative trajectories (with their flow arrows).

11. $\dfrac{dx}{dt} = 2y\,, \quad \dfrac{dy}{dt} = 2x$

12. $\dfrac{dx}{dt} = -8y\,, \quad \dfrac{dy}{dt} = 18x$

13. $\dfrac{dx}{dt} = (y-x)(y-1)\,, \quad \dfrac{dy}{dt} = (x-y)(x-1)$

14. $\dfrac{dx}{dt} = \dfrac{3}{y}\,, \quad \dfrac{dy}{dt} = \dfrac{2}{x}$

In Problems 15–18, find all critical points for the given system. Then use a software package to sketch the direction field in the phase plane and from this describe the stability of the critical points (i.e., compare with Figure 5.12).

15. $\dfrac{dx}{dt} = 2x + y + 3\,, \quad \dfrac{dy}{dt} = -3x - 2y - 4$

16. $\dfrac{dx}{dt} = -5x + 2y\,, \quad \dfrac{dy}{dt} = x - 4y$

17. $\dfrac{dx}{dt} = 2x + 13y\,, \quad \dfrac{dy}{dt} = -x - 2y$

18. $\dfrac{dx}{dt} = x(7 - x - 2y)\,, \quad \dfrac{dy}{dt} = y(5 - x - y)$

In Problems 19–24, convert the given second-order equation into a first-order system by setting $v = y'$. Then find all the critical points in the yv-plane. Finally, sketch (by hand or software) the direction fields, and describe the stability of the critical points (i.e., compare with Figure 5.12).

19. $\dfrac{d^2y}{dt^2} - y = 0$

20. $\dfrac{d^2y}{dt^2} + y = 0$

21. $\dfrac{d^2y}{dt^2} + y + y^5 = 0$

22. $\dfrac{d^2y}{dt^2} + y^3 = 0$

23. $y''(t) + y(t) - y(t)^4 = 0$

24. $y''(t) + y(t) - y(t)^3 = 0$

25. Using software, sketch the direction field in the phase plane for the system

$$dx/dt = y\,,$$
$$dy/dt = -x + x^3\,.$$

From the sketch, conjecture whether the solution passing through each given point is periodic:

(a) $(0.25, 0.25)$ **(b)** $(2, 2)$ **(c)** $(1, 0)$

26. Using software, sketch the direction field in the phase plane for the system

$$dx/dt = y\,,$$
$$dy/dt = -x - x^3\,.$$

From the sketch, conjecture whether all solutions of this system are bounded. Solve the related phase plane differential equation and confirm your conjecture.

27. Using software, sketch the direction field in the phase plane for the system

$$dx/dt = -2x + y\,,$$
$$dy/dt = -5x - 4y\,.$$

From the sketch, predict the asymptotic limit (as $t \to +\infty$) of the solution starting at $(1, 1)$.

28. Figure 5.16 displays some trajectories for the system

$$dx/dt = y\,,$$
$$dy/dt = -x + x^2\,.$$

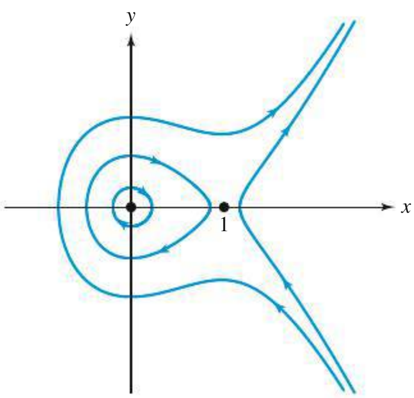


Figure 5.16 Phase plane for Problem 28

What types of critical points (compare Figure 5.12 on page 267) occur at $(0, 0)$ and $(1, 0)$?

29. Find the critical points and solve the related phase plane differential equation for the system

$$\frac{dx}{dt} = (x-1)(y-1)$$
$$\frac{dy}{dt} = y(y-1)\,.$$

Describe (without using computer software) the asymptotic behavior of trajectories (as $t \to \infty$) that start at **(a)** $(3, 2)$, **(b)** $(2, 1/2)$, **(c)** $(-2, 1/2)$, **(d)** $(3, -2)$.

30. A proof of Theorem 1, page 266, is outlined below. The goal is to show that $f(x^*, y^*) = g(x^*, y^*) = 0$. Justify each step.

(a) From the given hypotheses, deduce that $\lim_{t\to+\infty} x'(t) = f(x^*, y^*)$ and $\lim_{t\to+\infty} y'(t) = g(x^*, y^*)$.

(b) Suppose $f(x^*, y^*) > 0$. Then, by continuity, $x'(t) > f(x^*, y^*)/2$ for all large t (say, for $t \geq T$). Deduce from this that $x(t) > tf(x^*, y^*)/2 + C$ for $t > T$, where C is some constant.

(c) Conclude from part (b) that $\lim_{t\to+\infty} x(t) = +\infty$, contradicting the fact that this limit is the finite number x^*. Thus, $f(x^*, y^*)$ cannot be positive.

(d) Argue similarly that the supposition that $f(x^*, y^*) < 0$ also leads to a contradiction; hence, $f(x^*, y^*)$ must be zero.

(e) In the same manner, argue that $g(x^*, y^*)$ must be zero.

Therefore, $f(x^*, y^*) = g(x^*, y^*) = 0$, and (x^*, y^*) is a critical point.

31. Phase plane analysis provides a quick derivation of the energy integral lemma of Section 4.8 (page 201). By completing the following steps, prove that solutions of equations of the special form $y'' = f(y)$ satisfy

$$\frac{1}{2}(y')^2 - F(y) = \text{constant}\,,$$

where $F(y)$ is an antiderivative of $f(y)$.

(a) Set $v = y'$ and write $y'' = f(y)$ as an equivalent first-order system.

(b) Show that the solutions to the yv-phase plane equation for the system in part (a) satisfy $v^2/2 = F(y) + K$. Replacing v by y' then completes the proof.

32. Use the result of Problem 31 to prove that all solutions to the equation

$$y'' + y^3 = 0$$

remain bounded. [*Hint:* Argue that $y^4/4$ is bounded above by the constant appearing in Problem 31.]

33. A Problem of Current Interest. The motion of an iron bar attracted by the magnetic field produced by a parallel current wire and restrained by springs (see Figure 5.17) is governed by the equation

$$\frac{d^2x}{dt^2} = -x + \frac{1}{\lambda - x}\,, \quad \text{for} \quad -x_0 < x < \lambda\,,$$

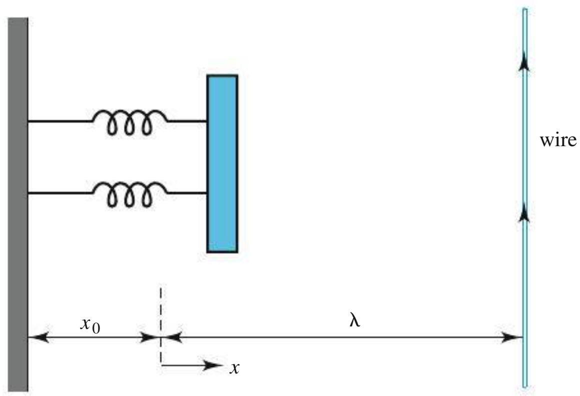


Figure 5.17 Bar restrained by springs and attracted by a parallel current

where the constants x_0 and λ are, respectively, the distances from the bar to the wall and to the wire *when the bar is at equilibrium (rest) with the current off.*

(a) Setting $v = dx/dt$, convert the second-order equation to an equivalent first-order system.

(b) Solve the related phase plane differential equation for the system in part (a) and thereby show that its solutions are given by

$$v = \pm\sqrt{C - x^2 - 2\ln(\lambda - x)}\,,$$

where C is a constant.

(c) Show that if $\lambda < 2$ there are no critical points in the xv-phase plane, whereas if $\lambda > 2$ there are two critical points. For the latter case, determine these critical points.

(d) Physically, the case $\lambda < 2$ corresponds to a current so high that the magnetic attraction completely overpowers the spring. To gain insight into this, use software to plot the phase plane diagrams for the system when $\lambda = 1$ and when $\lambda = 3$.

(e) From your phase plane diagrams in part (d), describe the possible motions of the bar when $\lambda = 1$ and when $\lambda = 3$, under various initial conditions.

34. Falling Object. The motion of an object moving vertically through the air is governed by the equation

$$\frac{d^2y}{dt^2} = -g - \frac{g}{V^2}\frac{dy}{dt}\left|\frac{dy}{dt}\right|,$$

where y is the upward vertical displacement and V is a constant called the terminal speed. Take $g = 32$ ft/sec^2 and $V = 50$ ft/sec. Sketch trajectories in the yv-phase plane for $-100 \le y \le 100$, $-100 \le v \le 100$, starting from $y = 0$ and $v = -75, -50, -25, 0, 25, 50$, and 75 ft/sec. Interpret the trajectories physically; why is V called the terminal speed?

35. Sticky Friction. An alternative for the damping friction model $F = -by'$ discussed in Section 4.1 is the "sticky friction" model. For a mass sliding on a surface as depicted in Figure 5.18, the contact friction is more complicated than simply $-by'$. We observe, for example, that even if the mass is displaced slightly off the equilibrium location $y = 0$, it may nonetheless remain stationary due to the fact that the spring force $-ky$ is insufficient to break the static friction's grip. If the maximum force that the friction can exert is denoted by μ, then a feasible model is given by

$$F_{\text{friction}} = \begin{cases} ky\,, & \text{if } |ky| < \mu \text{ and } y' = 0\,,\\ \mu\,\text{sign}(y)\,, & \text{if } |ky| \ge \mu \text{ and } y' = 0\,,\\ -\mu\,\text{sign}(y')\,, & \text{if } y' \ne 0\,.\end{cases}$$

(The function sign (s) is $+1$ when $s > 0$, -1 when $s < 0$, and 0 when $s = 0$.) The motion is governed by the equation

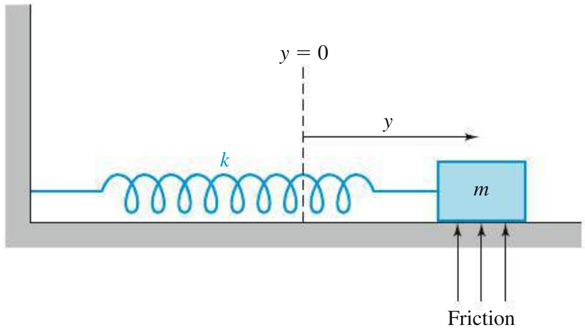


Figure 5.18 Mass–spring system with friction

$$m\frac{d^2y}{dt^2} = -ky + F_{\text{friction}}\,. \tag{16}$$

Thus, if the mass is at rest, friction *balances* the spring force if $|y| < \mu/k$ but simply *opposes* it with intensity μ if $|y| \ge \mu/k$. If the mass is moving, friction opposes the velocity with the same intensity μ.

(a) Taking $m = \mu = k = 1$, convert (16) into the first-order system

$$y' = v\,,$$
$$v' = \begin{cases} 0\,, & \text{if } |y| < 1 \text{ and } v = 0\,,\\ -y + \text{sign}(y)\,, & \text{if } |y| \ge 1 \text{ and } v = 0\,,\\ -y - \text{sign}(v)\,, & \text{if } v \ne 0\,.\end{cases} \tag{17}$$

(b) Form the phase plane equation for (17) when $v \ne 0$ and solve it to derive the solutions

$$v^2 + (y \pm 1)^2 = c\,,$$

where the plus sign prevails for $v > 0$ and the minus sign for $v < 0$.

(c) Identify the trajectories in the phase plane as two families of concentric semicircles. What is the center of the semicircles in the upper half-plane? The lower half-plane?

(d) What are the critical points for (17)?

(e) Sketch the trajectory in the phase plane of the mass released from rest at $y = 7.5$. At what value for y does the mass come to rest?

36. Rigid Body Nutation. Euler's equations describe the motion of the principal-axis components of the angular velocity of a freely rotating rigid body (such as a space station), as seen by an observer rotating with the body (the astronauts, for example). This motion is called *nutation.* If the angular velocity components are denoted by x, y, and z, then an example of Euler's equations is the three-dimensional autonomous system

$$dx/dt = yz\,,$$
$$dy/dt = -2xz\,,$$
$$dz/dt = xy\,.$$

The *trajectory* of a solution $x(t), y(t), z(t)$ to these equations is the curve generated by the points $\big(x(t), y(t), z(t)\big)$ in xyz-phase space as t varies over an interval I.

(a) Show that each trajectory of this system lies on the surface of a (possibly degenerate) sphere centered at the origin $(0, 0, 0)$. [*Hint:* Compute $\frac{d}{dt}(x^2 + y^2 + z^2)$.] What does this say about the magnitude of the angular velocity vector?

(b) Find all the critical points of the system, i.e., all points (x_0, y_0, z_0) such that $x(t) \equiv x_0$, $y(t) \equiv y_0$, $z(t) \equiv z_0$ is a solution. For such solutions, the angular velocity vector remains constant in the body system.

(c) Show that the trajectories of the system lie along the intersection of a sphere and an elliptic cylinder of the form $y^2 + 2x^2 = C$, for some constant C. [*Hint:* Consider the expression for dy/dx implied by Euler's equations.]

(d) Using the results of parts (b) and (c), argue that the trajectories of this system are *closed* curves. What does this say about the corresponding solutions?

(e) Figure 5.19 displays some typical trajectories for this system. Discuss the stability of the three critical points indicated on the positive axes.

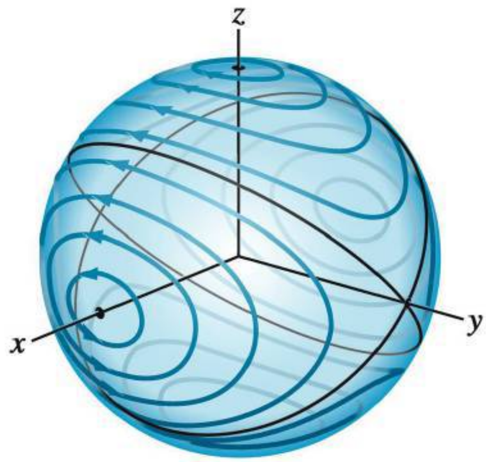


Figure 5.19 Trajectories for Euler's system

5.5 Applications to Biomathematics: Epidemic and Tumor Growth Models

In this section we are going to survey some issues in biological systems that have been successfully modeled by differential equations. We begin by reviewing the population models described in Sections 3.2 and 5.3.

In the *Malthusian model*, the rate of growth of a population $p(t)$ is proportional to the size of the existing population:

$$\textbf{(1)} \qquad \frac{dp}{dt} = kp \quad (k > 0)\ .$$

Cells that reproduce by splitting, such as amoebae and bacteria, are obvious biological examples of this type of growth. Equation (1) implies that a Malthusian population grows exponentially; there is no mechanism for constraining the growth. In Section 3.2 we saw that certain populations exhibit Malthusian growth over limited periods of time (as does compound interest).†

Inserting a negative growth rate,

$$\textbf{(2)} \qquad \frac{dp}{dt} = -kp\ ,$$

results in solutions that *decay* exponentially. Their average lifetime is $1/k$, and their half-life is $(\ln 2)/k$ (Problems 6 and 8). In animals, certain organs such as the kidney serve to cleanse

†Gordon E. Moore (1929–) has observed that the number of transistors on new integrated circuits produced by the electronics industry doubles every 24 months. "Moore's law" is commonly cited by industrialists.

the bloodstream of unwanted components (creatinine clearance, renal clearance), and their concentrations diminish exponentially. As a general rule, the body tends to dissipate ingested drugs in such a manner. (Of course, the most familiar physical instance of Malthusian disintegration is radioactive decay.) Note that if there are both growth and extinction processes, $dp/dt = k_+p - k_-p$ and the equation in (1) still holds with $k = k_+ - k_-$.

When there are two-party interactions occurring in the population that decrease the growth rate, such as competition for resources or violent crime, the *logistic model* might be applicable; it assumes that the extinction rate is proportional to the number of possible pairs in the population, $p(p-1)/2$:

$$\textbf{(3)} \qquad \frac{dp}{dt} = k_1p - k_2\frac{p(p-1)}{2} \quad \text{or, equivalently,} \quad \frac{dp}{dt} = -Ap(p-p_1)\,.^{\dagger}$$

Rodent, bird, and plant populations exhibit logistic growth rates due to social structure, territoriality, and competition for light and space, respectively. The logistic function

$$p(t) = \frac{p_0p_1}{p_0 + (p_1 - p_0)e^{-Ap_1t}}, \qquad p_0 := p(0)$$

was shown in Section 3.2 to be the solution of (3), and typical graphs of $p(t)$ were displayed there.

In Section 5.3 we observed that the *Volterra–Lotka model* for *two different* populations, a *predator* $x_2(t)$ and a *prey* $x_1(t)$, postulates a Malthusian growth rate for the prey and an extinction rate governed by x_1x_2, the number of possible pairings of one from each population,

$$\textbf{(4)} \qquad \frac{dx_1}{dt} = Ax_1 - Bx_1x_2\,,$$

while predators follow a Malthusian extinction rate and pairwise growth rate

$$\textbf{(5)} \qquad \frac{dx_2}{dt} = -Cx_2 + Dx_1x_2\,.$$

Volterra–Lotka dynamics have been observed in blood vessel growth (predator = new capillary tips; prey = chemoattractant), fish populations, and several animal–plant interactions.

Systems like (4)–(5) were studied in Section 5.3 with the aid of the Runge–Kutta algorithm. Now, armed with the insights of Section 5.4, we can further explore this model theoretically.

First, we perform a “reality check” by proving that the populations $x_1(t), x_2(t)$ in the Volterra–Lotka model never change sign. Separating (4) leads to

$$\frac{1}{x_1}\frac{dx_1}{dt} = \frac{d\ln x_1}{dt} = A - Bx_2\,,$$

while integrating from 0 to t results in

$$\textbf{(6)} \qquad x_1(t) = x_1(0)e^{\int_0^t\{A - Bx_2(\tau)\}\,d\tau}\,,$$

and the exponential factor is always positive. Thus $x_1(t)$ [and similarly $x_2(t)$] retains its initial sign (*negative* populations never arise).

---

†One might propose that the *growth* rate in animal populations is due to *two*–party interactions as well. (Wink, wink.) However, in monogamous societies, the number of pairs participating in procreation is proportional to $p/2$, leading to (1). A growth rate determined by *all* possible pairings $p(p-1)/2$ would indicate a highly hedonistic social order.

Example 1 Find and interpret the critical points for the Volterra–Lotka model (4)–(5).

Solution The system

$$\begin{aligned} \frac{dx_1}{dt} &= Ax_1 - Bx_1x_2 = -Bx_1\left(x_2 - \frac{A}{B}\right) = 0\,, \\ \frac{dx_2}{dt} &= -Cx_2 + Dx_1x_2 = Dx_2\left(x_1 - \frac{C}{D}\right) = 0 \end{aligned} \tag{7}$$

has the trivial solution $x_1(t) \equiv x_2(t) \equiv 0$, with an obvious interpretation in terms of populations. If all four coefficients A, B, C, and D are positive, there is also the more interesting solution

$$x_2(t) \equiv \frac{A}{B}\,, \quad x_1(t) \equiv \frac{C}{D}\,. \tag{8}$$

At these population levels, the growth and extinction rates for each species cancel. The direction field diagram in Figure 5.20 for the phase plane equation

$$\frac{dx_2}{dx_1} = \frac{-Cx_2 + Dx_1x_2}{Ax_1 - Bx_1x_2} = \frac{x_2}{x_1}\frac{-C + Dx_1}{A - Bx_2} \tag{9}$$

suggests that this equilibrium is a *center* (compare Figure 5.12 on page 267) with closed (periodic) neighboring trajectories, in accordance with the simulations in Section 5.3. However it is conceivable that some *spiral* trajectories might snake through the field pattern and approach the critical point asymptotically. A rather tricky argument in Problem 4 demonstrates that this is *not* the case. ◆

The SIR Epidemic Model. The SIR† model for an epidemic addresses the spread of diseases that are only contracted by contact with an infected individual; its victims, once recovered, are immune to further infection and are themselves noninfectious. So the members of a population of size N fall into three classes:

$S(t)$ = the number of susceptible individuals—that is, those who have not been infected; $s := S/N$ is the fraction of susceptibles.

$I(t)$ = the number of individuals who are currently infected, comprising a fraction $i := I/N$ of the population.

$R(t)$ = the number of individuals who have recovered from infection, comprising the fraction $r := R/N$.

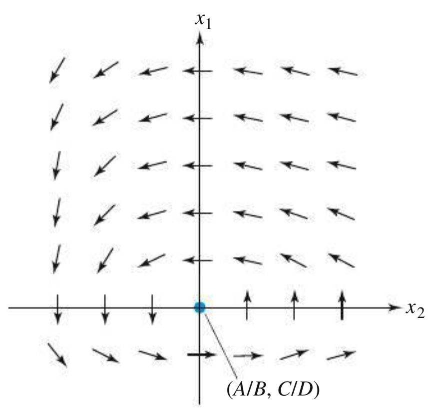


Figure 5.20 Typical direction field diagram for the Volterra–Lotka system

†Introduced by W. O. Kermack and A. G. McKendrick in "A Contribution to the Mathematical Theory of Epidemics," *Proc. Royal Soc. London*, Vol. A115 (1927): 700–721.

The classic SIR epidemic model assumes that on the average, an infectious individual encounters a people per unit time (usually per week). Thus, a total of aI people per week are contacted by infectees, but only a fraction $s = S/N$ of them are susceptible. So the susceptible population diminishes at a rate

(10) $$\frac{dS}{dt} = -saI \quad \text{or} \quad (\text{dividing by } N)$$

(11) $$\frac{ds}{dt} = -asi\,.$$

The parameter a is crucial in disease control. Crowded conditions, or high a, make it difficult to combat the spread of infection. Ideally, we would quarantine the infectees (low a) to fight the epidemic.

The infected population is (obviously) increased whenever a susceptible individual is infected. Additionally, infectees recover in a Malthusian-disintegrative manner over an average time of, say, $1/k$ weeks [recall (1)], so the infected population changes at a rate

(12) $$\frac{dI}{dt} = saI - kI = a\left(s - \frac{k}{a}\right)I \quad \text{or}$$

(13) $$\frac{di}{dt} = a\left(s - \frac{k}{a}\right)i\,.$$

And, of course, the population of recovered individuals increases whenever an infectee is healed:

(14) $$\frac{dR}{dt} = kI \quad \text{or} \quad \frac{dr}{dt} = ki\,.$$

With the SIR model, the total population count remains unchanged:

$$\frac{d(S + I + R)}{dt} = -saI + saI - kI + kI = 0\,.$$

Thus, any fatalities are tallied in the "recovered/noninfectious" population R.

Interestingly, equations (11) and (13) do not contain R or r, so they are suitable for phase plane analysis. In fact they constitute a Volterra–Lotka system with $A = 0$, $B = D = a$, and $C = k$. Because the coefficient A is zero, the critical point structure is different from that discussed in Example 1. Specifically, if $-asi$ in (11) is zero, then only $-ki$ remains on the right in (13), so $I(t) = i(t) \equiv 0$ is necessary and sufficient for a critical point, with S unrestricted. (Physically, this means the populations remain stable only if there are no carriers of the infection.)

Our earlier argument has shown that if $s(t)$ and $i(t)$ are initially positive, they remain so. As a result we conclude from (11) immediately that $s(t)$ decreases monotonically; as such, it has a limiting value $s(\infty)$ as $t \to \infty$. Does $i(t)$ have a limiting value also? If so, $\{s(\infty), i(\infty)\}$ would be a critical point by Theorem 1 of Section 5.4, page 266, and thus $i(\infty) = 0$.

To analyze $i(t)$ consider the phase plane equation for (11) and (13):

(15) $$\frac{di}{ds} = \frac{asi - ki}{-asi} = -1 + \frac{k}{as}\,,$$

which has solutions

(16) $$i = -s + \frac{k}{a}\ln s + K\,.$$

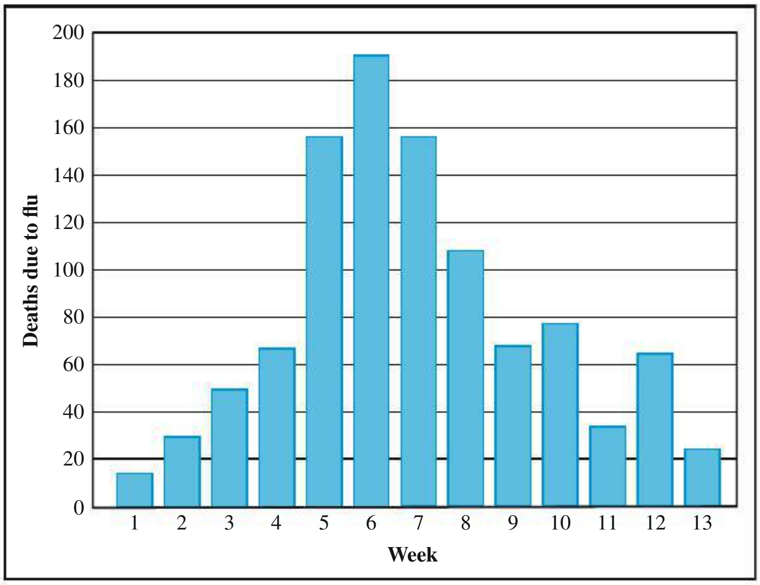


Figure 5.21 Mortality data for Hong Kong flu, New York City

From (16) we see that $s(\infty)$ cannot be zero; otherwise the right-hand side would eventually be negative, contradicting $i(t) > 0$. Therefore, (16) demonstrates that $i(t)$ *does* have a limiting value $i(\infty) = -s(\infty) + (k/a) \ln s(\infty) + K$. As noted, $i(\infty)$ must then be zero.

From (13) we further conclude that if $s(0)$ exceeds the "threshold value" k/a, the infected fraction $i(t)$ will initially increase $(di/dt > 0$ at $t = 0)$ before eventually dying out. The peak value of $i(t)$ occurs when $di/dt = 0 = a[s - (k/a)]i$, i.e., when $s(t)$ passes through the value k/a. In the jargon of epidemiology, this phenomenon defines an "epidemic." You will be directed in Problem 10 to show that if $s(0) \le k/a$, the infected population diminishes monotonically, and no epidemic develops.

Example 2† According to data issued by the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, the Hong Kong flu epidemic during the winter of 1968–1969 was responsible for 1035 deaths in New York City (population 7,900,000), according to the time chart in Figure 5.21. Analyze this data with the SIR model.

Solution Of course, we need to make some assumptions about the parameters. First of all, only a small percentage of people who contract Hong Kong flu perish, so let's assume that the chart reflects a scaled version of the infected population fraction $i(t)$. It is known that the recovery period for this flu is around 5 days, or $5/7$ week, so we try $k = 7/5 = 1.4$. And since the infectees spend much of their convalescence in bed, the average contact rate a is probably less than 1 person per day or 7 per week. The CDC estimated that the initial infected population $I(0)$ was about 10, so the initial data for (11), (13), and (14) are

$$s(0) = \frac{7{,}900{,}000 - 10}{7{,}900{,}000} \approx 0.9999987, \quad i(0) = \frac{10}{7{,}900{,}000} \approx 1.2658 \times 10^{-6}, \quad r(0) = 0 .$$

†We borrow liberally from "The SIR Model for Spread of Disease" by Duke University's David Smith and Lang Moore, *Journal of Online Mathematics and Its Applications*, The MAA Mathematical Sciences Digital Library, http://mathdl.maa.org/mathDL/4/?pa=content&sa=viewDocument&nodeId=479&bodyId=612, copyright 2000, CCP and the authors, published December, 2001. The article contains much interesting information about this epidemic.

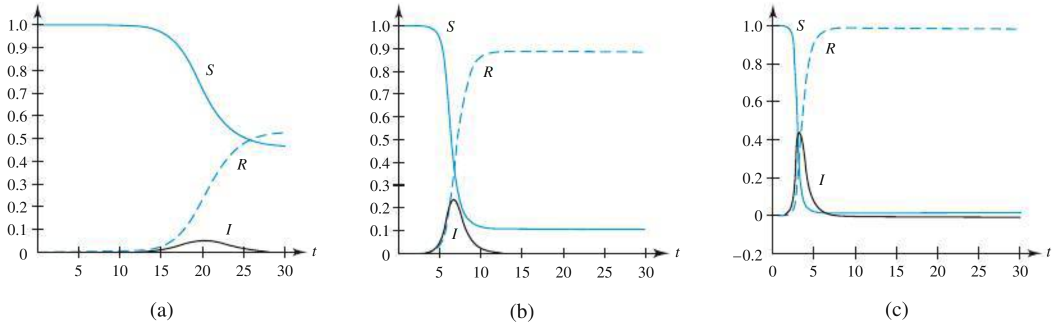


Figure 5.22 SIR simulations (a) $k = 1.4$, $a = 2.0$; (b) $k = 1.4$, $a = 3.5$; (c) $k = 1.4$, $a = 6.0$

Numerical simulations of this system are displayed in Figure 5.22.† The contact rate $a = 3.5$ per week generates an infection fraction curve that closely matches the mortality data's characteristics: time of peak and duration of epidemic. ◆

A Tumor Growth Model.‡ The observed growth of certain tumors can be explained by a model that is mathematically similar to the epidemic model. The total number of cells N in the tumor subdivides into a population P that proliferates by splitting (Malthusian growth) and a population Q that remains quiescent. However, the proliferating cells also can make a transition to the quiescent state, and this occurrence is modeled as a Malthusian-like decay with a "rate" $r(N)$ *that increases with the overall size of the tumor:*

$$\frac{dP}{dt} = cP - r(N)P \,, \tag{17}$$

$$\frac{dQ}{dt} = r(N)P \,. \tag{18}$$

Thus the total population N increases only when the proliferating cells split, as can be seen by adding the equations (17) and (18):

$$\frac{dN}{dt} = \frac{d(P+Q)}{dt} = cP \,. \tag{19}$$

We take (17) and (19) as the system for our analysis. The phase plane equation

$$\frac{dP}{dN} = \frac{cP - r(N)P}{cP} = 1 - \frac{r(N)}{c} \tag{20}$$

can be integrated, leading to a formula for P in terms of N

$$P = N - \frac{1}{c}\int r(N)\, dN + K \,. \tag{21}$$

†Appendix G describes various websites and commercial software that sketch direction fields and automate most of the differential equation algorithms discussed in this book.

‡The authors wish to thank Dr. Glenn Webb of Vanderbilt University for this application. See M. Gyllenberg and G. F. Webb, "Quiescence as an Explanation of Gompertz Tumor Growth," *Growth, Development, and Aging*, Vol. 53 (1989): 25–55.

Suppose the initial conditions are $P(0) = 1$, $Q(0) = 0$, and $N(0) = 1$ (a single proliferating cell). Then we can eliminate the nuisance constant K by taking the indefinite integral in (21) to run from 1 to N and evaluating at $t = 0$:

$$1 = 1 - \frac{1}{c}\int_1^1 r(N)\,dN + K \implies K = 0\,.$$

Insertion of (21) with $K = 0$ into (19) produces a differential equation for N alone:

$$\frac{dN}{dt} = cN - \int_1^N r(u)\,du\,. \tag{22}$$

Example 3 The *Gompertz law*

$$N(t) = e^{c(1-e^{-bt})/b} \tag{23}$$

has been observed experimentally for the growth of some tumors. Show that a transition rate $r(N)$ of the form $b(1 + \ln N)$ predicts Gompertzian growth.

Solution If the indicated integral of the rate $r(N)$ is carried out, (22) becomes

$$\frac{dN}{dt} = cN - b(N-1) - b(N\ln N - N + 1) = (c - b\ln N)\,N\,. \tag{24}$$

Dividing by N we obtain a linear differential equation for the function $\ln N$

$$\frac{d\ln N}{dt} = -b\ln N + c$$

whose solution, for the initial condition $N(0) = 1$, is found by the methods of Section 2.3 to be

$$\ln N(t) = \frac{c}{b}(1 - e^{-bt})\,,$$

confirming (23). ◆

Problem 9 invites the reader to show that if the growth rate is modeled as $r(N) = s(2N - 1)$, then the solution of (22) describes logistic growth. Typical curves for the Gompertz and logistic models are displayed in Figure 5.23. See also Figure 3.4 on page 98.

Other applications of differential equations to biomathematics appear in the discussions of artificial respiration (Project B, page 81) in Chapter 2, HIV infection (Project A, page 141) and aquaculture (Project B, page 144) in Chapter 3, and spread of staph infections (Project B) and the Ebola epidemic (Project F, page 314) in this chapter.

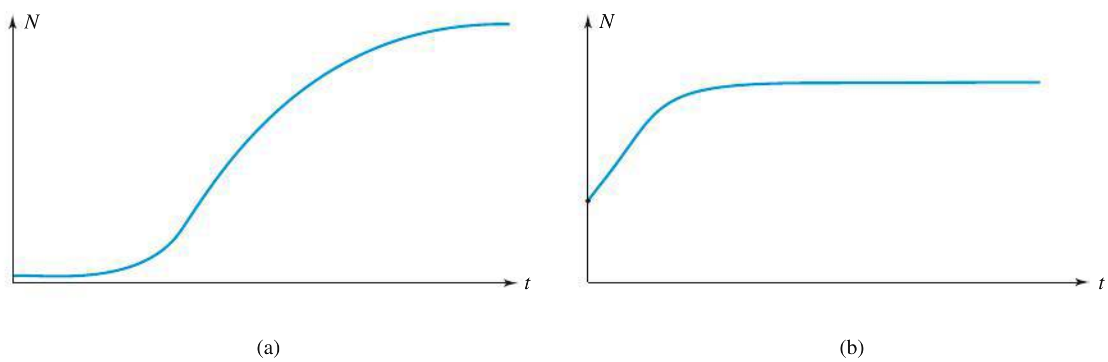


Figure 5.23 (a) Gompertz and (b) logistic curves

5.5 EXERCISES

1. Logistic Model. In Section 3.2 we discussed the logistic equation

$$\frac{dp}{dt} = Ap_1p - Ap^2, \qquad p(0) = p_0,$$

and its use in modeling population growth. A more general model might involve the equation

$$\frac{dp}{dt} = Ap_1p - Ap^r, \qquad p(0) = p_0, \tag{25}$$

where $r > 1$. To see the effect of changing the parameter r in (25), take $p_1 = 3, A = 1$, and $p_0 = 1$. Then use a numerical scheme such as Runge–Kutta with $h = 0.25$ to approximate the solution to (25) on the interval $0 \le t \le 5$ for $r = 1.5$, 2, and 3. What is the limiting population in each case? For $r > 1$, determine a general formula for the limiting population.

2. Radioisotopes and Cancer Detection. A radioisotope commonly used in the detection of breast cancer is technetium-99m. This radionuclide is attached to a chemical that upon injection into a patient accumulates at cancer sites. The isotope's radiation is then detected and the site located, using gamma cameras or other tomographic devices.

Technetium-99m decays radioactively in accordance with the equation $dy/dt = -ky$, with $k = 0.1155$/h. The short half-life of technetium-99m has the advantage that its radioactivity does not endanger the patient. A disadvantage is that the isotope must be manufactured in a cyclotron. Since hospitals are not equipped with cyclotrons, doses of technetium-99m have to be ordered in advance from medical suppliers.

Suppose a dosage of 5 millicuries (mCi) of technetium-99m is to be administered to a patient. Estimate the delivery time from production at the manufacturer to arrival at the hospital treatment room to be 24 hours and calculate the amount of the radionuclide that the hospital must order, to be able to administer the proper dosage.

3. Secretion of Hormones. The secretion of hormones into the blood is often a periodic activity. If a hormone is secreted on a 24-h cycle, then the rate of change of the level of the hormone in the blood may be represented by the initial value problem

$$\frac{dx}{dt} = \alpha - \beta \cos\frac{\pi t}{12} - kx, \quad x(0) = x_0,$$

where $x(t)$ is the amount of the hormone in the blood at time t, α is the average secretion rate, β is the amount of daily variation in the secretion, and k is a positive constant reflecting the rate at which the body removes the hormone from the blood. If $\alpha = \beta = 1, k = 2$, and $x_0 = 10$, solve for $x(t)$.

4. Prove that the critical point (8) of the Volterra–Lotka system is a center; that is, the neighboring trajectories are periodic. [*Hint:* Observe that (9) is separable and show that its solutions can be expressed as

$$[x_2^A e^{-Bx_2}] \cdot [x_1^C e^{-Dx_1}] = K.] \tag{26}$$

Prove that the maximum of the function $x^p e^{-qx}$ is $(p/qe)^p$, occurring at the unique value $x = p/q$ (see Figure 5.24), so the critical values (8) maximize the factors on the left in (26). Argue that if K takes the corresponding maximum value $(A/Be)^A(C/De)^C$, the critical point (8) is the (unique) solution of (26), and it cannot be an endpoint of any trajectory for (26) with a lower value of K.[†]

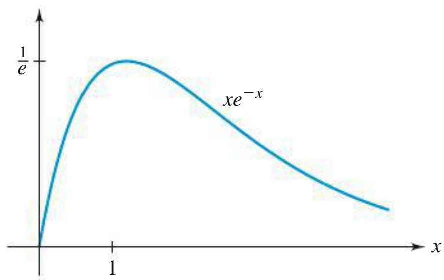


Figure 5.24 Graph of xe^{-x}

5. Suppose for a certain disease described by the SIR model it is determined that $a = 0.003$ and $b = 0.5$.

(a) In the *SI*-phase plane, sketch the trajectory corresponding to the initial condition that one person is infected and 700 persons are susceptible.

[†]In fact, the periodic fluctuations predicted by the Volterra–Lotka model were observed in fish populations by Lotka's son-in-law, Humberto D'Ancona.

(b) From your graph in part (a), estimate the peak number of infected persons. Compare this with the theoretical prediction $S = k/a \approx 167$ persons when the epidemic is at its peak.

6. Show that the half-life of solutions to (2)—that is, the time required for the solution to decay to one-half of its value—equals $(\ln 2)/k$.

7. Complete the solution of the tumor growth model for Example 3 on page 280 by finding $P(t)$ and $Q(t)$.

8. If $p(t)$ is a Malthusian population that diminishes according to (2), then $p(t_2) - p(t_1)$ is the number of individuals in the population whose lifetime lies between t_1 and t_2. Argue that the *average* lifetime of the population is given by the formula

$$\frac{\int_0^\infty t\left|\frac{dp(t)}{dt}\right|dt}{p(0)}$$

and show that this equals $1/k$.

9. Show that with the transition rate formula $r(N) = s(2N - 1)$, equation (22) takes the form of the equation for the logistic model (Section 3.2, equation (14)). Solve (22) for this case.

10. Prove that the infected population $I(t)$ in the SIR model does not increase if $S(0)$ is less than or equal to k/a.

11. An epidemic reported by the British Communicable Disease Surveillance Center in the *British Medical Journal* (March 4, 1978, p. 587) took place in a boarding school with 763 residents.† The statistics for the infected population are shown in the graph in Figure 5.25.

Assuming that the average duration of the infection is 2 days, use a numerical differential equation solver (see Appendix G) to try to reproduce the data. Take $S(0) = 762, I(0) = 1, R(0) = 0$ as initial conditions. Experiment with reasonable estimates for the average number of contacts per day by the infected students, who were confined to bed after the infection was detected. What value of this parameter seems to fit the curve best?

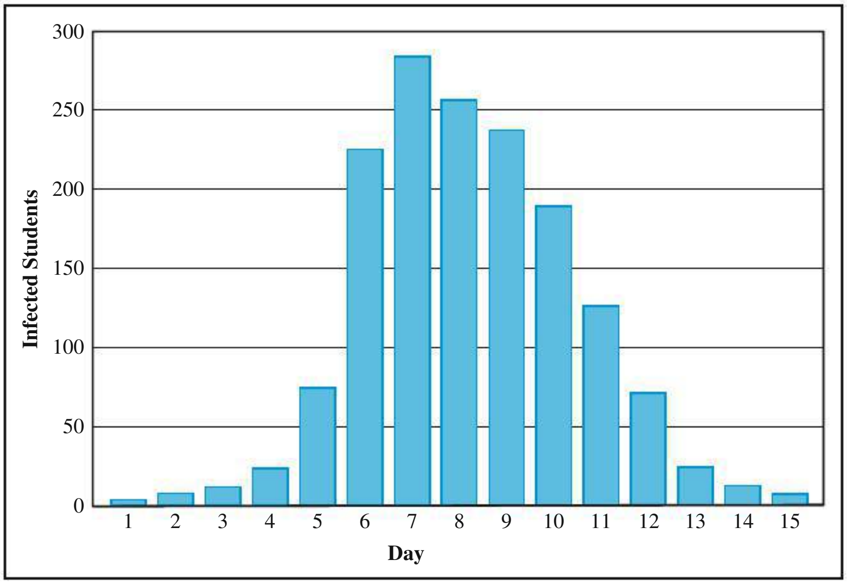


Figure 5.25 Flu data for Problem 11

†See also the discussion of this epidemic in *Mathematical Biology I, An Introduction*, by J. D. Murray (Springer-Verlag, New York, 2002), 325–326.

5.6 Coupled Mass–Spring Systems

In this section we extend the mass–spring model of Chapter 4 to include situations in which coupled springs connect two masses, both of which are free to move. The resulting motions can be very intriguing. For simplicity we'll neglect the effects of friction, gravity, and external forces. Let's analyze the following experiment.

Example 1 On a smooth horizontal surface, a mass $m_1 = 2$ kg is attached to a fixed wall by a spring with spring constant $k_1 = 4$ N/m. Another mass $m_2 = 1$ kg is attached to the first object by a spring with spring constant $k_2 = 2$ N/m. The objects are aligned horizontally so that the springs are their natural lengths (Figure 5.26). If both objects are displaced 3 m to the right of their equilibrium positions (Figure 5.27) and then released, what are the equations of motion for the two objects?

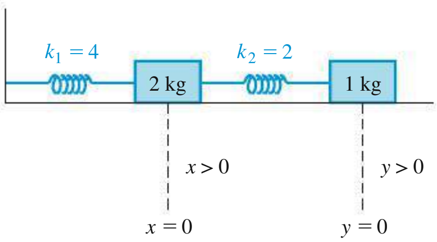


Figure 5.26 Coupled system at equilibrium

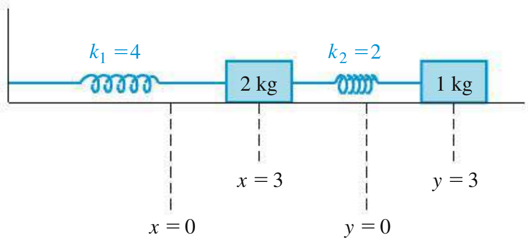


Figure 5.27 Coupled system at initial displacement

Solution From our assumptions, the only forces we need to consider are those due to the springs themselves. Recall that Hooke's law asserts that the force acting on an object due to a spring has magnitude proportional to the displacement of the spring from its natural length and has direction opposite to its displacement. That is, if the spring is either stretched or compressed, then it tries to return to its natural length.

Because each mass is free to move, we apply Newton's second law to each object. Let $x(t)$ denote the displacement (to the right) of the 2-kg mass from its equilibrium position and similarly, let $y(t)$ denote the corresponding displacement for the 1-kg mass. The 2-kg mass has a force F_1 acting on its left side due to one spring and a force F_2 acting on its right side due to the second spring. Referring to Figure 5.27 and applying Hooke's law, we see that

$$F_1 = -k_1 x\,, \qquad F_2 = +k_2(y - x)\,,$$

since $(y - x)$ is the net displacement of the second spring from its natural length. There is only one force acting on the 1-kg mass: the force due to the second spring, which is

$$F_3 = -k_2(y - x)\,.$$

Applying Newton's second law to these objects, we obtain the system

$$\text{(1)} \qquad \begin{aligned} m_1\frac{d^2x}{dt^2} &= F_1 + F_2 = -k_1x + k_2(y - x)\,, \\ m_2\frac{d^2y}{dt^2} &= F_3 = -k_2(y - x)\,, \end{aligned}$$

or

$$\text{(2)} \qquad \begin{aligned} m_1\frac{d^2x}{dt^2} + (k_1 + k_2)x - k_2y &= 0\,, \\ m_2\frac{d^2y}{dt^2} + k_2y - k_2x &= 0\,. \end{aligned}$$

In this problem, we know that $m_1 = 2$, $m_2 = 1$, $k_1 = 4$, and $k_2 = 2$. Substituting these values into system (2) yields

$$2\frac{d^2x}{dt^2} + 6x - 2y = 0\,, \tag{3}$$

$$\frac{d^2y}{dt^2} + 2y - 2x = 0\,. \tag{4}$$

We'll use the elimination method of Section 5.2 to solve (3)–(4). With $D \coloneqq d/dt$ we rewrite the system as

$$(2D^2 + 6)[x] - 2y = 0\,, \tag{5}$$

$$-2x + (D^2 + 2)[y] = 0\,. \tag{6}$$

Adding $(D^2 + 2)$ applied to equation (5) to 2 times equation (6) eliminates y:

$$[(D^2 + 2)(2D^2 + 6) - 4][x] = 0\,,$$

which simplifies to

$$2\frac{d^4x}{dt^4} + 10\frac{d^2x}{dt^2} + 8x = 0\,. \tag{7}$$

Notice that equation (7) is linear with constant coefficients. To solve it let's proceed as we did with linear second-order equations and try to find solutions of the form $x = e^{rt}$. Substituting e^{rt} in equation (7) gives

$$2(r^4 + 5r^2 + 4)e^{rt} = 0\,.$$

Thus, we get a solution to (7) when r satisfies the auxiliary equation

$$r^4 + 5r^2 + 4 = 0\,.$$

From the factorization $r^4 + 5r^2 + 4 = (r^2 + 1)(r^2 + 4)$, we see that the roots of the auxiliary equation are complex numbers i, $-i$, $2i$, $-2i$. Using Euler's formula, it follows that

$$z_1(t) = e^{it} = \cos t + i\sin t \qquad \text{and} \qquad z_2(t) = e^{2it} = \cos 2t + i\sin 2t$$

are complex-valued solutions to equation (7). To obtain real-valued solutions, we take the real and imaginary parts of $z_1(t)$ and $z_2(t)$. Thus, four real-valued solutions are

$$x_1(t) = \cos t\,, \qquad x_2(t) = \sin t\,, \qquad x_3(t) = \cos 2t\,, \qquad x_4(t) = \sin 2t\,,$$

and a general solution is

$$\boldsymbol{x(t) = a_1 \cos t + a_2 \sin t + a_3 \cos 2t + a_4 \sin 2t\,,} \tag{8}$$

where a_1, a_2, a_3, and a_4 are arbitrary constants.†

To obtain a formula for $y(t)$, we use equation (3) to express y in terms of x:

$$\begin{aligned} y(t) &= \frac{d^2x}{dt^2} + 3x \\ &= -a_1\cos t - a_2\sin t - 4a_3\cos 2t - 4a_4\sin 2t \\ &\quad + 3a_1\cos t + 3a_2\sin t + 3a_3\cos 2t + 3a_4\sin 2t\,, \end{aligned}$$

†A more detailed discussion of general solutions is given in Chapter 6.

and so

$$y(t) = 2a_1 \cos t + 2a_2 \sin t - a_3 \cos 2t - a_4 \sin 2t\,. \tag{9}$$

To determine the constants a_1, a_2, a_3, and a_4, let's return to the original problem. We were told that the objects were originally displaced 3 m to the right and then released. Hence,

$$x(0) = 3\,, \quad \frac{dx}{dt}(0) = 0\,; \quad y(0) = 3\,, \quad \frac{dy}{dt}(0) = 0\,. \tag{10}$$

On differentiating equations (8) and (9), we find

$$\frac{dx}{dt} = -a_1 \sin t + a_2 \cos t - 2a_3 \sin 2t + 2a_4 \cos 2t\,,$$
$$\frac{dy}{dt} = -2a_1 \sin t + 2a_2 \cos t + 2a_3 \sin 2t - 2a_4 \cos 2t\,.$$

Now, if we put $t = 0$ in the formulas for x, dx/dt, y, and dy/dt, the initial conditions (10) give the four equations

$$x(0) = a_1 + a_3 = 3\,, \qquad \frac{dx}{dt}(0) = a_2 + 2a_4 = 0\,,$$
$$y(0) = 2a_1 - a_3 = 3\,, \qquad \frac{dy}{dt}(0) = 2a_2 - 2a_4 = 0\,.$$

From this system, we find $a_1 = 2$, $a_2 = 0$, $a_3 = 1$, and $a_4 = 0$. Hence, the equations of motion for the two objects are

$$x(t) = 2\cos t + \cos 2t\,,$$
$$y(t) = 4\cos t - \cos 2t\,,$$

which are depicted in Figure 5.28. ◆

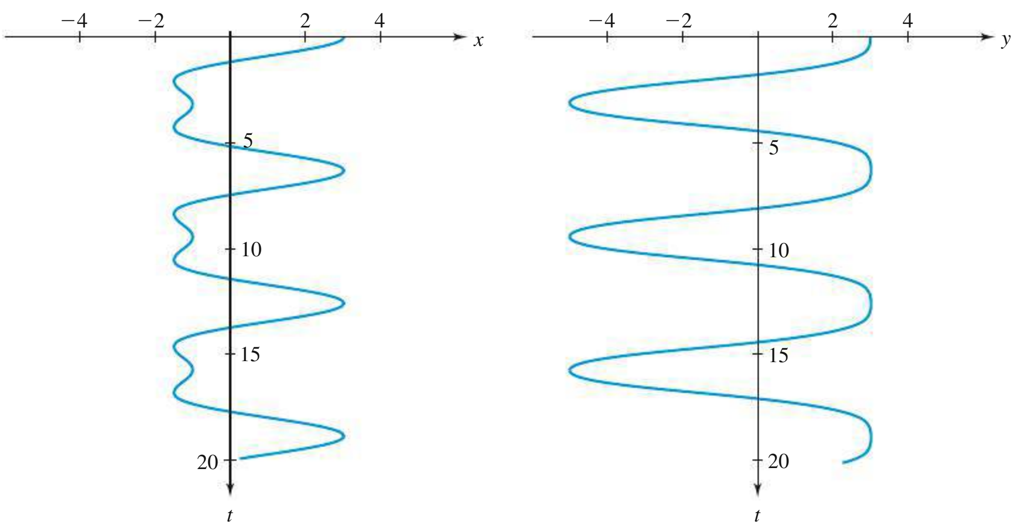


Figure 5.28 Graphs of the motion of the two masses in the coupled mass–spring system

The general solution pair (8), (9) that we have obtained is a combination of sinusoids oscillating at two different angular frequencies: 1 rad/sec and 2 rad/sec. These frequencies extend the notion of the *natural frequency* of the single (free, undamped) mass–spring oscillator (page 212, Section 4.9) and are called the **natural** (or **normal**) **angular frequencies**† of the system. A complex system consisting of more masses and springs would have many normal frequencies.

Notice that if the initial conditions were altered so that the constants a_3 and a_4 in (8) and (9) were zero, the motion would be a pure sinusoid oscillating at the single frequency 1 rad/sec. Similarly, if a_1 and a_2 were zero, only the 2 rad/sec oscillation would be "excited." Such solutions, wherein the entire motion is described by a single sinusoid, are called the **normal modes** of the system.‡ The normal modes in the following example are particularly easy to visualize because we will take all the masses and all the spring constants to be equal.

Example 2 Three identical springs with spring constant k and two identical masses m are attached in a straight line with the ends of the outside springs fixed (see Figure 5.29). Determine and interpret the normal modes of the system.

Solution We define the displacements from equilibrium, x and y, as in Example 1 on page 283. The equations expressing Newton's second law for the masses are quite analogous to (1), except for the effect of the third spring on the second mass:

$$mx'' = -kx + k(y - x) , \tag{11}$$

$$my'' = -k(y - x) - ky , \tag{12}$$

or

$$(mD^2 + 2k)[x] - ky = 0 ,$$
$$-kx + (mD^2 + 2k)[y] = 0 .$$

Eliminating y in the usual manner results in

$$\left[(mD^2 + 2k)^2 - k^2 \right][x] = 0 . \tag{13}$$

This has the auxiliary equation

$$(mr^2 + 2k)^2 - k^2 = (mr^2 + k)(mr^2 + 3k) = 0 ,$$

with roots $\pm i\sqrt{k/m}$, $\pm i\sqrt{3k/m}$. Setting $\omega := \sqrt{k/m}$, we get the following general solution to (13):

$$x(t) = C_1 \cos \omega t + C_2 \sin \omega t + C_3 \cos\left(\sqrt{3}\omega t\right) + C_4 \sin\left(\sqrt{3}\omega t\right) . \tag{14}$$

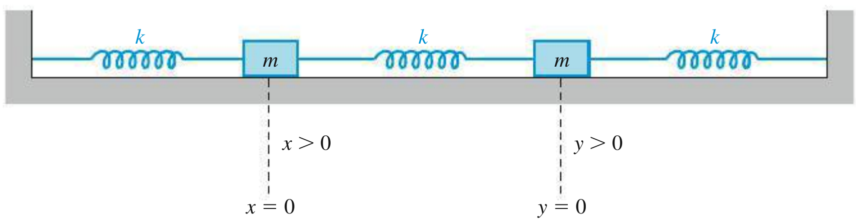


Figure 5.29 Coupled mass–spring system with fixed ends

†The study of the natural frequencies of oscillations of complex systems is known in engineering as **modal analysis.**

‡The normal modes are more naturally characterized in terms of *eigenvalues* (see Section 9.5).

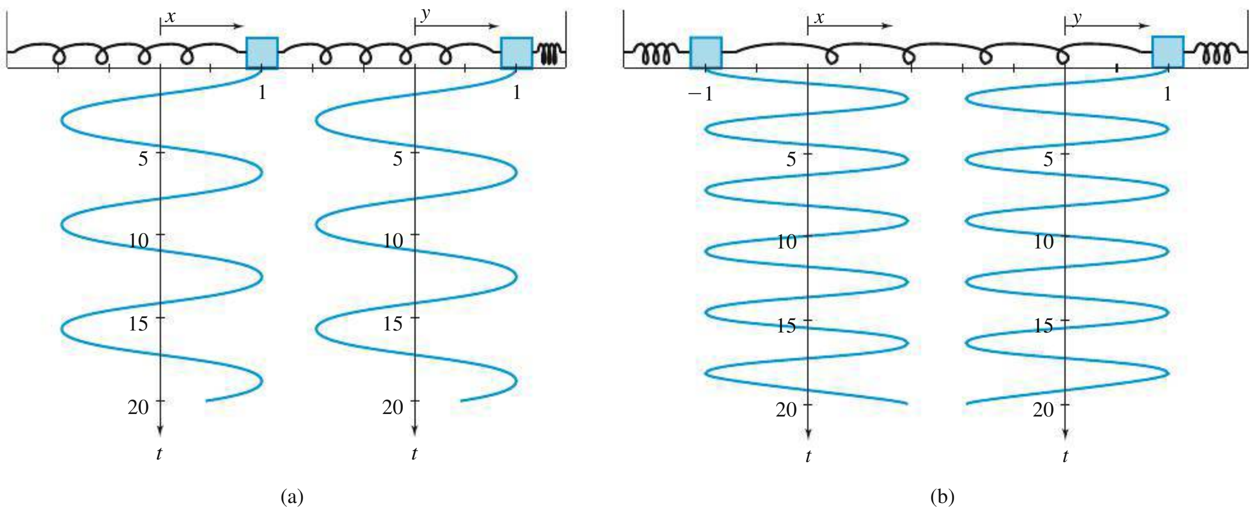


Figure 5.30 Normal modes for Example 2

To obtain $y(t)$, we solve for $y(t)$ in (11) and substitute $x(t)$ as given in (14). Upon simplifying, we get

$$y(t) = C_1 \cos \omega t + C_2 \sin \omega t - C_3 \cos\left(\sqrt{3}\omega t\right) - C_4 \sin\left(\sqrt{3}\omega t\right) . \tag{15}$$

From the formulas (14) and (15), we see that the normal angular frequencies are ω and $\sqrt{3}\omega$. Indeed, if $C_3 = C_4 = 0$, we have a solution where $y(t) \equiv x(t)$, oscillating at the angular frequency $\omega = \sqrt{k/m}$ rad/sec (equivalent to a frequency $\sqrt{k/m}/2\pi$ periods/sec). Now if $x(t) \equiv y(t)$ in Figure 5.29, the two masses are moving as if they were a single rigid body of mass $2m$, forced by a "double spring" with a spring constant given by $2k$. And indeed, according to equation (4) of Section 4.9 (page 212), we would expect such a system to oscillate at the angular frequency $\sqrt{2k/2m} = \sqrt{k/m}$ (!). This motion is depicted in Figure 5.30(a).

Similarly, if $C_1 = C_2 = 0$, we find the second normal mode where $y(t) = -x(t)$, so that in Figure 5.29 there are two mirror-image systems, each with mass m and a "spring and a half" with spring constant $k + 2k = 3k$. (The half-spring will be twice as stiff.) Section 4.9's equation (4) then predicts an angular oscillation frequency for each system of $\sqrt{3k/m} = \sqrt{3}\omega$, which again is consistent with (14) and (15). This motion is shown in Figure 5.30(b). ◆

5.6 EXERCISES

1. Two springs and two masses are attached in a straight line on a horizontal frictionless surface as illustrated in Figure 5.31. The system is set in motion by holding the mass m_2 at its equilibrium position and pulling the mass m_1 to the left of its equilibrium position a distance 1 m and then releasing both masses. Express Newton's law for the system and determine the equations of motion for the two masses if $m_1 = 1$ kg, $m_2 = 2$ kg, $k_1 = 4$ N/m, and $k_2 = 10/3$ N/m.

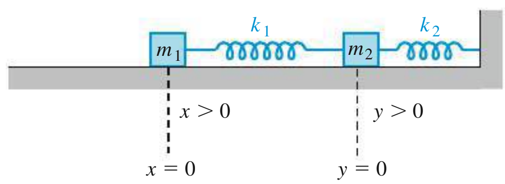


Figure 5.31 Coupled mass–spring system with one end free

2. Determine the equations of motion for the two masses described in Problem 1 if $m_1 = 1$ kg, $m_2 = 1$ kg, $k_1 = 3$ N/m, and $k_2 = 2$ N/m.

3. Four springs with the same spring constant and three equal masses are attached in a straight line on a horizontal frictionless surface as illustrated in Figure 5.32. Determine the normal frequencies for the system and describe the three normal modes of vibration.

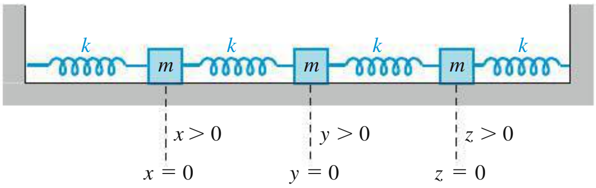


Figure 5.32 Coupled mass–spring system with three degrees of freedom

4. Two springs, two masses, and a dashpot are attached in a straight line on a horizontal frictionless surface as shown in Figure 5.33. The dashpot provides a damping force on mass m_2, given by $F = -by'$. Derive the system of differential equations for the displacements x and y.

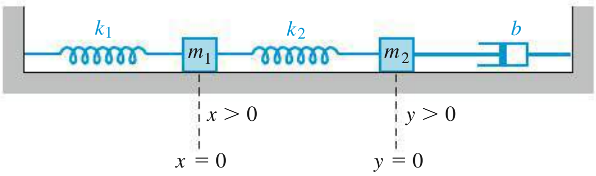


Figure 5.33 Coupled mass–spring system with one end damped

5. Two springs, two masses, and a dashpot are attached in a straight line on a horizontal frictionless surface as shown in Figure 5.34. The system is set in motion by holding the mass m_2 at its equilibrium position and pushing the mass m_1 to the left of its equilibrium position a distance 2 m and then releasing both masses. Determine the equations of motion for the two masses if $m_1 = m_2 = 1$ kg, $k_1 = k_2 = 1$ N/m, and $b = 1$ N-sec/m. [*Hint:* The dashpot damps both m_1 and m_2 with a force whose magnitude equals $b|y' - x'|$.]

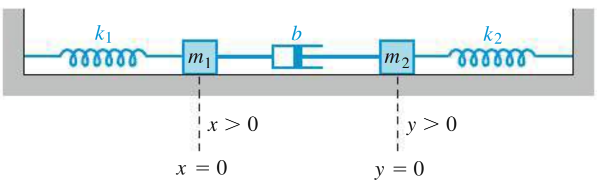


Figure 5.34 Coupled mass–spring system with damping between the masses

6. Referring to the coupled mass–spring system discussed in Example 1, suppose an external force $E(t) = 37\cos 3t$ is applied to the second object of mass 1 kg. The displacement functions $x(t)$ and $y(t)$ now satisfy the system

$$2x''(t) + 6x(t) - 2y(t) = 0\,, \tag{16}$$

$$y''(t) + 2y(t) - 2x(t) = 37\cos 3t\,. \tag{17}$$

(a) Show that $x(t)$ satisfies the equation

$$x^{(4)}(t) + 5x''(t) + 4x(t) = 37\cos 3t\,. \tag{18}$$

(b) Find a general solution $x(t)$ to equation (18). [*Hint:* Use undetermined coefficients with $x_p = A\cos 3t + B\sin 3t$.]

(c) Substitute $x(t)$ back into (16) to obtain a formula for $y(t)$.

(d) If both masses are displaced 2 m to the right of their equilibrium positions and then released, find the displacement functions $x(t)$ and $y(t)$.

7. Suppose the displacement functions $x(t)$ and $y(t)$ for a coupled mass–spring system (similar to the one discussed in Problem 6) satisfy the initial value problem

$$x''(t) + 5x(t) - 2y(t) = 0\,,$$
$$y''(t) + 2y(t) - 2x(t) = 3\sin 2t\,;$$
$$x(0) = x'(0) = 0\,,$$
$$y(0) = 1\,, \qquad y'(0) = 0\,.$$

Solve for $x(t)$ and $y(t)$.

8. A double pendulum swinging in a vertical plane under the influence of gravity (see Figure 5.35) satisfies the system

$$(m_1 + m_2)l_1^2\theta_1'' + m_2 l_1 l_2 \theta_2'' + (m_1 + m_2)l_1 g\theta_1 = 0\,,$$
$$m_2 l_2^2 \theta_2'' + m_2 l_1 l_2 \theta_1'' + m_2 l_2 g \theta_2 = 0\,,$$

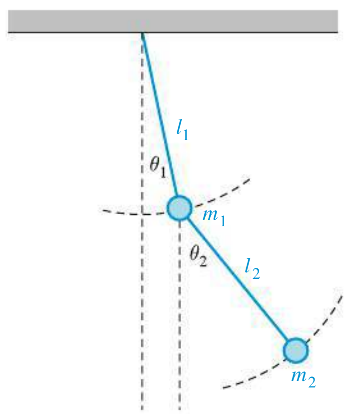


Figure 5.35 Double pendulum

when θ_1 and θ_2 are small angles. Solve the system when $m_1 = 3$ kg, $m_2 = 2$ kg, $l_1 = l_2 = 5$ m, $\theta_1(0) = \pi/6$, $\theta_2(0) = \theta_1'(0) = \theta_2'(0) = 0$.

9. The motion of a pair of identical pendulums coupled by a spring is modeled by the system

$$mx_1'' = -\frac{mg}{l}x_1 - k(x_1 - x_2)\ ,$$
$$mx_2'' = -\frac{mg}{l}x_2 + k(x_1 - x_2)$$

for small displacements (see Figure 5.36). Determine the two normal frequencies for the system.

10. Suppose the coupled mass–spring system of Problem 1 (Figure 5.31) is hung *vertically* from a support (with mass m_2 above m_1), as in Section 4.10, page 226.

(a) Argue that at equilibrium, the lower spring is stretched a distance l_1 from its natural length L_1, given by $l_1 = m_1 g/k_1$.

(b) Argue that at equilibrium, the upper spring is stretched a distance $l_2 = (m_1 + m_2)g/k_2$.

(c) Show that if x_1 and x_2 are redefined to be displacements from the equilibrium positions of the masses m_1 and m_2, then the equations of motion are identical with those derived in Problem 1.

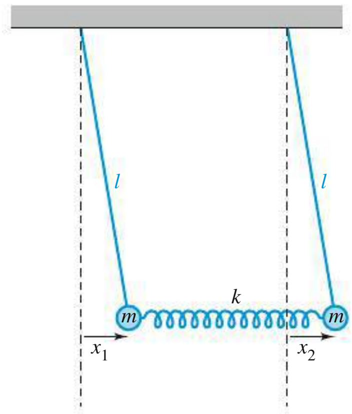


Figure 5.36 Coupled pendulums

5.7 Electrical Systems

The equations governing the voltage–current relations for the resistor, inductor, and capacitor were given in Section 3.5, together with Kirchhoff's laws that constrain how these quantities behave when the elements are electrically connected into a circuit. Now that we have the tools for solving linear equations of higher-order systems, we are in a position to analyze more complex electrical circuits.

Example 1 The series *RLC* circuit in Figure 5.37 on page 290 has a voltage source given by $E(t) = \sin 100t$ volts (V), a resistor of 0.02 ohms (Ω), an inductor of 0.001 henrys (H), and a capacitor of 2 farads (F). (These values are selected for numerical convenience; typical capacitance values are much smaller.) If the initial current and the initial charge on the capacitor are both zero, determine the current in the circuit for $t > 0$.

Solution Using the notation of Section 3.5, we have $L = 0.001$ H, $R = 0.02\ \Omega$, $C = 2$ F, and $E(t) = \sin 100t$. According to Kirchhoff's current law, the same current I passes through each circuit element. The current through the capacitor equals the instantaneous rate of change of its charge q:

$$(1) \qquad I = dq/dt\ .$$

From the physics equations in Section 3.5, we observe that the voltage drops across the capacitor (E_C), the resistor (E_R), and the inductor (E_L) are expressed as

$$(2) \qquad E_C = \frac{q}{C}\ , \qquad E_R = RI\ , \qquad E_L = L\frac{dI}{dt}\ .$$

Therefore, Kirchhoff's voltage law, which implies $E_L + E_R + E_C = E$, can be expressed as

$$(3) \qquad L\frac{dI}{dt} + RI + \frac{1}{C}q = E(t)\ .$$

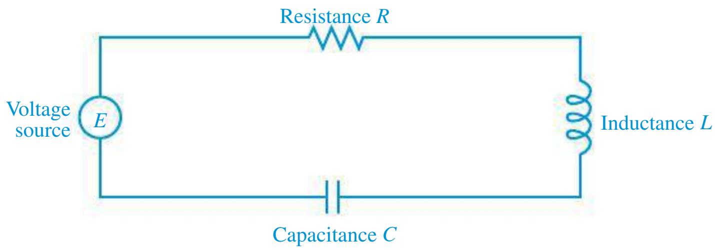


Figure 5.37 Schematic representation of an *RLC* series circuit

In most applications we will be interested in determining the current $I(t)$. If we differentiate (3) with respect to t and substitute I for dq/dt, we obtain

$$\textbf{(4)} \qquad L\frac{d^2I}{dt^2} + R\frac{dI}{dt} + \frac{1}{C}I = \frac{dE}{dt} .$$

After substitution of the given values this becomes

$$(0.001)\frac{d^2I}{dt^2} + (0.02)\frac{dI}{dt} + (0.5)I = 100\cos 100t ,$$

or, equivalently,

$$\textbf{(5)} \qquad \frac{d^2I}{dt^2} + 20\frac{dI}{dt} + 500I = 100{,}000\cos 100t .$$

The homogeneous equation associated with (5) has the auxiliary equation

$$r^2 + 20r + 500 = (r+10)^2 + (20)^2 = 0 ,$$

whose roots are $-10 \pm 20i$. Hence, the solution to the homogeneous equation is

$$\textbf{(6)} \qquad I_h(t) = C_1e^{-10t}\cos 20t + C_2e^{-10t}\sin 20t .$$

To find a particular solution for (5), we can use the method of undetermined coefficients. Setting

$$I_p(t) = A\cos 100t + B\sin 100t$$

and carrying out the procedure in Section 4.5, we ultimately find, to three decimals,

$$A = -10.080 , \qquad B = 2.122 .$$

Hence, a particular solution to (5) is given by

$$\textbf{(7)} \qquad I_p(t) = -10.080\cos 100t + 2.122\sin 100t .$$

Since $I = I_h + I_p$, we find from (6) and (7) that

$$\textbf{(8)} \qquad I(t) = e^{-10t}(C_1\cos 20t + C_2\sin 20t) - 10.080\cos 100t + 2.122\sin 100t .$$

To determine the constants C_1 and C_2, we need the values $I(0)$ and $I'(0)$. We were given $I(0) = q(0) = 0$. To find $I'(0)$, we substitute the values for L, R, and C into equation (3) and equate the two sides at $t = 0$. This gives

$$(0.001)I'(0) + (0.02)I(0) + (0.5)q(0) = \sin 0 .$$

Because $I(0) = q(0) = 0$, we find $I'(0) = 0$. Finally, using $I(t)$ in (8) and the initial conditions $I(0) = I'(0) = 0$, we obtain the system

$$I(0) = C_1 - 10.080 = 0\,,$$
$$I'(0) = -10C_1 + 20C_2 + 212.2 = 0\,.$$

Solving this system yields $C_1 = 10.080$ and $C_2 = -5.570$. Hence, the current in the *RLC* series circuit is

$$I(t) = e^{-10t}(10.080\cos 20t - 5.570\sin 20t) - 10.080\cos 100t + 2.122\sin 100t\,. \quad (9)$$ ◆

Observe that, as was the case with forced mechanical vibrations, the current in (9) is made up of two components. The first, I_h, is a **transient current** that tends to zero as $t \to +\infty$. The second,

$$I_p(t) = -10.080\cos 100t + 2.122\sin 100t\,,$$

is a sinusoidal **steady-state current** that remains.

It is straightforward to verify that the steady-state solution $I_p(t)$ that arises from the more general voltage source $E(t) = E_0 \sin\gamma t$ is

$$I_p(t) = \frac{E_0\sin(\gamma t + \theta)}{\sqrt{R^2 + [\gamma L - 1/(\gamma C)]^2}}\,, \quad (10)$$

where $\tan\theta = (1/C - L\gamma^2)/(\gamma R)$ (compare Section 4.10, page 221).

Example 2 At time $t = 0$, the charge on the capacitor in the electrical network shown in Figure 5.38 is 2 coulombs (C), while the current through the capacitor is zero. Determine the charge on the capacitor and the currents in the various branches of the network at any time $t > 0$.

Solution To determine the charge and currents in the electrical network, we begin by observing that the network consists of three closed circuits: loop 1 through the battery, resistor, and inductor; loop 2 through the battery, resistor, and capacitor; and loop 3 containing the capacitor and inductor. Taking advantage of Kirchhoff's current law, we denote the current passing through the battery and the resistor by I_1, the current through the inductor by I_2, and the current through the capacitor by I_3. For consistency of notation, we denote the charge on the capacitor by q_3. Hence, $I_3 = dq_3/dt$.

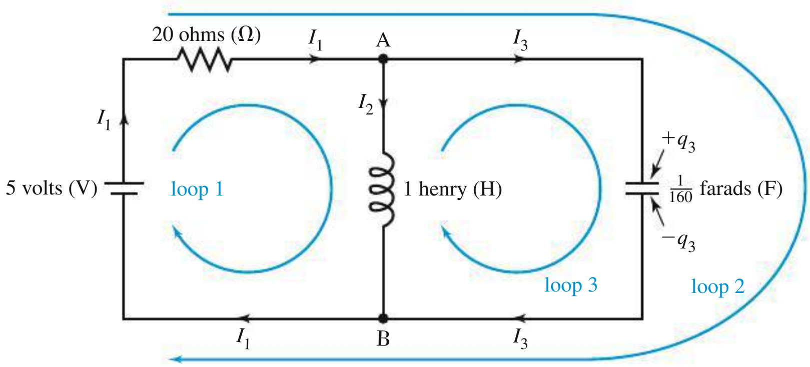


Figure 5.38 Schematic of an electrical network

As discussed at the beginning of this section, the voltage drop at a resistor is RI, at an inductor LdI/dt, and at a capacitor q/C. So, applying Kirchhoff's voltage law to the electrical network in Figure 5.38, we find for loop 1,

$$\underset{\text{(inductor)}}{\frac{dI_2}{dt}} + \underset{\text{(resistor)}}{20I_1} = \underset{\text{(battery)}}{5}\,; \tag{11}$$

for loop 2,

$$\underset{\text{(resistor)}}{20I_1} + \underset{\text{(capacitor)}}{160q_3} = \underset{\text{(battery)}}{5}\,; \tag{12}$$

and for loop 3,

$$\underset{\text{(inductor)}}{-\frac{dI_2}{dt}} + \underset{\text{(capacitor)}}{160q_3} = 0\,. \tag{13}$$

[The minus sign in (13) arises from taking a clockwise path around loop 3 so that the current passing through the inductor is $-I_2$.] Notice that these three equations are not independent: We can obtain equation (13) by subtracting (11) from (12). Hence, we have only two equations from which to determine the three unknowns I_1, I_2, and q_3. If we now apply Kirchhoff's current law to the two junction points in the network, we find at point A that $I_1 - I_2 - I_3 = 0$ and at point B that $I_2 + I_3 - I_1 = 0$. In both cases, we get

$$I_1 - I_2 - \frac{dq_3}{dt} = 0\,, \tag{14}$$

since $I_3 = dq_3/dt$. Assembling equations (11), (12), and (14) into a system, we have (with $D = d/dt$)

$$DI_2 + 20I_1 = 5\,, \tag{15}$$
$$20I_1 + 160q_3 = 5\,, \tag{16}$$
$$-I_2 + I_1 - Dq_3 = 0\,. \tag{17}$$

We solve these by the elimination method of Section 5.2. Using equation (16) to eliminate I_1 from the others, we are left with

$$DI_2 - 160q_3 = 0\,, \tag{18}$$
$$20I_2 + (20D + 160)q_3 = 5\,. \tag{19}$$

Elimination of I_2 then leads to

$$20D^2q_3 + 160Dq_3 + 3200q_3 = 0\,. \tag{20}$$

To obtain the initial conditions for the second-order equation (20), recall that at time $t = 0$, the charge on the capacitor is 2 coulombs and the current is zero. Hence,

$$q_3(0) = 2\,, \qquad \frac{dq_3}{dt}(0) = 0\,. \tag{21}$$

We can now solve the initial value problem (20)–(21) using the techniques of Chapter 4. Ultimately, we find

$$q_3(t) = 2e^{-4t}\cos 12t + \frac{2}{3}e^{-4t}\sin 12t\,,$$
$$I_3(t) = \frac{dq_3}{dt}(t) = -\frac{80}{3}e^{-4t}\sin 12t\,.$$

Next, to determine I_2, we substitute these expressions into (19) and obtain

$$\begin{aligned} I_2(t) &= \frac{1}{4} - \frac{dq_3}{dt}(t) - 8q_3(t) \\ &= \frac{1}{4} - 16e^{-4t}\cos 12t + \frac{64}{3}e^{-4t}\sin 12t\,. \end{aligned}$$

Finally, from $I_1 = I_2 + I_3$, we get

$$I_1(t) = \frac{1}{4} - 16e^{-4t}\cos 12t - \frac{16}{3}e^{-4t}\sin 12t\,. \quad ◆$$

Note that the differential equations that describe mechanical vibrations and *RLC* series circuits are essentially the same. And, in fact, there is a natural identification of the parameters m, b, and k for a mass–spring system with the parameters L, R, and C that describe circuits. This is illustrated in Table 5.3. Moreover, the terms *transient, steady-state, overdamped, critically damped, underdamped,* and *resonant frequency* described in Sections 4.9 and 4.10 apply to electrical circuits as well.

This analogy between a mechanical system and an electrical circuit extends to large-scale systems and circuits. An interesting consequence of this is the use of **analog simulation** and, in particular, analog computers to analyze mechanical systems. Large-scale mechanical systems are modeled by building a corresponding electrical system and then measuring the charges $q(t)$ and currents $I(t)$.

Although such analog simulations are important, *both* large-scale mechanical and electrical systems are currently modeled using digital computer simulation. This involves the numerical solution of the initial value problem governing the system. Still, the analogy between mechanical and electrical systems means that basically the same computer software can be used to analyze both systems.

TABLE 5.3 Analogy Between Mechanical and Electrical Systems

Mechanical Mass–Spring System with Damping		Electrical *RLC* Series Circuit	
$mx'' + bx' + kx = f(t)$		$Lq'' + Rq' + (1/C)q = E(t)$	
Displacement	x	Charge	q
Velocity	x'	Current	$q' = I$
Mass	m	Inductance	L
Damping constant	b	Resistance	R
Spring constant	k	(Capacitance)$^{-1}$	$1/C$
External force	$f(t)$	Voltage source	$E(t)$

5.7 EXERCISES

1. An *RLC* series circuit has a voltage source given by $E(t) = 20$ V, a resistor of 100 Ω, an inductor of 4 H, and a capacitor of 0.01 F. If the initial current is zero and the initial charge on the capacitor is 4 C, determine the current in the circuit for $t > 0$.
2. An *RLC* series circuit has a voltage source given by $E(t) = 40 \cos 2t$ V, a resistor of 2 Ω, an inductor of $1/4$ H, and a capacitor of $1/13$ F. If the initial current is zero and the initial charge on the capacitor is 3.5 C, determine the charge on the capacitor for $t > 0$.
3. An *RLC* series circuit has a voltage source given by $E(t) = 10 \cos 20t$ V, a resistor of 120 Ω, an inductor of 4 H, and a capacitor of $(2200)^{-1}$ F. Find the steady-state current (solution) for this circuit. What is the resonance frequency of the circuit?
4. An *LC* series circuit has a voltage source given by $E(t) = 30 \sin 50t$ V, an inductor of 2 H, and a capacitor of 0.02 F (but no resistor). What is the current in this circuit for $t > 0$ if at $t = 0$, $I(0) = q(0) = 0$?
5. An *RLC* series circuit has a voltage source of the form $E(t) = E_0 \cos \gamma t$ V, a resistor of 10 Ω, an inductor of 4 H, and a capacitor of 0.01 F. Sketch the frequency response curve for this circuit.
6. Show that when the voltage source in (4) is of the form $E(t) = E_0 \sin \gamma t$, then the steady-state solution I_p is as given in equation (10).
7. A mass–spring system with damping consists of a 7-kg mass, a spring with spring constant 3 N/m, a frictional component with damping constant 2 N-sec/m, and an external force given by $f(t) = 10 \cos 10t$ N. Using a 10-Ω resistor, construct an *RLC* series circuit that is the analog of this mechanical system in the sense that the two systems are governed by the same differential equation.
8. A mass–spring system with damping consists of a 16-lb weight, a spring with spring constant 64 lb/ft, a frictional component with damping constant 10 lb-sec/ft, and an external force given by $f(t) = 20 \cos 8t$ lb. Using an inductor of 0.01 H, construct an *RLC* series circuit that is the analog of this mechanical system.
9. Because of Euler's formula, $e^{i\theta} = \cos\theta + i \sin\theta$, it is often convenient to treat the voltage sources $E_0 \cos \gamma t$ and $E_0 \sin \gamma t$ simultaneously, using $E(t) = E_0 e^{i\gamma t}$. In this case, equation (3) becomes

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = E_0 e^{i\gamma t}, \quad (22)$$

where q is now complex (recall $I = q'$, $I' = q''$).

(a) Use the method of undetermined coefficients to show that the steady-state solution to (22) is

$$q_p(t) = \frac{E_0}{1/C - \gamma^2 L + i\gamma R} e^{i\gamma t}.$$

The technique is discussed in detail in Project F of Chapter 4, page 237.

(b) Now show that the steady-state current is

$$I_p(t) = \frac{E_0}{R + i[\gamma L - 1/(\gamma C)]} e^{i\gamma t}.$$

(c) Use the relation $\alpha + i\beta = \sqrt{\alpha^2 + \beta^2} e^{i\theta}$, where $\tan \theta = \beta/\alpha$, to show that I_p can be expressed in the form

$$I_p(t) = \frac{E_0}{\sqrt{R^2 + [\gamma L - 1/(\gamma C)]^2}} e^{i(\gamma t + \theta)},$$

where $\tan \theta = (1/C - L\gamma^2)/(\gamma R)$.

(d) The imaginary part of $e^{i\gamma t}$ is $\sin \gamma t$, so the imaginary part of the solution to (22) must be the solution to equation (3) for $E(t) = E_0 \sin \gamma t$. Verify that this is also the case for the current by showing that the imaginary part of I_p in part (c) is the same as that given in equation (10).

In Problems 10–13, find a system of differential equations and initial conditions for the currents in the networks given in the schematic diagrams (Figures 5.39–5.42 on pages 294–295). Assume that all initial currents are zero. Solve for the currents in each branch of the network.

10.

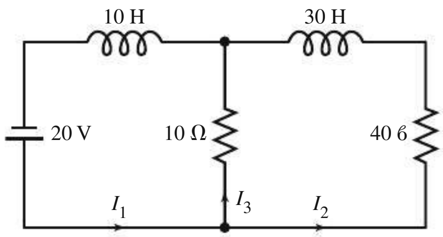


Figure 5.39 *RL* network for Problem 10

11.

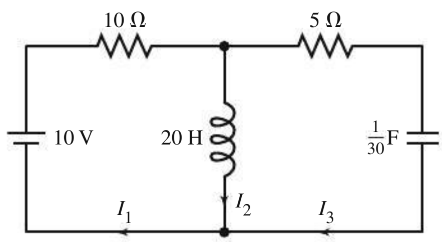


Figure 5.40 *RLC* network for Problem 11

12.

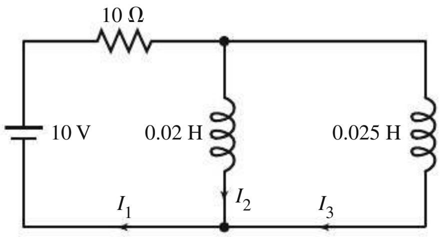


Figure 5.41 *RL* network for Problem 12

13.

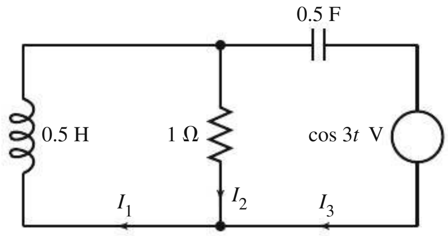


Figure 5.42 *RLC* network for Problem 13

5.8 Dynamical Systems, Poincaré Maps, and Chaos

In this section we take an excursion through an area of mathematics that has received a lot of attention both for the interesting mathematical phenomena being observed and for its application to fields such as meteorology, heat conduction, fluid mechanics, lasers, chemical reactions, and nonlinear circuits, among others. The area is that of nonlinear dynamical systems.[†]

A **dynamical system** is any system that allows one to determine (at least theoretically) the future states of the system given its present or past state. For example, the recursive formula (difference equation)

$$x_{n+1} = (1.05)x_n\,, \qquad n = 0, 1, 2, \ldots$$

is a dynamical system, since we can determine the next state, x_{n+1}, given the previous state, x_n. If we know x_0, then we can compute any future state $\left[\text{indeed, } x_{n+1} = x_0(1.05)^{n+1}\right]$.

Another example of a dynamical system is provided by the differential equation

$$\frac{dx}{dt} = -2x\,,$$

where the solution $x(t)$ specifies the state of the system at "time" t. If we know $x(t_0) = x_0$, then we can determine the state of the system at any future time $t > t_0$ by solving the initial value problem

$$\frac{dx}{dt} = -2x\,, \qquad x(t_0) = x_0\,.$$

Indeed, a simple calculation yields $x(t) = x_0e^{-2(t-t_0)}$ for $t \geq t_0$.

For a dynamical system defined by a differential equation, it is often helpful to work with a related dynamical system defined by a difference equation. For example, when we cannot express the solution to a differential equation using elementary functions, we can use a numerical technique such as the improved Euler's method or Runge–Kutta to approximate the solution to an initial value problem. This numerical scheme defines a new (but related) dynamical system that is often easier to study.

In Section 5.4, we used phase plane diagrams to study autonomous systems in the plane. Many important features of the system can be detected just by looking at these diagrams. For example, a closed trajectory corresponds to a periodic solution. The trajectories for *non*autonomous systems in

[†]For a more detailed study of dynamical systems, see *An Introduction to Chaotic Dynamical Systems*, by R. L. Devaney (Westview Press, 2003) and *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, by J. Guckenheimer and P. J. Holmes (Springer-Verlag, New York, 1983).

the phase plane are much more complicated to decipher. One technique that is helpful in this regard is the so-called **Poincaré map.** As we will see, these maps replace the study of a nonautonomous system with the study of a dynamical system defined by the location in the xv-plane $(v = dx/dt)$ of the solution at regularly spaced moments in time such as $t = 2\pi n$, where $n = 0, 1, 2, \ldots$. The advantage in using the Poincaré map will become clear when the method is applied to a nonlinear problem for which no explicit solution is known. In such a case, the trajectories are computed using a numerical scheme such as Runge–Kutta. Several software packages have options that will construct Poincaré maps for a given system.

To illustrate the Poincaré map, consider the equation

$$x''(t) + \omega^2 x(t) = F\cos t\,, \tag{1}$$

where F and ω are positive constants. We studied similar equations in Section 4.10 and found that a general solution for $\omega \neq 1$ is given by

$$x(t) = A\sin(\omega t + \phi) + \frac{F}{\omega^2 - 1}\cos t\,, \tag{2}$$

where the amplitude A and the phase angle ϕ are arbitrary constants. Since $v = x'$,

$$v(t) = \omega A\cos(\omega t + \phi) - \frac{F}{\omega^2 - 1}\sin t\,.$$

Because the forcing function $F\cos t$ is 2π-periodic, it is natural to seek 2π-periodic solutions to (1). For this purpose, we define the Poincaré map as

$$\begin{aligned} x_n &:= x(2\pi n) = A\sin(2\pi\omega n + \phi) + F/(\omega^2 - 1)\,, \\ v_n &:= v(2\pi n) = \omega A\cos(2\pi\omega n + \phi)\,, \end{aligned} \tag{3}$$

for $n = 0, 1, 2, \ldots$. In Figure 5.43 on page 297, we plotted the first 100 values of (x_n, v_n) in the xv-plane for different choices of ω. For simplicity, we have taken $A = F = 1$ and $\phi = 0$. These graphs are called **Poincaré sections.** We will interpret them shortly.

Now let's play the following game. We agree to ignore the fact that we already know the formula for $x(t)$ for all $t \geq 0$. We want to see what information about the solution we can glean just from the Poincaré section and the form of the differential equation.

Notice that the first two Poincaré sections in Figure 5.43, corresponding to $\omega = 2$ and 3, consist of a single point. This tells us that, starting with $t = 0$, every increment 2π of t returns us to the same point in the phase plane. This in turn implies that equation (1) has a 2π-periodic solution, which can be proved as follows: For $\omega = 2$, let $x(t)$ be the solution to (1) with $\big(x(0), v(0)\big) = (1/3, 2)$ and let $X(t) := x(t + 2\pi)$. Since the Poincaré section is just the point $(1/3, 2)$, we have $X(0) = x(2\pi) = 1/3$ and $X'(0) = x'(2\pi) = 2$. Thus, $x(t)$ and $X(t)$ have the same initial values at $t = 0$. Further, because $\cos t$ is 2π-periodic, we also have

$$X''(t) + \omega^2 X(t) = x''(t + 2\pi) + \omega^2 x(t + 2\pi) = \cos(t + 2\pi) = \cos t\,.$$

Consequently, $x(t)$ and $X(t)$ satisfy the same initial value problem. By the uniqueness theorems of Sections 4.2 and 4.5, these functions must agree on the interval $[0, \infty)$. Hence, $x(t) = X(t) = x(t + 2\pi)$ for all $t \geq 0$; that is, $x(t)$ is 2π-periodic. (The same reasoning works for $\omega = 3$.) With a similar argument, it follows from the Poincaré section for $\omega = 1/2$ that there is a solution of period 4π that alternates between the two points displayed in Figure 5.43(c) as t is incremented by 2π. For the case $\omega = 1/3$, we deduce that there is a solution of period 6π rotating among three points, and for $\omega = 1/4$, there is an 8π-periodic solution rotating among four points. We call these last three solutions **subharmonics.**

The case $\omega = \sqrt{2}$ is different. So far, in Figure 5.43(f), none of the points has repeated. Did we stop too soon? Will the points ever repeat? Here, the fact that $\sqrt{2}$ is irrational plays

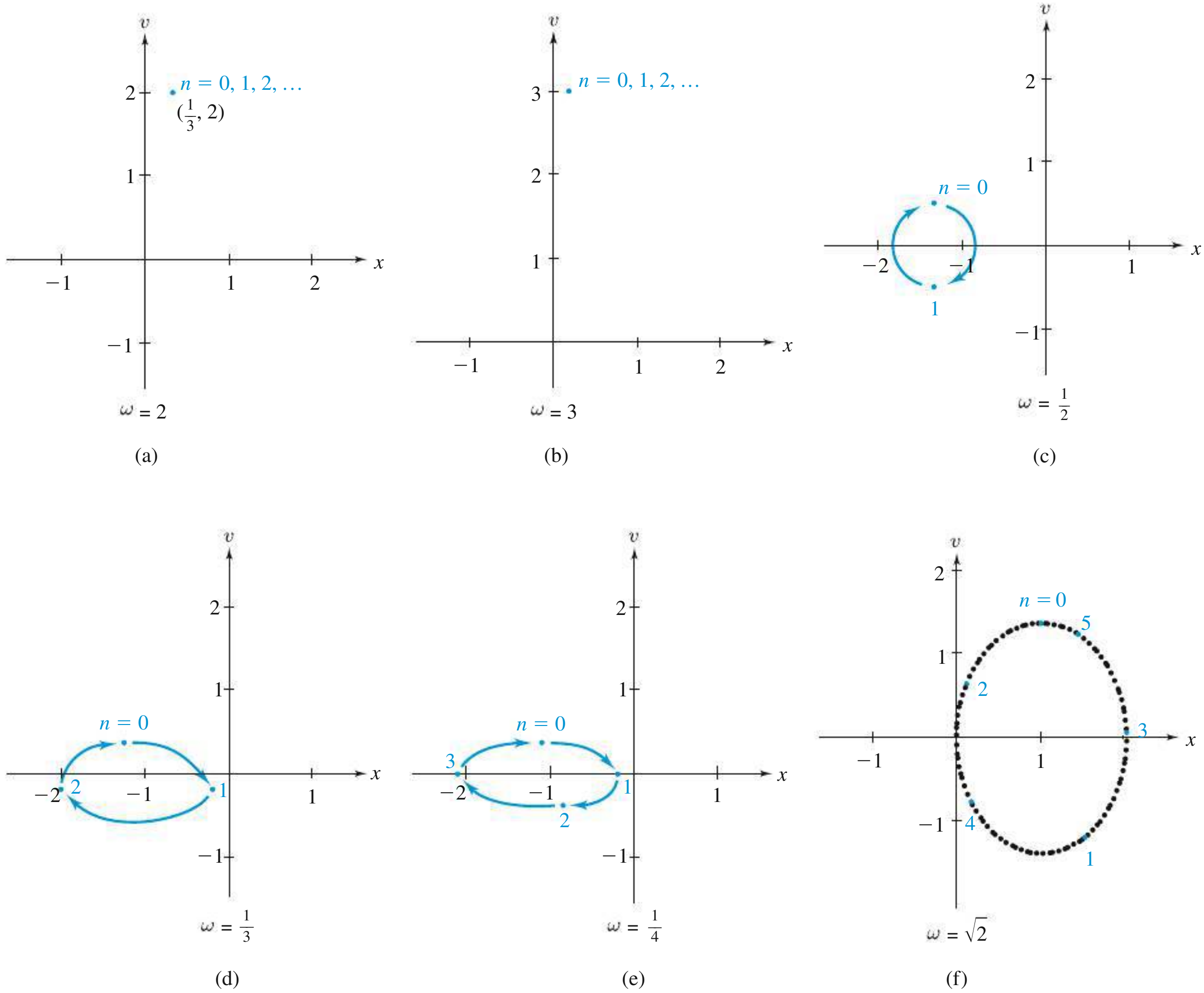


Figure 5.43 Poincaré sections for equation (1) for various values of ω

a crucial role. It turns out that every integer n yields a distinct point in the Poincaré section (see Problem 8). However, there is a pattern developing. The points all appear to lie on a simple curve, possibly an ellipse. To see that this is indeed the case, notice that when $\omega = \sqrt{2}$, $A = F = 1$, and $\phi = 0$, we have

$$x_n = \sin\left(2\sqrt{2}\pi n\right) + 1\,, \qquad v_n = \sqrt{2}\cos\left(2\sqrt{2}\pi n\right)\,, \qquad n = 0, 1, 2, \ldots .$$

It is then an easy computation to show that each (x_n, v_n) lies on the ellipse

$$(x-1)^2 + \frac{v^2}{2} = 1\,.$$

In our investigation of equation (1), we concentrated on 2π-periodic solutions because the forcing term $F\cos t$ has period 2π. [We observed subharmonics when $\omega = 1/2, 1/3$, and $1/4$—that is, solutions with periods $2(2\pi), 3(2\pi)$, and $4(2\pi)$.] When a damping term is introduced into the differential equation, the Poincaré map displays a different behavior. Recall that the solution will now be the sum of a transient and a steady-state term. For example, let's analyze the equation

$$x''(t) + bx'(t) + \omega^2 x(t) = F\cos t\,, \tag{4}$$

where b, F, and ω are positive constants.

When $b^2 < 4\omega^2$, the solution to (4) can be expressed as

$$x(t) = Ae^{-(b/2)t}\sin\left(\frac{\sqrt{4\omega^2 - b^2}}{2}t + \phi\right) + \frac{F}{\sqrt{(\omega^2-1)^2 + b^2}}\sin(t+\theta)\,, \tag{5}$$

where $\tan\theta = (\omega^2 - 1)/b$ and A and ϕ are arbitrary constants [compare equations (7) and (8) in Section 4.10]. The first term on the right-hand side of (5) is the transient and the second term, the steady-state solution. Let's construct the Poincaré map using $t = 2\pi n$, $n = 0, 1, 2, \ldots$. We will take $b = 0.22$, $\omega = A = F = 1$, and $\phi = 0$ to simplify the computations. Because $\tan\theta = (\omega^2 - 1)/b = 0$, we will take $\theta = 0$ as well. Then we have

$$\begin{aligned} x(2\pi n) &= x_n = e^{-0.22\pi n}\sin\left(\sqrt{0.9879}\,2\pi n\right)\,, \\ x'(2\pi n) &= v_n = -0.11e^{-0.22\pi n}\sin\left(\sqrt{0.9879}\,2\pi n\right) \\ &\quad + \sqrt{0.9879}\,e^{-0.22\pi n}\cos\left(\sqrt{0.9879}\,2\pi n\right) + \frac{1}{(0.22)}\,. \end{aligned}$$

The Poincaré section for $n = 0, 1, 2, \ldots, 10$ is shown in Figure 5.44 (black points). After just a few iterations, we observe that $x_n \approx 0$ and $v_n \approx 1/(0.22) \approx 4.545$; that is, the points of the Poincaré section are approaching a single point in the xv-plane (colored point). Thus, we might expect that there is a 2π-periodic solution corresponding to a particular choice of A and ϕ. [In this example, where we can explicitly represent the solution, we see that indeed a 2π-periodic solution arises when we take $A = 0$ in (5).]

There is an important difference between the Poincaré sections for equation (1) and those for equation (4). In Figure 5.43, the location of all of the points in (a)–(e) depends on the initial value selected (here $A = 1$ and $\phi = 0$). (See Problem 10.) However, in Figure 5.44, the first few points (black points) depend on the initial conditions, while the limit point (colored point) does *not* (see Problem 6). The latter behavior is typical for equations that have a "damping"

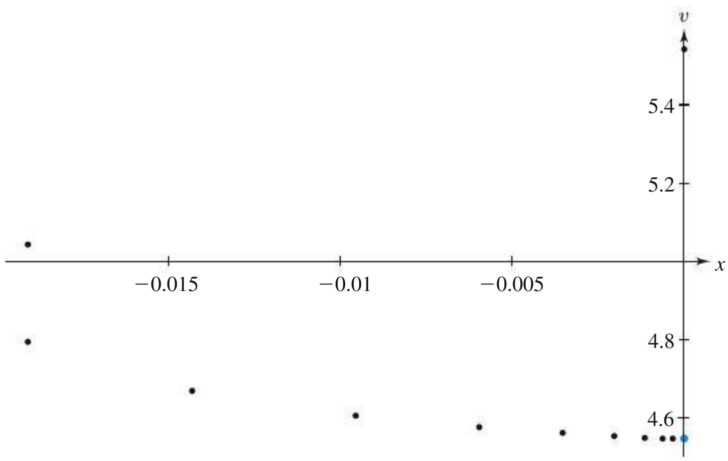
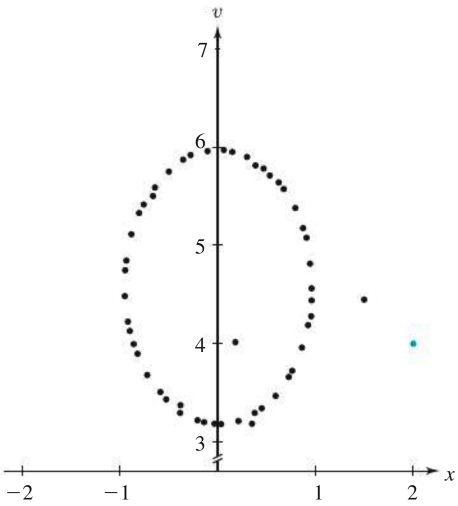
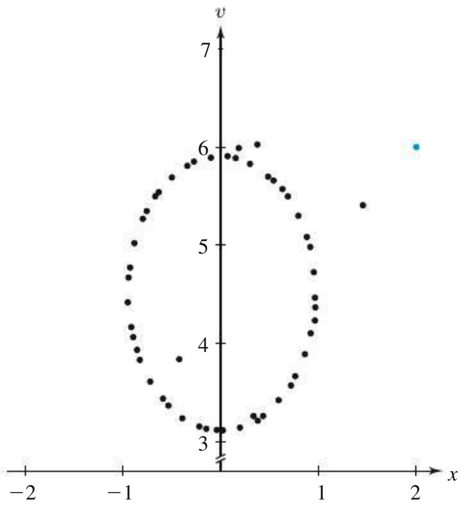


Figure 5.44 Poincaré section for equation (4) with $F = 1$, $b = 0.22$, and $\omega = 1$



(a) $x_0 = 2$, $v_0 = 4$



(b) $x_0 = 2$, $v_0 = 6$

Figure 5.45 Poincaré section for equation (6) with initial values x_0, v_0

term (i.e., $b > 0$); namely, the Poincaré section has a limit set† that is essentially independent of the initial conditions.

For equations with damping, the limit set may be more complicated than just a point. For example, the Poincaré map for the equation

$$x''(t) + (0.22)x'(t) + x(t) = \cos t + \cos\left(\sqrt{2}t\right) \tag{6}$$

has a limit set consisting of an ellipse (see Problem 11). This is illustrated in Figure 5.45 for the initial values $x_0 = 2$, $v_0 = 4$ and $x_0 = 2$, $v_0 = 6$.

So far we have seen limit sets for the Poincaré map that were either a single point or an ellipse—independent of the initial values. These particular limit sets are **attractors.** In general, an attractor is a set A with the property that there exists an open set‡ B containing A such that whenever the Poincaré map enters B, its points remain in B *and* the limit set of the Poincaré map is a subset of A. Further, we assume A has the *invariance property:* Whenever the Poincaré map starts at a point in A, it remains in A.

In the previous examples, the attractors of the dynamical system (Poincaré map) were easy to describe. In recent years, however, many investigators, working on a variety of applications, have encountered dynamical systems that do *not* behave orderly—their attractor sets are very complicated (not just isolated points or familiar geometric objects such as ellipses). The behavior of such systems is called **chaotic,** and the corresponding limit sets are referred to as **strange attractors.**

†The **limit set** for a map (x_n, v_n), $n = 1, 2, 3, \ldots$, is the set of points (p, q) such that $\lim_{k\to\infty}(x_{n_k}, v_{n_k}) = (p, q)$, where $n_1 < n_2 < n_3 < \ldots$ is some subsequence of the positive integers.

‡A set $B \subset \mathbf{R}^2$ is an **open set** if for each point $p \in B$ there is an open disk V containing p such that $V \subset B$.

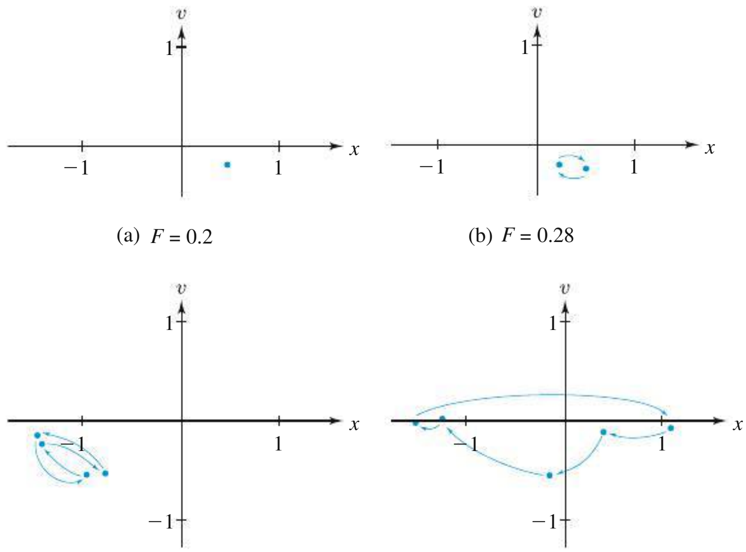


Figure 5.46 Poincaré sections for the Duffing equation (7) with $b = 0.3$ and $\gamma = 1.2$

To illustrate chaotic behavior and what is meant by a strange attractor, we discuss two nonlinear differential equations and a simple difference equation. First, let's consider the **forced Duffing equation**

$$x''(t) + bx'(t) - x(t) + x^3(t) = F \sin \gamma t \,. \tag{7}$$

We cannot express the solution to (7) in any explicit form, so we must obtain the Poincaré map by numerically approximating the solution to (7) for fixed initial values and then plot the approximations for $x(2\pi n/\gamma)$ and $v(2\pi n/\gamma) = x'(2\pi n/\gamma)$. (Because the forcing term $F \sin \gamma t$ has period $2\pi/\gamma$, we seek $2\pi/\gamma$-periodic solutions and subharmonics.) In Figure 5.46, we display the limit sets (attractors) when $b = 0.3$ and $\gamma = 1.2$ in the cases (a) $F = 0.2$, (b) $F = 0.28$, (c) $F = 0.29$, and (d) $F = 0.37$.

Notice that as the constant F increases, the Poincaré map changes character. When $F = 0.2$, the Poincaré section tells us that there is a $2\pi/\gamma$-periodic solution. For $F = 0.28$, there is a subharmonic of period $4\pi/\gamma$, and for $F = 0.29$ and 0.37, there are subharmonics with periods $8\pi/\gamma$ and $10\pi/\gamma$, respectively.

Things are dramatically different when $F = 0.5$: The solution is neither $2\pi/\gamma$-periodic nor subharmonic. The Poincaré section for $F = 0.5$ is illustrated in Figure 5.47 on page 301. This section was generated by numerically approximating the solution to (7) when $\gamma = 1.2$, $b = 0.3$, and $F = 0.5$, for fixed initial values.[†] Not all of the approximations $x(2\pi n/\gamma)$ and $v(2\pi n/\gamma)$ that were calculated are graphed; because of the presence of a transient solution, the first few points were omitted. It turns out that the plotted set is essentially independent of the initial values and has the property that once a point is in the set, all subsequent points lie in the set.

[†] *Historical Footnote:* When researchers first encountered these strange-looking Poincaré sections, they would check their computations using different computers and different numerical schemes [see Hénon and Heiles, "The Applicability of the Third Integral of Motion: Some Numerical Experiments," *Astronomical Journal*, Vol. 69 (1964): 75]. For special types of dynamical systems, such as the Hénon map, it can be shown that there exists a true trajectory that *shadows* the numerical trajectory [see M. Hammel, J. A. Yorke, and C. Grebogi, "Numerical Orbits of Chaotic Processes Represent True Orbits," *Bulletin American Mathematical Society*, Vol. 19 (1988): 466–469].

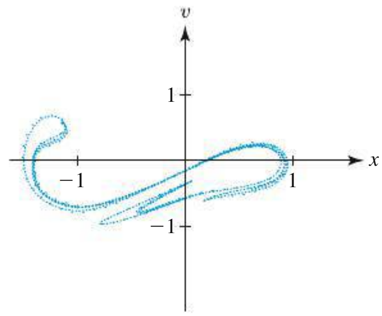


Figure 5.47 Poincaré section for the Duffing equation (7) with $b = 0.3$, $\gamma = 1.2$, and $F = 0.5$

Because of the complicated shape of the set, it is indeed a strange attractor. While the shape of the strange attractor does not depend on the initial values, the picture does change if we consider different sections; for example, $t = (2\pi n + \pi/2)/\gamma$, $n = 0, 1, 2, \ldots$ yields a different configuration.

Another example of a strange attractor occurs when we consider the **forced pendulum equation**

$$x''(t) + bx'(t) + \sin\big(x(t)\big) = F\cos t\,, \tag{8}$$

where the $x(t)$ term in (4) has been replaced by $\sin\big(x(t)\big)$. Here $x(t)$ is the angle between the pendulum and the vertical rest position, b is related to damping, and F represents the strength of the forcing function (see Figure 5.48). For $F = 2.7$ and $b = 0.22$, we have graphed in Figure 5.49 approximately 90,000 points in the Poincaré map. Since we cannot express the solution to (8) in any explicit form, the Poincaré map was obtained by numerically approximating the solution to (8) for fixed initial values and plotting the approximations for $x(2\pi n)$ and $v(2\pi n) = x'(2\pi n)$.

The Poincaré maps for the forced Duffing equation and for the forced pendulum equation not only illustrate the idea of a strange attractor; they also exhibit another peculiar behavior called **chaos.** Chaos occurs when small changes in the initial conditions lead to major changes in the behavior of the solution. Henri Poincaré described the situation as follows:

> It may happen that small differences in the initial conditions will produce very large ones in the final phenomena. A small error in the former produces an enormous error in the latter. Prediction becomes impossible . . . .

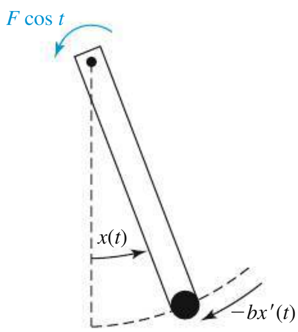


Figure 5.48 Forced damped pendulum

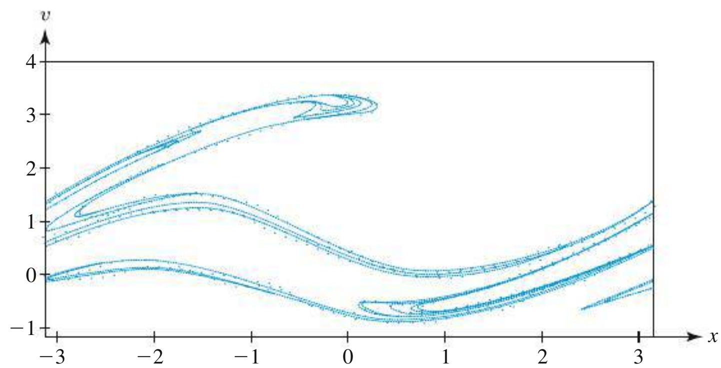


Figure 5.49 Poincaré section for the forced pendulum equation (8) with $b = 0.22$ and $F = 2.7$

In a physical experiment, we can never *exactly* (with infinite accuracy) reproduce the same initial conditions. Consequently, if the behavior is chaotic, even a slight difference in the initial conditions may lead to quite different values for the corresponding Poincaré map when n is large. Such behavior does not occur for solutions to either equation (4) or equation (1) (see Problems 6 and 7). However, two solutions to the Duffing equation (7) with $F = 0.5$ that correspond to two different but close initial values have Poincaré maps that do *not* remain close together. Although they both are attracted to the same set, their locations with respect to this set may be relatively far apart.

The phenomenon of chaos can also be illustrated by the following simple map. Let x_0 lie in $[0, 1)$ and define

(9) $$x_{n+1} = 2x_n \ (\text{mod } 1)\,,$$

where by (mod 1) we mean the decimal part of the number if it is greater than or equal to 1; that is

$$x_{n+1} = \begin{cases} 2x_n\,, & \text{for } 0 \le x_n < 1/2\,, \\ 2x_n - 1\,, & \text{for } 1/2 \le x_n < 1\,. \end{cases}$$

When $x_0 = 1/3$, we find

$$\begin{aligned} x_1 &= 2\cdot(1/3)\ (\text{mod } 1) = 2/3\,, \\ x_2 &= 2\cdot(2/3)\ (\text{mod } 1) = 4/3\ (\text{mod } 1) = 1/3\,, \\ x_3 &= 2\cdot(1/3)\ (\text{mod } 1) = 2/3\,, \\ x_4 &= 2\cdot(2/3)\ (\text{mod } 1) = 1/3\,, \quad \text{etc.} \end{aligned}$$

Written as a sequence, we get $\{1/3, 2/3, \overline{1/3, 2/3}, \ldots\}$, where the overbar denotes the repeated pattern.

What happens when we pick a starting value x_0 near $1/3$? Does the sequence cluster about $1/3$ and $2/3$ as does the mapping when $x_0 = 1/3$? For example, when $x_0 = 0.3$, we get the sequence

$$\{0.3, 0.6, 0.2, 0.4, 0.8, \overline{0.6, 0.2, 0.4, 0.8}, \ldots\}\,.$$

In Figure 5.50, we have plotted the values of x_n for $x_0 = 0.3$, 0.33, and 0.333. We have not plotted the first few terms, but only those that repeat. (This omission of the first few terms parallels the situation depicted in Figure 5.47 on page 301, where transient solutions arise.)

It is clear from Figure 5.50 that while the values for x_0 are getting closer to $1/3$, the corresponding maps are spreading out over the whole interval $[0, 1]$ and *not* clustering near $1/3$ and $2/3$. This behavior is chaotic, since the Poincaré maps for initial values near $1/3$ behave quite differently from the map for $x_0 = 1/3$. If we had selected x_0 to be irrational (which we can't do with a calculator), the sequence would *not* repeat and would be dense in $[0, 1]$.

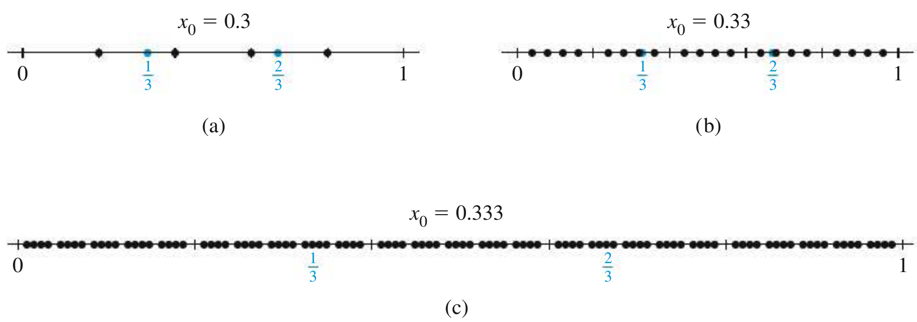


Figure 5.50 Plots of the map $x_{n+1} = 2x_n$ (mod 1) for $x_0 = 0.3$, 0.33, and 0.333

Systems that exhibit chaotic behavior arise in many applications. The challenge to engineers is to design systems that avoid this chaos and, instead, enjoy the property of **stability.** The topic of stable systems is discussed at length in Chapter 12.†

5.8 EXERCISES

A software package that supports the construction of Poincaré maps is required for the problems in this section.

1. Compute and graph the points of the Poincaré map with $t = 2\pi n$, $n = 0, 1, \ldots, 20$ for equation (1), taking $A = F = 1, \phi = 0$, and $\omega = 3/2$. Repeat, taking $\omega = 3/5$. Do you think the equation has a 2π-periodic solution for either choice of ω? A subharmonic solution?

2. Compute and graph the points of the Poincaré map with $t = 2\pi n$, $n = 0, 1, \ldots, 20$ for equation (1), taking $A = F = 1, \phi = 0$, and $\omega = 1/\sqrt{3}$. Describe the limit set for this system.

3. Compute and graph the points of the Poincaré map with $t = 2\pi n$, $n = 0, 1, \ldots, 20$ for equation (4), taking $A = F = 1$, $\phi = 0$, $\omega = 1$, and $b = -0.1$. What is happening to these points as $n \to \infty$?

4. Compute and graph the Poincaré map with $t = 2\pi n$, $n = 0, 1, \ldots, 20$ for equation (4), taking $A = F = 1$, $\phi = 0$, $\omega = 1$, and $b = 0.1$. Describe the attractor for this system.

5. Compute and graph the Poincaré map with $t = 2\pi n$, $n = 0, 1, \ldots, 20$ for equation (4), taking $A = F = 1$, $\phi = 0$, $\omega = 1/3$, and $b = 0.22$. Describe the attractor for this system.

6. Show that for $b > 0$, the Poincaré map for equation (4) is not chaotic by showing that as t gets large

$$x_n = x(2\pi n) \approx \frac{F}{\sqrt{(\omega^2 - 1)^2 + b^2}} \sin(2\pi n + \theta) ,$$

$$v_n = x'(2\pi n) \approx \frac{F}{\sqrt{(\omega^2 - 1)^2 + b^2}} \cos(2\pi n + \theta)$$

independent of the initial values $x_0 = x(0)$ and $v_0 = x'(0)$.

7. Show that the Poincaré map for equation (1) is not chaotic by showing that if (x_0, v_0) and (x_0^*, v_0^*) are two initial values that define the Poincaré maps $\{(x_n, v_n)\}$ and $\{(x_n^*, v_n^*)\}$, respectively, using the recursive formulas in (3), then one can make the distance between (x_n, v_n) and (x_n^*, v_n^*) small by making the distance between (x_0, v_0) and (x_0^*, v_0^*) small. [*Hint:* Let (A, ϕ) and (A^*, ϕ^*) be the polar coordinates of two points in the plane. From the law of cosines, it follows that the distance d between them is given by $d^2 = (A - A^*)^2 + 2AA^*[1 - \cos(\phi - \phi^*)]$.]

8. Consider the Poincaré maps defined in (3) with $\omega = \sqrt{2}$, $A = F = 1$, and $\phi = 0$. If this map were ever to repeat, then for two distinct positive integers n and m, $\sin(2\sqrt{2}\pi n) = \sin(2\sqrt{2}\pi m)$. Using basic properties of the sine function, show that this would imply that $\sqrt{2}$ is rational. It follows from this contradiction that the points of the Poincaré map do not repeat.

9. The doubling modulo 1 map defined by equation (9) exhibits some fascinating behavior. Compute the sequence obtained when

(a) $x_0 = k/7$ for $k = 1, 2, \ldots, 6$.

(b) $x_0 = k/15$ for $k = 1, 2, \ldots, 14$.

(c) $x_0 = k/2^j$, where j is a positive integer and $k = 1, 2, \ldots, 2^j - 1$.

Numbers of the form $k/2^j$ are called **dyadic numbers** and are dense in $[0, 1]$. That is, there is a dyadic number arbitrarily close to any real number (rational or irrational).

10. To show that the limit set of the Poincaré map given in (3) depends on the initial values, do the following:

(a) Show that when $\omega = 2$ or 3, the Poincaré map consists of the single point

$$(x, v) = \left(A \sin\phi + \frac{F}{\omega^2 - 1},\quad \omega A \cos\phi\right).$$

(b) Show that when $\omega = 1/2$, the Poincaré map alternates between the two points

$$\left(\frac{F}{\omega^2 - 1} \pm A \sin\phi\ ,\ \pm\ \omega A \cos\phi\right).$$

(c) Use the results of parts (a) and (b) to show that when $\omega =$ 2, 3, or $1/2$, the Poincaré map (3) depends on the initial values (x_0, v_0).

†All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems*, 7th ed.

11. To show that the limit set for the Poincaré map $x_n := x(2\pi n)$, $v_n := x'(2\pi n)$, where $x(t)$ is a solution to equation (6), is an ellipse and that this ellipse is the same for any initial values x_0, v_0, do the following:

(a) Argue that since the initial values affect only the transient solution to (6), the limit set for the Poincaré map is independent of the initial values.

(b) Now show that for n large,

$$x_n \approx a \sin(2\sqrt{2}\pi n + \psi)\,,$$
$$v_n \approx c + \sqrt{2}a \cos(2\sqrt{2}\pi n + \psi)\,,$$

where $a = \left(1 + 2(0.22)^2\right)^{-1/2}$, $c = (0.22)^{-1}$, and $\psi = \arctan\left\{-\left[(0.22)\sqrt{2}\right]^{-1}\right\}$.

(c) Use the result of part (b) to conclude that the ellipse

$$x^2 + \frac{(v-c)^2}{2} = a^2$$

contains the limit set of the Poincaré map.

12. Using a numerical scheme such as Runge–Kutta or a software package, calculate the Poincaré map for equation (7) when $b = 0.3$, $\gamma = 1.2$, and $F = 0.2$. (Notice that the closer you start to the limiting point, the sooner the transient part will die out.) Compare your map with Figure 5.46(a) on page 300. Redo for $F = 0.28$.

13. Redo Problem 12 with $F = 0.31$. What kind of behavior does the solution exhibit?

14. Redo Problem 12 with $F = 0.65$. What kind of behavior does the solution exhibit?

15. Chaos Machine. Chaos can be illustrated using a long ruler, a short ruler, a pin, and a tie tack (pivot). Construct the double pendulum as shown in Figure 5.51(a). The pendulum is set in motion by releasing it from a position such as the one shown in Figure 5.51(b). Repeatedly set the pendulum in motion, each time trying to release it from the same position. Record the number of times the short ruler flips over and the direction in which it was moving. If the pendulum was released in *exactly* the same position each time, then the motion would be the same. However, from your experiments you will observe that even beginning close to the same position leads to very different motions. This double pendulum exhibits chaotic behavior.

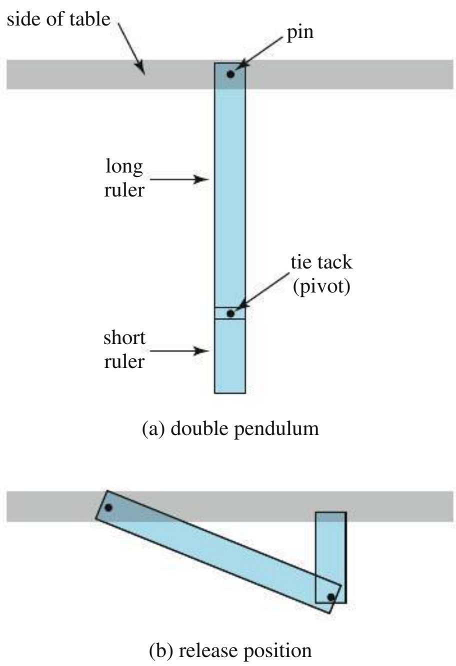


Figure 5.51 Double pendulum as a chaos machine

Chapter 5 Summary

Systems of differential equations arise in the study of biological systems, coupled mass–spring oscillators, electrical circuits, ecological models, and many other areas.

Linear systems with constant coefficients can be solved explicitly using a variant of the **Gauss elimination process.** For this purpose we begin by writing the system with operator notation, using $D := d/dt$, $D^2 := d^2/dt^2$, and so on. A system of two equations in two unknown functions then takes the form

$$L_1[x] + L_2[y] = f_1\,, \qquad L_3[x] + L_4[y] = f_2\,,$$

where L_1, L_2, L_3, and L_4 are polynomial expressions in D. Applying L_4 to the first equation, L_2 to the second, and subtracting, we get a single (typically higher-order) equation in $x(t)$, namely,

$$(L_4L_1 - L_2L_3)[x] = L_4[f_1] - L_2[f_2]\,.$$

We then solve this constant coefficient equation for $x(t)$. Similarly, we can eliminate x from the system to obtain a single equation for $y(t)$, which we can also solve. This procedure introduces some extraneous constants, but by substituting the expressions for x and y back into one of the original equations, we can determine the relationships among these constants.

A preliminary step for the application of numerical algorithms for solving systems or single equations of higher order is to rewrite them as an equivalent system of first-order equations in **normal form:**

$$\begin{aligned} x_1'(t) &= f_1(t, x_1, x_2, \ldots, x_m)\,, \\ x_2'(t) &= f_2(t, x_1, x_2, \ldots, x_m)\,, \\ &\vdots \\ x_m'(t) &= f_m(t, x_1, x_2, \ldots, x_m)\,. \end{aligned} \tag{1}$$

For example, by setting $v = y'$, we can rewrite the second-order equation $y'' = f(t, y, y')$ as the normal system

$$\begin{aligned} y' &= v\,, \\ v' &= f(t, y, v)\,. \end{aligned}$$

The normal system (1) has the outward appearance of a vectorized version of a single first-order equation, and as such it suggests how to generalize numerical algorithms such as those of Euler and Runge–Kutta.

A technique for studying the qualitative behavior of solutions to the **autonomous system**

$$\frac{dx}{dt} = f(x, y)\,, \qquad \frac{dy}{dt} = g(x, y) \tag{2}$$

is phase plane analysis. We begin by finding the **critical points** of (2)—namely, points (x_0, y_0) where

$$f(x_0, y_0) = 0 \quad \text{and} \quad g(x_0, y_0) = 0\,.$$

The corresponding constant solution pairs $x(t) \equiv x_0$, $y(t) \equiv y_0$ are called **equilibrium solutions** to (2). We then sketch the direction field for the **related phase plane differential equation**

$$\frac{dy}{dx} = \frac{g(x, y)}{f(x, y)} \tag{3}$$

with appropriate direction arrows (oriented by the sign of dx/dt or dy/dt). From this we can usually conjecture qualitative features of the solutions and of the critical points, such as stability and asymptotic behavior. Software is typically used to visualize the solution curves to (3), which provide the highways along which the **trajectories** of the system (2) travel.

Nonautonomous systems can be studied by considering a **Poincaré map** for the system. A Poincaré map can be used to detect periodic and subharmonic solutions and to study systems whose solutions exhibit chaotic behavior.

REVIEW PROBLEMS FOR CHAPTER 5

In Problems 1–4, find a general solution $x(t)$, $y(t)$ for the given system.

1. $x' + y'' + y = 0\,,$
$x'' + y' = 0$

2. $x' = x + 2y\,,$
$y' = -4x - 3y$

3. $2x' - y' = y + 3x + e^t\,,$
$3y' - 4x' = y - 15x + e^{-t}$

4. $x'' + x - y'' = 2e^{-t}\,,$
$x'' - x + y'' = 0$

In Problems 5 and 6, solve the given initial value problem.

5. $x' = z - y\,;\quad x(0) = 0\,,$
$y' = z\,;\quad y(0) = 0\,,$
$z' = z - x\,;\quad z(0) = 2$

6. $x' = y + z\,;\quad x(0) = 2\,,$
$y' = x + z\,;\quad y(0) = 2\,,$
$z' = x + y\,;\quad z(0) = -1$

7. For the interconnected tanks problem of Section 5.1, page 241, suppose that instead of pure water being fed into tank A, a brine solution with concentration 0.2 kg/L is used; all other data remain the same. Determine the mass of salt in each tank at time t if the initial masses are $x_0 = 0.1$ kg and $y_0 = 0.3$ kg.

In Problems 8–11, write the given higher-order equation or system in an equivalent normal form (compare Section 5.3).

8. $2y'' - ty' + 8y = \sin t$

9. $3y''' + 2y' - e^t y = 5$

10. $x'' - x + y = 0\,,$
$x' - y + y'' = 0$

11. $x''' + y' + y'' = t\,,$
$x'' - x' + y''' = 0$

In Problems 12 and 13, solve the related phase plane equation for the given system. Then sketch by hand several representative trajectories (with their flow arrows) and describe the stability of the critical points (i.e., compare with Figure 5.12, page 267).

12. $x' = y - 2\,,$
$y' = 2 - x$

13. $x' = 4 - 4y\,,$
$y' = -4x$

14. Find all the critical points and determine the phase plane solution curves for the system

$$\frac{dx}{dt} = \sin x \cos y\,,$$
$$\frac{dy}{dt} = \cos x \sin y\,.$$

In Problems 15 and 16, sketch some typical trajectories for the given system, and by comparing with Figure 5.12, page 267, identify the type of critical point at the origin.

15. $x' = -2x - y\,,$
$y' = 3x - y$

16. $x' = -x + 2y\,,$
$y' = x + y$

17. In the electrical circuit of Figure 5.52, take $R_1 = R_2 = 1\ \Omega$, $C = 1$ F, and $L = 1$ H. Derive three equations for the unknown currents I_1, I_2, and I_3 by writing Kirchhoff's voltage law for loops 1 and 2, and Kirchhoff's current law for the top juncture. Find the general solution.

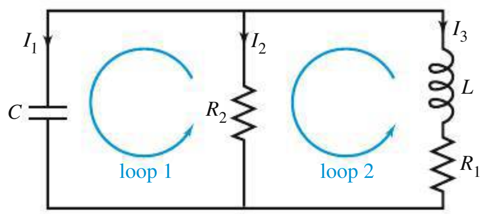


Figure 5.52 Electrical circuit for Problem 17

18. In the coupled mass–spring system depicted in Figure 5.26, page 283, take each mass to be 1 kg and let $k_1 = 8$ N/m, while $k_2 = 3$ N/m. What are the natural angular frequencies of the system? What is the general solution?

Projects for Chapter 5

A Designing a Landing System for Interplanetary Travel

Courtesy of Alfred Clark, Jr., Professor Emeritus, University of Rochester, Rochester, NY

You are a second-year Starfleet Academy Cadet aboard the U.S.S. *Enterprise* on a continuing study of the star system Glia. The object of study on the present expedition is the large airless planet Glia-4. A class 1 sensor probe of mass m is to be sent to the planet's surface to collect data. The probe has a modifiable landing system so that it can be used on planets of different gravity. The system consists of a linear spring (force $= -kx$, where x is displacement), a nonlinear spring (force $= -ax^3$), and a shock-damper (force $= -b\dot{x}$),† all in parallel. Figure 5.53 shows a schematic of the system. During the landing process, the probe's thrusters are used to create a constant rate of descent. The velocity at impact varies; the symbol V_L is used to denote the largest velocity likely to happen in practice. At the instant of impact, (1) the thrust is turned off, and (2) the suspension springs are at their unstretched natural length.

(a) Let the displacement x be measured from the unstretched length of the springs and be taken negative downward (i.e., compression gives a negative x). Show that the equation governing the oscillations after impact is

$$m\ddot{x} + b\dot{x} + kx + ax^3 = -mg\ .$$

(b) The probe has a mass $m = 1220$ kg. The linear spring is permanently installed and has a stiffness $k = 35{,}600$ N/m. The gravity on the surface of Glia-4 is $g = 17.5$ m/sec^2. The nonlinear spring is removable; an appropriate spring must be chosen for each mission. These nonlinear springs are made of coralidium, a rare and difficult-to-fabricate alloy. Therefore, the *Enterprise* stocks only four different strengths: $a = 150{,}000$, $300{,}000$, $450{,}000$, and $600{,}000$ N/m^3. Determine which springs give a compression as close as

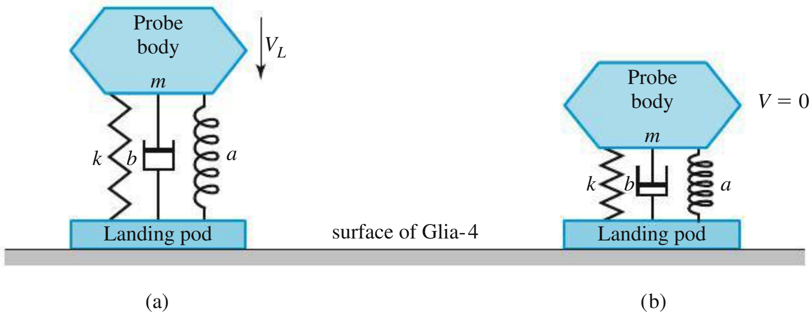


Figure 5.53 Schematic of the probe landing system. (a) The system at the instant of impact. The springs are not stretched or compressed, the thrusters have been turned off, and the velocity is V_L downward. (b) The probe has reached a state of rest on the surface, and the springs are compressed enough to support the weight. Between states (a) and (b), the probe oscillates relative to the landing pod.

†The symbol $\dot{x}$ denotes dx/dt.

possible to 0.3 m without exceeding 0.3 m, when the ship is resting on the surface of Glia-4. (The limit of 0.3 m is imposed by unloading clearance requirements.)

(c) The other adjustable component on the landing system is the linear shock-damper, which may be adjusted in increments of $\Delta b = 500$ N-sec/m, from a low value of 1000 N-sec/m to a high value of 10,000 N-sec/m. It is desirable to make b as small as possible because a large b produces large forces at impact. However, if b is too small, there is some danger that the probe will rebound after impact. To minimize the chance of this, find the smallest value of b such that the springs are always in compression during the oscillations after impact. Use a minimum impact velocity $V_L = 5$ m/sec downward. To find this value of b, you will need to use a software package to integrate the differential equation.

B Spread of Staph Infections in Hospitals—Part I

Courtesy of Joanna Wares, University of Richmond, and
Glenn Webb, Vanderbilt University

Methicillin-resistant *Staphylococcus aureus* (MRSA), commonly referred to as staph, is a bacterium that causes serious infections in humans and is resistant to treatment with the widely used antibiotic methicillin. MRSA has traditionally been a problem inside hospitals, where elderly patients or patients with compromised immune systems could more easily contract the bacteria and develop bloodstream infections. MRSA is implicated in a large percentage of hospital fatalities, causing more deaths per year than AIDS. Recently, a genetically different strain of MRSA has been found in the community at large. The new strain (CA-MRSA) is able to infect healthy and young people, which the traditional strain (HA-MRSA) rarely does. As CA-MRSA appears in the community, it is inevitably being spread into hospitals. Some studies suggest that CA-MRSA will overtake HA-MRSA in the hospital, which would increase the severity of the problem and likely cause more deaths per year.

To predict whether or not CA-MRSA will overtake HA-MRSA, a compartmental model has been developed by mathematicians in collaboration with medical professionals (see references [1], [2] on page 310). This model classifies all patients in the hospital into three groups:

- $H(t)$ = patients colonized with the traditional hospital strain, HA-MRSA.
- $C(t)$ = patients colonized with the community strain, CA-MRSA.
- $S(t)$ = susceptible patients, those not colonized with either strain.

The *parameters* of the model are

- β_C = the rate (per day) at which CA-MRSA is transmitted between patients.
- β_H = the rate (per day) at which HA-MRSA is transmitted between patients.
- δ_C = the rate (per day) at which patients who are colonized with CA-MRSA exit the hospital by death or discharge.
- δ_H = the rate (per day) at which patients who are colonized with HA-MRSA exit the hospital by death or discharge.
- δ_S = the rate (per day) at which susceptible patients exit the hospital by death or discharge.
- α_C = the rate (per day) at which patients who are colonized with CA-MRSA successfully undergo decolonization measures.
- α_H = the rate (per day) at which patients who are colonized with HA-MRSA successfully undergo decolonization measures.
- N = the total number of patients in the hospital.
- Λ = the rate (per day) at which patients enter the hospital.

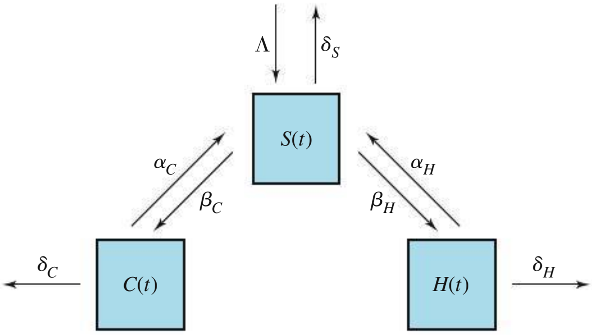


Figure 5.54 A diagram of how patients transit between the compartments

Patients move between compartments as they become colonized or decolonized (see Figure 5.54). This type of model is typically known as a SIS (susceptible-infected-susceptible) model, in which patients who become colonized can become susceptible again and colonized again (there is no immunity).

The transition between states is described by the following system of differential equations:

$$\frac{dS}{dt} = \underbrace{\Lambda}_{\text{entrance rate}} - \underbrace{\frac{\beta_H S(t)H(t)}{N}}_{\text{acquire HA-MRSA}} - \underbrace{\frac{\beta_C S(t)C(t)}{N}}_{\text{acquire CA-MRSA}} + \underbrace{\alpha_H H(t)}_{\text{HA-MRSA decolonized}} + \underbrace{\alpha_C C(t)}_{\text{CA-MRSA decolonized}} - \underbrace{\delta_S S(t)}_{\text{exit hospital}}$$

$$\frac{dH}{dt} = \underbrace{\frac{\beta_H S(t)H(t)}{N}}_{\text{from S}} - \underbrace{\alpha_H H(t)}_{\text{decolonized}} - \underbrace{\delta_H H(t)}_{\text{exit hospital}}$$

$$\frac{dC}{dt} = \underbrace{\frac{\beta_C S(t)C(t)}{N}}_{\text{from S}} - \underbrace{\alpha_C C(t)}_{\text{decolonized}} - \underbrace{\delta_C C(t)}_{\text{exit hospital}}.$$

If we assume that the hospital is always full, we can conserve the system by letting $\Lambda = \delta_S S(t) + \delta_H H(t) + \delta_C C(t)$. In this case $S(t) + C(t) + H(t) = N$ for all t (assuming you start with a population of size N).

(a) Show that this assumption simplifies the above system of equations to

$$\text{(1)} \quad \begin{aligned} \frac{dH}{dt} &= (\beta_H/N)(N - C - H)H - (\delta_H + \alpha_H)H\,, \\ \frac{dC}{dt} &= (\beta_C/N)(N - C - H)C - (\delta_C + \alpha_C)C\,. \end{aligned}$$

S is then determined by the equation $S(t) = N - H(t) - C(t)$.

Parameter values obtained from the Beth Israel Deaconess Medical Center are given in Table 5.4 on page 310. Plug these values into the model and then complete the following problems.

(b) Find the three equilibria (critical points) of the system (1).

(c) Using a computer, sketch the direction field for the system (1).

(d) Which trajectory configuration exists near each critical point (node, spiral, saddle, or center)? What do they represent in terms of how many patients are susceptible, colonized with HA-MRSA, and colonized with CA-MRSA over time?

(e) Examining the direction field, do you think CA-MRSA will overtake HA-MRSA in the hospital?

Further discussion of this model appears in Project E of Chapter 12.†

†All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems*, 7th ed.

TABLE 5.4 **Parameter Values for the Transmission Dynamics of Community-Acquired and Hospital-Acquired Methicillin-Resistant *Staphylococcus aureus* Colonization (CA-MRSA and HA-MRSA)**

Parameter	Symbol	Baseline Value
Total number of patients	N	**400**
Length of stay		
Susceptible	$1/\delta_S$	5 days
Colonized CA-MRSA	$1/\delta_C$	7 days
Colonized HA-MRSA	$1/\delta_H$	5 days
Transmission rate per susceptible patient to		
Colonized CA-MRSA per colonized CA-MRSA	β_C	0.45 per day
Colonized HA-MRSA per colonized HA-MRSA	β_H	0.4 per day
Decolonization rate per colonized patient per day per length of stay		
CA-MRSA	α_C	0.1 per day
HA-MRSA	α_H	0.1 per day

References

1. D'Agata, E. M. C., Webb, G. F., Pressley, J. 2010. "Rapid emergence of co-colonization with community-acquired and hospital-acquired methicillin-resistant *Staphylococcus aureus* strains in the hospital setting". *Mathematical Modelling of Natural Phenomena* 5(3): 76–93.
2. Pressley, J., D'Agata, E. M. C., Webb, G. F. 2010. "The effect of co-colonization with community-acquired and hospital-acquired methicillin-resistant *Staphylococcus aureus* strains on competitive exclusion". *Journal of Theoretical Biology* 265(3): 645–656.

C Things That Bob

Courtesy of Richard Bernatz, Department of Mathematics, Luther College

The motion of various-shaped objects that bob in a pool of water can be modeled by a second-order differential equation derived from Newton's second law of motion, $F = ma$. The forces acting on the object include the force due to gravity, a frictional force due to the motion of the object in the water, and a buoyant force based on **Archimedes' principle:** An object that is completely or partially submerged in a fluid is acted on by an upward (buoyant) force equal to the weight of the water it displaces.

(a) The first step is to write down the governing differential equation. The dependent variable is the depth z of the object's lowest point in the water. Take z to be negative downward so that $z = -1$ means 1 ft of the object has submerged. Let $V(z)$ be the submerged volume of the object, m be the mass of the object, ρ be the density of water (in pounds per cubic foot), g be the acceleration due to gravity, and γ_w be the coefficient of friction for water. Assuming that the frictional force is proportional to the vertical velocity of the object, write down the governing second-order ODE.

(b) For the time being, neglect the effect of friction and assume the object is a cube measuring L feet on a side. Write down the governing differential equation for this case. Next, designate $z = l$ (a negative number) to be the depth of submersion such that the buoyant force is equal and opposite the gravitational force. Introduce a new variable, ζ, that gives the displacement of the object from its equilibrium position l (that is, $z = \zeta + l$). You can now write the ODE in a more familiar form. [*Hint:* Recall the mass–spring system and the equilibrium case.] Now you should recognize the type of solution for this problem. What is the natural frequency?

(c) In this task you consider the effect of friction. The bobbing object is a cube, 1 ft on a side, that weighs 32 lb. Let $\gamma_w = 3$ lb-sec/ft, $\rho = 62.57$ lb/ft^3, and suppose the object is initially placed on the surface of the water. Solve the governing ODE by hand to find the general solution. Next, find the particular solution for the case in which the cube is initially placed on the surface of the water and is given no initial velocity. Provide a plot of the position of the object as a function of time t.

(d) In this step of the project, you develop a numerical solution to the same problem presented in part (c). The numerical solution will be useful (indeed necessary) for subsequent parts of the project. This case provides a trial to verify that your numerical solution is correct. Go back to the initial ODE you developed in part (a). Using parameter values given in part (c), solve the initial value problem for the cube starting on the surface with no initial velocity. To solve this problem numerically, you will have to write the second-order ODE as a system of two first-order ODEs, one for vertical position z and one for vertical velocity w. Plot your results for vertical position as a function of time t for the first 3 or 4 sec and compare with the analytical solution you found in part (c). Are they in close agreement? What might you have to do in order to compare these solutions? Provide a plot of both your analytical and numerical solutions on the same graph.

(e) Suppose a sphere of radius R is allowed to bob in the water. Derive the governing second-order equation for the sphere using Archimedes' principle and allowing for friction due to its motion in the water. Suppose a sphere weighs 32 lb, has a radius of $1/2$ ft, and $\gamma_w = 3.0$ lb-sec/ft. Determine the limiting value of the position of the sphere without solving the ODE. Next, solve numerically the governing ODE for the velocity and position of the sphere as a function of time for a sphere placed on the surface of the water. You will need to write the governing second-order ODE as a system of two ODEs, one for velocity and one for position. What is the limiting position of the sphere for your solution? Does it agree with the equilibrium solution you found above? How does it compare with the equilibrium position of the cube? If it is different, explain why.

(f) Suppose the sphere in part (d) is a volleyball. Calculate the position of the sphere as a function of time t for the first 3 sec if the ball is submerged so that its lowest point is 5 ft under water. Will the ball leave the water? How high will it go? Next, calculate the ball's trajectory for initial depths lower than 5 ft and higher than 5 ft. Provide plots of velocity and position for each case and comment on what you see. Specifically, comment on the relationship between the initial depth of the ball and the maximum height the ball eventually attains.

You might consider taking a volleyball into a swimming pool to gather real data in order to verify and improve on your model. If you do so, report the data you found and explain how you used it for verification and improvement of your model.

D Hamiltonian Systems

The problems in this project explore the **Hamiltonian**[†] **formulation** of the laws of motion of a system and its phase plane implications. This formulation replaces Newton's second law $F = ma = my''$ and is based on three mathematical manipulations:

[†] *Historical Footnote:* Sir William Rowan Hamilton (1805–1865) was an Irish mathematical physicist. Besides his work in mechanics, he invented quaternions and discovered the anticommutative law for vector products.

(i) It is presumed that the force $F(t, y, y')$ depends only on y and has an antiderivative $-V(y)$, that is, $F = F(y) = -dV(y)/dy$.
(ii) The velocity variable y' is replaced throughout by the momentum $p = my'$ (so $y' = p/m$).
(iii) The **Hamiltonian** of the system is defined as

$$H = H(y, p) = \frac{p^2}{2m} + V(y) .$$

(a) Express Newton's law $F = my''$ as an equivalent first-order system in the manner prescribed in Section 5.3.

(b) Show that this system is equivalent to **Hamilton's equations**

$$\text{(2)} \quad \frac{dy}{dt} = \frac{\partial H}{\partial p} \quad \left(= \frac{p}{m}\right),$$

$$\text{(3)} \quad \frac{dp}{dt} = -\frac{\partial H}{\partial y} \quad \left(= -\frac{dV}{dy}\right).$$

(c) Using Hamilton's equations and the chain rule, show that the Hamiltonian remains constant along the solution curves:

$$\frac{d}{dt}H(y, p) = 0 .$$

In the formula for the Hamiltonian function $H(y, p)$, the first term, $p^2/(2m) = m(y')^2/2$, is the **kinetic energy** of the mass. By analogy, then, the second term $V(y)$ is known as the **potential energy** of the mass, and the Hamiltonian is the total energy. The total (mechanical)† energy is constant—hence "conserved"—when the forces $F(y)$ do not depend on time t or velocity y'; such forces are called **conservative.** The energy integral lemma of Section 4.8 (page 203) is simply an alternate statement of the conservation of energy.

Hamilton's formulation for mechanical systems and the conservation of energy principle imply that the phase plane trajectories of conservative systems lie on the curves where the Hamiltonian $H(y, p)$ is constant, and plotting these curves may be considerably easier than solving for the trajectories directly (which, in turn, is easier than solving the original system!).

(d) For the mass–spring oscillator of Section 4.1, the spring force is given by $F = -ky$ (where k is the spring constant). Find the Hamiltonian, express Hamilton's equations, and show that the phase plane trajectories $H(y, p) =$ constant for this system are the ellipses given by $p^2/(2m) + ky^2/2 =$ constant. See Figure 5.14, page 269.

The **damping** force $-by'$ considered in Section 4.1 is not conservative, of course. Physically speaking, we know that damping drains the energy from a system until it grinds to a halt at an equilibrium point. In the phase plane, we can qualitatively describe the trajectory as continuously migrating to successively lower constant-energy orbits; stable centers become asymptotically stable spiral points when damping is taken into consideration.

(e) The second Hamiltonian equation (3), which effectively states $p' = my'' = F$, has to be changed to

$$p' = -\frac{\partial H}{\partial y} - by' = -\frac{\partial H}{\partial y} - \frac{bp}{m}$$

when damping is present. Show that the Hamiltonian decreases along trajectories in this case (for $b > 0$):

$$\frac{d}{dt}H(y, p) = -b\left(\frac{p}{m}\right)^2 = -b(y')^2 .$$

†Physics states that when all forms of energy, such as heat and radiation, are taken into account, energy is conserved even when the forces are not conservative.

(f) The force on a mass–spring system suspended vertically in a gravitational field was shown in Section 4.10 (page 226) to be $F = -ky + mg$. Derive the Hamiltonian and sketch the phase plane trajectories. Sketch the trajectories when damping is present.

(g) As indicated in Section 4.8 (page 207), the Duffing spring force is modeled by $F = -y - y^3$. Derive the Hamiltonian and sketch the phase plane trajectories. Sketch the trajectories when damping is present.

(h) For the pendulum system studied in Section 4.8, Example 8, the force is given by (cf. Figure 4.18, page 208)

$$F = -\ell mg \sin\theta = -\frac{\partial}{\partial\theta}(-\ell mg \cos\theta) = -\frac{\partial}{\partial\theta}V(\theta)$$

(where ℓ is the length of the pendulum). For angular variables, the Hamiltonian formulation dictates expressing the *angular velocity variable* θ' *in terms of the angular momentum* $p = m\ell^2\theta'$; the kinetic energy, mass $\times$ velocity$^2/2$, is expressed as $m(\ell\theta')^2/2 = p^2/(2m\ell^2)$. Derive the Hamiltonian for the pendulum and sketch the phase plane trajectories. Sketch the trajectories when damping is present.

(i) The Coulomb force field is a force that varies as the reciprocal square of the distance from the origin: $F = k/y^2$. The force is *attractive* if $k < 0$ and *repulsive* if $k > 0$. Sketch the phase plane trajectories for this motion. Sketch the trajectories when damping is present.

(j) For an attractive Coulomb force field, what is the *escape velocity* for a particle situated at a position y? That is, what is the minimal (outward-directed) velocity required for the trajectory to reach $y = \infty$?

E Cleaning Up the Great Lakes

A simple mathematical model that can be used to determine the time it would take to clean up the Great Lakes can be developed using a multiple compartmental analysis approach.† In particular, we can view each lake as a tank that contains a liquid in which is dissolved a particular pollutant (DDT, phosphorus, mercury). Schematically, we view the lakes as consisting of five tanks connected as indicated in Figure 5.55.

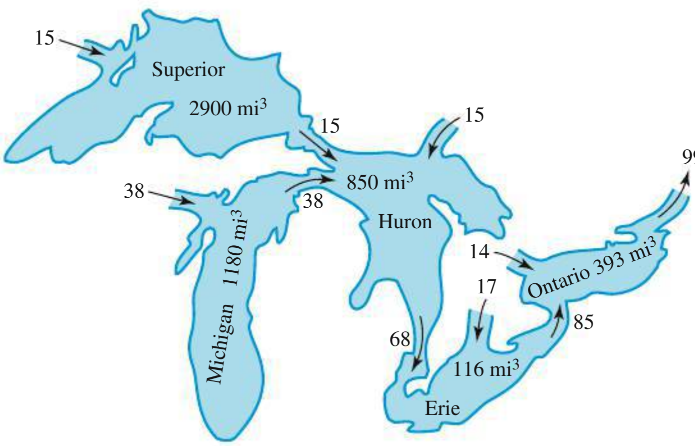


Figure 5.55 Compartmental model of the Great Lakes with flow rates (mi^3/yr) and volumes (mi^3)

†For a detailed discussion of this model, see *An Introduction to Mathematical Modeling* by Edward A. Bender (Dover Publications, New York, 2000), Chapter 8.

For our model, we make the following assumptions:

1. The volume of each lake remains constant.
2. The flow rates are constant throughout the year.
3. When a liquid enters the lake, perfect mixing occurs and the pollutants are uniformly distributed.
4. Pollutants are dissolved in the water and enter or leave by inflow or outflow of solution.

Before using this model to obtain estimates on the cleanup times for the lakes, we consider some simpler models:

(a) Use the outflow rates given in Figure 5.55 to determine the time it would take to "drain" each lake. This gives a lower bound on how long it would take to remove all the pollutants.

(b) A better estimate is obtained by assuming that each lake is a separate tank with *only* clean water flowing in. Use this approach to determine how long it would take the pollution level in each lake to be reduced to 50% of its original level. How long would it take to reduce the pollution to 5% of its original level?

(c) Finally, to take into account the fact that pollution from one lake flows into the next lake in the chain, use the entire multiple compartment model given in Figure 5.55 to determine when the pollution level in each lake has been reduced to 50% of its original level, assuming pollution has ceased (that is, inflows not from a lake are clean water). Assume that all the lakes initially have the same pollution concentration p. How long would it take for the pollution to be reduced to 5% of its original level?

F The 2014-2015 Ebola Epidemic

Courtesy of Glenn Webb, Vanderbilt University

We develop a model for the 2014–2015 Ebola epidemic in West Africa, which began in the spring of 2014 and spread though the countries of Sierra Leone, Liberia, and Guinea. Initially there was great concern that the epidemic might develop with great severity, and even spread globally.[†] Instead, by early 2015 the epidemic had subsided in these West African countries, and in January 2016, the World Health Organization (WHO) declared that the epidemic was contained in all three countries. The reasons that Ebola subsided are complex, but a key role was increased identification and isolation of infected cases, and contact tracing of these cases to identify additional cases.

Our model incorporates the principal features of contact tracing, namely, the number of contacts per identified infectious case, the likelihood that a traced contact is infectious, and the efficiency of the isolation of contact-traced individuals. The model consists of a system of ordinary differential equations for the compartments of the epidemic population, and is based on an earlier version of the model.[‡] The model incorporates the unique features of the Ebola outbreaks in this region. These features include the rates of transmission to susceptibles from both infectious cases and improperly handled deceased cases, the rates of reporting cases, and the rates of recovery and mortality of unreported cases.

[†]H. Nishiura and G. Chowell, "Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014." *Eurosurveil.*, Vol. 19 no. 36 (Sept. 11, 2014).

[‡]C. Browne, X. Huo, P. Magal. M. Seydi, O. Seydi, G. Webb, "A model of the 2014 Ebola epidemic in West Africa with contact tracing," *PLOS Currents Outbreaks*, published online, January 30, 2015.

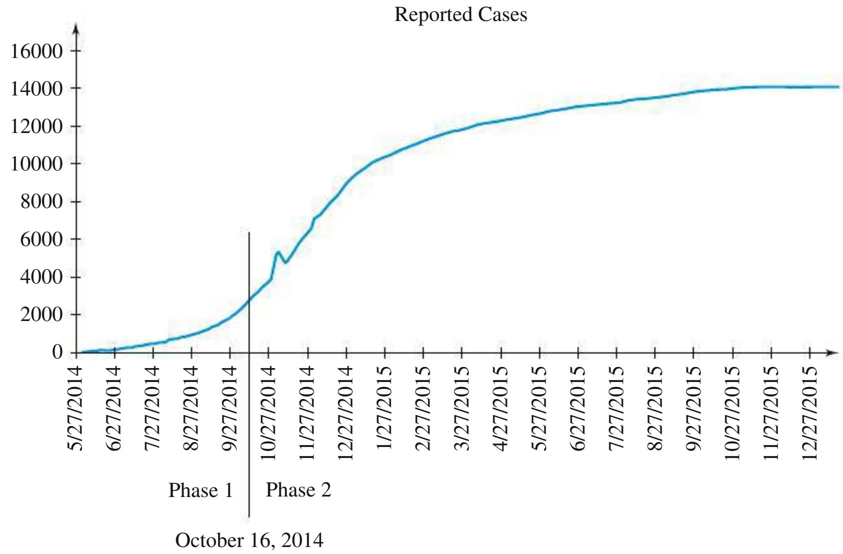


Figure 5.56 WHO Ebola Situation Data: reported cases. (Center for Disease Control and Prevention, Previous Case Counts: http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html)

The model extends the SIR paradigm of Section 5.5, by refining the population classes into susceptibles $S(t)$ (those who have not been affected), exposed $E(t)$ (incubating cases), $I(t)$ (infectious cases), and contaminated deceased $C(t)$ (improperly handled corpses of cases). The successful containment of the epidemic was due to measures isolating infected individuals and contaminated corpses, implemented around October 16, 2014. The WHO data in Figure 5.56 traces the chronology of the number of *reported* cases $R(t)$; the inflection point clearly highlights the initiation of isolation and contact tracing measures.

Figure 5.57 on page 316 schematically depicts the transitions between the classifications, as quantified by the subsequent system of differential equations:

TABLE 5.5 Ebola Model Parameters in Sierra Leone

Parameter	Description
$N = 6 \times 10^6$	Population of Sierra Leone (assumed to be constant)
$\beta = 0.321$	Transmission rate exclusive of improper handling of deceased cases
$\varepsilon = 0.002$	Transmission rate due to improper handling of deceased cases
$\kappa = 0$ (1st phase) $\kappa = 8$ (2nd phase)	Average number of contacts traced per reported and hospitalized infectious case
$\alpha = 0.08$ (1st phase) $\alpha = 0.145$ (2nd phase)	Rate of reporting and hospitalization of infectious cases not resulting from contact tracing
$\pi = 0.5$	Probability a contact-traced infected individual is isolated without causing a new case
$\omega = 0.1$	Probability a contact-traced individual is infected
$\gamma = 0.0333$	Rate of recovery of unreported infectious cases
$V = 0.125$	Rate of mortality of unreported infectious cases
$\sigma = 1/9$	Average incubation period (denominator)
$\psi = 1/5$ (1st phase) $\psi = 1/2$ (2nd phase)	Average number of days before proper handling of deceased unreported cases (denominator)

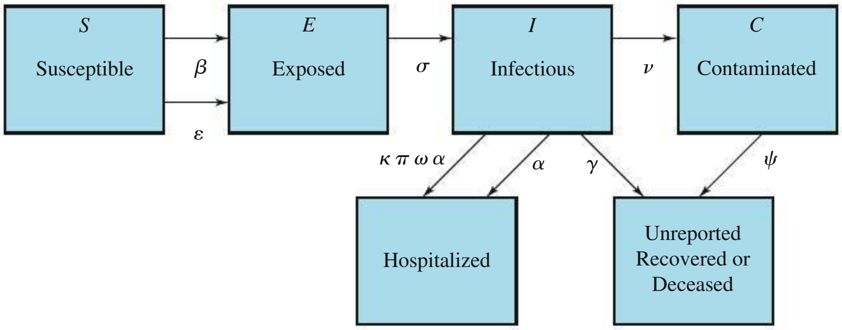


Figure 5.57 Schematic of Ebola transitions

$$\frac{dS(t)}{dt} = \underbrace{-\beta S(t)\frac{I(t)}{N}}_{\text{susceptible-infected interaction}} \quad \underbrace{-\varepsilon S(t)\frac{C(t)}{N}}_{\text{susceptible-contaminated interaction}}$$

$$\frac{dE(t)}{dt} = \beta S(t)\frac{I(t)}{N} + \varepsilon S(t)\frac{C(t)}{N} \quad \underbrace{-\sigma E(t)}_{\text{incubation}}$$

$$\frac{dI(t)}{dt} = \sigma E(t) \quad \underbrace{-\nu I(t)}_{\text{mortality, unreported infected}} \quad \underbrace{-(\alpha + \gamma + \alpha\kappa\pi\omega)I(t)}_{\text{hospitalization, recovery, or isolation by contact tracing}}$$

$$\frac{dC(t)}{dt} = \nu I(t) \quad \underbrace{-\psi C(t)}_{\text{proper handling of unreported corpses}}$$

$$\frac{dR(t)}{dt} = \alpha I(t) + \psi C(t)$$

By trial and error and numerical experimentation, the following parameter values were found to give a good fit, to the WHO data, of the solution of this system. The initial values were taken to be $S(0) = N$, $E(0) = 20$, $I(0) = 20$, $C(0) = 10$. Note the change in the values of κ, α, and ψ on the 142nd day, reflecting the initiation of contact tracing.

(a) Use a software package to solve the system for the time interval indicated in Figure 5.56 on page 315. Plot the number of reported cases $R(t)$. Note the high quality of the fit of the simulations to the WHO data.

(b) Plot the number of persons affected by the epidemic, $6 \times 10^6 - S(t)$. You should see that about 42% of all cases were *unreported* on October 16, 2014, but this fell to 10% by January, 2016.

G Phase-Locked Loops

Today's high-speed communications technology requires circuitry that will measure or match incoming radio-frequency oscillations in the gigahertz range. Mathematically, that means one must determine the frequency ω in a sinusoidal signal $A \sin(\omega t + \alpha)$ when ω is in the neighborhood of $2\pi \times 10^9$ rad/s. In theory this can be accomplished by taking three measurements of the sinusoid within a quarter-period. However, obtaining three accurate measurements of a signal in an interval of 10^{-10} seconds is impractical. The *phase-locked loop* accomplishes the task in hardware, using a voltage-controlled oscillator (VCO) to synthesize another sinusoid $B \cos\theta(t)$ whose *phase*, $\theta(t)$, matches that of the incoming signal $(\omega t + \alpha)$. The frequency $d\theta(t)/dt$ of the VCO output is linearly related to the voltage applied to the VCO (say, $d\theta(t)/dt = E + Fv(t)$), and it matches ω once synchronization is achieved. Thus one can determine the incoming frequency by reading off the VCO voltage.

A diagram of the phase-locked loop circuit is given in Figure 5.58. In brief: the mixer combines the signals $A \sin(\omega t + \alpha)$ and $B \cos\theta(t)$ to create a *control signal* $\frac{AB}{2}\sin[\omega t + \alpha - \theta(t)]$ whose phase equals the *difference* between the phase of the incoming signal and the phase of the VCO output.† This control signal provides a measure of the lack of synchronization; it is zero when $\theta(t) = \omega t + \alpha(\pm 2\pi n)$. Thus to help $\theta(t)$ catch up to $\omega t + \alpha$, the circuit adjusts the VCO voltage $v(t)$ so as to increase the output frequency, $d\theta(t)/dt$, in proportion to a weighted combination of the control signal and its derivative:

$$\textbf{(4)} \qquad \frac{d}{dt}\frac{d\theta(t)}{dt} = \frac{d^2\theta}{dt^2} = \frac{d(E+Fv)}{dt}$$
$$= K_1\frac{AB}{2}\sin[\omega t + \alpha - \theta(t)] + K_2\frac{d}{dt}\left\{\frac{AB}{2}\sin[\omega t + \alpha - \theta(t)]\right\}.$$

We are going to employ the phase plane to analyze how the phase-locked loop approaches synchronicity when it is controlled by equation (4).

(a) Convert (4) into an autonomous equation by introducing the *error* $e(t) := \omega t + \alpha - \theta(t)$ and deriving

$$\textbf{(5)} \qquad \frac{d^2e}{dt^2} = -\frac{d^2\theta}{dt^2} = -K_1\frac{AB}{2}\sin e - K_2\frac{AB}{2}(\cos e)\frac{de}{dt}.$$

(b) Put equation (5) into normal form (two first-order differential equations in $x_1 := e(t)$ and $x_2 = de/dt$) and find all the critical points.

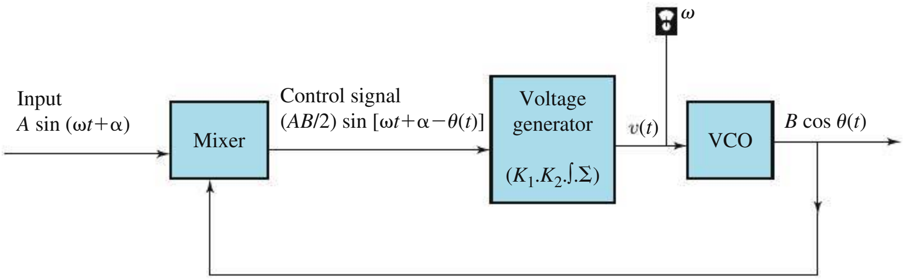


Figure 5.58 Phase-locked loop

†A classic trigonometric identity shows that multiplication of the incoming and VCO signals yields the control signal plus a high-frequency oscillation, which is easily filtered out.

(c) Use software to sketch the direction field in this system's phase plane, with $K_1 \frac{AB}{2} = 4$, $K_2 \frac{AB}{2} = 1$. (The duplicitous nomenclature is unintended: *phase plane* does not refer to the *phase* of the sinusoid.) Trace a few trajectories to get a feel for how the phase-locked loop performs. Which critical points are stable, and which are unstable? Are all the stable critical points suitable for synchronization?

(d) Apply the mass–spring analogy (Section 4.8) to equation (5). The mass m is one, the spring force is $-K_1 \frac{AB}{2} \sin e$ instead of $-ke$, and the damping force is $-K_2 \frac{AB}{2} (\cos e) \frac{de}{dt}$ instead of $-b \frac{de}{dt}$. If e is small, the mass–spring parameters and the phase-locked loop parameters are approximately the same (recall Project D, Chapter 4). Sketch the direction field for the (normal form of the) damped mass–spring system

$$\frac{d^2e}{dt^2} = -b \frac{de}{dt} - ke$$

with $b = 1$, $k = 4$, and compare with that of the phase-locked loop. Explain why the number of equilibria is different, and why the nature of the phase-locked loop's "forces" renders certain of its equilibria to be unstable.

CHAPTER 9

Matrix Methods for Linear Systems

9.1 Introduction

In this chapter we return to the analysis of systems of differential equations. When the equations in the system are *linear*, matrix algebra provides a compact notation for expressing the system. In fact, the notation itself suggests new and elegant ways of characterizing the solution properties, as well as novel, efficient techniques for explicitly obtaining solutions.

In Chapter 5 we analyzed physical situations wherein two fluid tanks containing brine solutions were interconnected and pumped so as ultimately to deplete the salt content in each tank. By accounting for the various influxes and outfluxes of brine, a system of differential equations for the salt contents $(x(t)$ and $y(t))$ of each tank was derived; a typical model is

$$\text{(1)} \qquad \begin{aligned} dx/dt &= -4x + 2y \,, \\ dy/dt &= 4x - 4y \,. \end{aligned}$$

Express this system in matrix notation as a single equation.

The right-hand side of the first member of (1) possesses a mathematical structure that is familiar from vector calculus; namely, it is the **dot product**† of two vectors:

$$\text{(2)} \qquad -4x + 2y = [-4 \quad 2] \cdot [x \quad y] \,.$$

Similarly, the second right-hand side in (1) is the dot product

$$4x - 4y = [4 \quad -4] \cdot [x \quad y] \,.$$

The frequent occurrence in mathematics of arrays of dot products, such as evidenced in the system (1), led to the development of **matrix algebra**, a mathematical discipline whose basic operation—the matrix product—is the arrangement of a set of dot products according to the following plan:

$$\begin{bmatrix} -4 & 2 \\ 4 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} [-4 \quad 2] \cdot [x \quad y] \\ [4 \quad -4] \cdot [x \quad y] \end{bmatrix} = \begin{bmatrix} -4x + 2y \\ 4x - 4y \end{bmatrix} .$$

In general, the product of a **matrix**—i.e., an m by n rectangular array of numbers—and a *column vector* is defined to be the collection of dot products of the *rows* of the matrix with the

†Recall that the dot product of two vectors **u** and **v** equals the length of **u** times the length of **v** times the cosine of the angle between **u** and **v**. However, it is more conveniently computed from the *components* of **u** and **v** by the "inner product" indicated in equation (2).

vector, arranged as a column vector:

$$\begin{bmatrix} \textbf{row \# 1} \\ \textbf{row \# 2} \\ \vdots \\ \textbf{row \# m} \end{bmatrix} \begin{bmatrix} \\ \mathbf{v} \\ \\ \end{bmatrix} = \begin{bmatrix} [\textbf{row \# 1}]\cdot\mathbf{v} \\ [\textbf{row \# 2}]\cdot\mathbf{v} \\ \vdots \\ [\textbf{row \# m}]\cdot\mathbf{v} \end{bmatrix},$$

where the vector $\mathbf{v}$ has n components; the dot product of two n-dimensional vectors is computed in the obvious way:

$$[a_1 \quad a_2 \ \cdots \ a_n]\cdot[x_1 \quad x_2 \ \cdots \ x_n] = a_1x_1 + a_2x_2 + \cdots + a_nx_n .$$

Using the notation for the matrix product, we can write the system (1) for the interconnected tanks as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -4 & 2 \\ 4 & -4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

The following example demonstrates a four-dimensional implementation of this notation. Note that the coefficients in the linear system need not be constants.

Example 1 Express the system

$$\begin{aligned} x_1' &= 2x_1 + t^2x_2 + (4t + e^t)x_4 , \\ x_2' &= (\sin t)x_2 + (\cos t)x_3 , \\ x_3' &= x_1 + x_2 + x_3 + x_4 , \\ x_4' &= 0 \end{aligned} \qquad (3)$$

as a matrix equation.

Solution We express the right-hand side of the first member of (3) as the dot product

$$2x_1 + t^2x_2 + (4t + e^t)x_4 = [2 \quad t^2 \quad 0 \quad (4t + e^t)]\cdot[x_1 \quad x_2 \quad x_3 \quad x_4] .$$

The other dot products are similarly identified, and the matrix form is given by

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \end{bmatrix} = \begin{bmatrix} 2 & t^2 & 0 & (4t + e^t) \\ 0 & \sin t & \cos t & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \ \blacklozenge$$

In general, if a system of differential equations is expressed as

$$\begin{aligned} x_1' &= a_{11}(t)x_1 + a_{12}(t)x_2 + \cdots + a_{1n}(t)x_n \\ x_2' &= a_{21}(t)x_1 + a_{22}(t)x_2 + \cdots + a_{2n}(t)x_n \\ &\vdots \\ x_n' &= a_{n1}(t)x_1 + a_{n2}(t)x_2 + \cdots + a_{nn}(t)x_n , \end{aligned}$$

it is said to be a linear homogeneous system in **normal form**.† The matrix formulation of such

†The normal form was defined for general systems in Section 5.3, page 256.

a system is then

$$\mathbf{x}' = \mathbf{A}\mathbf{x} ,$$

where $\mathbf{A}$ is the **coefficient matrix**

$$\mathbf{A} = \mathbf{A}(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & & \vdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{bmatrix}$$

and $\mathbf{x}$ is the solution vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} .$$

Note that we have used $\mathbf{x}'$ to denote the vector of derivatives

$$\mathbf{x}' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}' = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix} .$$

Example 2 Express the differential equation for the undamped, unforced mass–spring oscillator (recall Section 4.1, page 152)

(4) $\quad my'' + ky = 0$

as an equivalent system of first-order equations in normal form, expressed in matrix notation.

Solution We have to express the *second* derivative, y'', as a *first* derivative in order to formulate (4) as a first-order system. This is easy; the acceleration y'' is the derivative of the *velocity* $v = y'$, so (4) becomes

(5) $\quad mv' + ky = 0 .$

The first-order system is then assembled by identifying v with y', and appending it to (5):

$$\begin{aligned} y' &= v \\ mv' &= -ky . \end{aligned}$$

To put this system in normal form and express it as a matrix equation, we need to divide the second equation by the mass m:

$$\begin{bmatrix} y \\ v \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ -k/m & 0 \end{bmatrix} \begin{bmatrix} y \\ v \end{bmatrix} . \quad \blacklozenge$$

In general, the customary way to write an nth-order linear homogeneous differential equation

$$a_n(t)y^{(n)} + a_{n-1}(t)y^{(n-1)} + \cdots + a_1(t)y' + a_0(t)y = 0$$

as an equivalent system in normal form is to define the first $(n-1)$ derivatives of y (including y, the zeroth derivative, itself) to be new unknowns:

$$\begin{aligned} x_1(t) &= y(t)\,, \\ x_2(t) &= y'(t)\,, \\ &\vdots \\ x_n(t) &= y^{(n-1)}(t)\,. \end{aligned}$$

Then the system consists of the identification of $x_j(t)$ as the derivative of $x_{j-1}(t)$, together with the original differential equation expressed in these variables (and divided by $a_n(t)$):

$$\begin{aligned} x_1' &= x_2\,, \\ x_2' &= x_3\,, \\ &\vdots \\ x_{n-1}' &= x_n\,, \\ x_n' &= -\frac{a_0(t)}{a_n(t)}x_1 - \frac{a_1(t)}{a_n(t)}x_2 - \cdots - \frac{a_{n-1}(t)}{a_n(t)}x_n\,. \end{aligned}$$

For systems of two or more higher-order differential equations, the same procedure is applied to each unknown function in turn; an example will make this clear.

Example 3 The coupled mass–spring oscillator depicted in Figure 5.26 on page 283 was shown to be governed by the system

$$\begin{aligned} 2\frac{d^2x}{dt^2} + 6x - 2y &= 0\,, \\ \frac{d^2y}{dt^2} + 2y - 2x &= 0\,. \end{aligned} \tag{6}$$

Write (6) in matrix notation.

Solution We introduce notation for the lower-order derivatives:

$$x_1 = x\,, \qquad x_2 = x'\,, \qquad x_3 = y\,, \qquad x_4 = y'\,. \tag{7}$$

In these variables, the system (6) states

$$\begin{aligned} 2x_2' + 6x_1 - 2x_3 &= 0\,, \\ x_4' + 2x_3 - 2x_1 &= 0\,. \end{aligned} \tag{8}$$

The normal form is then

$$\begin{aligned} x_1' &= x_2\,, \\ x_2' &= -3x_1 + x_3\,, \\ x_3' &= x_4\,, \\ x_4' &= 2x_1 - 2x_3 \end{aligned}$$

or in matrix notation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & -2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \quad ◆$$

9.1 EXERCISES

In Problems 1–6, express the given system of differential equations in matrix notation.

1. $x' = 7x + 2y\,,$
$y' = 3x - 2y$

2. $x' = y\,,$
$y' = -x$

3. $x' = x + y + z\,,$
$y' = 2z - x\,,$
$z' = 4y$

4. $x_1' = x_1 - x_2 + x_3 - x_4\,,$
$x_2' = x_1 + x_4\,,$
$x_3' = \sqrt{\pi}x_1 - x_3\,,$
$x_4' = 0$

5. $x' = (\sin t)x + e^t y\,,$
$y' = (\cos t)x + (a + bt^3)y$

6. $x_1' = (\cos 2t)x_1\,,$
$x_2' = (\sin 2t)x_2\,,$
$x_3' = x_1 - x_2$

In Problems 7–10, express the given higher-order differential equation as a matrix system in normal form.

7. The damped mass–spring oscillator equation $my'' + by' + ky = 0$

8. Legendre's equation $(1 - t^2)y'' - 2ty' + 2y = 0$

9. The Airy equation $y'' - ty = 0$

10. Bessel's equation $y'' + \frac{1}{t}y' + \left(1 - \frac{n^2}{t^2}\right)y = 0$

In Problems 11–13, express the given system of higher-order differential equations as a matrix system in normal form.

11. $x'' + 3x + 2y = 0\,,$
$y'' - 2x = 0$

12. $x'' + 3x' - y' + 2y = 0\,,$
$y'' + x' + 3y' + y = 0$

13. $x'' - 3x' + t^2y - (\cos t)x = 0\,,$
$y''' + y'' - tx' + y' + e^t x = 0$

9.2 Review 1: Linear Algebraic Equations

Here and in the next section we review some basic facts concerning linear algebraic systems and matrix algebra that will be useful in solving linear systems of differential equations in normal form. Readers competent in these areas may proceed to Section 9.4.

A set of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1\,, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2\,, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

(where the a_{ij}'s and b_i's are given constants) is called a *linear system of n algebraic equations* in the n unknowns $x_1, x_2, \ldots, x_n$. The procedure for solving the system using elimination methods is well known. Herein we describe a particularly convenient implementation of the method called the *Gauss–Jordan elimination algorithm.*[†] The basic idea of this formulation is to use the first equation to eliminate x_1 in all the other equations; then use the second equation to eliminate x_2 in all the others; and so on. If all goes well, the resulting system will be "uncoupled," and the values of the unknowns $x_1, x_2, \ldots, x_n$ will be apparent. A short example will make this clear.

[†]The Gauss–Jordan algorithm is neither the fastest nor the most accurate computer algorithm for solving a linear system of algebraic equations, but for solutions executed by hand it has many pedagogical advantages. Usually it is much faster than *Cramer's rule*, described in Appendix D.

Example 1 Solve the system

$$\begin{aligned} 2x_1 + 6x_2 + 8x_3 &= 16\,, \\ 4x_1 + 15x_2 + 19x_3 &= 38\,, \\ 2x_1 \phantom{{}+15x_2} + 3x_3 &= 6\,. \end{aligned}$$

Solution By subtracting 2 times the first equation from the second, we eliminate x_1 from the latter. Similarly, x_1 is eliminated from the third equation by subtracting 1 times the first equation from it:

$$\begin{aligned} 2x_1 + 6x_2 + 8x_3 &= 16\,, \\ 3x_2 + 3x_3 &= 6\,, \\ -6x_2 - 5x_3 &= -10\,. \end{aligned}$$

Next we subtract multiples of the second equation from the first and third to eliminate x_2 in them; the appropriate multiples are 2 and -2, respectively:

$$\begin{aligned} 2x_1 \phantom{{}+3x_2} + 2x_3 &= 4\,, \\ 3x_2 + 3x_3 &= 6\,, \\ x_3 &= 2\,. \end{aligned}$$

Finally, we eliminate x_3 from the first two equations by subtracting multiples (2 and 3, respectively) of the third equation:

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

The system is now uncoupled; i.e., we can solve each equation separately:

$$x_1 = 0\,, \quad x_2 = 0\,, \quad x_3 = 2\,. \quad ◆$$

Two complications can disrupt the straightforward execution of the Gauss–Jordan algorithm. The first occurs when the impending variable to be eliminated (say, x_j) does not occur in the jth equation. The solution is usually obvious; we employ one of the subsequent equations to eliminate x_j. Example 2 illustrates this maneuver.

Example 2 Solve the system

$$\begin{aligned} x_1 + 2x_2 + 4x_3 + x_4 &= 0\,, \\ -x_1 - 2x_2 - 2x_3 \phantom{{}+2x_4} &= 1\,, \\ -2x_1 - 4x_2 - 8x_3 + 2x_4 &= 4\,, \\ x_1 + 4x_2 + 2x_3 \phantom{{}+2x_4} &= -3\,. \end{aligned}$$

Solution The first unknown x_1 is eliminated from the last three equations by subtracting multiples of the first equation:

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

Now, we cannot use the second equation to eliminate the second unknown because x_2 is not present. The next equation that *does* contain x_2 is the fourth, so we switch the second and fourth equation:

$$\begin{aligned} x_1 + 2x_2 + 4x_3 + x_4 &= 0\,, \\ 2x_2 - 2x_3 - x_4 &= -3\,, \\ 4x_4 &= 4\,, \\ 2x_3 + x_4 &= 1\,, \end{aligned}$$

and proceed to eliminate x_2:

$$\begin{aligned} x_1 \qquad + 6x_3 + 2x_4 &= 3\,, \\ 2x_2 - 2x_3 - x_4 &= -3\,, \\ 4x_4 &= 4\,, \\ 2x_3 + x_4 &= 1\,. \end{aligned}$$

To eliminate x_3, we have to switch again,

$$\begin{aligned} x_1 \qquad + 6x_3 + 2x_4 &= 3\,, \\ 2x_2 - 2x_3 - x_4 &= -3\,, \\ 2x_3 + x_4 &= 1\,, \\ 4x_4 &= 4\,, \end{aligned}$$

and eliminate, in turn, x_3 and x_4. This gives

$$\begin{aligned} x_1 \qquad\qquad - x_4 &= 0\,, \\ 2x_2 \qquad\qquad &= -2\,, \\ 2x_3 + x_4 &= 1\,, \\ 4x_4 &= 4\,, \end{aligned} \qquad \text{and} \qquad \begin{aligned} x_1 \qquad\qquad &= 1\,, \\ 2x_2 \qquad\quad &= -2\,, \\ 2x_3 \qquad &= 0\,, \\ 4x_4 &= 4\,. \end{aligned}$$

The solution to the uncoupled equations is

$$x_1 = 1\,, \quad x_2 = -1\,, \quad x_3 = 0\,, \quad x_4 = 1\,. \quad ◆$$

The other complication that can disrupt the Gauss–Jordan algorithm is much more profound. What if, when we are "scheduled" to eliminate the unknown x_j, it is absent from *all* of the subsequent equations? The first thing to do is to move on to the elimination of the *next* unknown x_{j+1}, as demonstrated in Example 3.

Example 3 Apply the Gauss–Jordan algorithm to the system

$$\begin{aligned} 2x_1 + 4x_2 + x_3 &= 8\,, \\ 2x_1 + 4x_2 \qquad &= 6\,, \\ -4x_1 - 8x_2 + x_3 &= -10\,. \end{aligned} \tag{1}$$

Solution Elimination of x_1 proceeds as usual:

$$\begin{aligned} 2x_1 + 4x_2 + x_3 &= 8\,, \\ -x_3 &= -2\,, \\ 3x_3 &= 6\,. \end{aligned}$$

Now since x_2 is absent from the second *and* third equations, we use the second equation to eliminate x_3 :

$$
\begin{aligned}
2x_1 + 4x_2 \qquad &= 6\,, \\
-x_3 &= -2\,, \\
0 &= 0\,.
\end{aligned} \tag{2}
$$

How do we interpret the system (2)? The final equation contains no information, of course, and we ignore it.† The second equation implies that $x_3 = 2$.

The first equation implies that $x_1 = 3 - 2x_2$, but there is no equation for x_2. Evidently, x_2 is a "free" variable, and we can assign *any* value to it—as long as we take x_1 to be $3 - 2x_2$. Thus (1) has an infinite number of solutions, and a convenient way of characterizing them is

$$x_1 = 3 - 2s\,, \quad x_2 = s\,, \quad x_3 = 2\,; \quad -\infty < s < \infty\,.$$

We remark that an equivalent solution can be obtained by treating x_1 as the free variable, say $x_1 = s$, and taking $x_2 = (3 - s)/2$, $x_3 = 2$. ◆

The final example is contrived to demonstrate all the features that we have encountered.

Example 4 Find all solutions to the system

$$
\begin{aligned}
x_1 - x_2 + 2x_3 + 2x_4 &= 0\,, \\
2x_1 - 2x_2 + 4x_3 + 3x_4 &= 1\,, \\
3x_1 - 3x_2 + 6x_3 + 9x_4 &= -3\,, \\
4x_1 - 4x_2 + 8x_3 + 8x_4 &= 0\,.
\end{aligned}
$$

Solution We use the first equation to eliminate x_1:

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

Now, both x_2 *and* x_3 are absent from all subsequent equations, so we use the second equation to eliminate x_4.

$$
\begin{aligned}
x_1 - x_2 + 2x_3 \qquad &= 2\,, \\
-x_4 &= 1\,, \\
0 &= 0\,, \\
0 &= 0\,.
\end{aligned}
$$

There are no constraints on either x_2 or x_3; thus we take them to be free variables and characterize the solutions by

$$x_1 = 2 + s - 2t\,, \quad x_2 = s\,, \quad x_3 = t\,, \quad x_4 = -1, \quad -\infty < s, t < \infty\,. \ ◆$$

In closing, we note that if the execution of the Gauss–Jordan algorithm results in a display of the form $0 = 1$ (or $0 = k$, where $k \neq 0$), the original system has no solutions; it is *inconsistent.* This is explored in Problem 12.

†The occurrence of the identity $0 = 0$ in the Gauss–Jordan algorithm implies that one of the original equations was *redundant.* In this case you may observe that the final equation in (1) can be derived by subtracting 3 times the second equation from the first.

9.2 EXERCISES

In Problems 1–11, find all solutions to the system using the Gauss–Jordan elimination algorithm.

1. $$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 6\,, \\ 2x_1 + x_2 + x_3 &= 6\,, \\ x_1 + x_2 + 3x_3 &= 6 \end{aligned}$$

2. $$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 1\,, \\ x_1 \qquad\qquad + x_4 &= 0\,, \\ 2x_1 + 2x_2 - x_3 + x_4 &= 0\,, \\ x_1 + 2x_2 - x_3 + x_4 &= 0 \end{aligned}$$

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

4. $$\begin{aligned} x_3 + x_4 &= 0\,, \\ x_1 + x_2 + x_3 + x_4 &= 1\,, \\ 2x_1 - x_2 + x_3 + 2x_4 &= 0\,, \\ 2x_1 - x_2 + x_3 + x_4 &= 0 \end{aligned}$$

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

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

7. $$\begin{aligned} -x_1 + 3x_2 &= 0\,, \\ -3x_1 + 9x_2 &= 0 \end{aligned}$$

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

9. $$\begin{aligned} (1-i)x_1 + 2x_2 &= 0\,, \\ -x_1 - (1+i)x_2 &= 0 \end{aligned}$$

10. $$\begin{aligned} x_1 + x_2 + x_3 &= i\,, \\ 2x_1 + 3x_2 - ix_3 &= 0\,, \\ x_1 + 2x_2 + x_3 &= i \end{aligned}$$

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

12. Use the Gauss–Jordan elimination algorithm to show that the following systems of equations are inconsistent. That is, demonstrate that the existence of a solution would imply a mathematical contradiction.

(a) $$\begin{aligned} 2x_1 - x_2 &= 2\,, \\ -6x_1 + 3x_2 &= 4 \end{aligned}$$

(b) $$\begin{aligned} 2x_1 \qquad + x_3 &= -1\,, \\ -3x_1 + x_2 + 4x_3 &= 1\,, \\ -x_1 + x_2 + 5x_3 &= 1 \end{aligned}$$

13. Use the Gauss–Jordan elimination algorithm to show that the following system of equations has a unique solution for $r = 2$, but an infinite number of solutions for $r = 1$.

$$\begin{aligned} 2x_1 - 3x_2 &= rx_1\,, \\ x_1 - 2x_2 &= rx_2 \end{aligned}$$

14. Use the Gauss–Jordan elimination algorithm to show that the following system of equations has a unique solution for $r = -1$, but an infinite number of solutions for $r = 2$.

$$\begin{aligned} x_1 + 2x_2 - x_3 &= rx_1\,, \\ x_1 \qquad + x_3 &= rx_2\,, \\ 4x_1 - 4x_2 + 5x_3 &= rx_3 \end{aligned}$$

9.3 Review 2: Matrices and Vectors

A **matrix** is a rectangular array of numbers arranged in rows and columns. An $m \times n$ matrix—that is, a matrix with m rows and n columns—is usually denoted by

$$\mathbf{A} := \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix},$$

where the element in the ith row and jth column is a_{ij}. The notation $[a_{ij}]$ is also used to designate $\mathbf{A}$. The matrices we will work with usually consist of real numbers, but in certain instances we allow complex-number entries.

Some matrices of special interest are **square matrices**, which have the same number of rows and columns; **diagonal matrices**, which are square matrices with only zero entries off the main diagonal (that is, $a_{ij} = 0$ if $i \neq j$); and (column) **vectors**, which are $n \times 1$ matrices. For example, if

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

then $\mathbf{A}$ is a square matrix, $\mathbf{B}$ is a diagonal matrix, and $\mathbf{x}$ is a vector. An $m \times n$ matrix whose entries are all zero is called a **zero matrix** and is denoted by $\mathbf{0}$. For consistency, we denote matrices by boldfaced capitals, such as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{I}$, $\mathbf{X}$, and $\mathbf{Y}$, and reserve boldfaced lowercase letters, such as $\mathbf{c}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, for vectors.

Algebra of Matrices

Matrix Addition and Scalar Multiplication. The operations of matrix addition and scalar multiplication are very straightforward. Addition is performed by adding corresponding elements:

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

Formally, the *sum* of two $m \times n$ matrices is given by

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

(The sole novelty here is that addition is not defined for two matrices whose dimensions m, n differ.)

To multiply a matrix by a scalar (number), we simply multiply each element in the matrix by the number:

$$3\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \end{bmatrix}.$$

In other words, $r\mathbf{A} = r[a_{ij}] = [ra_{ij}]$. The notation $-\mathbf{A}$ stands for $(-1)\mathbf{A}$.

Properties of Matrix Addition and Scalar Multiplication. Matrix addition and scalar multiplication are nothing more than mere bookkeeping, and the usual algebraic properties hold. If $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are $m \times n$ matrices and r, s are scalars, then

$$\begin{aligned} &\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}, && \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}, \\ &\mathbf{A} + \mathbf{0} = \mathbf{A}, && \mathbf{A} + (-\mathbf{A}) = \mathbf{0}, \\ &r(\mathbf{A} + \mathbf{B}) = r\mathbf{A} + r\mathbf{B}, && (r + s)\mathbf{A} = r\mathbf{A} + s\mathbf{A}, \\ &r(s\mathbf{A}) = (rs)\mathbf{A} = s(r\mathbf{A}) . \end{aligned}$$

Matrix Multiplication. The matrix product is what makes matrix algebra interesting and useful. We indicated in Section 9.1 that the product of a matrix $\mathbf{A}$ and a column vector $\mathbf{x}$ is the column vector composed of dot products of the rows of $\mathbf{A}$ with $\mathbf{x}$:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 2 \\ 4 \cdot 1 + 5 \cdot 0 + 6 \cdot 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 16 \end{bmatrix}.$$

More generally, the product of two matrices **A** and **B** is formed by taking the array of dot products of the *rows* of the first "factor" **A** with the *columns* of the second factor **B**; the dot product of the ith row of **A** with the jth column of **B** is written as the ijth entry of the product **AB**:

$$\begin{bmatrix} 1 & 0 & 1 \\ 3 & -1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & x \\ -1 & -1 & y \\ 4 & 1 & z \end{bmatrix} = \begin{bmatrix} 1+0+4 & 2+0+1 & x+0+z \\ 3+1+8 & 6+1+2 & 3x-y+2z \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 3 & x+z \\ 12 & 9 & 3x-y+2z \end{bmatrix}.$$

Note that **AB** is only defined when the number of columns of **A** matches the number of rows of **B**. A useful formula for the product of an $m \times n$ matrix **A** and an $n \times p$ matrix **B** is

$$\mathbf{AB} := [c_{ij}], \qquad \text{where} \quad c_{ij} := \sum_{k=1}^{n} a_{ik} b_{kj}.$$

The dot product of the ith row of **A** and the jth column of **B** is seen in the "sum of products" expression for c_{ij}.

Since **AB** is computed in terms of the *rows* of the first factor and the *columns* of the second factor, it should not be surprising that, in general, **AB** does not equal **BA** (matrix multiplication does not *commute*):

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}, \qquad \text{but} \qquad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}.$$

In fact, the dimensions of **A** and **B** may render one or the other of these products undefined:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}; \qquad \begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ not defined}.$$

By the same token, one might not expect $(\mathbf{AB})\mathbf{C}$ to equal $\mathbf{A}(\mathbf{BC})$, since in $(\mathbf{AB})\mathbf{C}$ we take dot products with the *columns* of **B**, whereas in $\mathbf{A}(\mathbf{BC})$ we employ the *rows* of **B**. So it is a pleasant surprise that *this complication does not arise*, and the "parenthesis grouping" rules are the customary ones:

Properties of Matrix Multiplication

$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$	(Associativity)
$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$	(Distributivity)
$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$	(Distributivity)
$(r\mathbf{A})\mathbf{B} = r(\mathbf{AB}) = \mathbf{A}(r\mathbf{B})$	(Associativity)

To summarize, the algebra of matrices proceeds much like the standard algebra of numbers, *except that we should never presume that we can switch the order of matrix factors.* (If you think $(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + \mathbf{B}^2 + 2\mathbf{AB}$, what error have you made?)

Matrices as Linear Operators. Let **A** be an $m \times n$ matrix and let **x** and **y** be $n \times 1$ vectors. Then **Ax** is an $m \times 1$ vector, and so we can think of multiplication by **A** as defining an operator that maps $n \times 1$ vectors into $m \times 1$ vectors. A consequence of the distributivity and associativity properties is that multiplication by **A** defines a **linear operator**, since $\mathbf{A}(\mathbf{x}+\mathbf{y}) = \mathbf{Ax} + \mathbf{Ay}$ and $\mathbf{A}(r\mathbf{x}) = r\mathbf{Ax}$. Moreover, if **A** is an $m \times n$ matrix and **B** is an $n \times p$ matrix, then the $m \times p$ matrix **AB** defines a linear operator that is the composition of the linear operator defined by **B**

with the linear operator defined by $\mathbf{A}$. That is, $(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx})$, where $\mathbf{x}$ is a $p \times 1$ vector.

Examples of linear operations are

(i) stretching or contracting the components of a vector by constant factors;
(ii) rotating a vector through some angle about a fixed axis;
(iii) reflecting a vector in a plane mirror.

The Matrix Formulation of Linear Algebraic Systems. Matrix algebra was developed to provide a convenient tool for expressing and analyzing linear algebraic systems. Note that the set of equations

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 1\,, \\ x_1 + 3x_2 + 2x_3 &= -1\,, \\ x_1 + x_3 &= 0 \end{aligned}$$

can be written using the matrix product

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}. \tag{1}$$

In general, we express the linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1\,, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2\,, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

in matrix notation as $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is the *coefficient matrix,* $\mathbf{x}$ is the vector of unknowns, and $\mathbf{b}$ is the vector of constants occurring on the right-hand side:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

If $\mathbf{b} = \mathbf{0}$, the system $\mathbf{Ax} = \mathbf{b}$ is said to be *homogeneous* (analogous to the nomenclature of Section 4.2).

Matrix Transpose. The matrix obtained from $\mathbf{A}$ by interchanging its rows and columns is called the **transpose** of $\mathbf{A}$ and is denoted by $\mathbf{A}^T$. For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 6 \\ -1 & 2 & -1 \end{bmatrix}, \quad \text{then}$$

$$\mathbf{A}^T = \begin{bmatrix} 1 & -1 \\ 2 & 2 \\ 6 & -1 \end{bmatrix}.$$

In general, we have $[a_{ij}]^T = [b_{ij}]$, where $b_{ij} = a_{ji}$. Properties of the transpose are explored in Problem 7.

Matrix Identity. There is a "multiplicative identity" in matrix algebra, namely, a square diagonal matrix $\mathbf{I}$ with ones down the main diagonal. Multiplying $\mathbf{I}$ on the right or left by any other matrix (with compatible dimensions) reproduces the latter matrix:

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

(The notation $\mathbf{I}_n$ is used if it is convenient to specify the dimensions, $n \times n$, of the identity matrix.)

Matrix Inverse. Some *square* matrices $\mathbf{A}$ can be paired with other (square) matrices $\mathbf{B}$ having the property that $\mathbf{BA} = \mathbf{I}$:

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

When this happens, it can be shown that

(i) $\mathbf{B}$ is the *unique* matrix satisfying $\mathbf{BA} = \mathbf{I}$, and
(ii) $\mathbf{B}$ also satisfies $\mathbf{AB} = \mathbf{I}$.

In such a case, we say that $\mathbf{B}$ is the **inverse** of $\mathbf{A}$ and write $\mathbf{B} = \mathbf{A}^{-1}$.

Not every matrix possesses an inverse; the zero matrix $\mathbf{0}$, for example, can never satisfy the equation $\mathbf{0B} = \mathbf{I}$. A matrix that has no inverse is said to be **singular.**

If we know an inverse for the coefficient matrix $\mathbf{A}$ in a system of linear equations $\mathbf{Ax} = \mathbf{b}$, the solution can be calculated directly by computing $\mathbf{A}^{-1}\mathbf{b}$, as the following derivation shows:

$$\mathbf{Ax} = \mathbf{b} \text{ implies } \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \text{ implies } \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}\,.$$

Using (2), for example, we can solve equation (1) quite efficiently:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & -1 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{3}{2} & 1 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{5}{2} \\ \frac{1}{2} \\ -\frac{5}{2} \end{bmatrix}.$$

On the other hand, the coefficient matrix for any *inconsistent* system has no inverse. For example, the coefficient matrices

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

for the inconsistent systems of Problem 12, Exercises 9.2 (page 504), are necessarily singular.

When $\mathbf{A}^{-1}$ is known, solving $\mathbf{Ax} = \mathbf{b}$ by multiplying $\mathbf{b}$ by $\mathbf{A}^{-1}$ is certainly easier than applying the Gauss–Jordan algorithm of the previous section†. So it appears advantageous to be able to find matrix inverses. Some inverses can be obtained directly from the interpretation of the matrix as a linear operator. For example, the inverse of a matrix that rotates a vector is

†(When the effort to compute the inverse is accounted for, Gauss–Jordan emerges as the winner.)

the matrix that rotates it in the opposite direction. A matrix that performs a mirror reflection is its own inverse (what do you get if you reflect twice?). But in general one must employ an algorithm to compute a matrix inverse. The underlying strategy for this algorithm is based on the observation that if $\mathbf{X}$ denotes the inverse of $\mathbf{A}$, then $\mathbf{X}$ must satisfy the equation $\mathbf{AX} = \mathbf{I}$; finding $\mathbf{X}$ amounts to solving n linear systems of equations for the columns $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ of $\mathbf{X}$:

$$\mathbf{Ax}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{Ax}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \cdots, \quad \mathbf{Ax}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

The implementation of this operation is neatly executed by the following variation of the Gauss–Jordan algorithm.

Finding the Inverse of a Matrix. By a **row operation**, we mean any one of the following:

(a) Interchanging two rows of the matrix
(b) Multiplying a row of the matrix by a nonzero scalar
(c) Adding a scalar multiple of one row of the matrix to another row.

If the $n \times n$ matrix $\mathbf{A}$ has an inverse, then $\mathbf{A}^{-1}$ can be determined by performing row operations on the $n \times 2n$ matrix $[\mathbf{A} \,\vdots\, \mathbf{I}]$ obtained by writing $\mathbf{A}$ and $\mathbf{I}$ side by side. In particular, we perform row operations on the matrix $[\mathbf{A} \,\vdots\, \mathbf{I}]$ until the first n rows and columns form the identity matrix; that is, the new matrix is $[\mathbf{I} \,\vdots\, \mathbf{B}]$. Then $\mathbf{A}^{-1} = \mathbf{B}$. We remark that if this procedure fails to produce a matrix of the form $[\mathbf{I} \,\vdots\, \mathbf{B}]$, then $\mathbf{A}$ has no inverse.

Example 1 Find the inverse of $\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 1 \end{bmatrix}$.

Solution We first form the matrix $[\mathbf{A} \,\vdots\, \mathbf{I}]$ and *row-reduce* the matrix to $[\mathbf{I} \,\vdots\, \mathbf{A}^{-1}]$. Computing, we find the following:

The matrix $[\mathbf{A} \,\vdots\, \mathbf{I}]$

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

Subtract the first row from the second and third to obtain

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

Add 2 times the second row to the third row to obtain

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

Subtract 2 times the second row from the first to obtain

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

Multiply the third row by $1/2$ to obtain

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

Add the third row to the first and then subtract the third row from the second to obtain

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

The matrix shown in color is $\mathbf{A}^{-1}$. [Compare equation (2).] ◆

It is convenient to have an expression for the inverse of a generic 2×2 matrix. The following formula is easily verified by mental arithmetic:

(3) $$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{if } ad - bc \neq 0\,.$$

The denominator in (3), whose nonvanishing is the crucial condition for the existence of the inverse, is known as the determinant.

Determinants. The **determinant** of a 2×2 matrix $\mathbf{A}$, denoted $\det \mathbf{A}$ or $|\mathbf{A}|$, is defined by

$$\det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

The determinants of higher-order square matrices can be defined recursively in terms of lower-order determinants, using the concept of the *minor*; the **minor** of a particular entry is the determinant of the submatrix formed when that entry's row and column are deleted. Then the determinant of an $n \times n$ matrix equals the alternating-sign sum of the products of the entries of the first row with their minors. For a 3×3 matrix $\mathbf{A}$ this looks like

$$\det \mathbf{A} := \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

For example,

$$\begin{vmatrix} 1 & 2 & 1 \\ 0 & 3 & 5 \\ 2 & 1 & -1 \end{vmatrix} = 1 \begin{vmatrix} 3 & 5 \\ 1 & -1 \end{vmatrix} - 2 \begin{vmatrix} 0 & 5 \\ 2 & -1 \end{vmatrix} + 1 \begin{vmatrix} 0 & 3 \\ 2 & 1 \end{vmatrix}$$

$$= 1(-3-5) - 2(0-10) + 1(0-6) = 6\,.$$

For a 4×4 matrix, we have

$$\begin{vmatrix} 4 & 3 & 2 & -6 \\ -2 & 1 & 2 & 1 \\ 3 & 0 & 3 & 5 \\ 5 & 2 & 1 & -1 \end{vmatrix} = 4 \begin{vmatrix} 1 & 2 & 1 \\ 0 & 3 & 5 \\ 2 & 1 & -1 \end{vmatrix} - 3 \begin{vmatrix} -2 & 2 & 1 \\ 3 & 3 & 5 \\ 5 & 1 & -1 \end{vmatrix}$$

$$+ 2 \begin{vmatrix} -2 & 1 & 1 \\ 3 & 0 & 5 \\ 5 & 2 & -1 \end{vmatrix} - (-6) \begin{vmatrix} -2 & 1 & 2 \\ 3 & 0 & 3 \\ 5 & 2 & 1 \end{vmatrix}.$$

As we have shown, the first minor is 6; and the others are computed the same way, resulting in

$$\begin{vmatrix} 4 & 3 & 2 & -6 \\ -2 & 1 & 2 & 1 \\ 3 & 0 & 3 & 5 \\ 5 & 2 & 1 & -1 \end{vmatrix} = 4(6) - 3(60) + 2(54) - (-6)(36) = 168 .$$

Although higher-order determinants can be calculated similarly, a more practical way to evaluate them involves the row-reduction of the matrix to upper triangular form. Here we will deal mainly with low-order determinants, and direct the reader to a linear algebra text for further discussion.[†]

Determinants have a geometric interpretation: $\det \mathbf{A}$ is the volume (in n-dimensional space) of the parallelepiped whose edges are given by the column vectors of $\mathbf{A}$. But their chief value lies in the role they play in the following theorem, which summarizes many of the results from linear algebra that we shall need, and in Cramer's rule, described in Appendix D.

Matrices and Systems of Equations

Theorem 1. Let $\mathbf{A}$ be an $n \times n$ matrix. The following statements are equivalent:

(a) $\mathbf{A}$ is singular (does not have an inverse).

(b) The determinant of $\mathbf{A}$ is zero.

(c) $\mathbf{Ax} = \mathbf{0}$ has nontrivial solutions ($\mathbf{x} \neq \mathbf{0}$).

(d) The columns (rows) of $\mathbf{A}$ form a linearly dependent set.

In part (d), the statement that the n columns of $\mathbf{A}$ are linearly dependent means that there exist scalars $c_1, \dots, c_n$, *not all zero*, such that

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = \mathbf{0} ,$$

where $\mathbf{a}_j$ is the vector forming the jth column of $\mathbf{A}$.

If $\mathbf{A}$ is a singular square matrix (so $\det \mathbf{A} = 0$), then $\mathbf{Ax} = \mathbf{0}$ has infinitely many solutions. Indeed, Theorem 1 asserts that there is a vector $\mathbf{x}_0 \neq \mathbf{0}$ such that $\mathbf{Ax}_0 = \mathbf{0}$, and we can get infinitely many other solutions by multiplying $\mathbf{x}_0$ by any scalar, i.e., taking $\mathbf{x} = c\mathbf{x}_0$. Furthermore, $\mathbf{Ax} = \mathbf{b}$ either has no solutions or it has infinitely many of them of the form

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h ,$$

where $\mathbf{x}_p$ is a *particular* solution to $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x}_h$ is any of the infinity of solutions to $\mathbf{Ax} = \mathbf{0}$ (see Problem 15). The resemblance of this situation to that of solving nonhomogeneous linear differential equations should be quite apparent.

To illustrate, in Example 3 of Section 9.2 (page 502) we saw that the system

$$\begin{bmatrix} 2 & 4 & 1 \\ 2 & 4 & 0 \\ -4 & -8 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ -10 \end{bmatrix}$$

has solutions

$$x_1 = 3 - 2s , \qquad x_2 = s , \qquad x_3 = 2 ; \quad -\infty < s < \infty .$$

[†]Your authors' personal favorite is *Fundamentals of Matrix Analysis with Applications*, by Edward Barry Saff and Arthur David Snider (John Wiley & Sons, Hoboken, New Jersey, 2016).

Writing these in matrix notation, we can identify the vectors $\mathbf{x}_p$ and $\mathbf{x}_h$ mentioned above:

$$\mathbf{x} = \begin{bmatrix} 3-2s \\ s \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + s\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} = \mathbf{x}_p + \mathbf{x}_h \,.$$

Note further that the determinant of $\mathbf{A}$ is indeed zero,

$$\det \mathbf{A} = 2\begin{vmatrix} 4 & 0 \\ -8 & 1 \end{vmatrix} - 4\begin{vmatrix} 2 & 0 \\ -4 & 1 \end{vmatrix} + 1\begin{vmatrix} 2 & 4 \\ -4 & -8 \end{vmatrix} = 2\cdot 4 - 4\cdot 2 + 1\cdot 0 = 0\,,$$

and that the linear dependence of the columns of $\mathbf{A}$ is exhibited by the identity

$$-2\begin{bmatrix} 2 \\ 2 \\ -4 \end{bmatrix} + 1\begin{bmatrix} 4 \\ 4 \\ -8 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

If $\mathbf{A}$ is a nonsingular square matrix (i.e., $\mathbf{A}$ has an inverse and $\det \mathbf{A} \neq 0$), then the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has $\mathbf{x} = \mathbf{0}$ as its only solution. More generally, when $\det \mathbf{A} \neq 0$, the system $\mathbf{Ax} = \mathbf{b}$ has a unique solution (namely, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$).

Calculus of Matrices

If we allow the entries $a_{ij}(t)$ in a matrix $\mathbf{A}(t)$ to be functions of the variable t, then $\mathbf{A}(t)$ is a **matrix function of** t. Similarly, if the entries $x_i(t)$ of a vector $\mathbf{x}(t)$ are functions of t, then $\mathbf{x}(t)$ is a **vector function of** t.

These matrix and vector functions have a calculus much like that of real-valued functions. A matrix $\mathbf{A}(t)$ is said to be **continuous at** t_0 if each entry $a_{ij}(t)$ is continuous at t_0. Moreover, $\mathbf{A}(t)$ is **differentiable at** t_0 if each entry $a_{ij}(t)$ is differentiable at t_0, and we write

$$\text{(4)} \qquad \frac{d\mathbf{A}}{dt}(t_0) = \mathbf{A}'(t_0) := [a'_{ij}(t_0)]\,.$$

Similarly, we define

$$\text{(5)} \qquad \int_a^b \mathbf{A}(t)\,dt := \left[\int_a^b a_{ij}(t)\,dt\right].$$

Example 2 Let $\mathbf{A}(t) = \begin{bmatrix} t^2+1 & \cos t \\ e^t & 1 \end{bmatrix}$.

Find: **(a)** $\mathbf{A}'(t)$. **(b)** $\displaystyle\int_0^1 \mathbf{A}(t)dt$.

Solution Using formulas (4) and (5), we compute

$$\textbf{(a)}\ \mathbf{A}'(t) = \begin{bmatrix} 2t & -\sin t \\ e^t & 0 \end{bmatrix}. \qquad \textbf{(b)}\ \int_0^1 \mathbf{A}(t)dt = \begin{bmatrix} \frac{4}{3} & \sin 1 \\ e-1 & 1 \end{bmatrix}. \ \blacklozenge$$

Example 3 Show that $\mathbf{x}(t) = \begin{bmatrix} \cos\omega t \\ \sin\omega t \end{bmatrix}$ is a solution of the matrix differential equation $\mathbf{x}' = \mathbf{Ax}$, where

$$\mathbf{A} = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}.$$

Solution We simply verify that $\mathbf{x}'(t)$ and $\mathbf{Ax}(t)$ are the same vector function:

$$\mathbf{x}'(t) = \begin{bmatrix} -\omega \sin \omega t \\ \omega \cos \omega t \end{bmatrix}; \qquad \mathbf{Ax} = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix} \begin{bmatrix} \cos \omega t \\ \sin \omega t \end{bmatrix} = \begin{bmatrix} -\omega \sin \omega t \\ \omega \cos \omega t \end{bmatrix}. \ \blacklozenge$$

The basic properties of differentiation are valid for matrix functions.

Differentiation Formulas for Matrix Functions

$$\frac{d}{dt}(\mathbf{CA}) = \mathbf{C}\frac{d\mathbf{A}}{dt} \qquad (\mathbf{C} \text{ a constant matrix}).$$

$$\frac{d}{dt}(\mathbf{A}+\mathbf{B}) = \frac{d\mathbf{A}}{dt} + \frac{d\mathbf{B}}{dt}.$$

$$\frac{d}{dt}(\mathbf{AB}) = \mathbf{A}\frac{d\mathbf{B}}{dt} + \frac{d\mathbf{A}}{dt}\mathbf{B}.$$

In the last formula, the order in which the matrices are written is very important because, as we have emphasized, matrix multiplication does not always commute.

9.3 EXERCISES

1. Let $\mathbf{A} := \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}$ and $\mathbf{B} := \begin{bmatrix} -1 & 0 \\ 2 & -3 \end{bmatrix}$.

Find: **(a)** $\mathbf{A} + \mathbf{B}$. **(b)** $3\mathbf{A} - \mathbf{B}$.

2. Let $\mathbf{A} := \begin{bmatrix} 2 & 0 & 5 \\ 2 & 1 & 1 \end{bmatrix}$ and $\mathbf{B} := \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & -2 \end{bmatrix}$.

Find: **(a)** $\mathbf{A} + \mathbf{B}$. **(b)** $7\mathbf{A} - 4\mathbf{B}$.

3. Let $\mathbf{A} := \begin{bmatrix} 2 & 4 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{B} := \begin{bmatrix} -1 & 3 \\ 5 & 2 \end{bmatrix}$.

Find: **(a)** $\mathbf{AB}$. **(b)** $\mathbf{A}^2 = \mathbf{AA}$. **(c)** $\mathbf{B}^2 = \mathbf{BB}$.

4. Let $\mathbf{A} := \begin{bmatrix} 2 & 1 \\ 0 & 4 \\ -1 & 3 \end{bmatrix}$ and $\mathbf{B} := \begin{bmatrix} 1 & 1 & -1 \\ 0 & 3 & 1 \end{bmatrix}$.

Find: **(a)** $\mathbf{AB}$. **(b)** $\mathbf{BA}$.

5. Let $\mathbf{A} := \begin{bmatrix} 1 & -2 \\ 2 & -3 \end{bmatrix}$, $\mathbf{B} := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, and

$\mathbf{C} := \begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix}$.

Find: **(a)** $\mathbf{AB}$. **(b)** $\mathbf{AC}$. **(c)** $\mathbf{A}(\mathbf{B} + \mathbf{C})$.

6. Let $\mathbf{A} := \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$, $\mathbf{B} := \begin{bmatrix} 0 & 3 \\ 1 & 2 \end{bmatrix}$, and

$\mathbf{C} := \begin{bmatrix} 1 & -4 \\ 1 & 1 \end{bmatrix}$.

Find: **(a)** $\mathbf{AB}$. **(b)** $(\mathbf{AB})\mathbf{C}$. **(c)** $(\mathbf{A} + \mathbf{B})\mathbf{C}$.

7. **(a)** Show that if $\mathbf{u}$ and $\mathbf{v}$ are each $n \times 1$ column vectors, then the matrix product $\mathbf{u}^T\mathbf{v}$ is the same as the dot product $\mathbf{u} \cdot \mathbf{v}$.

(b) Let $\mathbf{v}$ be a 3×1 column vector with $\mathbf{v}^T = [2 \quad 3 \quad 5]$. Show that, for $\mathbf{A}$ as given in Example 1 (page 509), $(\mathbf{Av})^T = \mathbf{v}^T\mathbf{A}^T$.

(c) Does $(\mathbf{Av})^T = \mathbf{v}^T\mathbf{A}^T$ hold for every $m \times n$ matrix $\mathbf{A}$ and $n \times 1$ vector $\mathbf{v}$?

(d) Does $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ hold for every pair of matrices $\mathbf{A}, \mathbf{B}$ such that both matrix products are defined? Justify your answer.

8. Let $\mathbf{A} := \begin{bmatrix} 2 & -1 \\ -3 & 4 \end{bmatrix}$ and $\mathbf{B} := \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$.

Verify that $\mathbf{AB} \neq \mathbf{BA}$.

In Problems 9–14, use the method of Example 1 to compute the inverse of the given matrix, if it exists. For Problems 9 and 10, confirm your answer by comparison with formula (3).

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

10. $\begin{bmatrix} 4 & 1 \\ 5 & 9 \end{bmatrix}$

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

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

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

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

15. Prove that if $\mathbf{x}_p$ satisfies $\mathbf{A}\mathbf{x}_p = \mathbf{b}$, then every solution to the nonhomogeneous system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is of the form $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_h$ is a solution to the corresponding homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$.

16. Let $\mathbf{A} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$.

(a) Show that $\mathbf{A}$ is singular.

(b) Show that $\mathbf{A}\mathbf{x} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}$ has no solutions.

(c) Show that $\mathbf{A}\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix}$ has infinitely many solutions.

In Problems 17–20, find the matrix function $\mathbf{X}^{-1}(t)$ whose value at t is the inverse of the given matrix $\mathbf{X}(t)$.

17. $\mathbf{X}(t) = \begin{bmatrix} e^t & e^{4t} \\ e^t & 4e^{4t} \end{bmatrix}$

18. $\mathbf{X}(t) = \begin{bmatrix} \sin 2t & \cos 2t \\ 2\cos 2t & -2\sin 2t \end{bmatrix}$

19. $\mathbf{X}(t) = \begin{bmatrix} e^t & e^{-t} & e^{2t} \\ e^t & -e^{-t} & 2e^{2t} \\ e^t & e^{-t} & 4e^{2t} \end{bmatrix}$

20. $\mathbf{X}(t) = \begin{bmatrix} e^{3t} & 1 & t \\ 3e^{3t} & 0 & 1 \\ 9e^{3t} & 0 & 0 \end{bmatrix}$

In Problems 21–26, evaluate the given determinant.

21. $\begin{vmatrix} 4 & 3 \\ -1 & 2 \end{vmatrix}$

22. $\begin{vmatrix} 12 & 8 \\ 3 & 2 \end{vmatrix}$

23. $\begin{vmatrix} 1 & 0 & 0 \\ 3 & 1 & 2 \\ 1 & 5 & -2 \end{vmatrix}$

24. $\begin{vmatrix} 1 & 0 & 2 \\ 0 & 3 & -1 \\ -1 & 2 & 1 \end{vmatrix}$

25. $\begin{vmatrix} 1 & 4 & 3 \\ -1 & -1 & 2 \\ 4 & 5 & 2 \end{vmatrix}$

26. $\begin{vmatrix} 1 & 4 & 4 \\ 3 & 0 & -3 \\ 1 & 6 & 2 \end{vmatrix}$

In Problems 27–29, determine the values of r for which $\det(\mathbf{A} - r\mathbf{I}) = 0$.

27. $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$

28. $\mathbf{A} = \begin{bmatrix} 3 & 3 \\ 2 & 4 \end{bmatrix}$

29. $\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

30. Illustrate the equivalence of the assertions (a)–(d) in Theorem 1 (page 511) for the matrix

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

as follows.

(a) Show that the row-reduction procedure applied to $[\mathbf{A} \,\vdots\, \mathbf{I}]$ fails to produce the inverse of $\mathbf{A}$.

(b) Calculate $\det \mathbf{A}$.

(c) Determine a nontrivial solution $\mathbf{x}$ to $\mathbf{A}\mathbf{x} = \mathbf{0}$.

(d) Find scalars c_1, c_2, and c_3, *not all zero*, so that $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3 = \mathbf{0}$, where $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_3$ are the columns of $\mathbf{A}$.

In Problems 31 and 32, find $d\mathbf{x}/dt$ for the given vector functions.

31. $\mathbf{x}(t) = \begin{bmatrix} e^{3t} \\ 2e^{3t} \\ -e^{3t} \end{bmatrix}$

32. $\mathbf{x}(t) = \begin{bmatrix} e^{-t}\sin 3t \\ 0 \\ -e^{-t}\sin 3t \end{bmatrix}$

In Problems 33 and 34, find $d\mathbf{X}/dt$ for the given matrix functions.

33. $\mathbf{X}(t) = \begin{bmatrix} e^{5t} & 3e^{2t} \\ -2e^{5t} & -e^{2t} \end{bmatrix}$

34. $\mathbf{X}(t) = \begin{bmatrix} \sin 2t & \cos 2t & e^{-2t} \\ -\sin 2t & 2\cos 2t & 3e^{-2t} \\ 3\sin 2t & \cos 2t & e^{-2t} \end{bmatrix}$

In Problems 35 and 36, verify that the given vector function satisfies the given system.

35. $\mathbf{x}' = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}\mathbf{x}, \quad \mathbf{x}(t) = \begin{bmatrix} e^{3t} \\ 2e^{3t} \end{bmatrix}$

36. $\mathbf{x}' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\mathbf{x}, \quad \mathbf{x}(t) = \begin{bmatrix} 0 \\ e^t \\ -3e^t \end{bmatrix}$

In Problems 37 and 38, verify that the given matrix function satisfies the given matrix differential equation.

37. $\mathbf{X}' = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}\mathbf{X}, \quad \mathbf{X}(t) = \begin{bmatrix} e^{2t} & e^{3t} \\ -e^{2t} & -2e^{3t} \end{bmatrix}$

38. $\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & -2 \\ 0 & -2 & 3 \end{bmatrix}\mathbf{X},$

$\mathbf{X}(t) = \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^t & e^{5t} \\ 0 & e^t & -e^{5t} \end{bmatrix}$

In Problems 39 and 40, the matrices $\mathbf{A}(t)$ and $\mathbf{B}(t)$ are given. Find

(a) $\int \mathbf{A}(t)dt$. **(b)** $\int_0^1 \mathbf{B}(t)dt$. **(c)** $\frac{d}{dt}[\mathbf{A}(t)\mathbf{B}(t)]$.

39. $\mathbf{A}(t) = \begin{bmatrix} t & e^t \\ 1 & e^t \end{bmatrix}, \quad \mathbf{B}(t) = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}$

40. $\mathbf{A}(t) = \begin{bmatrix} 1 & e^{-2t} \\ 3 & e^{-2t} \end{bmatrix}, \quad \mathbf{B}(t) = \begin{bmatrix} e^{-t} & e^{-t} \\ -e^{-t} & 3e^{-t} \end{bmatrix}$

41. An $n \times n$ matrix $\mathbf{A}$ is called **symmetric** if $\mathbf{A}^T = \mathbf{A}$; that is, if $a_{ij} = a_{ji}$, for all $i, j = 1, \ldots, n$. Show that if $\mathbf{A}$ is an $n \times n$ matrix, then $\mathbf{A} + \mathbf{A}^T$ is a symmetric matrix.

42. Let $\mathbf{A}$ be an $m \times n$ matrix. Show that $\mathbf{A}^T\mathbf{A}$ is a symmetric $n \times n$ matrix and $\mathbf{A}\mathbf{A}^T$ is a symmetric $m \times m$ matrix (see Problem 41).

43. The **inner product** of two vectors is a generalization of the dot product, for vectors with complex entries. It is defined by

$$(\mathbf{x}, \mathbf{y}) := \sum_{i=1}^{n} x_i \bar{y}_i , \qquad \text{where}$$

$\mathbf{x} = \text{col}\,(x_1, x_2, \ldots, x_n)$, $\mathbf{y} = \text{col}\,(y_1, y_2, \ldots, y_n)$ are complex vectors and the overbar denotes complex conjugation.

(a) Show that $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\bar{\mathbf{y}}$, where $\bar{\mathbf{y}} = \text{col}\,(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n)$.

(b) Prove that for any $n \times 1$ vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ and any complex number λ, we have

$$(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})} ,$$
$$(\mathbf{x}, \mathbf{y} + \mathbf{z}) = (\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{z}) ,$$
$$(\lambda\mathbf{x}, \mathbf{y}) = \lambda(\mathbf{x}, \mathbf{y}) , \quad (\mathbf{x}, \lambda\mathbf{y}) = \bar{\lambda}(\mathbf{x}, \mathbf{y}) .$$

9.4 Linear Systems in Normal Form

In keeping with the introduction presented in Section 9.1, we say that a system of n linear differential equations is in **normal form** if it is expressed as

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t) , \tag{1}$$

where $\mathbf{x}(t) = \text{col}\big(x_1(t), \ldots, x_n(t)\big)$, $\mathbf{f}(t) = \text{col}\big(f_1(t), \ldots, f_n(t)\big)$, and $\mathbf{A}(t) = [a_{ij}(t)]$ is an $n \times n$ matrix. As with a scalar linear differential equation, a system is called **homogeneous** when $\mathbf{f}(t) \equiv \mathbf{0}$; otherwise, it is called **nonhomogeneous**. When the elements of $\mathbf{A}$ are all constants, the system is said to have **constant coefficients**. Recall that an nth-order linear differential equation

$$y^{(n)}(t) + p_{n-1}(t)y^{(n-1)}(t) + \cdots + p_0(t)y(t) = g(t) \tag{2}$$

can be rewritten as a first-order system in normal form using the substitution $x_1(t) := y(t)$, $x_2(t) := y'(t), \ldots, x_n(t) := y^{(n-1)}(t)$; indeed, equation (2) is equivalent to $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t)$, where $\mathbf{x}(t) = \text{col}\big(x_1(t), \ldots, x_n(t)\big)$, $\mathbf{f}(t) := \text{col}\big(0, \ldots, 0, g(t)\big)$, and

$$\mathbf{A}(t) := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -p_0(t) & -p_1(t) & -p_2(t) & \cdots & -p_{n-2}(t) & -p_{n-1}(t) \end{bmatrix} .$$

The theory for systems in normal form parallels very closely the theory of linear differential equations presented in Chapters 4 and 6. In many cases the proofs for scalar linear differential equations carry over to normal systems with appropriate modifications. Conversely, results for normal systems apply to scalar linear equations since, as we showed, any scalar linear equation can be expressed as a normal system. This is the case with the existence and uniqueness theorems for linear differential equations.

The **initial value problem** for the normal system (1) is the problem of finding a differentiable vector function $\mathbf{x}(t)$ that satisfies the system on an interval I and also satisfies the **initial condition** $\mathbf{x}(t_0) = \mathbf{x}_0$, where t_0 is a given point of I and $\mathbf{x}_0 = \text{col}(x_{1,0}, \ldots, x_{n,0})$ is a given vector.

Existence and Uniqueness

Theorem 2. If $\mathbf{A}(t)$ and $\mathbf{f}(t)$ are continuous on an open interval I that contains the point t_0, then for any choice of the initial vector $\mathbf{x}_0$, there exists a unique solution $\mathbf{x}(t)$ on the whole interval I to the initial value problem

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t) , \qquad \mathbf{x}(t_0) = \mathbf{x}_0 .$$

We give a proof of this result in Chapter 13[†] and obtain as corollaries the existence and uniqueness theorems for second-order equations (Theorem 4, Section 4.5, page 182) and higher-order linear equations (Theorem 1, Section 6.1, page 319).

If we rewrite system (1) as $\mathbf{x}' - \mathbf{A}\mathbf{x} = \mathbf{f}$ and define the operator $L[\mathbf{x}] := \mathbf{x}' - \mathbf{A}\mathbf{x}$, then we can express system (1) in the operator form $L[\mathbf{x}] = \mathbf{f}$. Here the operator L maps vector functions into vector functions. Moreover, L is a *linear* operator in the sense that for any scalars a, b and differentiable vector functions $\mathbf{x}$, $\mathbf{y}$, we have

$$L[a\mathbf{x} + b\mathbf{y}] = aL[\mathbf{x}] + bL[\mathbf{y}] .$$

The proof of this linearity follows from the properties of matrix multiplication (see Problem 27).

As a consequence of the linearity of L, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are solutions to the *homogeneous* system $\mathbf{x}' = \mathbf{A}\mathbf{x}$, or $L[\mathbf{x}] = \mathbf{0}$ in operator notation, then any linear combination of these vectors, $c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$, is also a solution. Moreover, we will see that if the solutions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent, then *every* solution to $L[\mathbf{x}] = \mathbf{0}$ can be expressed as $c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$ for an appropriate choice of the constants $c_1, \ldots, c_n$.

Linear Dependence of Vector Functions

Definition 1. The m vector functions $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are said to be **linearly dependent on an interval** I if there exist constants $c_1, \ldots, c_m$, not all zero, such that

$$c_1\mathbf{x}_1(t) + \cdots + c_m\mathbf{x}_m(t) = \mathbf{0} \tag{3}$$

for all t in I. If the vectors are not linearly dependent, they are said to be **linearly independent on** I.

Example 1 Show that the vector functions $\mathbf{x}_1(t) = \text{col}(e^t, 0, e^t)$, $\mathbf{x}_2(t) = \text{col}(3e^t, 0, 3e^t)$, and $\mathbf{x}_3(t) = \text{col}(t, 1, 0)$ are linearly dependent on $(-\infty, \infty)$.

Solution Notice that $\mathbf{x}_2$ is just 3 times $\mathbf{x}_1$ and therefore $3\mathbf{x}_1(t) - \mathbf{x}_2(t) + 0 \cdot \mathbf{x}_3(t) = \mathbf{0}$ for all t. Hence, $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are linearly dependent on $(-\infty, \infty)$. ◆

Example 2 Show that

$$\mathbf{x}_1(t) = \begin{bmatrix} t \\ |t| \end{bmatrix} , \quad \mathbf{x}_2(t) = \begin{bmatrix} |t| \\ t \end{bmatrix}$$

are linearly independent on $(-\infty, \infty)$.

[†]All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems*, 7th ed.

Solution Note that at every instant t_0, the column vector $\mathbf{x}_1(t_0)$ is a multiple of $\mathbf{x}_2(t_0)$; indeed, $\mathbf{x}_1(t_0) = \mathbf{x}_2(t_0)$ for $t_0 \geq 0$, and $\mathbf{x}_1(t_0) = -\mathbf{x}_2(t_0)$ for $t_0 \leq 0$. Nonetheless, the vector functions are not dependent, because the c's in condition (3) are not allowed to change with t; for $t < 0$, the equation $c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) = \mathbf{0}$ implies $c_1 - c_2 = 0$, but for $t > 0$ it implies $c_1 + c_2 = 0$. Thus $c_1 = c_2 = 0$ and the functions are independent. ◆

Example 3 Show that the vector functions $\mathbf{x}_1(t) = \text{col}(e^{2t}, 0, e^{2t})$, $\mathbf{x}_2(t) = \text{col}(e^{2t}, e^{2t}, -e^{2t})$, and $\mathbf{x}_3(t) = \text{col}(e^{t}, 2e^{t}, e^{t})$ are linearly independent on $(-\infty, \infty)$.

Solution To prove independence, we *assume* c_1, c_2, and c_3 are constants for which

$$c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + c_3\mathbf{x}_3(t) = \mathbf{0}$$

holds at every t in $(-\infty, \infty)$ and show that this forces $c_1 = c_2 = c_3 = 0$. In particular, when $t = 0$ we obtain

$$c_1\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + c_3\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \mathbf{0}\,,$$

which is equivalent to the system of linear equations

$$\begin{aligned} c_1 + c_2 + c_3 &= 0\,, \\ c_2 + 2c_3 &= 0\,, \\ c_1 - c_2 + c_3 &= 0\,. \end{aligned} \tag{4}$$

Either by solving (4) or by checking that the determinant of its coefficients is nonzero (recall Theorem 1 on page 511), we can verify that (4) has only the trivial solution $c_1 = c_2 = c_3 = 0$. Therefore the vector functions $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are linearly independent on $(-\infty, \infty)$ (in fact, on any interval containing $t = 0$). ◆

As Example 3 illustrates, if $\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_n(t)$ are n vector functions, each having n components, we can establish their linear independence on an interval I if we can find *one point* t_0 in I where the determinant

$$\det[\mathbf{x}_1(t_0) \ldots \mathbf{x}_n(t_0)]$$

is not zero. Because of the analogy with scalar equations, we call this determinant the **Wronskian**.

Wronskian

Definition 2. The **Wronskian** of n vector functions $\mathbf{x}_1(t) = \text{col}(x_{1,1}, \ldots, x_{n,1}), \ldots,$ $\mathbf{x}_n(t) = \text{col}(x_{1,n}, \ldots, x_{n,n})$ is defined to be the function

$$W[\mathbf{x}_1, \ldots, \mathbf{x}_n](t) := \begin{vmatrix} x_{1,1}(t) & x_{1,2}(t) & \cdots & x_{1,n}(t) \\ x_{2,1}(t) & x_{2,2}(t) & \cdots & x_{2,n}(t) \\ \vdots & \vdots & & \vdots \\ x_{n,1}(t) & x_{n,2}(t) & \cdots & x_{n,n}(t) \end{vmatrix}.$$

We see that n vector functions are linearly independent on an interval if their Wronskian is nonzero *at any point in the interval*. But now we show that if these functions happen to be independent solutions to a homogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is an $n \times n$ matrix of continuous functions, then the Wronskian is *never* zero on I. For suppose to the contrary that $W(t_0) = 0$. Then by Theorem 1 the vanishing of the determinant implies that the column vectors $\mathbf{x}_1(t_0), \mathbf{x}_2(t_0), \ldots, \mathbf{x}_n(t_0)$ are linearly dependent. Thus there exist scalars $c_1, \ldots, c_n$ *not all zero*, such that at t_0

$$c_1\mathbf{x}_1(t_0) + \cdots + c_n\mathbf{x}_n(t_0) = \mathbf{0} .$$

However, $c_1\mathbf{x}_1(t) + \cdots + c_n\mathbf{x}_n(t)$ and the vector function $\mathbf{z}(t) \equiv \mathbf{0}$ are both solutions to $\mathbf{x}' = \mathbf{A}\mathbf{x}$ on I, and they agree at the point t_0. So these solutions must be identical on I according to the existence-uniqueness theorem (Theorem 2, page 516). That is,

$$c_1\mathbf{x}_1(t) + \cdots + c_n\mathbf{x}_n(t) = \mathbf{0}$$

for *all* t in I. But this contradicts the given information that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent on I. We have shown that $W(t_0) \neq 0$, and since t_0 is an arbitrary point, it follows that $W(t) \neq 0$ for all $t \in I$.

The preceding argument has two important implications that parallel the scalar case. First, the Wronskian of *solutions* to $\mathbf{x}' = \mathbf{A}\mathbf{x}$ *is either identically zero or never zero on* I (see also Problem 33). Second, a set of n *solutions* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to $\mathbf{x}' = \mathbf{A}\mathbf{x}$ on I *is linearly independent on* I *if and only if their Wronskian is never zero on* I. With these facts in hand, we can imitate the proof given for the scalar case in Section 6.1 (Theorem 2, page 322) to obtain the following representation theorem for the solutions to $\mathbf{x}' = \mathbf{A}\mathbf{x}$.

Representation of Solutions (Homogeneous Case)

Theorem 3. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be n linearly independent solutions to the homogeneous system

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) \tag{5}$$

on the interval I, where $\mathbf{A}(t)$ is an $n \times n$ matrix function continuous on I. Then every solution to (5) on I can be expressed in the form

$$\mathbf{x}(t) = c_1\mathbf{x}_1(t) + \cdots + c_n\mathbf{x}_n(t) , \tag{6}$$

where $c_1, \ldots, c_n$ are constants.

A set of solutions $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ that are linearly independent on I or, equivalently, whose Wronskian does not vanish on I, is called a **fundamental solution set** for (5) on I. The linear combination in (6), written with arbitrary constants, is referred to as a **general solution** to (5).

If we take the vectors in a fundamental solution set and let them form the columns of a matrix $\mathbf{X}(t)$, that is,

$$\mathbf{X}(t) = [\mathbf{x}_1(t)\,\mathbf{x}_2(t)\ldots\mathbf{x}_n(t)] = \begin{bmatrix} x_{1,1}(t) & x_{1,2}(t) & \cdots & x_{1,n}(t) \\ x_{2,1}(t) & x_{2,2}(t) & \cdots & x_{2,n}(t) \\ \vdots & \vdots & & \vdots \\ x_{n,1}(t) & x_{n,2}(t) & \cdots & x_{n,n}(t) \end{bmatrix},$$

then the matrix $\mathbf{X}(t)$ is called a **fundamental matrix** for (5). We can use it to express the general solution (6) as

$$\mathbf{x}(t) = \mathbf{X}(t)\mathbf{c} ,$$

where $\mathbf{c} = \text{col}(c_1, \ldots, c_n)$ is an arbitrary constant vector. Since $\det \mathbf{X} = W[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ is never zero on I, it follows from Theorem 1 on page 511 that $\mathbf{X}(t)$ is invertible for every t in I.

Example 4 Verify that the set

$$S = \left\{ \begin{bmatrix} e^{2t} \\ e^{2t} \\ e^{2t} \end{bmatrix}, \begin{bmatrix} -e^{-t} \\ 0 \\ e^{-t} \end{bmatrix}, \begin{bmatrix} 0 \\ e^{-t} \\ -e^{-t} \end{bmatrix} \right\}$$

is a fundamental solution set for the system

$$\text{(7)} \qquad \mathbf{x}'(t) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \mathbf{x}(t)$$

on the interval $(-\infty, \infty)$ and find a fundamental matrix for (7). Also determine a general solution for (7).

Solution Substituting the first vector in the set S into the right-hand side of (7) gives

$$\mathbf{Ax} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} e^{2t} \\ e^{2t} \\ e^{2t} \end{bmatrix} = \begin{bmatrix} 2e^{2t} \\ 2e^{2t} \\ 2e^{2t} \end{bmatrix} = \mathbf{x}'(t) .$$

Hence this vector satisfies system (7) for all t. Similar computations verify that the remaining vectors in S are also solutions to (7) on $(-\infty, \infty)$. For us to show that S is a fundamental solution set, it is enough to observe that the Wronskian

$$W(t) = \begin{vmatrix} e^{2t} & -e^{-t} & 0 \\ e^{2t} & 0 & e^{-t} \\ e^{2t} & e^{-t} & -e^{-t} \end{vmatrix} = e^{2t} \begin{vmatrix} 0 & e^{-t} \\ e^{-t} & -e^{-t} \end{vmatrix} + e^{-t} \begin{vmatrix} e^{2t} & e^{-t} \\ e^{2t} & -e^{-t} \end{vmatrix} = -3$$

is never zero.

A fundamental matrix $\mathbf{X}(t)$ for (7) is just the matrix we used to compute the Wronskian; that is,

$$\text{(8)} \qquad \mathbf{X}(t) \coloneqq \begin{bmatrix} e^{2t} & -e^{-t} & 0 \\ e^{2t} & 0 & e^{-t} \\ e^{2t} & e^{-t} & -e^{-t} \end{bmatrix} .$$

A general solution to (7) can now be expressed as

$$\mathbf{x}(t) = \mathbf{X}(t)\mathbf{c} = c_1 \begin{bmatrix} e^{2t} \\ e^{2t} \\ e^{2t} \end{bmatrix} + c_2 \begin{bmatrix} -e^{-t} \\ 0 \\ e^{-t} \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ e^{-t} \\ -e^{-t} \end{bmatrix} . \quad ◆$$

It is easy to check that the fundamental matrix in (8) satisfies the equation

$$\mathbf{X}'(t) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \mathbf{X}(t) ;$$

indeed, this is equivalent to showing that $\mathbf{x}' = \mathbf{Ax}$ for each column $\mathbf{x}$ in S. In general, a fundamental matrix for a system $\mathbf{x}' = \mathbf{Ax}$ satisfies the corresponding **matrix differential equation** $\mathbf{X}' = \mathbf{AX}$.

Another consequence of the linearity of the operator L defined by $L[\mathbf{x}] \coloneqq \mathbf{x}' - \mathbf{Ax}$ is the **superposition principle** for linear systems. It states that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are solutions, respectively, to the nonhomogeneous systems

$$L[\mathbf{x}] = \mathbf{g}_1 \text{ and } L[\mathbf{x}] = \mathbf{g}_2 ,$$

then $c_1\mathbf{x}_1 + c_2\mathbf{x}_2$ is a solution to

$$L[\mathbf{x}] = c_1\mathbf{g}_1 + c_2\mathbf{g}_2 .$$

Using the superposition principle and the representation theorem for homogeneous systems, we can prove the following theorem.

Representation of Solutions (Nonhomogeneous Case)

Theorem 4. If $\mathbf{x}_p$ is a particular solution to the nonhomogeneous system

(9) $$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t)$$

on the interval I and $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a fundamental solution set on I for the corresponding homogeneous system $\mathbf{x}(t) = \mathbf{A}(t)\mathbf{x}(t)$, then every solution to (9) on I can be expressed in the form

(10) $$\mathbf{x}(t) = \mathbf{x}_p(t) + c_1\mathbf{x}_1(t) + \cdots + c_n\mathbf{x}_n(t) ,$$

where $c_1, \ldots, c_n$ are constants.

The proof of this theorem is almost identical to the proofs of Theorem 4 in Section 4.5 (page 182) and Theorem 4 in Section 6.1 (page 325). We leave the proof as an exercise.

The linear combination of $\mathbf{x}_p, \mathbf{x}_1, \ldots, \mathbf{x}_n$ in (10) written with arbitrary constants $c_1, \ldots, c_n$ is called a **general solution** of (9). This general solution can also be expressed as $\mathbf{x} = \mathbf{x}_p + \mathbf{X}\mathbf{c}$, where $\mathbf{X}$ is a fundamental matrix for the homogeneous system and $\mathbf{c}$ is an arbitrary constant vector.

We now summarize the results of this section as they apply to the problem of finding a general solution to a system of n linear first-order differential equations in normal form.

Approach to Solving Normal Systems

1. To determine a general solution to the $n \times n$ homogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x}$:

(a) Find a fundamental solution set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ that consists of n linearly independent solutions to the homogeneous system.

(b) Form the linear combination

$$\mathbf{x} = \mathbf{X}\mathbf{c} = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n ,$$

where $\mathbf{c} = \operatorname{col}(c_1, \ldots, c_n)$ is any constant vector and $\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_n]$ is the fundamental matrix, to obtain a general solution.

2. To determine a general solution to the nonhomogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}$:

(a) Find a particular solution $\mathbf{x}_p$ to the nonhomogeneous system.

(b) Form the sum of the particular solution and the general solution $\mathbf{X}\mathbf{c} = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$ to the corresponding homogeneous system in part 1,

$$\mathbf{x} = \mathbf{x}_p + \mathbf{X}\mathbf{c} = \mathbf{x}_p + c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n ,$$

to obtain a general solution to the given system.

We devote the rest of this chapter to methods for finding fundamental solution sets for homogeneous systems and particular solutions for nonhomogeneous systems.

9.4 EXERCISES

In Problems 1–4, write the given system in the matrix form $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}$.

1. $x'(t) = 3x(t) - y(t) + t^2\,,$
   $y'(t) = -x(t) + 2y(t) + e^t$

2. $r'(t) = 2r(t) + \sin t\,,$
   $\theta'(t) = r(t) - \theta(t) + 1$

3. $\dfrac{dx}{dt} = t^2x - y - z + t\,,$
   $\dfrac{dy}{dt} = e^t z + 5\,,$
   $\dfrac{dz}{dt} = tx - y + 3z - e^t$

4. $\dfrac{dx}{dt} = x + y + z\,,$
   $\dfrac{dy}{dt} = 2x - y + 3z\,,$
   $\dfrac{dz}{dt} = x + 5z$

In Problems 5–8, rewrite the given scalar equation as a first-order system in normal form. Express the system in the matrix form $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}$.

5. $y''(t) - 3y'(t) - 10y(t) = \sin t$

6. $x''(t) + x(t) = t^2$

7. $\dfrac{d^4w}{dt^4} + w = t^2$

8. $\dfrac{d^3y}{dt^3} - \dfrac{dy}{dt} + y = \cos t$

In Problems 9–12, write the given system as a set of scalar equations.

9. $\mathbf{x}' = \begin{bmatrix} 5 & 0 \\ -2 & 4 \end{bmatrix}\mathbf{x} + e^{-2t}\begin{bmatrix} 2 \\ -3 \end{bmatrix}$

10. $\mathbf{x}' = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}\mathbf{x} + e^{t}\begin{bmatrix} t \\ 1 \end{bmatrix}$

11. $\mathbf{x}' = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & 5 \\ 0 & 5 & 1 \end{bmatrix}\mathbf{x} + e^{t}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

12. $\mathbf{x}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 2 \end{bmatrix}\mathbf{x} + t\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$

In Problems 13–19, determine whether the given vector functions are linearly dependent (LD) or linearly independent (LI) on the interval $(-\infty, \infty)$.

13. $\begin{bmatrix} t \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix}$

14. $\begin{bmatrix} te^{-t} \\ e^{-t} \end{bmatrix}, \begin{bmatrix} e^{-t} \\ e^{-t} \end{bmatrix}$

15. $e^t\begin{bmatrix} 1 \\ 5 \end{bmatrix}, e^t\begin{bmatrix} -3 \\ -15 \end{bmatrix}$

16. $\begin{bmatrix} \sin t \\ \cos t \end{bmatrix}, \begin{bmatrix} \sin 2t \\ \cos 2t \end{bmatrix}$

17. $e^{2t}\begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}, e^{2t}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, e^{3t}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

18. $\begin{bmatrix} \sin t \\ \cos t \end{bmatrix}, \begin{bmatrix} \sin t \\ \sin t \end{bmatrix}, \begin{bmatrix} \cos t \\ \cos t \end{bmatrix}$

19. $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} t \\ 0 \\ t \end{bmatrix}, \begin{bmatrix} t^2 \\ 0 \\ t^2 \end{bmatrix}$

20. Let
$$\mathbf{x}_1 = \begin{bmatrix} \cos t \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} \sin t \\ \cos t \\ \cos t \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} \cos t \\ \sin t \\ \cos t \end{bmatrix}.$$
   (a) Compute the Wronskian.
   (b) Are these vector functions linearly independent on $(-\infty, \infty)$?
   (c) Is there a first-order homogeneous linear system for which these functions are solutions?

In Problems 21–24, the given vector functions are solutions to a system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$. *Determine whether they form a fundamental solution set. If they do, find a fundamental matrix for the system and give a general solution.*

21. $\mathbf{x}_1 = e^{2t}\begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \mathbf{x}_2 = e^{2t}\begin{bmatrix} -2 \\ 4 \end{bmatrix}$

22. $\mathbf{x}_1 = e^{-t}\begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad \mathbf{x}_2 = e^{4t}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

23. $\mathbf{x}_1 = \begin{bmatrix} e^{-t} \\ 2e^{-t} \\ e^{-t} \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} e^{t} \\ 0 \\ e^{t} \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} e^{3t} \\ -e^{3t} \\ 2e^{3t} \end{bmatrix}$

24. $\mathbf{x}_1 = \begin{bmatrix} e^{t} \\ e^{t} \\ e^{t} \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} \sin t \\ \cos t \\ -\sin t \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} -\cos t \\ \sin t \\ \cos t \end{bmatrix}$

25. Verify that the vector functions
$$\mathbf{x}_1 = \begin{bmatrix} e^{t} \\ e^{t} \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} e^{-t} \\ 3e^{-t} \end{bmatrix}$$
   are solutions to the homogeneous system
$$\mathbf{x}' = \mathbf{A}\mathbf{x} = \begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}\mathbf{x}\,,$$
   on $(-\infty, \infty)$, and that
$$\mathbf{x}_p = \frac{3}{2}\begin{bmatrix} te^{t} \\ te^{t} \end{bmatrix} - \frac{1}{4}\begin{bmatrix} e^{t} \\ 3e^{t} \end{bmatrix} + \begin{bmatrix} t \\ 2t \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
   is a particular solution to the nonhomogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}(t)$, where $\mathbf{f}(t) = \operatorname{col}(e^t, t)$. Find a general solution to $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}(t)$.

26. Verify that the vector functions
$$\mathbf{x}_1 = \begin{bmatrix} e^{3t} \\ 0 \\ e^{3t} \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} -e^{3t} \\ e^{3t} \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} -e^{-3t} \\ -e^{-3t} \\ e^{-3t} \end{bmatrix}$$

are solutions to the homogeneous system

$$\mathbf{x}' = \mathbf{A}\mathbf{x} = \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \mathbf{x}$$

on $(-\infty, \infty)$, and that

$$\mathbf{x}_p = \begin{bmatrix} 5t+1 \\ 2t \\ 4t+2 \end{bmatrix}$$

is a particular solution to $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}(t)$, where $\mathbf{f}(t) = \text{col}(-9t, 0, -18t)$. Find a general solution to $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}(t)$.

27. Prove that the operator defined by $L[\mathbf{x}] \coloneqq \mathbf{x}' - \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is an $n \times n$ matrix function and $\mathbf{x}$ is an $n \times 1$ differentiable vector function, is a linear operator.

28. Let $\mathbf{X}(t)$ be a fundamental matrix for the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$. Show that $\mathbf{x}(t) = \mathbf{X}(t)\mathbf{X}^{-1}(t_0)\mathbf{x}_0$ is the solution to the initial value problem $\mathbf{x}' = \mathbf{A}\mathbf{x}$, $\mathbf{x}(t_0) = \mathbf{x}_0$.

In Problems 29–30, verify that $\mathbf{X}(t)$ is a fundamental matrix for the given system and compute $\mathbf{X}^{-1}(t)$. Use the result of Problem 28 to find the solution to the given initial value problem.

29. $\mathbf{x}' = \begin{bmatrix} 0 & 6 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \mathbf{x}, \quad \mathbf{x}(0) = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix};$

$$\mathbf{X}(t) = \begin{bmatrix} 6e^{-t} & -3e^{-2t} & 2e^{3t} \\ -e^{-t} & e^{-2t} & e^{3t} \\ -5e^{-t} & e^{-2t} & e^{3t} \end{bmatrix}$$

30. $\mathbf{x}' = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \mathbf{x}, \quad \mathbf{x}(0) = \begin{bmatrix} 3 \\ -1 \end{bmatrix};$

$$\mathbf{X}(t) = \begin{bmatrix} e^{-t} & e^{5t} \\ -e^{-t} & e^{5t} \end{bmatrix}$$

31. Show that

$$\begin{vmatrix} t^2 & t|t| \\ 2t & 2|t| \end{vmatrix} \equiv 0$$

on $(-\infty, \infty)$, but that the two vector functions

$$\begin{bmatrix} t^2 \\ 2t \end{bmatrix}, \quad \begin{bmatrix} t|t| \\ 2|t| \end{bmatrix}$$

are linearly independent on $(-\infty, \infty)$.

32. Abel's Formula. If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are any n solutions to the $n \times n$ system $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$, then Abel's formula gives a representation for the Wronskian $W(t) \coloneqq W[\mathbf{x}_1, \ldots, \mathbf{x}_n](t)$. Namely,

$$W(t) = W(t_0)\exp\left(\int_{t_0}^{t} \{a_{11}(s) + \cdots + a_{nn}(s)\}\, ds\right),$$

where $a_{11}(s), \ldots, a_{nn}(s)$ are the main diagonal elements of $\mathbf{A}(s)$. Prove this formula in the special case when $n = 3$. [*Hint:* Follow the outline in Problem 30 of Exercises 6.1, page 327.]

33. Using Abel's formula (Problem 32), confirm that the Wronskian of n solutions to $\mathbf{x}' = \mathbf{A}\mathbf{x}$ on the interval I is either identically zero on I or never zero on I.

34. Prove that a fundamental solution set for the homogeneous system $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$ always exists on an interval I, provided $\mathbf{A}(t)$ is continuous on I. [*Hint:* Use the existence and uniqueness theorem (Theorem 2) and make judicious choices for $\mathbf{x}_0$.]

35. Prove Theorem 3 on the representation of solutions of the homogeneous system.

36. Prove Theorem 4 on the representation of solutions of the nonhomogeneous system.

37. To illustrate the connection between a higher-order equation and the equivalent first-order system, consider the equation

$$y'''(t) - 6y''(t) + 11y'(t) - 6y(t) = 0\,. \tag{11}$$

(a) Show that $\{e^t, e^{2t}, e^{3t}\}$ is a fundamental solution set for (11).

(b) Using the definition in Section 6.1, compute the Wronskian of $\{e^t, e^{2t}, e^{3t}\}$.

(c) Setting $x_1 = y, x_2 = y', x_3 = y''$, show that equation (11) is equivalent to the first-order system

$$\mathbf{x}' = \mathbf{A}\mathbf{x}\,, \tag{12}$$

where

$$\mathbf{A} \coloneqq \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{bmatrix}.$$

(d) The substitution used in part (c) suggests that

$$S \coloneqq \left\{ \begin{bmatrix} e^t \\ e^t \\ e^t \end{bmatrix}, \begin{bmatrix} e^{2t} \\ 2e^{2t} \\ 4e^{2t} \end{bmatrix}, \begin{bmatrix} e^{3t} \\ 3e^{3t} \\ 9e^{3t} \end{bmatrix} \right\}$$

is a fundamental solution set for system (12). Verify that this is the case.

(e) Compute the Wronskian of S. How does it compare with the Wronskian computed in part (b)?

38. Define $\mathbf{x}_1(t), \mathbf{x}_2(t)$, and $\mathbf{x}_3(t)$, for $-\infty < t < \infty$, by

$$\mathbf{x}_1(t) = \begin{bmatrix} \sin t \\ \sin t \\ 0 \end{bmatrix}, \quad \mathbf{x}_2(t) = \begin{bmatrix} \sin t \\ 0 \\ \sin t \end{bmatrix}, \quad \mathbf{x}_3(t) = \begin{bmatrix} 0 \\ \sin t \\ \sin t \end{bmatrix}.$$

(a) Show that for the three scalar functions *in each individual row* there are nontrivial linear combinations that sum to zero for all t.

(b) Show that, nonetheless, the three vector functions are linearly independent. (No single nontrivial combination works *for each row, for all t*.)

(c) Calculate the Wronskian $W[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3](t)$.

(d) Is there a linear third-order homogeneous differential equation system having $\mathbf{x}_1(t), \mathbf{x}_2(t)$, and $\mathbf{x}_3(t)$ as solutions?

9.5 Homogeneous Linear Systems with Constant Coefficients

In this section we discuss a procedure for obtaining a general solution for the homogeneous system

$$\textbf{(1)} \qquad \mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)\,,$$

where $\mathbf{A}$ is a (real) *constant* $n \times n$ matrix. The general solution we seek will be defined for all t because the elements of $\mathbf{A}$ are just constant functions, which are continuous on $(-\infty, \infty)$ (recall Theorem 2, page 516). In Section 9.4 we showed that a general solution to (1) can be constructed from a fundamental solution set consisting of n linearly independent solutions to (1). Thus our goal is to find n such vector solutions.

In Chapter 4 we were successful in solving homogeneous linear equations with constant coefficients by guessing that the equation had a solution of the form e^{rt}. Because any scalar linear equation can be expressed as a system, it is reasonable to expect system (1) to have solutions of the form

$$\mathbf{x}(t) = e^{rt}\mathbf{u}\,,$$

where r is a constant and $\mathbf{u}$ is a constant vector, both of which must be determined. Substituting $e^{rt}\mathbf{u}$ for $\mathbf{x}(t)$ in (1) gives

$$re^{rt}\mathbf{u} = \mathbf{A}e^{rt}\mathbf{u} = e^{rt}\mathbf{A}\mathbf{u}\,.$$

Canceling the factor e^{rt} and rearranging terms, we find that

$$\textbf{(2)} \qquad (\mathbf{A} - r\mathbf{I})\mathbf{u} = \mathbf{0}\,,$$

where $r\mathbf{I}$ denotes the diagonal matrix with r's along its main diagonal.

The preceding calculation shows that $\mathbf{x}(t) = e^{rt}\mathbf{u}$ is a solution to (1) if and only if r and $\mathbf{u}$ satisfy equation (2). Since the trivial case, $\mathbf{u} = \mathbf{0}$, is of no help in finding linearly independent solutions to (1), we require that $\mathbf{u} \neq \mathbf{0}$. Such vectors are given a special name, as follows.

Eigenvalues and Eigenvectors

Definition 3. Let $\mathbf{A} = [a_{ij}]$ be an $n \times n$ constant matrix. The **eigenvalues** of $\mathbf{A}$ are those (real or complex) numbers r for which $(\mathbf{A} - r\mathbf{I})\mathbf{u} = \mathbf{0}$ has at least one nontrivial (real or complex) solution $\mathbf{u}$. The corresponding nontrivial solutions $\mathbf{u}$ are called the **eigenvectors** of $\mathbf{A}$ associated with r.

As stated in Theorem 1 of Section 9.3, a linear homogeneous system of n *algebraic* equations in n unknowns has a nontrivial solution if and only if the determinant of its coefficients is zero. Hence, a necessary and sufficient condition for (2) to have a nontrivial solution is that

$$\textbf{(3)} \qquad |\mathbf{A} - r\mathbf{I}| = 0\,.$$

Expanding the determinant of $\mathbf{A} - r\mathbf{I}$ in terms of its cofactors, we find that it is an nth-degree polynomial in r; that is,

$$\textbf{(4)} \qquad |\mathbf{A} - r\mathbf{I}| = p(r)\,.$$

Therefore, *finding the eigenvalues of a matrix* $\mathbf{A}$ *is equivalent to finding the zeros of the polynomial* $p(r)$. Equation (3) is called the **characteristic equation** of $\mathbf{A}$, and $p(r)$ in (4) is the **characteristic polynomial** of $\mathbf{A}$. The characteristic equation plays a role for systems similar to the role played by the auxiliary equation for scalar equations.

Many commercially available software packages can be used to compute the eigenvalues and eigenvectors for a given matrix. Three such packages are MATLAB®, available from The MathWorks, Inc.; MATHEMATICA®, available from Wolfram Research; and MAPLESOFT®, available from Waterloo Maple Inc. Although you are encouraged to make use of such packages, the examples and most exercises in this text can be easily carried out without them. Those exercises for which a computer package is desirable are flagged with the icon .

Example 1 Find the eigenvalues and eigenvectors of the matrix

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

Solution The characteristic equation for $\mathbf{A}$ is

$$|\mathbf{A} - r\mathbf{I}| = \begin{vmatrix} 2-r & -3 \\ 1 & -2-r \end{vmatrix} = (2-r)(-2-r) + 3 = r^2 - 1 = 0 .$$

Hence the eigenvalues of $\mathbf{A}$ are $r_1 = 1$, $r_2 = -1$. To find the eigenvectors corresponding to $r_1 = 1$, we must solve $(\mathbf{A} - r_1\mathbf{I})\mathbf{u} = \mathbf{0}$. Substituting for $\mathbf{A}$ and r_1 gives

$$\begin{bmatrix} 1 & -3 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{5}$$

Notice that this matrix equation is equivalent to the single scalar equation $u_1 - 3u_2 = 0$. Therefore, the solutions to (5) are obtained by assigning an arbitrary value for u_2 (say, $u_2 = s$) and setting $u_1 = 3u_2 = 3s$. Consequently, the eigenvectors associated with $r_1 = 1$ can be expressed as

$$\mathbf{u}_1 = s\begin{bmatrix} 3 \\ 1 \end{bmatrix}. \tag{6}$$

For $r_2 = -1$, the equation $(\mathbf{A} - r_2\mathbf{I})\mathbf{u} = \mathbf{0}$ becomes

$$\begin{bmatrix} 3 & -3 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Solving, we obtain $u_1 = s$ and $u_2 = s$, with s arbitrary. Therefore, the eigenvectors associated with the eigenvalue $r_2 = -1$ are

$$\mathbf{u}_2 = s\begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{7}$$ ◆

We remark that in the above example the collection (6) of all eigenvectors associated with $r_1 = 1$ forms a one-dimensional subspace when the zero vector is adjoined. The same is true for $r_2 = -1$. These subspaces are called **eigenspaces.**

Example 2 Find the eigenvalues and eigenvectors of the matrix

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

Solution The characteristic equation for $\mathbf{A}$ is

$$|\mathbf{A} - r\mathbf{I}| = \begin{vmatrix} 1-r & 2 & -1 \\ 1 & -r & 1 \\ 4 & -4 & 5-r \end{vmatrix} = 0,$$

which simplifies to $(r-1)(r-2)(r-3) = 0$. Hence, the eigenvalues of $\mathbf{A}$ are $r_1 = 1$, $r_2 = 2$, and $r_3 = 3$. To find the eigenvectors corresponding to $r_1 = 1$, we set $r = 1$ in $(\mathbf{A} - r\mathbf{I})\mathbf{u} = \mathbf{0}$. This gives

$$\begin{bmatrix} 0 & 2 & -1 \\ 1 & -1 & 1 \\ 4 & -4 & 4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{8}$$

Using elementary row operations (Gaussian elimination), we see that (8) is equivalent to the two equations

$$\begin{aligned} u_1 - u_2 + u_3 &= 0, \\ 2u_2 - u_3 &= 0. \end{aligned}$$

Thus, we can obtain the solutions to (8) by assigning an arbitrary value to u_2 (say, $u_2 = s$), solving $2u_2 - u_3 = 0$ for u_3 to get $u_3 = 2s$, and then solving $u_1 - u_2 + u_3 = 0$ for u_1 to get $u_1 = -s$. Hence, the eigenvectors associated with $r_1 = 1$ are

$$\mathbf{u}_1 = s\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}. \tag{9}$$

For $r_2 = 2$, we solve

$$\begin{bmatrix} -1 & 2 & -1 \\ 1 & -2 & 1 \\ 4 & -4 & 3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

in a similar fashion to obtain the eigenvectors

$$\mathbf{u}_2 = s\begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}. \tag{10}$$

Finally, for $r_3 = 3$, we solve

$$\begin{bmatrix} -2 & 2 & -1 \\ 1 & -3 & 1 \\ 4 & -4 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and get the eigenvectors

$$\mathbf{u}_3 = s\begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix}. \tag{11}$$ ◆

Let's return to the problem of finding a general solution to a homogeneous system of differential equations. We have already shown that $e^{rt}\mathbf{u}$ is a solution to (1) if r is an eigenvalue and $\mathbf{u}$ a corresponding eigenvector. The question is: Can we obtain n linearly independent solutions to the homogeneous system by finding all the eigenvalues and eigenvectors of $\mathbf{A}$? The answer is yes, *if* $\mathbf{A}$ *has n linearly independent eigenvectors.*

n Linearly Independent Eigenvectors

Theorem 5. Suppose the $n \times n$ constant matrix $\mathbf{A}$ has n linearly independent eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. Let r_i be the eigenvalue† corresponding to $\mathbf{u}_i$. Then

(12) $$\{e^{r_1 t}\mathbf{u}_1, e^{r_2 t}\mathbf{u}_2, \ldots, e^{r_n t}\mathbf{u}_n\}$$

is a fundamental solution set (and $\mathbf{X}(t) = [e^{r_1 t}\mathbf{u}_1 \; e^{r_2 t}\mathbf{u}_2 \; \cdots \; e^{r_n t}\mathbf{u}_n]$ is a fundamental matrix) on $(-\infty, \infty)$ for the homogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x}$. Consequently, a general solution of $\mathbf{x}' = \mathbf{A}\mathbf{x}$ is

(13) $$\mathbf{x}(t) = c_1 e^{r_1 t}\mathbf{u}_1 + c_2 e^{r_2 t}\mathbf{u}_2 + \cdots + c_n e^{r_n t}\mathbf{u}_n ,$$

where $c_1, \ldots, c_n$ are arbitrary constants.

Proof. As we have seen, the vector functions listed in (12) are solutions to the homogeneous system. Moreover, their Wronskian is

$$W(t) = \det[e^{r_1 t}\mathbf{u}_1, \ldots, e^{r_n t}\mathbf{u}_n] = e^{(r_1 + \cdots + r_n)t}\det[\mathbf{u}_1, \ldots, \mathbf{u}_n] .$$

Since the eigenvectors are assumed to be linearly independent, it follows from Theorem 1 in Section 9.3 that $\det[\mathbf{u}_1, \ldots, \mathbf{u}_n]$ is not zero. Hence the Wronskian $W(t)$ is never zero. This shows that (12) is a fundamental solution set, and consequently a general solution is given by (13). ◆

An application of Theorem 5 is given in the next example.

Example 3 Find a general solution of

(14) $$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) \text{, where } \mathbf{A} = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix} .$$

Solution In Example 1 we showed that the matrix $\mathbf{A}$ has eigenvalues $r_1 = 1$ and $r_2 = -1$. Taking, say, $s = 1$ in equations (6) and (7), we get the corresponding eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

Because $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent, it follows from Theorem 5 that a general solution to (14) is

(15) $$\mathbf{x}(t) = c_1 e^{t}\begin{bmatrix} 3 \\ 1 \end{bmatrix} + c_2 e^{-t}\begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$ ◆

†The eigenvalues $r_1, \ldots, r_n$ may be real or complex and need not be distinct. In this section the cases we discuss have real eigenvalues. We consider complex eigenvalues in Section 9.6.

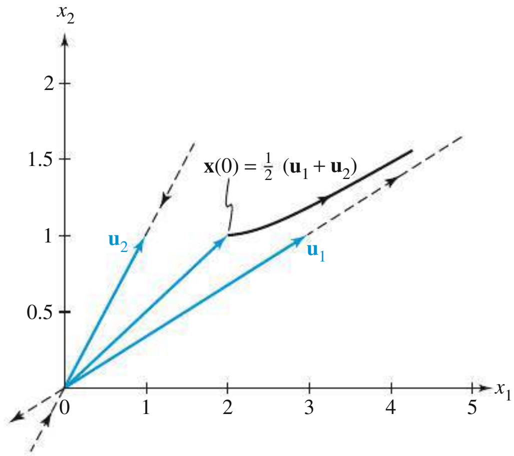


Figure 9.1 Trajectories of solutions for Example 3

If we sum the vectors on the right-hand side of equation (15) and then write out the expressions for the components of $\mathbf{x}(t) = \operatorname{col}\big(x_1(t), x_2(t)\big)$, we get

$$x_1(t) = 3c_1e^t + c_2e^{-t}\,,$$
$$x_2(t) = c_1e^t + c_2e^{-t}\,.$$

This is the familiar form of a general solution for a system, as discussed in Section 5.2.

Example 3 nicely illustrates the geometric role played by the eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$. If the initial vector $\mathbf{x}(0)$ is a scalar multiple of $\mathbf{u}_1$(i.e., $\mathbf{x}(0) = c_1\mathbf{u}_1$), then the vector solution to the system, $\mathbf{x}(t) = c_1e^t\mathbf{u}_1$, will always have the same or opposite direction as $\mathbf{u}_1$. That is, it will lie along the straight line determined by $\mathbf{u}_1$ (see Figure 9.1). Furthermore, the trajectory of this solution, as t increases, will tend to infinity, since the corresponding eigenvalue $r_1 = 1$ is positive (observe the e^t term). A similar assertion holds if the initial vector is a scalar multiple of $\mathbf{u}_2$, except that since $r_2 = -1$ is negative, the trajectory $\mathbf{x}(t) = c_2e^{-t}\mathbf{u}_2$ will approach the origin as t increases (because of e^{-t}). For an initial vector that involves both $\mathbf{u}_1$ and $\mathbf{u}_2$, such as $\mathbf{x}(0) = \frac{1}{2}(\mathbf{u}_1 + \mathbf{u}_2)$, the resulting trajectory is a blend of the above motions, with the contribution due to the larger eigenvalue $r_1 = 1$ dominating as t increases; see Figure 9.1.

The straight-line trajectories in the x_1x_2-plane (the phase plane), then, point along the directions of the eigenvectors of the matrix $\mathbf{A}$. (See Section 5.4, Figure 5.11, page 266, for example.)

A useful property of eigenvectors that concerns their linear independence is stated in the next theorem.

Linear Independence of Eigenvectors

Theorem 6. If $r_1, \ldots, r_m$ are *distinct* eigenvalues for the matrix $\mathbf{A}$ and $\mathbf{u}_i$ is an eigenvector associated with r_i, then $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are linearly independent.

Proof. Let's first treat the case $m = 2$. Suppose, to the contrary, that $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly dependent so that

$$\mathbf{u}_1 = c\mathbf{u}_2 \tag{16}$$

for some constant c. Multiplying both sides of (16) by $\mathbf{A}$ and using the fact that $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors with corresponding eigenvalues r_1 and r_2, we obtain

$$r_1\mathbf{u}_1 = cr_2\mathbf{u}_2\,. \tag{17}$$

Next we multiply (16) by r_2 and then subtract from (17) to get

$$(r_1 - r_2)\mathbf{u}_1 = \mathbf{0}\,.$$

Since $\mathbf{u}_1$ is not the zero vector, we must have $r_1 = r_2$. But this violates the assumption that the eigenvalues are distinct! Hence $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent.

The cases $m > 2$ follow by induction. The details of the proof are left as Problem 48. ◆

Combining Theorems 5 and 6, we get the following corollary.

n Distinct Eigenvalues

Corollary 1. If the $n \times n$ constant matrix $\mathbf{A}$ has n distinct eigenvalues $r_1, \ldots, r_n$ and $\mathbf{u}_i$ is an eigenvector associated with r_i, then

$$\{e^{r_1 t}\mathbf{u}_1, \ldots, e^{r_n t}\mathbf{u}_n\}$$

is a fundamental solution set for the homogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x}$.

Example 4 Solve the initial value problem

$$\mathbf{x}'(t) = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{bmatrix}\mathbf{x}(t)\,, \qquad \mathbf{x}(0) = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}. \tag{18}$$

Solution In Example 2 we showed that the 3×3 coefficient matrix $\mathbf{A}$ has the three distinct eigenvalues $r_1 = 1$, $r_2 = 2$, and $r_3 = 3$. If we set $s = 1$ in equations (9), (10), and (11), we obtain the corresponding eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}, \qquad \mathbf{u}_3 = \begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix},$$

whose linear independence is guaranteed by Theorem 6. Hence, a general solution to (18) is

$$\begin{aligned} \mathbf{x}(t) &= c_1 e^{t}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} + c_2 e^{2t}\begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} + c_3 e^{3t}\begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} -e^{t} & -2e^{2t} & -e^{3t} \\ e^{t} & e^{2t} & e^{3t} \\ 2e^{t} & 4e^{2t} & 4e^{3t} \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}. \end{aligned} \tag{19}$$

To satisfy the initial condition in (18), we solve

$$\mathbf{x}(0) = \begin{bmatrix} -1 & -2 & -1 \\ 1 & 1 & 1 \\ 2 & 4 & 4 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$$

and find that $c_1 = 0$, $c_2 = 1$, and $c_3 = -1$. Inserting these values into (19) gives the desired solution. ◆

There is a special class of $n \times n$ matrices that *always* have n linearly independent eigenvectors. These are the real symmetric matrices.

Real Symmetric Matrices

Definition 4. A **real symmetric matrix A** is a matrix with real entries that satisfies $\mathbf{A}^T = \mathbf{A}$.

Taking the transpose of a matrix interchanges its rows and columns. Doing this is equivalent to "flipping" the matrix about its main diagonal. Consequently, $\mathbf{A}^T = \mathbf{A}$ if and only if $\mathbf{A}$ is symmetric about its main diagonal.

If $\mathbf{A}$ is an $n \times n$ real symmetric matrix, it is known[†] that all its eigenvalues are real and that there always exist n linearly independent eigenvectors. In such a case, Theorem 5 applies and a general solution to $\mathbf{x}' = \mathbf{A}\mathbf{x}$ is given by (13).

Example 5 Find a general solution of

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)\,, \text{ where } \mathbf{A} = \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}. \tag{20}$$

Solution $\mathbf{A}$ is symmetric, so we are assured that $\mathbf{A}$ has three linearly independent eigenvectors. To find them, we first compute the characteristic equation for $\mathbf{A}$:

$$|\mathbf{A} - r\mathbf{I}| = \begin{vmatrix} 1-r & -2 & 2 \\ -2 & 1-r & 2 \\ 2 & 2 & 1-r \end{vmatrix} = -(r-3)^2(r+3) = 0\,.$$

Thus the eigenvalues of $\mathbf{A}$ are $r_1 = r_2 = 3$ and $r_3 = -3$.

Notice that the eigenvalue $r = 3$ has multiplicity 2 when considered as a root of the characteristic equation. Therefore, we must find *two* linearly independent eigenvectors associated with $r = 3$. Substituting $r = 3$ in $(\mathbf{A} - r\mathbf{I})\mathbf{u} = \mathbf{0}$ gives

$$\begin{bmatrix} -2 & -2 & 2 \\ -2 & -2 & 2 \\ 2 & 2 & -2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This system is equivalent to the single equation $-u_1 - u_2 + u_3 = 0$, so we can obtain its solutions by assigning an arbitrary value to u_2, say $u_2 = v$, and an arbitrary value to u_3, say $u_3 = s$. Solving for u_1, we find $u_1 = u_3 - u_2 = s - v$. Therefore, the eigenvectors associated with $r_1 = r_2 = 3$ can be expressed as

$$\mathbf{u} = \begin{bmatrix} s - v \\ v \\ s \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + v\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

[†]See *Fundamentals of Matrix Analysis with Applications*, by Edward Barry Saff and Arthur David Snider (John Wiley & Sons, Hoboken, New Jersey, 2016).

By first taking $s = 1, v = 0$ and then taking $s = 0, v = 1$, we get the two linearly independent eigenvectors

$$\textbf{(21)} \qquad \mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

For $r_3 = -3$, we solve

$$(\mathbf{A} + 3\mathbf{I})\mathbf{u} = \begin{bmatrix} 4 & -2 & 2 \\ -2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

to obtain the eigenvectors $\operatorname{col}(-s, -s, s)$. Taking $s = 1$ gives

$$\mathbf{u}_3 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.$$

Since the eigenvectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ are linearly independent, a general solution to (20) is

$$\mathbf{x}(t) = c_1 e^{3t} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 e^{3t} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + c_3 e^{-3t} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}. \quad \blacklozenge$$

If a matrix **A** *is not symmetric, it is possible for* **A** *to have a repeated eigenvalue but not to have two linearly independent corresponding eigenvectors.* In particular, the matrix

$$\textbf{(22)} \qquad \mathbf{A} = \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix}$$

has the repeated eigenvalue $r_1 = r_2 = -1$, but Problem 35 shows that all the eigenvectors associated with $r = -1$ are of the form $\mathbf{u} = s \operatorname{col}(1, 2)$. Consequently, no two eigenvectors are linearly independent.

A procedure for finding a general solution in such a case is illustrated in Problems 35–40, but the underlying theory is deferred to Section 9.8, where we discuss the matrix exponential.

A final note. If an $n \times n$ matrix **A** has n linearly independent eigenvectors $\mathbf{u}_i$ with eigenvalues r_i, a little inspection reveals that property (2) is expressed columnwise by the equation

$$\textbf{(23)} \qquad \begin{bmatrix} & & \\ & \mathbf{A} & \\ & & \end{bmatrix} \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & r_n \end{bmatrix}$$

or $\mathbf{AU} = \mathbf{UD}$, where **U** is the matrix whose column vectors are eigenvectors and **D** is a diagonal matrix whose diagonal entries are the eigenvalues. Since **U**'s columns are independent, **U** is invertible and we can write

$$\textbf{(24)} \qquad \mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1} \text{ or } \mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U},$$

and we say that **A** is *diagonalizable*. [In this context equation (24) expresses a *similarity transformation*.] Because the argument that leads from (2) to (23) to (24) can be reversed, we have a new characterization: *An n × n matrix has n linearly independent eigenvectors if, and only if, it is diagonalizable.*

9.5 EXERCISES

In Problems 1–8, find the eigenvalues and eigenvectors of the given matrix.

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

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

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

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

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

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

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

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

In Problems 9 and 10, some of the eigenvalues of the given matrix are complex. Find all the eigenvalues and eigenvectors.

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

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

In Problems 11–16, find a general solution of the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ for the given matrix $\mathbf{A}$.

11. $\mathbf{A} = \begin{bmatrix} -1 & \frac{3}{4} \\ -5 & 3 \end{bmatrix}$

12. $\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 12 & 1 \end{bmatrix}$

13. $\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 0 & 3 \\ 2 & 3 & 0 \end{bmatrix}$

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

15. $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 2 & 1 & 2 \end{bmatrix}$

16. $\mathbf{A} = \begin{bmatrix} -7 & 0 & 6 \\ 0 & 5 & 0 \\ 6 & 0 & 2 \end{bmatrix}$

17. Consider the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, $t \geq 0$, with

$$\mathbf{A} = \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}.$$

(a) Show that the matrix $\mathbf{A}$ has eigenvalues $r_1 = 2$ and $r_2 = -2$ with corresponding eigenvectors $\mathbf{u}_1 = \text{col}(\sqrt{3}, 1)$ and $\mathbf{u}_2 = \text{col}(1, -\sqrt{3})$.

(b) Sketch the trajectory of the solution having initial vector $\mathbf{x}(0) = -\mathbf{u}_1$.

(c) Sketch the trajectory of the solution having initial vector $\mathbf{x}(0) = \mathbf{u}_2$.

(d) Sketch the trajectory of the solution having initial vector $\mathbf{x}(0) = \mathbf{u}_2 - \mathbf{u}_1$.

18. Consider the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, $t \geq 0$, with

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

(a) Show that the matrix $\mathbf{A}$ has eigenvalues $r_1 = -1$ and $r_2 = -3$ with corresponding eigenvectors $\mathbf{u}_1 = \text{col}(1, 1)$ and $\mathbf{u}_2 = \text{col}(1, -1)$.

(b) Sketch the trajectory of the solution having initial vector $\mathbf{x}(0) = \mathbf{u}_1$.

(c) Sketch the trajectory of the solution having initial vector $\mathbf{x}(0) = -\mathbf{u}_2$.

(d) Sketch the trajectory of the solution having initial vector $\mathbf{x}(0) = \mathbf{u}_1 - \mathbf{u}_2$.

In Problems 19–24, find a fundamental matrix for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ for the given matrix $\mathbf{A}$.

19. $\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 8 & 1 \end{bmatrix}$

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

21. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 8 & -14 & 7 \end{bmatrix}$

22. $\mathbf{A} = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 3 & 3 & -1 \end{bmatrix}$

23. $\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 7 \end{bmatrix}$

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

25. Using matrix algebra techniques, find a general solution of the system

$$\begin{aligned} x' &= x + 2y - z, \\ y' &= x + z, \\ z' &= 4x - 4y + 5z. \end{aligned}$$

26. Using matrix algebra techniques, find a general solution of the system

$$\begin{aligned} x' &= 3x - 4y, \\ y' &= 4x - 7y. \end{aligned}$$

In Problems 27–30, use a linear algebra software package such as MATLAB®, MAPLESOFT®, *or* MATHEMATICA® *to compute the required eigenvalues and eigenvectors and then give a fundamental matrix for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ for the given matrix $\mathbf{A}$.*

27. $\mathbf{A} = \begin{bmatrix} 0 & 1.1 & 0 \\ 0 & 0 & 1.3 \\ 0.9 & 1.1 & -6.9 \end{bmatrix}$

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

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

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

In Problems 31–34, solve the given initial value problem.

31. $\mathbf{x}'(t) = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}\mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

32. $\mathbf{x}'(t) = \begin{bmatrix} 6 & -3 \\ 2 & 1 \end{bmatrix}\mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} -10 \\ -6 \end{bmatrix}$

33. $\mathbf{x}'(t) = \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}\mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix}$

34. $\mathbf{x}'(t) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} -1 \\ 4 \\ 0 \end{bmatrix}$

35. **(a)** Show that the matrix

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

has the repeated eigenvalue $r = -1$ and that all the eigenvectors are of the form $\mathbf{u} = s\,\mathrm{col}(1, 2)$.

(b) Use the result of part (a) to obtain a nontrivial solution $\mathbf{x}_1(t)$ to the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$.

(c) To obtain a second linearly independent solution to $\mathbf{x}' = \mathbf{A}\mathbf{x}$, try $\mathbf{x}_2(t) = te^{-t}\mathbf{u}_1 + e^{-t}\mathbf{u}_2$. [*Hint:* Substitute $\mathbf{x}_2$ into the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ and derive the relations

$$(\mathbf{A} + \mathbf{I})\mathbf{u}_1 = \mathbf{0}, \qquad (\mathbf{A} + \mathbf{I})\mathbf{u}_2 = \mathbf{u}_1.$$

Since $\mathbf{u}_1$ must be an eigenvector, set $\mathbf{u}_1 = \mathrm{col}(1, 2)$ and solve for $\mathbf{u}_2$.]

(d) What is $(\mathbf{A} + \mathbf{I})^2\mathbf{u}_2$? (In Section 9.8, $\mathbf{u}_2$ will be identified as a *generalized eigenvector*.)

36. Use the method discussed in Problem 35 to find a general solution to the system

$$\mathbf{x}'(t) = \begin{bmatrix} 5 & -3 \\ 3 & -1 \end{bmatrix}\mathbf{x}(t).$$

37. **(a)** Show that the matrix

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

has the repeated eigenvalue $r = 2$ with multiplicity 3 and that all the eigenvectors of $\mathbf{A}$ are of the form $\mathbf{u} = s\,\mathrm{col}(1, 0, 0)$.

(b) Use the result of part (a) to obtain a solution to the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ of the form $\mathbf{x}_1(t) = e^{2t}\mathbf{u}_1$.

(c) To obtain a second linearly independent solution to $\mathbf{x}' = \mathbf{A}\mathbf{x}$, try $\mathbf{x}_2(t) = te^{2t}\mathbf{u}_1 + e^{2t}\mathbf{u}_2$. [*Hint:* Show that $\mathbf{u}_1$ and $\mathbf{u}_2$ must satisfy

$$(\mathbf{A} - 2\mathbf{I})\mathbf{u}_1 = \mathbf{0}, \qquad (\mathbf{A} - 2\mathbf{I})\mathbf{u}_2 = \mathbf{u}_1.]$$

(d) To obtain a third linearly independent solution to $\mathbf{x}' = \mathbf{A}\mathbf{x}$, try

$$\mathbf{x}_3(t) = \frac{t^2}{2}e^{2t}\mathbf{u}_1 + te^{2t}\mathbf{u}_2 + e^{2t}\mathbf{u}_3.$$

[*Hint:* Show that $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ must satisfy

$$(\mathbf{A} - 2\mathbf{I})\mathbf{u}_1 = \mathbf{0}, \qquad (\mathbf{A} - 2\mathbf{I})\mathbf{u}_2 = \mathbf{u}_1,$$
$$(\mathbf{A} - 2\mathbf{I})\mathbf{u}_3 = \mathbf{u}_2.]$$

(e) Show that $(\mathbf{A} - 2\mathbf{I})^2\mathbf{u}_2 = (\mathbf{A} - 2\mathbf{I})^3\mathbf{u}_3 = \mathbf{0}$.

38. Use the method discussed in Problem 37 to find a general solution to the system

$$\mathbf{x}'(t) = \begin{bmatrix} 3 & -2 & 1 \\ 2 & -1 & 1 \\ -4 & 4 & 1 \end{bmatrix}\mathbf{x}(t).$$

39. **(a)** Show that the matrix

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

has the repeated eigenvalue $r = 1$ of multiplicity 3 and that all the eigenvectors of $\mathbf{A}$ are of the form $\mathbf{u} = s\,\mathrm{col}(-1, 1, 0) + v\,\mathrm{col}(-1, 0, 1)$.

(b) Use the result of part (a) to obtain two linearly independent solutions to the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ of the form

$\mathbf{x}_1(t) = e^t\mathbf{u}_1$ and $\mathbf{x}_2(t) = e^t\mathbf{u}_2$.

(c) To obtain a third linearly independent solution to $\mathbf{x}' = \mathbf{A}\mathbf{x}$, try $\mathbf{x}_3(t) = te^t\mathbf{u}_3 + e^t\mathbf{u}_4$. [*Hint:* Show that $\mathbf{u}_3$ and $\mathbf{u}_4$ must satisfy

$$(\mathbf{A} - \mathbf{I})\mathbf{u}_3 = \mathbf{0}, \qquad (\mathbf{A} - \mathbf{I})\mathbf{u}_4 = \mathbf{u}_3.$$

Choose $\mathbf{u}_3$, an eigenvector of $\mathbf{A}$, so that you can solve for $\mathbf{u}_4$.]

(d) What is $(\mathbf{A} - \mathbf{I})^2\mathbf{u}_4$?

40. Use the method discussed in Problem 39 to find a general solution to the system

$$\mathbf{x}'(t) = \begin{bmatrix} 1 & 3 & -2 \\ 0 & 7 & -4 \\ 0 & 9 & -5 \end{bmatrix}\mathbf{x}(t).$$

41. Use the substitution $x_1 = y$, $x_2 = y'$ to convert the linear equation $ay'' + by' + cy = 0$, where a, b, and c are constants, into a normal system. Show that the characteristic equation for this system is the same as the auxiliary equation for the original equation.

42. **(a)** Show that the Cauchy–Euler equation $at^2y'' + bty' + cy = 0$ can be written as a **Cauchy–Euler system**

$$tx' = Ax \quad (25)$$

with a constant coefficient matrix $\mathbf{A}$, by setting $x_1 = y/t$ and $x_2 = y'$.

(b) Show that for $t > 0$ any system of the form (25) with $\mathbf{A}$ an $n \times n$ constant matrix has nontrivial solutions of the form $\mathbf{x}(t) = t^r\mathbf{u}$ if and only if r is an eigenvalue of $\mathbf{A}$ and $\mathbf{u}$ is a corresponding eigenvector.

In Problems 43 and 44, use the result of Problem 42 to find a general solution of the given system.

43. $t\mathbf{x}'(t) = \begin{bmatrix} 1 & 3 \\ -1 & 5 \end{bmatrix}\mathbf{x}(t), \quad t > 0$

44. $t\mathbf{x}'(t) = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}\mathbf{x}(t), \quad t > 0$

45. **Mixing Between Interconnected Tanks.** Two tanks, each holding 50 L of liquid, are interconnected by pipes with liquid flowing from tank A into tank B at a rate of 4 L/min and from tank B into tank A at 1 L/min (see Figure 9.2). The liquid inside each tank is kept well stirred. Pure water flows into tank A at a rate of 3 L/min, and the solution flows out of tank B at 3 L/min. If, initially, tank A contains 2.5 kg of salt and tank B contains no salt (only water), determine the mass of salt in each tank at time $t \geq 0$. Graph on the same axes the two quantities $x_1(t)$ and $x_2(t)$, where $x_1(t)$ is the mass of salt in tank A and $x_2(t)$ is the mass in tank B.

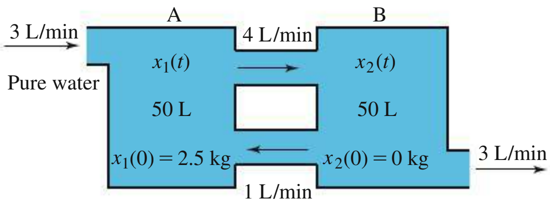


Figure 9.2 Mixing problem for interconnected tanks

46. **Mixing with a Common Drain.** Two tanks, each holding 1 L of liquid, are connected by a pipe through which liquid flows from tank A into tank B at a rate of $3 - \alpha$ L/min $(0 < \alpha < 3)$. The liquid inside each tank is kept well stirred. Pure water flows into tank A at a rate of 3 L/min. Solution flows out of tank A at α L/min and out of tank B at $3 - \alpha$ L/min. If, initially, tank B contains no salt (only water) and tank A contains 0.1 kg of salt, determine the mass of salt in each tank at time $t \geq 0$. How does the mass of salt in tank A depend on the choice of α? What is the maximum mass of salt in tank B? (See Figure 9.3.)

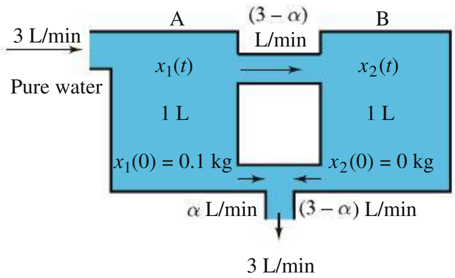


Figure 9.3 Mixing problem for a common drain, $0 < \alpha < 3$

47. To find a general solution to the system

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

proceed as follows:

(a) Use a numerical root-finding procedure to approximate the eigenvalues.

(b) If r is an eigenvalue, then let $\mathbf{u} = \operatorname{col}(u_1, u_2, u_3)$ be an eigenvector associated with r. To solve for $\mathbf{u}$, assume $u_1 = 1$. (If not u_1, then either u_2 or u_3 may be chosen to be 1. Why?) Now solve the system

$$(\mathbf{A} - r\mathbf{I})\begin{bmatrix} 1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

for u_2 and u_3. Use this procedure to find approximations for three linearly independent eigenvectors for $\mathbf{A}$.

(c) Use these approximations to give a general solution to the system.

48. To complete the proof of Theorem 6, page 527, assume the induction hypothesis that $\mathbf{u}_1, \ldots, \mathbf{u}_k$, $2 \leq k$, are linearly independent.

(a) Show that if

$$c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + c_{k+1}\mathbf{u}_{k+1} = \mathbf{0},$$

then

$$c_1(r_1 - r_{k+1})\mathbf{u}_1 + \cdots + c_k(r_k - r_{k+1})\mathbf{u}_k = \mathbf{0}.$$

(b) Use the result of part (a) and the induction hypothesis to conclude that $\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}$ are linearly independent. The theorem follows by induction.

49. **Stability.** A homogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ with constant coefficients is **stable** if it has a fundamental matrix whose entries all remain bounded as $t \to +\infty$. (It will follow from Theorem 9 in Section 9.8 that if one fundamental matrix of the system has this property, then all fundamental matrices for the system do.) Otherwise, the system is **unstable**. A stable system is **asymptotically stable** if all solutions approach the zero solution as $t \to +\infty$. Stability is discussed in more detail in Chapter 12.†

†All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems*, 7th ed.

(a) Show that if $\mathbf{A}$ has all distinct real eigenvalues, then $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ is stable if and only if all eigenvalues are nonpositive.
(b) Show that if $\mathbf{A}$ has all distinct real eigenvalues, then $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ is asymptotically stable if and only if all eigenvalues are negative.
(c) Argue that in parts (a) and (b), we can replace "has distinct real eigenvalues" by "is symmetric" and the statements are still true.

50. In an ice tray, the water level in any particular ice cube cell will change at a rate proportional to the *difference* between that cell's water level and the level in the adjacent cells.

(a) Argue that a reasonable differentiable equation model for the water levels x, y, and z in the simplified three-cell tray depicted in Figure 9.4 is given by

$$x' = y - x\,, \quad y' = x + z - 2y\,, \quad z' = y - z\,.$$

(b) Use eigenvectors to solve this system for the initial conditions $x(0) = 3$, $y(0) = z(0) = 0$.

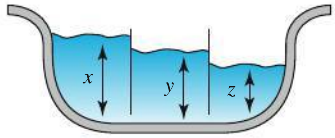


Figure 9.4 Ice tray

9.6 Complex Eigenvalues

In the previous section, we showed that the homogeneous system

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)\,, \tag{1}$$

where $\mathbf{A}$ is a constant $n \times n$ matrix, has a solution of the form $\mathbf{x}(t) = e^{rt}\mathbf{u}$ if and only if r is an eigenvalue of $\mathbf{A}$ and $\mathbf{u}$ is a corresponding eigenvector. In this section we show how to obtain two real vector solutions to system (1) when $\mathbf{A}$ is real and has a pair† of complex conjugate eigenvalues $\alpha + i\beta$ and $\alpha - i\beta$.

Suppose $r_1 = \alpha + i\beta(\alpha$ and β real numbers) is an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\mathbf{z} = \mathbf{a} + i\mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are real constant vectors. We first observe that the complex conjugate of $\mathbf{z}$, namely $\overline{\mathbf{z}} := \mathbf{a} - i\mathbf{b}$, is an eigenvector associated with the eigenvalue $r_2 = \alpha - i\beta$. To see this, note that taking the complex conjugate of $(\mathbf{A} - r_1\mathbf{I})\mathbf{z} = \mathbf{0}$ yields $(\mathbf{A} - \overline{r}_1\mathbf{I})\overline{\mathbf{z}} = \mathbf{0}$ because the conjugate of the product is the product of the conjugates and $\mathbf{A}$ and $\mathbf{I}$ have real entries $(\overline{\mathbf{A}} = \mathbf{A}, \overline{\mathbf{I}} = \mathbf{I})$. Since $r_2 = \overline{r}_1$, we see that $\overline{\mathbf{z}}$ is an eigenvector associated with r_2. Therefore, two linearly independent complex vector solutions to (1) are

$$\mathbf{w}_1(t) = e^{r_1 t}\mathbf{z} = e^{(\alpha + i\beta)t}(\mathbf{a} + i\mathbf{b})\,, \tag{2}$$

$$\mathbf{w}_2(t) = e^{r_2 t}\,\overline{\mathbf{z}} = e^{(\alpha - i\beta)t}(\mathbf{a} - i\mathbf{b})\,. \tag{3}$$

As in Section 4.3, where we handled complex roots to the auxiliary equation, let's use one of these complex solutions and Euler's formula to obtain two real vector solutions. With the aid of Euler's formula, we rewrite $\mathbf{w}_1(t)$ as

$$\begin{aligned}\mathbf{w}_1(t) &= e^{\alpha t}(\cos\beta t + i\sin\beta t)(\mathbf{a} + i\mathbf{b})\\ &= e^{\alpha t}\{(\cos\beta t\,\mathbf{a} - \sin\beta t\,\mathbf{b}) + i(\sin\beta t\,\mathbf{a} + \cos\beta t\,\mathbf{b})\}\,.\end{aligned}$$

We have thereby expressed $\mathbf{w}_1(t)$ in the form $\mathbf{w}_1(t) = \mathbf{x}_1(t) + i\mathbf{x}_2(t)$, where $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are the two *real* vector functions

$$\mathbf{x}_1(t) := e^{\alpha t}\cos\beta t\,\mathbf{a} - e^{\alpha t}\sin\beta t\,\mathbf{b}\,, \tag{4}$$

$$\mathbf{x}_2(t) := e^{\alpha t}\sin\beta t\,\mathbf{a} + e^{\alpha t}\cos\beta t\,\mathbf{b}\,. \tag{5}$$

†Recall that the complex roots of a polynomial equation with real coefficients must occur in complex conjugate pairs.

Since $\mathbf{w}_1(t)$ is a solution to (1), then

$$\begin{aligned} \mathbf{w}_1'(t) &= \mathbf{A}\mathbf{w}_1(t)\,, \\ \mathbf{x}_1' + i\mathbf{x}_2' &= \mathbf{A}\mathbf{x}_1 + i\mathbf{A}\mathbf{x}_2\,. \end{aligned}$$

Equating the real and imaginary parts yields

$$\mathbf{x}_1'(t) = \mathbf{A}\mathbf{x}_1(t) \quad \text{and} \quad \mathbf{x}_2'(t) = \mathbf{A}\mathbf{x}_2(t)\,.$$

Hence, $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are real vector solutions to (1) associated with the complex conjugate eigenvalues $\alpha \pm i\beta$. Because $\mathbf{a}$ and $\mathbf{b}$ are not both the zero vector, it can be shown that $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are linearly independent vector functions on $(-\infty, \infty)$ (see Problem 15).

Let's summarize our findings.

Complex Eigenvalues

If the real matrix $\mathbf{A}$ has complex conjugate eigenvalues $\alpha \pm i\beta$ with corresponding eigenvectors $\mathbf{a} \pm i\mathbf{b}$, then two linearly independent real vector solutions to $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ are

(6) $$e^{\alpha t}\cos\beta t\,\mathbf{a} - e^{\alpha t}\sin\beta t\,\mathbf{b}\,,$$

(7) $$e^{\alpha t}\sin\beta t\,\mathbf{a} + e^{\alpha t}\cos\beta t\,\mathbf{b}\,.$$

Example 1 Find a general solution of

(8) $$\mathbf{x}'(t) = \begin{bmatrix} -1 & 2 \\ -1 & -3 \end{bmatrix}\mathbf{x}(t)\,.$$

Solution The characteristic equation for $\mathbf{A}$ is

$$|\mathbf{A} - r\mathbf{I}| = \begin{vmatrix} -1-r & 2 \\ -1 & -3-r \end{vmatrix} = r^2 + 4r + 5 = 0\,.$$

Hence, $\mathbf{A}$ has eigenvalues $r = -2 \pm i$.

To find a general solution, we need only find an eigenvector associated with the eigenvalue $r = -2 + i$. Substituting $r = -2 + i$ into $(\mathbf{A} - r\mathbf{I})\mathbf{z} = \mathbf{0}$ gives

$$\begin{bmatrix} 1-i & 2 \\ -1 & -1-i \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The solutions can be expressed as $z_1 = 2s$ and $z_2 = (-1+i)s$, with s arbitrary. Hence, the eigenvectors associated with $r = -2 + i$ are $\mathbf{z} = s\,\mathrm{col}(2, -1+i)$. Taking $s = 1$ gives the eigenvector

$$\mathbf{z} = \begin{bmatrix} 2 \\ -1+i \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} + i\begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

We have found that $\alpha = -2, \beta = 1$, and $\mathbf{z} = \mathbf{a} + i\mathbf{b}$ with $\mathbf{a} = \mathrm{col}(2, -1)$, and $\mathbf{b} = \mathrm{col}(0, 1)$, so a general solution to (8) is

$$\begin{aligned} \mathbf{x}(t) &= c_1\left\{e^{-2t}\cos t\begin{bmatrix} 2 \\ -1 \end{bmatrix} - e^{-2t}\sin t\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} \\ &\quad + c_2\left\{e^{-2t}\sin t\begin{bmatrix} 2 \\ -1 \end{bmatrix} + e^{-2t}\cos t\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} \end{aligned}$$

(9) $$\mathbf{x}(t) = c_1\begin{bmatrix} 2e^{-2t}\cos t \\ -e^{-2t}(\cos t + \sin t) \end{bmatrix} + c_2\begin{bmatrix} 2e^{-2t}\sin t \\ e^{-2t}(\cos t - \sin t) \end{bmatrix}. \quad \blacklozenge$$

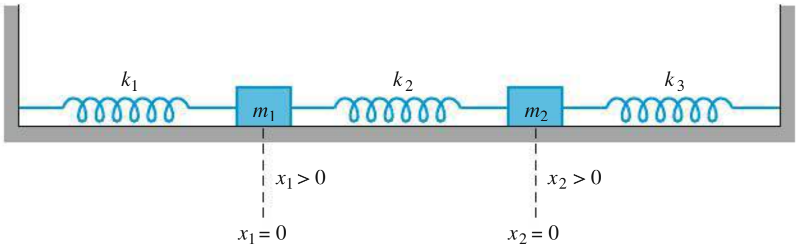


Figure 9.5 Coupled mass–spring system with fixed ends

Complex eigenvalues occur in modeling coupled mass–spring systems. For example, the motion of the mass–spring system illustrated in Figure 9.5 is governed by the second-order system

$$\textbf{(10)}\qquad \begin{aligned} m_1x_1'' &= -k_1x_1 + k_2(x_2 - x_1)\ , \\ m_2x_2'' &= -k_2(x_2 - x_1) - k_3x_2\ , \end{aligned}$$

where x_1 and x_2 represent the displacements of the masses m_1 and m_2 to the right of their equilibrium positions and k_1, k_2, k_3 are the spring constants of the three springs (see the discussion in Section 5.6). If we introduce the new variables $y_1 := x_1, y_2 := x_1', y_3 := x_2, y_4 := x_2'$, then we can rewrite the system in the normal form

$$\textbf{(11)}\qquad \mathbf{y}'(t) = \mathbf{A}\mathbf{y}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -(k_1+k_2)/m_1 & 0 & k_2/m_1 & 0 \\ 0 & 0 & 0 & 1 \\ k_2/m_2 & 0 & -(k_2+k_3)/m_2 & 0 \end{bmatrix} \mathbf{y}(t)\ .$$

For such a system, it turns out that $\mathbf{A}$ has only imaginary eigenvalues and they occur in complex conjugate pairs: $\pm i\beta_1$, $\pm i\beta_2$. Hence, any solution will consist of sums of sine and cosine functions. The frequencies of these functions

$$f_1 := \frac{\beta_1}{2\pi} \quad \text{and} \quad f_2 := \frac{\beta_2}{2\pi}$$

are called the **normal** or **natural frequencies** of the system (β_1 and β_2 are the **angular frequencies** of the system).

In some engineering applications, the only information that is required about a particular device is a knowledge of its normal frequencies; one must ensure that they are far from the frequencies that occur naturally in the device's operating environment (so that no resonances will be excited).

Example 2 Determine the normal frequencies for the coupled mass–spring system governed by system (11) when $m_1 = m_2 = 1$ kg, $k_1 = 1$ N/m, $k_2 = 2$ N/m, and $k_3 = 3$ N/m.

Solution To find the eigenvalues of $\mathbf{A}$, we must solve the characteristic equation

$$|\mathbf{A} - r\mathbf{I}| = \begin{vmatrix} -r & 1 & 0 & 0 \\ -3 & -r & 2 & 0 \\ 0 & 0 & -r & 1 \\ 2 & 0 & -5 & -r \end{vmatrix} = r^4 + 8r^2 + 11 = 0\ .$$

From the quadratic formula we find $r^2 = -4 \pm \sqrt{5}$, so the four eigenvalues of $\mathbf{A}$ are $\pm i\sqrt{4-\sqrt{5}}$ and $\pm i\sqrt{4+\sqrt{5}}$. Hence, the two normal frequencies for this system are

$$\frac{\sqrt{4-\sqrt{5}}}{2\pi} \approx 0.211 \quad \text{and} \quad \frac{\sqrt{4+\sqrt{5}}}{2\pi} \approx 0.397 \text{ cycles per second.} \quad ◆$$

9.6 EXERCISES

In Problems 1–4, find a general solution of the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ for the given matrix $\mathbf{A}$.

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

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

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

In Problems 5–9, find a fundamental matrix for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ for the given matrix $\mathbf{A}$.

5. $\mathbf{A} = \begin{bmatrix} -1 & -2 \\ 8 & -1 \end{bmatrix}$ **6.** $\mathbf{A} = \begin{bmatrix} -2 & -2 \\ 4 & 2 \end{bmatrix}$

7. $\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

8. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -13 & 4 \end{bmatrix}$

9. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$

In Problems 10–12, use a linear algebra software package to compute the required eigenvalues and eigenvectors for the given matrix $\mathbf{A}$ and then give a fundamental matrix for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$.

10. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 13 & -4 & -12 & 4 \end{bmatrix}$

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

12. $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -29 & -4 \end{bmatrix}$

In Problems 13 and 14, find the solution to the given system that satisfies the given initial condition.

13. $\mathbf{x}'(t) = \begin{bmatrix} -3 & -1 \\ 2 & -1 \end{bmatrix} \mathbf{x}(t)$,

(a) $\mathbf{x}(0) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$ **(b)** $\mathbf{x}(\pi) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(c) $\mathbf{x}(-2\pi) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ **(d)** $\mathbf{x}(\pi/2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

14. $\mathbf{x}'(t) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \mathbf{x}(t)$,

(a) $\mathbf{x}(0) = \begin{bmatrix} -2 \\ 2 \\ -1 \end{bmatrix}$ **(b)** $\mathbf{x}(-\pi) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

15. Show that $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ given by equations (4) and (5) are linearly independent on $(-\infty, \infty)$, provided $\beta \neq 0$ and $\mathbf{a}$ and $\mathbf{b}$ are not both the zero vector.

16. Show that $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ given by equations (4) and (5) can be obtained as linear combinations of $\mathbf{w}_1(t)$ and $\mathbf{w}_2(t)$ given by equations (2) and (3). [*Hint:* Show that

$$\mathbf{x}_1(t) = \frac{\mathbf{w}_1(t) + \mathbf{w}_2(t)}{2}, \qquad \mathbf{x}_2(t) = \frac{\mathbf{w}_1(t) - \mathbf{w}_2(t)}{2i}.\Big]$$

In Problems 17 and 18, use the results of Problem 42 in Exercises 9.5 to find a general solution to the given Cauchy–Euler system for $t > 0$.

17. $t\mathbf{x}'(t) = \begin{bmatrix} -1 & -1 & 0 \\ 2 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \mathbf{x}(t)$

18. $t\mathbf{x}'(t) = \begin{bmatrix} -1 & -1 \\ 9 & -1 \end{bmatrix} \mathbf{x}(t)$

19. For the coupled mass–spring system governed by system (10), assume $m_1 = m_2 = 1$ kg, $k_1 = k_2 = 2$ N/m, and $k_3 = 3$ N/m. Determine the normal frequencies for this coupled mass–spring system.

20. For the coupled mass–spring system governed by system (10), assume $m_1 = m_2 = 1$ kg, $k_1 = k_2 = k_3 = 1$ N/m, and assume initially that $x_1(0) = 0$ m, $x_1'(0) = 0$ m/sec, $x_2(0) = 2$ m, and $x_2'(0) = 0$ m/sec. Using matrix algebra techniques, solve this initial value problem.

21. ***RLC* Network.** The currents in the *RLC* network given by the schematic diagram in Figure 9.6 are governed by the following equations:

$$4I_2'(t) + 52q_1(t) = 10\,,$$
$$13I_3(t) + 52q_1(t) = 10\,,$$
$$I_1(t) = I_2(t) + I_3(t)\,,$$

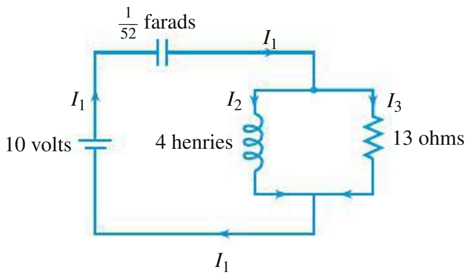


Figure 9.6 *RLC* network for Problem 21

where $q_1(t)$ is the charge on the capacitor, $I_1(t) = q_1'(t)$, and initially $q_1(0) = 0$ coulombs and $I_1(0) = 0$ amps. Solve for the currents I_1, I_2, and I_3. [*Hint:* Differentiate the first two equations, eliminate I_1, and form a normal system with $x_1 = I_2$, $x_2 = I_2'$, and $x_3 = I_3$.]

22. ***RLC* Network.** The currents in the *RLC* network given by the schematic diagram in Figure 9.7 are governed by the following equations:

$$50I_1'(t) + 80I_2(t) = 160\,,$$
$$50I_1'(t) + 800q_3(t) = 160\,,$$
$$I_1(t) = I_2(t) + I_3(t)\,,$$

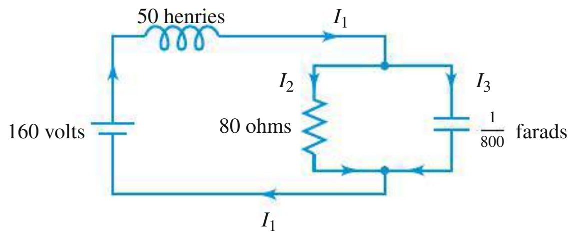


Figure 9.7 *RLC* network for Problem 22

where $q_3(t)$ is the charge on the capacitor, $I_3(t) = q_3'(t)$, and initially $q_3(0) = 0.5$ coulombs and $I_3(0) = 0$ amps. Solve for the currents I_1, I_2, and I_3. [*Hint:* Differentiate the first two equations, use the third equation to eliminate I_3, and form a normal system with $x_1 = I_1$, $x_2 = I_1'$, and $x_3 = I_2$.]

23. **Stability.** In Problem 49 of Exercises 9.5, (page 542), we discussed the notion of stability and asymptotic stability for a linear system of the form $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$. Assume that $\mathbf{A}$ has all distinct eigenvalues (real or complex).

(a) Show that the system is stable if and only if all the eigenvalues of $\mathbf{A}$ have nonpositive real part.

(b) Show that the system is asymptotically stable if and only if all the eigenvalues of $\mathbf{A}$ have negative real part.

24. **(a)** For Example 1, page 535, verify that

$$\mathbf{x}(t) = c_1\begin{bmatrix} -e^{-2t}\cos t + e^{-2t}\sin t \\ e^{-2t}\cos t \end{bmatrix} + c_2\begin{bmatrix} -e^{-2t}\sin t - e^{-2t}\cos t \\ e^{-2t}\sin t \end{bmatrix}$$

is another general solution to equation (8).

(b) How can the general solution of part (a) be directly obtained from the general solution derived in (9) on page 535?

9.7 Nonhomogeneous Linear Systems

The techniques discussed in Chapters 4 and 6 for finding a particular solution to the nonhomogeneous equation $y'' + p(x)y' + q(x)y = g(x)$ have natural extensions to nonhomogeneous linear systems.

Undetermined Coefficients

The method of undetermined coefficients can be used to find a particular solution to the nonhomogeneous linear system

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$$

when $\mathbf{A}$ is an $n \times n$ *constant* matrix and the entries of $\mathbf{f}(t)$ are polynomials, exponential functions, sines and cosines, or finite sums and products of these functions. We can use the procedure box in Section 4.5 (page 184) and reproduced at the back of the book as a *guide* in choosing the form of a particular solution $\mathbf{x}_p(t)$. Some exceptions are discussed in the exercises (see Problems 25–28).

Example 1 Find a general solution of

(1) $$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + t\mathbf{g}\,, \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{g} = \begin{bmatrix} -9 \\ 0 \\ -18 \end{bmatrix}.$$

Solution In Example 5 in Section 9.5, page 529, we found that a general solution to the corresponding homogeneous system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ is

(2) $$\mathbf{x}_h(t) = c_1 e^{3t}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 e^{3t}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + c_3 e^{-3t}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.$$

Since the entries in $\mathbf{f}(t) \coloneqq t\mathbf{g}$ are just linear functions of t, we are inclined to seek a particular solution of the form

$$\mathbf{x}_p(t) = t\mathbf{a} + \mathbf{b} = t\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix},$$

where the constant vectors $\mathbf{a}$ and $\mathbf{b}$ are to be determined. Substituting this expression for $\mathbf{x}_p(t)$ into system (1) yields

$$\mathbf{a} = \mathbf{A}(t\mathbf{a} + \mathbf{b}) + t\mathbf{g}\,,$$

which can be written as

$$t(\mathbf{A}\mathbf{a} + \mathbf{g}) + (\mathbf{A}\mathbf{b} - \mathbf{a}) = \mathbf{0}\,.$$

Setting the "coefficients" of this vector polynomial equal to zero yields the two systems

(3) $$\mathbf{A}\mathbf{a} = -\mathbf{g}\,,$$

(4) $$\mathbf{A}\mathbf{b} = \mathbf{a}\,.$$

By Gaussian elimination or by using a linear algebra software package, we can solve (3) for $\mathbf{a}$ and we find $\mathbf{a} = \mathrm{col}(5, 2, 4)$. Next we substitute for $\mathbf{a}$ in (4) and solve for $\mathbf{b}$ to obtain $\mathbf{b} = \mathrm{col}(1, 0, 2)$. Hence a particular solution for (1) is

(5) $$\mathbf{x}_p(t) = t\mathbf{a} + \mathbf{b} = t\begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 5t + 1 \\ 2t \\ 4t + 2 \end{bmatrix}.$$

A general solution for (1) is $\mathbf{x}(t) = \mathbf{x}_h(t) + \mathbf{x}_p(t)$, where $\mathbf{x}_h(t)$ is given in (2) and $\mathbf{x}_p(t)$ in (5). ◆

In the preceding example, the nonhomogeneous term $\mathbf{f}(t)$ was a vector polynomial. If, instead, $\mathbf{f}(t)$ has the form

$$\mathbf{f}(t) = \mathrm{col}(1, t, \sin t)\,,$$

then, using the superposition principle, we would seek a particular solution of the form

$$\mathbf{x}_p(t) = t\mathbf{a} + \mathbf{b} + (\sin t)\mathbf{c} + (\cos t)\mathbf{d}\,.$$

Similarly, if

$$\mathbf{f}(t) = \mathrm{col}(t, e^t, t^2),$$

we would take

$$\mathbf{x}_p(t) = t^2\mathbf{a} + t\mathbf{b} + \mathbf{c} + e^t\mathbf{d}.$$

Of course, we must modify our guess, should one of the terms be a solution to the corresponding homogeneous system. If this is the case, the annihilator method [equations (15) and (16) of Section 6.3, page 337] would appear to suggest that for a nonhomogeneity $\mathbf{f}(t)$ of the form $e^{rt}t^m\mathbf{g}$, where r is an eigenvalue of $\mathbf{A}$, m is a nonnegative integer, and $\mathbf{g}$ is a constant vector, a particular solution of $\mathbf{x}' = \mathbf{Ax} + \mathbf{f}$ can be found in the form

$$\mathbf{x}_p(t) = e^{rt}\{t^{m+s}\mathbf{a}_{m+s} + t^{m+s-1}\mathbf{a}_{m+s-1} + \cdots + t\mathbf{a}_1 + \mathbf{a}_0\},$$

for a suitable choice of s. We omit the details.

Variation of Parameters

In Section 4.6 we discussed the method of variation of parameters for a general constant-coefficient second-order linear equation. Simply put, the idea is that if a general solution to the homogeneous equation has the form $x_h(t) = c_1x_1(t) + c_2x_2(t)$, where $x_1(t)$ and $x_2(t)$ are linearly independent solutions to the homogeneous equation, then a particular solution to the nonhomogeneous equation would have the form $x_p(t) = v_1(t)x_1(t) + v_2(t)x_2(t)$, where $v_1(t)$ and $v_2(t)$ are certain functions of t. A similar idea can be used for systems.

Let $\mathbf{X}(t)$ be a fundamental matrix for the homogeneous system

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t), \tag{6}$$

where now *the entries of* $\mathbf{A}$ *may be any continuous functions of* t. Because a general solution to (6) is given by $\mathbf{X}(t)\mathbf{c}$, where $\mathbf{c}$ is a constant $n \times 1$ vector, we seek a particular solution to the nonhomogeneous system

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t) \tag{7}$$

of the form

$$\mathbf{x}_p(t) = \mathbf{X}(t)\boldsymbol{v}(t), \tag{8}$$

where $\boldsymbol{v}(t) = \mathrm{col}\big(v_1(t), \ldots, v_n(t)\big)$ is a vector function of t to be determined.

To derive a formula for $\boldsymbol{v}(t)$, we first differentiate (8) using the matrix version of the product rule to obtain

$$\mathbf{x}_p'(t) = \mathbf{X}(t)\boldsymbol{v}'(t) + \mathbf{X}'(t)\boldsymbol{v}(t).$$

Substituting the expressions for $\mathbf{x}_p(t)$ and $\mathbf{x}_p'(t)$ into (7) yields

$$\mathbf{X}(t)\boldsymbol{v}'(t) + \mathbf{X}'(t)\boldsymbol{v}(t) = \mathbf{A}(t)\mathbf{X}(t)\boldsymbol{v}(t) + \mathbf{f}(t). \tag{9}$$

Since $\mathbf{X}(t)$ satisfies the matrix equation $\mathbf{X}'(t) = \mathbf{A}(t)\mathbf{X}(t)$, equation (9) becomes

$$\mathbf{X}\boldsymbol{v}' + \mathbf{AX}\boldsymbol{v} = \mathbf{AX}\boldsymbol{v} + \mathbf{f},$$

$$\mathbf{X}\boldsymbol{v}' = \mathbf{f}.$$

Multiplying both sides of the last equation by $\mathbf{X}^{-1}(t)$ [which exists since the columns of $\mathbf{X}(t)$ are linearly independent] gives

$$\boldsymbol{v}'(t) = \mathbf{X}^{-1}(t)\mathbf{f}(t).$$

Integrating, we obtain

$$\boldsymbol{v}(t) = \int \mathbf{X}^{-1}(t)\mathbf{f}(t)dt\,.$$

Hence, a particular solution to (7) is

$$\mathbf{x}_p(t) = \mathbf{X}(t)\boldsymbol{v}(t) = \mathbf{X}(t)\int \mathbf{X}^{-1}(t)\mathbf{f}(t)dt\,. \tag{10}$$

Combining (10) with the solution $\mathbf{X}(t)\mathbf{c}$ to the homogeneous system yields the following general solution to (7):

$$\mathbf{x}(t) = \mathbf{X}(t)\mathbf{c} + \mathbf{X}(t)\int \mathbf{X}^{-1}(t)\mathbf{f}(t)dt\,. \tag{11}$$

The elegance of the derivation of the variation of parameters formula (10) for systems becomes evident when one compares it with the more lengthy derivations for the scalar case in Sections 4.6 and 6.4.

Given an initial value problem of the form

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t)\,, \qquad \mathbf{x}(t_0) = \mathbf{x}_0\,, \tag{12}$$

we can use the initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ to solve for $\mathbf{c}$ in (11). Expressing $\mathbf{x}(t)$ using a definite integral, we have

$$\mathbf{x}(t) = \mathbf{X}(t)\mathbf{c} + \mathbf{X}(t)\int_{t_0}^{t} \mathbf{X}^{-1}(s)\mathbf{f}(s)ds\,.$$

From the initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$, we find

$$\mathbf{x}_0 = \mathbf{x}(t_0) = \mathbf{X}(t_0)\mathbf{c} + \mathbf{X}(t_0)\int_{t_0}^{t_0} \mathbf{X}^{-1}(s)\mathbf{f}(s)ds = \mathbf{X}(t_0)\mathbf{c}\,.$$

Solving for $\mathbf{c}$, we have $\mathbf{c} = \mathbf{X}^{-1}(t_0)\mathbf{x}_0$. Thus, the solution to (12) is given by the formula

$$\mathbf{x}(t) = \mathbf{X}(t)\mathbf{X}^{-1}(t_0)\mathbf{x}_0 + \mathbf{X}(t)\int_{t_0}^{t} \mathbf{X}^{-1}(s)\mathbf{f}(s)\,ds\,. \tag{13}$$

To apply the variation of parameters formulas, we first must determine a fundamental matrix $\mathbf{X}(t)$ for the homogeneous system. In the case when the coefficient matrix $\mathbf{A}$ is constant, we have discussed methods for finding $\mathbf{X}(t)$. However, if the entries of $\mathbf{A}$ depend on t, the determination of $\mathbf{X}(t)$ may be extremely difficult (entailing, perhaps, a matrix power series!).

Example 2 Find the solution to the initial value problem

$$\mathbf{x}'(t) = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} e^{2t} \\ 1 \end{bmatrix}, \qquad \mathbf{x}(0) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}. \tag{14}$$

Solution In Example 3 in Section 9.5, we found two linearly independent solutions to the corresponding homogeneous system; namely,

$$\mathbf{x}_1(t) = \begin{bmatrix} 3e^t \\ e^t \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2(t) = \begin{bmatrix} e^{-t} \\ e^{-t} \end{bmatrix}.$$

Hence a fundamental matrix for the homogeneous system is

$$\mathbf{X}(t) = \begin{bmatrix} 3e^t & e^{-t} \\ e^t & e^{-t} \end{bmatrix}.$$

Although the solution to (14) can be found via the method of undetermined coefficients, we shall find it directly from formula (13). For this purpose, we need $\mathbf{X}^{-1}(t)$. By formula (3) of Section 9.3 (page 510):

$$\mathbf{X}^{-1}(t) = \begin{bmatrix} \frac{1}{2}e^{-t} & -\frac{1}{2}e^{-t} \\ -\frac{1}{2}e^t & \frac{3}{2}e^t \end{bmatrix}.$$

Substituting into formula (13), we obtain the solution

$$\begin{aligned}
\mathbf{x}(t) &= \begin{bmatrix} 3e^t & e^{-t} \\ e^t & e^{-t} \end{bmatrix}\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{bmatrix}\begin{bmatrix} -1 \\ 0 \end{bmatrix} \\
&\quad + \begin{bmatrix} 3e^t & e^{-t} \\ e^t & e^{-t} \end{bmatrix}\int_0^t \begin{bmatrix} \frac{1}{2}e^{-s} & -\frac{1}{2}e^{-s} \\ -\frac{1}{2}e^s & \frac{3}{2}e^s \end{bmatrix}\begin{bmatrix} e^{2s} \\ 1 \end{bmatrix} ds \\
&= \begin{bmatrix} -\frac{3}{2}e^t + \frac{1}{2}e^{-t} \\ -\frac{1}{2}e^t + \frac{1}{2}e^{-t} \end{bmatrix} + \begin{bmatrix} 3e^t & e^{-t} \\ e^t & e^{-t} \end{bmatrix}\int_0^t \begin{bmatrix} \frac{1}{2}e^s - \frac{1}{2}e^{-s} \\ -\frac{1}{2}e^{3s} + \frac{3}{2}e^s \end{bmatrix} ds \\
&= \begin{bmatrix} -\frac{3}{2}e^t + \frac{1}{2}e^{-t} \\ -\frac{1}{2}e^t + \frac{1}{2}e^{-t} \end{bmatrix} + \begin{bmatrix} 3e^t & e^{-t} \\ e^t & e^{-t} \end{bmatrix}\begin{bmatrix} \frac{1}{2}e^t + \frac{1}{2}e^{-t} - 1 \\ \frac{3}{2}e^t - \frac{1}{6}e^{3t} - \frac{4}{3} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{9}{2}e^t - \frac{5}{6}e^{-t} + \frac{4}{3}e^{2t} + 3 \\ -\frac{3}{2}e^t - \frac{5}{6}e^{-t} + \frac{1}{3}e^{2t} + 2 \end{bmatrix}. \quad \blacklozenge
\end{aligned}$$

9.7 EXERCISES

In Problems 1–6, use the method of undetermined coefficients to find a general solution to the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$, where $\mathbf{A}$ and $\mathbf{f}(t)$ are given.

1. $\mathbf{A} = \begin{bmatrix} 6 & 1 \\ 4 & 3 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} -11 \\ -5 \end{bmatrix}$

2. $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} -t-1 \\ -4t-2 \end{bmatrix}$

3. $\mathbf{A} = \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} 2e^t \\ 4e^t \\ -2e^t \end{bmatrix}$

4. $\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} -4\cos t \\ -\sin t \end{bmatrix}$

5. $\mathbf{A} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} e^{2t} \\ \sin t \\ t \end{bmatrix}$

6. $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{f}(t) = e^{-2t}\begin{bmatrix} t \\ 3 \end{bmatrix}$

In Problems 7–10, use the method of undetermined coefficients to determine only the form of a particular solution for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$, *where* $\mathbf{A}$ *and* $\mathbf{f}(t)$ *are given.*

7. $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} \sin 3t \\ t \end{bmatrix}$

8. $\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 2 & 2 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} t^2 \\ t+1 \end{bmatrix}$

9. $\mathbf{A} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} e^{2t} \\ \sin t \\ t \end{bmatrix}$

10. $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} te^{-t} \\ 3e^{-t} \end{bmatrix}$

In Problems 11–16, use the variation of parameters formula (11) to find a general solution of the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$, *where* $\mathbf{A}$ *and* $\mathbf{f}(t)$ *are given.*

11. $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

12. $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

13. $\mathbf{A} = \begin{bmatrix} 8 & -4 \\ 4 & -2 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} t^{-2}/2 \\ t^{-2} \end{bmatrix}$

14. $\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} t^2 \\ 1 \end{bmatrix}$

15. $\mathbf{A} = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} t^{-1} \\ 4+2t^{-1} \end{bmatrix}$

16. $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} 8\sin t \\ 0 \end{bmatrix}$

In Problems 17–20, use the variation of parameters formulas (11) and possibly a linear algebra software package to find a general solution of the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$, *where* $\mathbf{A}$ *and* $\mathbf{f}(t)$ *are given.*

17. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} 3e^{t} \\ -e^{t} \\ -e^{t} \end{bmatrix}$

18. $\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & -1 & 2 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} 0 \\ e^{t} \\ e^{t} \end{bmatrix}$

19. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} t \\ 0 \\ e^{-t} \\ t \end{bmatrix}$

20. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 8 & -4 & -2 & -1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} e^{t} \\ 0 \\ 1 \\ 0 \end{bmatrix}$

In Problems 21 and 22, find the solution to the given system that satisfies the given initial condition.

21. $\mathbf{x}'(t) = \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} e^{t} \\ -e^{t} \end{bmatrix}$,

(a) $\mathbf{x}(0) = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$ **(b)** $\mathbf{x}(1) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(c) $\mathbf{x}(5) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ **(d)** $\mathbf{x}(-1) = \begin{bmatrix} -4 \\ 5 \end{bmatrix}$

22. $\mathbf{x}'(t) = \begin{bmatrix} 0 & 2 \\ 4 & -2 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} 4t \\ -4t-2 \end{bmatrix}$,

(a) $\mathbf{x}(0) = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$ **(b)** $\mathbf{x}(2) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

23. Using matrix algebra techniques and the method of undetermined coefficients, find a general solution for

$$x''(t) + y'(t) - x(t) + y(t) = -1\,,$$
$$x'(t) + y'(t) - x(t) = t^2\,.$$

Compare your solution with the solution in Example 4 in Section 5.2, page 247.

24. Using matrix algebra techniques and the method of undetermined coefficients, solve the initial value problem

$$x'(t) - 2y(t) = 4t\,, \qquad x(0) = 4\,;$$
$$y'(t) + 2y(t) - 4x(t) = -4t - 2\,, \qquad y(0) = -5\,.$$

Compare your solution with the solution in Example 1 in Section 7.10, page 412.

25. To find a general solution to the system

$$\mathbf{x}'(t) = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}\mathbf{x}(t) + \mathbf{f}(t)\,, \quad \text{where} \quad \mathbf{f}(t) = \begin{bmatrix} e^{t} \\ 0 \end{bmatrix},$$

proceed as follows:

(a) Find a fundamental solution set for the corresponding homogeneous system.

(b) The obvious choice for a particular solution would be a vector function of the form $\mathbf{x}_p(t) = e^{t}\mathbf{a}$; however, the homogeneous system has a solution of this form. The next choice would be $\mathbf{x}_p(t) = te^{t}\mathbf{a}$. Show that this choice does *not* work.

(c) For systems, multiplying by t is not always sufficient. The proper guess is

$$\mathbf{x}_p(t) = te^{t}\mathbf{a} + e^{t}\mathbf{b}\,.$$

Use this guess to find a particular solution of the given system.

(d) Use the results of parts (a) and (c) to find a general solution of the given system.

26. For the system of Problem 25, we found that a proper guess for a particular solution is $\mathbf{x}_p(t) = te^t\mathbf{a} + e^t\mathbf{b}$. In some cases $\mathbf{a}$ or $\mathbf{b}$ may be zero.

(a) Find a particular solution for the system of Problem 25 if $\mathbf{f}(t) = \text{col}(3e^t, 6e^t)$.

(b) Find a particular solution for the system of Problem 25 if $\mathbf{f}(t) = \text{col}(e^t, e^t)$.

27. Find a general solution of the system

$$\mathbf{x}'(t) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} -1 \\ -1 - e^{-t} \\ -2e^{-t} \end{bmatrix}.$$

[*Hint:* Try $\mathbf{x}_p(t) = e^{-t}\mathbf{a} + te^{-t}\mathbf{b} + \mathbf{c}$.]

28. Find a particular solution for the system

$$\mathbf{x}'(t) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} -3 \\ 1 \end{bmatrix}.$$

[*Hint:* Try $\mathbf{x}_p(t) = t\mathbf{a} + \mathbf{b}$.]

In Problems 29 and 30, find a general solution to the given Cauchy–Euler system for $t > 0$. (See Problem 42 in Exercises 9.5, page 533.) Remember to express the system in the form $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t)$ before using the variation of parameters formula.

29. $t\mathbf{x}'(t) = \begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} t^{-1} \\ 1 \end{bmatrix}$

30. $t\mathbf{x}'(t) = \begin{bmatrix} 4 & -3 \\ 8 & -6 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} t \\ 2t \end{bmatrix}$

31. Use the variation of parameters formula (10) to derive a formula for a particular solution y_p to the scalar equation $y'' + p(t)y' + q(t)y = g(t)$ in terms of two linearly independent solutions $y_1(t), y_2(t)$ of the corresponding homogeneous equation. Show that your answer agrees with the formulas derived in Section 4.6. [*Hint:* First write the scalar equation in system form.]

32. Conventional Combat Model. A simplistic model of a pair of conventional forces in combat yields the following system:

$$\mathbf{x}' = \begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}\mathbf{x} + \begin{bmatrix} p \\ q \end{bmatrix},$$

where $\mathbf{x} = \text{col}(x_1, x_2)$. The variables $x_1(t)$ and $x_2(t)$ represent the strengths of opposing forces at time t. The terms $-ax_1$ and $-dx_2$ represent the *operational loss rates*, and the terms $-bx_2$ and $-cx_1$ represent the *combat loss rates* for the troops x_1 and x_2, respectively. The constants p and q represent the respective rates of reinforcement. Let $a = 1$, $b = 4$, $c = 3$, $d = 2$, and $p = q = 5$. By solving the appropriate initial value problem, determine which forces will win if

(a) $x_1(0) = 20$, $\quad x_2(0) = 20$.
(b) $x_1(0) = 21$, $\quad x_2(0) = 20$.
(c) $x_1(0) = 20$, $\quad x_2(0) = 21$.

33. *RL* Network. The currents in the *RL* network given by the schematic diagram in Figure 9.8 are governed by the following equations:

$$2I_1'(t) + 90I_2(t) = 9,$$
$$I_3'(t) + 30I_4(t) - 90I_2(t) = 0,$$
$$60I_5(t) - 30I_4(t) = 0,$$
$$I_1(t) = I_2(t) + I_3(t),$$
$$I_3(t) = I_4(t) + I_5(t).$$

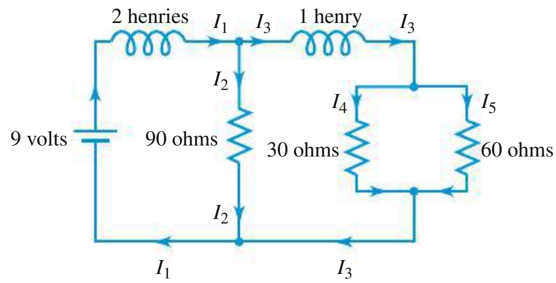


Figure 9.8 *RL* network for Problem 33

Assume the currents are initially zero. Solve for the five currents $I_1, \ldots, I_5$. [*Hint:* Eliminate all unknowns except I_2 and I_5, and form a normal system with $x_1 = I_2$ and $x_2 = I_5$.]

34. Mixing Problem. Two tanks A and B, each holding 50 L of liquid, are interconnected by pipes. The liquid flows from tank A into tank B at a rate of 4 L/min and from B into A at a rate of 1 L/min (see Figure 9.9). The liquid inside each tank is kept well stirred. A brine solution that has a concentration of 0.2 kg/L of salt flows into tank A at a rate of 4 L/min. A brine solution that has a concentration of 0.1 kg/L of salt flows into tank B at a rate of 1 L/min. The solutions flow out of the system from both tanks—from tank A at 1 L/min and from tank B at 4 L/min. If, initially, tank A contains pure water and tank B contains 0.5 kg of salt, determine the mass of salt in each tank at time $t \geq 0$. After several minutes have elapsed, which tank has the higher concentration of salt? What is its limiting concentration?

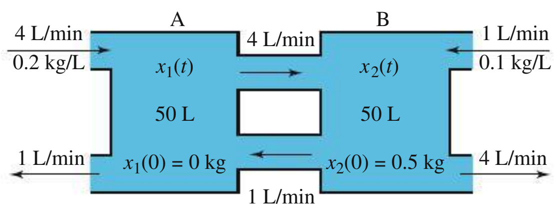


Figure 9.9 Mixing problem for interconnected tanks

9.8 The Matrix Exponential Function

In this chapter we have developed various ways to extend techniques for scalar differential equations to systems. In this section we take a substantial step further by showing that with the right notation, the formulas for solving normal systems with constant coefficients are identical to the formulas for solving first-order equations with constant coefficients. For example, we know that a general solution to the equation $x'(t) = ax(t)$, where a is a constant, is $x(t) = ce^{at}$. Analogously, we show that a general solution to the normal system

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)\,, \tag{1}$$

where $\mathbf{A}$ is a constant $n \times n$ matrix, is $\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{c}$. Our first task is to define the matrix exponential $e^{\mathbf{A}t}$.

If $\mathbf{A}$ is a constant $n \times n$ matrix, we define $e^{\mathbf{A}t}$ by taking the series expansion for e^{at} and replacing a by $\mathbf{A}$; that is,

$$e^{\mathbf{A}t} := \mathbf{I} + \mathbf{A}t + \mathbf{A}^2\frac{t^2}{2!} + \cdots + \mathbf{A}^n\frac{t^n}{n!} + \cdots. \tag{2}$$

(Note that we also replace 1 by $\mathbf{I}$.) By the right-hand side of (2), we mean the $n \times n$ matrix whose elements are power series with coefficients given by the corresponding entries in the matrices $\mathbf{I}, \mathbf{A}, \mathbf{A}^2/2!, \ldots$.

If $\mathbf{A}$ is a diagonal matrix, then the computation of $e^{\mathbf{A}t}$ is straightforward. For example, if

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

then

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \qquad \mathbf{A}^3 = \begin{bmatrix} -1 & 0 \\ 0 & 8 \end{bmatrix}, \qquad \ldots, \qquad \mathbf{A}^n = \begin{bmatrix} (-1)^n & 0 \\ 0 & 2^n \end{bmatrix},$$

and so

$$e^{\mathbf{A}t} = \sum_{n=0}^{\infty} \mathbf{A}^n \frac{t^n}{n!} = \begin{bmatrix} \displaystyle\sum_{n=0}^{\infty} (-1)^n \frac{t^n}{n!} & 0 \\ 0 & \displaystyle\sum_{n=0}^{\infty} 2^n \frac{t^n}{n!} \end{bmatrix} = \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{2t} \end{bmatrix}.$$

More generally, if $\mathbf{A}$ is an $n \times n$ *diagonal* matrix with $r_1, r_2, \ldots, r_n$ down its main diagonal, then $e^{\mathbf{A}t}$ is the diagonal matrix with $e^{r_1 t}, e^{r_2 t}, \ldots, e^{r_n t}$ down its main diagonal (see Problem 26). If $\mathbf{A}$ is not a diagonal matrix, the computation of $e^{\mathbf{A}t}$ is more involved. We deal with this important problem later in this section.

It can be shown that the series (2) converges for all t and has many of the same properties[†] as the scalar exponential e^{at}.

[†]For proofs of these and other properties of the matrix exponential function, see *Fundamentals of Matrix Analysis with Applications*, by Edward Barry Saff and Arthur David Snider (John Wiley & Sons, Hoboken, New Jersey, 2016), Chapter 7. See also the amusing articles, "Nineteen Dubious Ways to Compute the Exponential of a Matrix," by Cleve Moler and Charles van Loan, *SIAM Review*, Vol. 20, No. 4 (Oct. 1978), and "Nineteen . . . Matrix, Twenty-Five Years Later," ibid., Vol. 45, No. 1 (Jan. 2003).

Properties of the Matrix Exponential Function

Theorem 7. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ constant matrices and r, s, and t be real (or complex) numbers. Then,

(a) $e^{\mathbf{A}0} = e^{\mathbf{0}} = \mathbf{I}$.

(b) $e^{\mathbf{A}(t+s)} = e^{\mathbf{A}t}e^{\mathbf{A}s}$.

(c) $(e^{\mathbf{A}t})^{-1} = e^{-\mathbf{A}t}$.

(d) $e^{(\mathbf{A}+\mathbf{B})t} = e^{\mathbf{A}t}e^{\mathbf{B}t}$, provided that $\mathbf{AB} = \mathbf{BA}$.

(e) $e^{r\mathbf{I}t} = e^{rt}\mathbf{I}$.

Property (c) has profound implications. First, it asserts that for any matrix $\mathbf{A}$, the matrix $e^{\mathbf{A}t}$ *has* an inverse for all t. Moreover, this inverse is obtained by simply replacing t by $-t$. (Note that (c) follows from (a) and (b) with $s = -t$.) In applying property (d) (the law of exponents), one must exercise care because of the stipulation that the matrices $\mathbf{A}$ and $\mathbf{B}$ commute (see Problem 25).

Another important property of the matrix exponential arises from the fact that we can differentiate the series in (2) term by term. This gives

$$\begin{aligned}\frac{d}{dt}(e^{\mathbf{A}t}) &= \frac{d}{dt}\left(\mathbf{I} + \mathbf{A}t + \mathbf{A}^2\frac{t^2}{2} + \cdots + \mathbf{A}^n\frac{t^n}{n!} + \cdots\right)\\ &= \mathbf{A} + \mathbf{A}^2t + \mathbf{A}^3\frac{t^2}{2} + \cdots + \mathbf{A}^n\frac{t^{n-1}}{(n-1)!} + \cdots\\ &= \mathbf{A}\left[\mathbf{I} + \mathbf{A}t + \mathbf{A}^2\frac{t^2}{2} + \cdots + \mathbf{A}^{n-1}\frac{t^{n-1}}{(n-1)!} + \cdots\right].\end{aligned}$$

Hence,

$$\frac{d}{dt}(e^{\mathbf{A}t}) = \mathbf{A}e^{\mathbf{A}t},$$

and so $e^{\mathbf{A}t}$ *is a solution to the matrix differential equation* $\mathbf{X}' = \mathbf{AX}$. Since $e^{\mathbf{A}t}$ is invertible [property (c)], it follows that the columns of $e^{\mathbf{A}t}$ are linearly independent solutions to system (1). Combining these facts we have the following.

e^{At} Is a Fundamental Matrix

Theorem 8. If $\mathbf{A}$ is an $n \times n$ constant matrix, then the columns of the matrix exponential $e^{\mathbf{A}t}$ form a fundamental solution set for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$. Therefore, $e^{\mathbf{A}t}$ is a fundamental matrix for the system, and a general solution is $\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{c}$.

If a fundamental matrix $\mathbf{X}(t)$ for the system $\mathbf{x}' = \mathbf{Ax}$ has somehow been determined, it is easy to compute $e^{\mathbf{A}t}$, as the next theorem describes.

Relationship Between Fundamental Matrices

Theorem 9. Let $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ be two fundamental matrices for the same system $\mathbf{x}' = \mathbf{Ax}$. Then there exists a constant matrix $\mathbf{C}$ such that $\mathbf{Y}(t) = \mathbf{X}(t)\mathbf{C}$ for all t. In particular,

$$e^{\mathbf{A}t} = \mathbf{X}(t)\mathbf{X}(0)^{-1}. \tag{3}$$

Proof. Since $\mathbf{X}(t)$ is a fundamental matrix, every column of $\mathbf{Y}(t)$ can be expressed as $\mathbf{X}(t)\mathbf{c}$ for a suitable constant vector $\mathbf{c}$, so column-by-column we have

$$\left[\quad \mathbf{Y}(t) \quad\right] = \left[\quad \mathbf{X}(t) \quad\right]\begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \mathbf{X}(t)\mathbf{C}\,.$$

If we choose $\mathbf{Y}(t) = e^{\mathbf{A}t} = \mathbf{X}(t)\mathbf{C}$, then (3) follows by setting $t = 0$. ◆

If the $n \times n$ matrix $\mathbf{A}$ has n linearly independent eigenvectors $\mathbf{u}_i$, then Theorem 5 of Section 9.5 (page 526) provides us with $\mathbf{X}(t)$ and (3) gives us

$$e^{\mathbf{A}t} = [e^{r_1 t}\mathbf{u}_1 \; e^{r_2 t}\mathbf{u}_2 \; \cdots \; e^{r_n t}\mathbf{u}_n][\mathbf{u}_1 \; \mathbf{u}_2 \; \cdots \; \mathbf{u}_n]^{-1}\,. \tag{4}$$

Are there any other ways that we can compute $e^{\mathbf{A}t}$? As we observed, if $\mathbf{A}$ is a diagonal matrix, then we simply exponentiate the diagonal elements of $\mathbf{A}t$ to obtain $e^{\mathbf{A}t}$. Also, if $\mathbf{A}$ is a **nilpotent** matrix, that is, $\mathbf{A}^k = \mathbf{0}$ for some positive integer k, then the series for $e^{\mathbf{A}t}$ has only a finite number of terms—it "truncates"—since $\mathbf{A}^k = \mathbf{A}^{k+1} = \cdots = \mathbf{0}$. In such cases, $e^{\mathbf{A}t}$ reduces to

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \cdots + \mathbf{A}^{k-1}\frac{t^{k-1}}{(k-1)!} + \mathbf{0} + \mathbf{0} + \cdots = \mathbf{I} + \mathbf{A}t + \cdots + \mathbf{A}^{k-1}\frac{t^{k-1}}{(k-1)!}\,.$$

Thus $e^{\mathbf{A}t}$ can be calculated in finite terms if $\mathbf{A}$ is diagonal or nilpotent. Can we take this any further? Yes; a consequence of the Cayley–Hamilton theorem† is that when the characteristic polynomial for $\mathbf{A}$ has the form $p(r) = (r_1 - r)^n$, then $(r_1\mathbf{I} - \mathbf{A})^n = \mathbf{0} = (-1)^n(\mathbf{A} - r_1\mathbf{I})^n$. So if $\mathbf{A}$ has only one (multiple) eigenvalue r_1, then $\mathbf{A} - r_1\mathbf{I}$ is nilpotent, and we exploit that by writing $\mathbf{A} = r_1\mathbf{I} + \mathbf{A} - r_1\mathbf{I}$:

$$e^{\mathbf{A}t} = e^{r_1\mathbf{I}t}e^{(\mathbf{A}-r_1\mathbf{I})t} = e^{r_1 t}\left[\mathbf{I} + (\mathbf{A} - r_1\mathbf{I})t + \cdots + (\mathbf{A} - r_1\mathbf{I})^{n-1}\frac{t^{n-1}}{(n-1)!}\right].$$

Example 1 Find the fundamental matrix $e^{\mathbf{A}t}$ for the system

$$\mathbf{x}' = \mathbf{A}\mathbf{x}\,, \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ -2 & -2 & -1 \end{bmatrix}.$$

Solution We begin by computing the characteristic polynomial for $\mathbf{A}$:

$$p(r) = |\mathbf{A} - r\mathbf{I}| = \begin{vmatrix} 2-r & 1 & 1 \\ 1 & 2-r & 1 \\ -2 & -2 & -1-r \end{vmatrix} = -r^3 + 3r^2 - 3r + 1 = -(r-1)^3\,.$$

Thus, $r = 1$ is an eigenvalue of $\mathbf{A}$ with multiplicity 3. By the Cayley–Hamilton theorem, $(\mathbf{A} - \mathbf{I})^3 = \mathbf{0}$, and so

$$e^{\mathbf{A}t} = e^t e^{(\mathbf{A}-\mathbf{I})t} = e^t\left\{\mathbf{I} + (\mathbf{A} - \mathbf{I})t + (\mathbf{A} - \mathbf{I})^2\frac{t^2}{2}\right\}. \tag{5}$$

†The Cayley–Hamilton theorem states that a matrix satisfies its own characteristic equation, that is, $p(\mathbf{A}) = \mathbf{0}$. For a discussion of this theorem, see *Fundamentals of Matrix Analysis with Applications*, by Edward Barry Saff and Arthur David Snider (John Wiley & Sons, Hoboken, New Jersey, 2016), Section 6.3.

Computing, we find

$$\mathbf{A}-\mathbf{I}=\begin{bmatrix}1&1&1\\1&1&1\\-2&-2&-2\end{bmatrix}\quad\text{and}\quad(\mathbf{A}-\mathbf{I})^2=\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix}.$$

Substituting into (5) yields

$$e^{\mathbf{A}t}=e^t\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}+te^t\begin{bmatrix}1&1&1\\1&1&1\\-2&-2&-2\end{bmatrix}=\begin{bmatrix}e^t+te^t&te^t&te^t\\te^t&e^t+te^t&te^t\\-2te^t&-2te^t&e^t-2te^t\end{bmatrix}. \quad\blacklozenge \tag{6}$$

We are not through with nilpotency yet. What if we have a nonzero vector $\mathbf{u}$, an exponent m, and a scalar r satisfying $(\mathbf{A}-r\mathbf{I})^m\mathbf{u}=\mathbf{0}$, so that $\mathbf{A}-r\mathbf{I}$ is "nilpotent when restricted to $\mathbf{u}$"? Such vectors are given a (predictable) name.

Generalized Eigenvectors

Definition 5. Let $\mathbf{A}$ be a square matrix. A nonzero vector $\mathbf{u}$ satisfying

$$(\mathbf{A}-r\mathbf{I})^m\mathbf{u}=\mathbf{0} \tag{7}$$

for some scalar r and some positive integer m is called a **generalized eigenvector** associated with r.

[Note that r must be an eigenvalue of $\mathbf{A}$, since the final *nonzero* vector in the list $\mathbf{u}$, $(\mathbf{A}-r\mathbf{I})\mathbf{u}$, $(\mathbf{A}-r\mathbf{I})^2\mathbf{u}, \dots, (\mathbf{A}-r\mathbf{I})^{m-1}\mathbf{u}$ is a "regular" eigenvector.]

A valuable feature of generalized eigenvectors $\mathbf{u}$ is that we can compute $e^{\mathbf{A}t}\mathbf{u}$ in finite terms *without knowing* $e^{\mathbf{A}t}$, because

$$\begin{aligned}e^{\mathbf{A}t}\mathbf{u}&=e^{r\mathbf{I}t}e^{(\mathbf{A}-r\mathbf{I})t}\mathbf{u}\\&=e^{rt}\left[\mathbf{I}\mathbf{u}+t(\mathbf{A}-r\mathbf{I})\mathbf{u}+\cdots+\frac{t^{m-1}}{(m-1)!}(\mathbf{A}-r\mathbf{I})^{m-1}\mathbf{u}+\frac{t^m}{m!}(\mathbf{A}-r\mathbf{I})^m\mathbf{u}+\cdots\right]\\&=e^{rt}\left[\mathbf{u}+t(\mathbf{A}-r\mathbf{I})\mathbf{u}+\cdots+\frac{t^{m-1}}{(m-1)!}(\mathbf{A}-r\mathbf{I})^{m-1}\mathbf{u}+\quad\mathbf{0}\quad+\cdots\right].\end{aligned} \tag{8}$$

Moreover, $e^{\mathbf{A}t}\mathbf{u}$ is a solution to the system $\mathbf{x}'=\mathbf{A}\mathbf{x}$ (recall Theorem 8). Hence if we can find n generalized eigenvectors $\mathbf{u}_i$ for the $n\times n$ matrix $\mathbf{A}$ that are linearly independent, the corresponding solutions $\mathbf{x}_i(t)=e^{\mathbf{A}t}\mathbf{u}_i$ will form a fundamental solution set and can be assembled into a fundamental matrix $\mathbf{X}(t)$. (Since the $\mathbf{x}_i$'s are solutions that reduce to the linearly independent $\mathbf{u}_i$'s at $t=0$, they are always linearly independent.) Finally, we get the matrix exponential by applying (3) from Theorem 9:

$$e^{\mathbf{A}t}=\mathbf{X}(t)\mathbf{X}(0)^{-1}=[e^{\mathbf{A}t}\mathbf{u}_1\ e^{\mathbf{A}t}\mathbf{u}_2\ \cdots\ e^{\mathbf{A}t}\mathbf{u}_n][\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n]^{-1}, \tag{9}$$

computed in a finite number of steps.

Of course, any (regular) eigenvector is a generalized eigenvector (corresponding to $m=1$), and if $\mathbf{A}$ has a full set of n linearly independent eigenvectors, then (9) is simply the representation (4). But what if $\mathbf{A}$ is *defective*, that is, possesses fewer than n linearly independent eigenvectors? Luckily, the primary decomposition theorem in advanced linear algebra* guarantees that when the characteristic polynomial of $\mathbf{A}$ is

$$p(r) = (r_1 - r)^{m_1}(r_2 - r)^{m_2}\cdots(r_k - r)^{m_k}, \tag{10}$$

where the r_i's are the distinct eigenvalues of $\mathbf{A}$ and m_i is the multiplicity of the eigenvalue r_i, then for each i there exist m_i linearly independent generalized eigenvectors satisfying

$$(\mathbf{A} - r_i\mathbf{I})^{m_i}\mathbf{u} = \mathbf{0}. \tag{11}$$

Furthermore, the conglomeration of these $n = m_1 + m_2 + \cdots + m_k$ generalized eigenvectors is linearly independent.†

We're home! The following scheme will always yield a fundamental solution set, for *any* square matrix $\mathbf{A}$.

Solving $\mathbf{x}' = \mathbf{A}\mathbf{x}$

To obtain a fundamental solution set for $\mathbf{x}' = \mathbf{A}\mathbf{x}$ for any constant square matrix $\mathbf{A}$:

(a) Compute the characteristic polynomial $p(r) = |\mathbf{A} - r\mathbf{I}|$.

(b) Find the zeros of $p(r)$ and express it as $p(r) = (r_1 - r)^{m_1}(r_2 - r)^{m_2}\cdots(r_k - r)^{m_k}$, where $r_1, r_2, \ldots, r_k$ are the distinct zeros (i.e., eigenvalues) and $m_1, m_2, \ldots, m_k$ are their multiplicities.

(c) For each eigenvalue r_i find m_i linearly independent generalized eigenvectors by applying the Gauss–Jordan algorithm to the system $(\mathbf{A} - r_i\mathbf{I})^{m_i}\mathbf{u} = \mathbf{0}$.

(d) Form $n = m_1 + m_2 + \cdots + m_k$ linearly independent solutions to $\mathbf{x}' = \mathbf{A}\mathbf{x}$ by computing

$$\mathbf{x}(t) := e^{\mathbf{A}t}\mathbf{u} = e^{rt}\left[\mathbf{u} + t(\mathbf{A} - r\mathbf{I})\mathbf{u} + \frac{t^2}{2!}(\mathbf{A} - r\mathbf{I})^2\mathbf{u} + \cdots\right]$$

for each generalized eigenvector $\mathbf{u}$ found in part (c) and corresponding eigenvalue r. If r has multiplicity m, this series terminates after m or fewer terms.

We can then, if desired, assemble the fundamental matrix $\mathbf{X}(t)$ from the n solutions and obtain the matrix exponential $e^{\mathbf{A}t}$ using (9).‡

Example 2 Find the fundamental matrix $e^{\mathbf{A}t}$ for the system

$$\mathbf{x}' = \mathbf{A}\mathbf{x}, \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \tag{12}$$

Solution We begin by finding the characteristic polynomial for $\mathbf{A}$:

$$p(r) = |\mathbf{A} - r\mathbf{I}| = \begin{vmatrix} 1-r & 0 & 0 \\ 1 & 3-r & 0 \\ 0 & 1 & 1-r \end{vmatrix} = -(r-1)^2(r-3).$$

*See *Fundamentals of Matrix Analysis with Applications*, by Edward Barry Saff and Arthur David Snider (John Wiley & Sons, Hoboken, New Jersey, 2016), Section 7.3.

†Some of the generalized eigenvectors satisfy (11) with a lower exponent, but that will not affect the calculation.

‡With defective matrices this algorithm may prove ineffective for computer implementation; rounding effects inevitably foil the machine's ability to detect multiple eigenvalues.

Hence, the eigenvalues of $\mathbf{A}$ are $r = 1$ with multiplicity 2 and $r = 3$ with multiplicity 1.

Since $r = 1$ has multiplicity 2, we must determine two linearly independent associated generalized eigenvectors satisfying $(\mathbf{A} - \mathbf{I})^2\mathbf{u} = \mathbf{0}$. From

$$(\mathbf{A} - \mathbf{I})^2\mathbf{u} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 4 & 0 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

we find $u_2 = s$, $u_1 = -2u_2 = -2s$, and $u_3 = v$, where s and v are arbitrary.

Taking $s = 0$ and $v = 1$, we obtain the generalized eigenvector $\mathbf{u}_1 = \operatorname{col}(0, 0, 1)$. The corresponding solution to (12) is

$$\textbf{(13)} \qquad \mathbf{x}_1(t) = e^t\{\mathbf{u}_1 + t(\mathbf{A} - \mathbf{I})\mathbf{u}_1\} = e^t\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + te^t\begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ e^t \end{bmatrix}$$

($\mathbf{u}_1$ is, in fact, a regular eigenvector).

Next we take $s = 1$ and $v = 0$ to derive the second linearly independent generalized eigenvector $\mathbf{u}_2 = \operatorname{col}(-2, 1, 0)$ and (linearly independent) solution

$$\textbf{(14)} \qquad \mathbf{x}_2(t) = e^{\mathbf{A}t}\mathbf{u}_2 = e^t\{\mathbf{u}_2 + t(\mathbf{A} - \mathbf{I})\mathbf{u}_2\}$$
$$= e^t\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + te^t\begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$
$$= e^t\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + te^t\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -2e^t \\ e^t \\ te^t \end{bmatrix}.^{\dagger}$$

For the eigenvalue $r = 3$, we solve $(\mathbf{A} - 3\mathbf{I})\mathbf{u} = \mathbf{0}$, that is,

$$\begin{bmatrix} -2 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & -2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

to obtain the eigenvector $\mathbf{u}_3 = \operatorname{col}(0, 2, 1)$. Hence, a third linearly independent solution to (12) is

$$\textbf{(15)} \qquad \mathbf{x}_3(t) = e^{3t}\mathbf{u}_3 = e^{3t}\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2e^{3t} \\ e^{3t} \end{bmatrix}.$$

The matrix $\mathbf{X}(t)$ whose columns are the vectors $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$, and $\mathbf{x}_3(t)$,

$$\mathbf{X}(t) = \begin{bmatrix} 0 & -2e^t & 0 \\ 0 & e^t & 2e^{3t} \\ e^t & te^t & e^{3t} \end{bmatrix},$$

†Note that $\mathbf{x}_2(t)$ has been expressed in the format discussed in Problem 35 of Exercises 9.5, page 532.

is a fundamental matrix for (12). Setting $t = 0$ and then computing $\mathbf{X}^{-1}(0)$, we find

$$\mathbf{X}(0) = \begin{bmatrix} 0 & -2 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{X}^{-1}(0) = \begin{bmatrix} -\frac{1}{4} & -\frac{1}{2} & 1 \\ -\frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}.$$

It now follows from formula (3) that

$$e^{\mathbf{A}t} = \mathbf{X}(t)\mathbf{X}^{-1}(0) = \begin{bmatrix} 0 & -2e^t & 0 \\ 0 & e^t & 2e^{3t} \\ e^t & te^t & e^{3t} \end{bmatrix} \begin{bmatrix} -\frac{1}{4} & -\frac{1}{2} & 1 \\ -\frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} e^t & 0 & 0 \\ -\frac{1}{2}e^t + \frac{1}{2}e^{3t} & e^{3t} & 0 \\ -\frac{1}{4}e^t - \frac{1}{2}te^t + \frac{1}{4}e^{3t} & -\frac{1}{2}e^t + \frac{1}{2}e^{3t} & e^t \end{bmatrix}. \quad \blacklozenge$$

Use of the fundamental matrix $e^{\mathbf{A}t}$ simplifies many computations. For example, the properties $e^{\mathbf{A}t}e^{-\mathbf{A}s} = e^{\mathbf{A}(t-s)}$ and $(e^{\mathbf{A}t_0})^{-1} = e^{-\mathbf{A}t_0}$ enable us to rewrite the variation of parameters formula (13) in Section 9.7 (page 541) in a simpler form. Namely, the solution to the initial value problem $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}(t)$, $\mathbf{x}(t_0) = \mathbf{x}_0$ is given by

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}_0 + \int_{t_0}^{t} e^{\mathbf{A}(t-s)}\mathbf{f}(s)\,ds\,, \tag{16}$$

which is a system version of the formula for the solution to the scalar initial value problem $x' = ax + f(t)$, $x(t_0) = x_0$.

In closing, we remark that the software packages for eigenvalue computation listed in Section 9.5 (page 524) also contain subroutines for computing the matrix exponential.

9.8 EXERCISES

In Problems 1–6, (a) show that the given matrix $\mathbf{A}$ *satisfies* $(\mathbf{A} - r\mathbf{I})^k = \mathbf{0}$ *for some number* r *and some positive integer* k *and (b) use this fact to determine the matrix* $e^{\mathbf{A}t}$. *[Hint: Compute the characteristic polynomial and use the Cayley–Hamilton theorem.]*

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

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

3. $\mathbf{A} = \begin{bmatrix} 2 & 1 & -1 \\ -3 & -1 & 1 \\ 9 & 3 & -4 \end{bmatrix}$

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

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

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

In Problems 7–10, determine $e^{\mathbf{A}t}$ by first finding a fundamental matrix $\mathbf{X}(t)$ for $\mathbf{x}' = \mathbf{A}\mathbf{x}$ and then using formula (3).

7. $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

8. $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$

9. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}$

10. $\mathbf{A} = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}$

In Problems 11 and 12, determine $e^{\mathbf{A}t}$ by using generalized eigenvectors to find a fundamental matrix and then using formula (3).

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

12. $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$

In Problems 13–16 determine $e^{\mathbf{A}t}$, using a linear software package to find the eigenvectors.

13. $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & -3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix}$

14. $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix}$

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

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

In Problems 17–20, use the generalized eigenvectors of $\mathbf{A}$ to find a general solution to the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, where $\mathbf{A}$ is given.

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

18. $\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$

19. $\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 1 & 1 & 2 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

20. $\mathbf{A} = \begin{bmatrix} -1 & -8 & 1 \\ -1 & -3 & 2 \\ -4 & -16 & 7 \end{bmatrix}$

21. Use the results of Problem 5 to find the solution to the initial value problem

$$\mathbf{x}'(t) = \begin{bmatrix} -2 & 0 & 0 \\ 4 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix}\mathbf{x}(t), \qquad \mathbf{x}(0) = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

22. Use your answer to Problem 12 to find the solution to the initial value problem

$$\mathbf{x}'(t) = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}\mathbf{x}(t), \qquad \mathbf{x}(0) = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}.$$

23. Use the results of Problem 3 and the variation of parameters formula (16) to find the solution to the initial value problem

$$\mathbf{x}'(t) = \begin{bmatrix} 2 & 1 & -1 \\ -3 & -1 & 1 \\ 9 & 3 & -4 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} 0 \\ t \\ 0 \end{bmatrix},$$

$$\mathbf{x}(0) = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}.$$

24. Use your answer to Problem 9 and the variation of parameters formula (16) to find the solution to the initial value problem

$$\mathbf{x}'(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} 0 \\ 0 \\ t \end{bmatrix},$$

$$\mathbf{x}(0) = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

25. Let

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

(a) Show that $\mathbf{AB} \neq \mathbf{BA}$.

(b) Show that property (d) in Theorem 7 does not hold for these matrices. That is, show that $e^{(\mathbf{A}+\mathbf{B})t} \neq e^{\mathbf{A}t}e^{\mathbf{B}t}$.

26. Let $\mathbf{A}$ be a diagonal $n \times n$ matrix with entries $r_1, \ldots, r_n$ down its main diagonal. To compute $e^{\mathbf{A}t}$, proceed as follows:

(a) Show that $\mathbf{A}^k$ is the diagonal matrix with entries $r_1^k, \ldots, r_n^k$ down its main diagonal.

(b) Use the result of part (a) and the defining equation (2) to show that $e^{\mathbf{A}t}$ is the diagonal matrix with entries $e^{r_1 t}, \ldots, e^{r_n t}$ down its main diagonal.

27. In Problems 35–40 of Exercises 9.5, page 532, some ad hoc formulas were invoked to find general solutions to the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ when $\mathbf{A}$ had repeated eigenvalues.

Using the generalized eigenvector procedure outlined on page 548, justify the ad hoc formulas proposed in

(a) Problem 35, Exercises 9.5.
(b) Problem 37, Exercises 9.5.
(c) Problem 39, Exercises 9.5.

28. Let

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

(a) Find a general solution to $\mathbf{x}' = \mathbf{A}\mathbf{x}$.

(b) Determine which initial conditions $\mathbf{x}(0) = \mathbf{x}_0$ yield a solution $\mathbf{x}(t) = \text{col}\left(x_1(t), x_2(t), x_3(t)\right)$ that remains bounded for all $t \geq 0$; that is, satisfies

$$||\mathbf{x}(t)|| := \sqrt{x_1^2(t) + x_2^2(t) + x_3^2(t)} \leq M$$

for some constant M and all $t \geq 0$.

29. For the matrix **A** in Problem 28, solve the initial value problem

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \sin(2t)\begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix},$$

$$\mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Chapter 9 Summary

In this chapter we discussed the theory of linear systems in normal form and presented methods for solving such systems. The theory and methods are natural extensions of the development for second-order and higher-order linear equations. The important properties and techniques are as follows.

Homogeneous Normal Systems

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$$

The $n \times n$ matrix function $\mathbf{A}(t)$ is assumed to be continuous on an interval I.

Fundamental Solution Set: $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. A set of n vector solutions $\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)$ of the homogeneous system on the interval I form a **fundamental solution set** provided they are linearly independent on I or, equivalently, their **Wronskian**

$$W[\mathbf{x}_1, \ldots, \mathbf{x}_n](t) := \det[\mathbf{x}_1, \ldots, \mathbf{x}_n] = \begin{vmatrix} x_{1,1}(t) & x_{1,2}(t) & \cdots & x_{1,n}(t) \\ x_{2,1}(t) & x_{2,2}(t) & \cdots & x_{2,n}(t) \\ \vdots & \vdots & & \vdots \\ x_{n,1}(t) & x_{n,2}(t) & \cdots & x_{n,n}(t) \end{vmatrix}$$

is never zero on I.

Fundamental Matrix: $\mathbf{X}(t)$. An $n \times n$ matrix function $\mathbf{X}(t)$ whose column vectors form a fundamental solution set for the homogeneous system is called a **fundamental matrix**. The determinant of $\mathbf{X}(t)$ is the Wronskian of the fundamental solution set. Since the Wronskian is never zero on the interval I, then $\mathbf{X}^{-1}(t)$ exists for t in I.

General Solution to Homogeneous System: $\mathbf{Xc} = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$. If $\mathbf{X}(t)$ is a fundamental matrix whose column vectors are $\mathbf{x}_1, \ldots, \mathbf{x}_n$, then a general solution to the homogeneous system is

$$\mathbf{x}(t) = \mathbf{X}(t)\mathbf{c} = c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \cdots + c_n\mathbf{x}_n(t) ,$$

where $\mathbf{c} = \operatorname{col}(c_1, \ldots, c_n)$ is an arbitrary constant vector.

Homogeneous Systems with Constant Coefficients. The form of a general solution for a homogeneous system with constant coefficients depends on the eigenvalues and eigenvectors of the $n \times n$ constant matrix $\mathbf{A}$. An **eigenvalue** of $\mathbf{A}$ is a number r such that the system $\mathbf{Au} = r\mathbf{u}$ has a nontrivial solution $\mathbf{u}$, called an **eigenvector** of $\mathbf{A}$ associated with the eigenvalue r. Finding the eigenvalues of A is equivalent to finding the roots of the **characteristic equation**

$$|\mathbf{A} - r\mathbf{I}| = 0 .$$

The corresponding eigenvectors are found by solving the system $(\mathbf{A} - r\mathbf{I})\mathbf{u} = \mathbf{0}$.

If the matrix $\mathbf{A}$ has n linearly independent eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and r_i is the eigenvalue corresponding to the eigenvector $\mathbf{u}_i$, then

$$\{e^{r_1 t}\mathbf{u}_1, e^{r_2 t}\mathbf{u}_2, \ldots, e^{r_n t}\mathbf{u}_n\}$$

is a fundamental solution set for the homogeneous system. A class of matrices that always has n linearly independent eigenvectors is the set of **symmetric** matrices—that is, matrices that satisfy $\mathbf{A} = \mathbf{A}^T$.

If $\mathbf{A}$ has complex conjugate eigenvalues $\alpha \pm i\beta$ and associated eigenvectors $\mathbf{z} = \mathbf{a} \pm i\mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are real vectors, then two linearly independent real vector solutions to the homogeneous system are

$$e^{\alpha t}\cos\beta t\,\mathbf{a} - e^{\alpha t}\sin\beta t\,\mathbf{b} , \qquad e^{\alpha t}\sin\beta t\,\mathbf{a} + e^{\alpha t}\cos\beta t\,\mathbf{b} .$$

When $\mathbf{A}$ has a repeated eigenvalue r of multiplicity m, then it is possible that $\mathbf{A}$ does *not* have n linearly independent eigenvectors. However, associated with r are m linearly independent **generalized eigenvectors** that can be used to generate m linearly independent solutions to the homogeneous system (see page 548 under "Generalized Eigenvectors").

Nonhomogeneous Normal Systems

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t)$$

The $n \times n$ matrix function $\mathbf{A}(t)$ and the vector function $\mathbf{f}(t)$ are assumed continuous on an interval I.

General Solution to Nonhomogeneous System: $\mathbf{x}_p + \mathbf{Xc}$. If $\mathbf{x}_p(t)$ is any particular solution for the nonhomogeneous system and $\mathbf{X}(t)$ is a fundamental matrix for the associated homogeneous system, then a general solution for the nonhomogeneous system is

$$\mathbf{x}(t) = \mathbf{x}_p(t) + \mathbf{X}(t)\mathbf{c} = \mathbf{x}_p(t) + c_1\,\mathbf{x}_1(t) + \cdots + c_n\,\mathbf{x}_n(t) ,$$

where $\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)$ are the column vectors of $\mathbf{X}(t)$ and $\mathbf{c} = \operatorname{col}(c_1, \ldots, c_n)$ is an arbitrary constant vector.

Undetermined Coefficients. If the nonhomogeneous term $\mathbf{f}(t)$ is a vector whose components are polynomials, exponential or sinusoidal functions, and $\mathbf{A}$ is a constant matrix, then one can use an extension of the method of undetermined coefficients to decide the form of a particular solution to the nonhomogeneous system.

Variation of Parameters: $\mathbf{X}(t)\boldsymbol{v}(t)$. Let $\mathbf{X}(t)$ be a fundamental matrix for the homogeneous system. A particular solution to the nonhomogeneous system is given by the **variation of parameters** formula

$$\mathbf{x}_p(t) = \mathbf{X}(t)\boldsymbol{v}(t) = \mathbf{X}(t)\int \mathbf{X}^{-1}(t)\mathbf{f}(t)\,dt\,.$$

Matrix Exponential Function

If $\mathbf{A}$ is a constant $n \times n$ matrix, then the matrix exponential function

$$e^{\mathbf{A}t} := \mathbf{I} + \mathbf{A}t + \mathbf{A}^2\frac{t^2}{2!} + \cdots + \mathbf{A}^n\frac{t^n}{n!} + \cdots$$

is a fundamental matrix for the homogeneous system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$. The matrix exponential has some of the same properties as the scalar exponential e^{at}. In particular,

$$e^{\mathbf{0}} = \mathbf{I}\,, \qquad e^{\mathbf{A}(t+s)} = e^{\mathbf{A}t}e^{\mathbf{A}s}\,, \qquad (e^{\mathbf{A}t})^{-1} = e^{-\mathbf{A}t}\,.$$

If $(\mathbf{A} - r\mathbf{I})^k = \mathbf{0}$ for some r and k, then the series for $e^{\mathbf{A}t}$ has only a finite number of terms:

$$e^{\mathbf{A}t} = e^{rt}\left\{\mathbf{I} + (\mathbf{A} - r\mathbf{I})t + \cdots + (\mathbf{A} - r\mathbf{I})^{k-1}\frac{t^{k-1}}{(k-1)!}\right\}.$$

The matrix exponential function $e^{\mathbf{A}t}$ can also be computed from any fundamental matrix $\mathbf{X}(t)$ via the formula

$$e^{\mathbf{A}t} = \mathbf{X}(t)\mathbf{X}^{-1}(0)\,.$$

Generalized Eigenvectors

If an eigenvalue r_i of a constant $n \times n$ matrix $\mathbf{A}$ has multiplicity m_i there exist m_i linearly independent generalized eigenvectors $\mathbf{u}$ satisfying $(\mathbf{A} - r_i\mathbf{I})^{m_i}\mathbf{u} = \mathbf{0}$. Each such $\mathbf{u}$ determines a solution to the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ of the form

$$\mathbf{x}(t) = e^{r_i t}\left[\mathbf{I} + (\mathbf{A} - r_i\mathbf{I})t + \cdots + (\mathbf{A} - r_i\mathbf{I})^{m_i-1}\frac{t^{m_i-1}}{(m_i-1)!}\right]\mathbf{u}$$

and the totality of all such solutions is linearly independent on $(-\infty, \infty)$ and forms a fundamental solution set for the system.

REVIEW PROBLEMS FOR CHAPTER 9

In Problems 1–4, find a general solution for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, where $\mathbf{A}$ is given.

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

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

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

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

In Problems 5 and 6, find a fundamental matrix for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, where $\mathbf{A}$ is given.

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

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

In Problems 7–10, find a general solution for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$, where $\mathbf{A}$ and $\mathbf{f}(t)$ are given.

7. $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$

8. $\mathbf{A} = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} e^{4t} \\ 3e^{4t} \end{bmatrix}$

9. $\mathbf{A} = \begin{bmatrix} 2 & 1 & -1 \\ -3 & -1 & 1 \\ 9 & 3 & -4 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} t \\ 0 \\ 1 \end{bmatrix}$

10. $\mathbf{A} = \begin{bmatrix} 2 & -2 & 3 \\ 0 & 3 & 2 \\ 0 & -1 & 2 \end{bmatrix}, \quad \mathbf{f}(t) = \begin{bmatrix} e^{-t} \\ 2 \\ 1 \end{bmatrix}$

In Problems 11 and 12, solve the given initial value problem.

11. $\mathbf{x}'(t) = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}\mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

12. $\mathbf{x}'(t) = \begin{bmatrix} 2 & 1 \\ -4 & 2 \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} te^{2t} \\ e^{2t} \end{bmatrix}, \quad \mathbf{x}(0) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

In Problems 13 and 14, find a general solution for the Cauchy–Euler system $t\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, where $\mathbf{A}$ is given.

13. $\mathbf{A} = \begin{bmatrix} 0 & 3 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 0 \end{bmatrix}$ **14.** $\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}$

In Problems 15 and 16, find the fundamental matrix $e^{\mathbf{A}t}$ for the system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, where $\mathbf{A}$ is given.

15. $\mathbf{A} = \begin{bmatrix} 4 & 2 & 3 \\ 2 & 1 & 2 \\ -1 & 2 & 0 \end{bmatrix}$ **16.** $\mathbf{A} = \begin{bmatrix} 0 & 1 & 4 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$

17. For each of the following, determine if the statement made is *True* or *False*.

(a) Every $n \times n$ matrix of real numbers has n linearly independent eigenvectors.

(b) If $\mathbf{X}(t)$ is a fundamental matrix for $\mathbf{x}' = \mathbf{A}\mathbf{x}$, where $\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 8 & 4 \end{bmatrix}$, then the product $\mathbf{X}(t)\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$ is also a fundamental matrix for this system.

(c) The vector functions $\begin{bmatrix} 0 \\ 1 \\ e^t \end{bmatrix}, \begin{bmatrix} 0 \\ t \\ te^t \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2e^t \end{bmatrix}$ are linearly dependent on $(-\infty, \infty)$.

(d) If $\mathbf{A}$ is an $n \times n$ matrix of real numbers and $\mathbf{x}^*(t)$ satisfies the initial value problem $\mathbf{x}' = \mathbf{A}\mathbf{x}$, $\mathbf{x}(5) = \mathbf{0}$, then $\mathbf{x}^*(t) = \mathbf{0}$ for all t.

(e) If $\mathbf{A}$ is a 3×3 matrix of real numbers whose eigenvalues are 2, 2, and 1, then $\mathbf{A}$ has two linearly independent generalized eigenvectors that correspond to the eigenvalue 2.

(f) The Wronskian $W[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n](t)$ of n linearly independent $n \times 1$ vector functions $\mathbf{x}_i$ on an open interval I is never zero on I.

(g) If a 3×3 matrix of real numbers has eigenvalues $2 + 3i$, $2 - 3i$, and 5, then the matrix has three linearly independent eigenvectors.

TECHNICAL WRITING EXERCISES FOR CHAPTER 9

1. Explain how the theory of homogeneous linear differential equations (as described in Section 6.1) follows from the theory of linear systems in normal form (as described in Section 9.4).

2. Discuss the similarities and differences between the method for finding solutions to a linear constant-coefficient differential equation (see Chapters 4 and 6) and the method for finding solutions to a linear system in normal form that has constant coefficients (see Sections 9.5 and 9.6).

3. Explain how the variation of parameters formulas for linear second-order equations derived in Section 4.6 follow from the formulas derived in Section 9.7 for linear systems in normal form.

4. Explain how you would define the matrix functions $\sin \mathbf{A}t$ and $\cos \mathbf{A}t$, where $\mathbf{A}$ is a constant $n \times n$ matrix. How are these functions related to the matrix exponential and how are they connected to the solutions of the system $\mathbf{x}'' + \mathbf{A}^2\mathbf{x} = \mathbf{0}$? Your discussion should include the cases when $\mathbf{A}$ is

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

Projects for Chapter 9

A Uncoupling Normal Systems

The easiest normal systems to solve are systems of the form

$$\mathbf{x}'(t) = \mathbf{D}\mathbf{x}(t)\,, \tag{1}$$

where $\mathbf{D}$ is an $n \times n$ diagonal matrix. Such a system actually consists of n uncoupled equations

$$x_i'(t) = d_{ii}x_i(t)\,, \qquad i = 1, \ldots, n\,, \tag{2}$$

whose solution is

$$x_i(t) = c_i e^{d_{ii}t}\,,$$

where the c_i's are arbitrary constants. This raises the following question: When can we *uncouple* a normal system?

To answer this question, we need the following result from linear algebra. An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ has n linearly independent eigenvectors $\mathbf{p}_1, \ldots, \mathbf{p}_n$. Moreover, if $\mathbf{P}$ is the matrix whose columns are $\mathbf{p}_1, \ldots, \mathbf{p}_n$, then

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}\,, \tag{3}$$

where $\mathbf{D}$ is the diagonal matrix whose entry d_{ii} is the eigenvalue associated with the vector $\mathbf{p}_i$.

(a) Use the above result to show that the system

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)\,, \tag{4}$$

where $\mathbf{A}$ is an $n \times n$ diagonalizable matrix, is equivalent to an uncoupled system

$$\mathbf{y}'(t) = \mathbf{D}\mathbf{y}(t)\,, \tag{5}$$

where $\mathbf{y} = \mathbf{P}^{-1}\mathbf{x}$ and $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

(b) Solve system (5).

(c) Use the results of parts (a) and (b) to show that a general solution to (4) is given by

$$\mathbf{x}(t) = c_1 e^{d_{11}t}\mathbf{p}_1 + c_2 e^{d_{22}t}\mathbf{p}_2 + \cdots + c_n e^{d_{nn}t}\mathbf{p}_n\,.$$

(d) Use the procedure discussed in parts (a)–(c) to obtain a general solution for the system

$$\mathbf{x}'(t) = \begin{bmatrix} 5 & -4 & 4 \\ 1 & 0 & 1 \\ -1 & 2 & 1 \end{bmatrix}\mathbf{x}(t)\,.$$

Specify $\mathbf{P}$, $\mathbf{D}$, $\mathbf{P}^{-1}$, and $\mathbf{y}$.

B Matrix Laplace Transform Method

The Laplace transform method for solving systems of linear differential equations with constant coefficients was discussed in Chapter 7. To apply the procedure for equations given in matrix form, we first extend the definition of the Laplace operator $\mathscr{L}$ to a column vector of functions $\mathbf{x} = \operatorname{col}(x_1(t), \ldots, x_n(t))$ by taking the transform of each component:

$$\mathscr{L}\{\mathbf{x}\}(s) := \operatorname{col}\big(\mathscr{L}\{x_1\}(s), \ldots, \mathscr{L}\{x_n\}(s)\big) .$$

With this notation, the vector analogue of the important property relating the Laplace transform of the derivative of a function (see Theorem 4, Chapter 7, page 362) becomes

$$\mathscr{L}\{\mathbf{x}'\}(s) = s\mathscr{L}\{\mathbf{x}\}(s) - \mathbf{x}(0) . \tag{6}$$

Now suppose we are given the initial value problem

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{f}(t) \ , \qquad \mathbf{x}(0) = \mathbf{x}_0 , \tag{7}$$

where $\mathbf{A}$ is a constant $n \times n$ matrix. Let $\hat{\mathbf{x}}(s)$ denote the Laplace transform of $\mathbf{x}(t)$ and $\hat{\mathbf{f}}(s)$ denote the transform of $\mathbf{f}(t)$. Then, taking the transform of the system and using the relation (6), we get

$$\mathscr{L}\{\mathbf{x}'\} = \mathscr{L}\{\mathbf{A}\mathbf{x} + \mathbf{f}\} ,$$

$$s\hat{\mathbf{x}} - \mathbf{x}_0 = \mathbf{A}\hat{\mathbf{x}} + \hat{\mathbf{f}} .$$

Next we collect the $\hat{\mathbf{x}}$ terms and solve for $\hat{\mathbf{x}}$ by premultiplying by $(s\mathbf{I} - \mathbf{A})^{-1}$:

$$(s\mathbf{I} - \mathbf{A})\hat{\mathbf{x}} = \hat{\mathbf{f}} + \mathbf{x}_0 \ ,$$

$$\hat{\mathbf{x}} = (s\mathbf{I} - \mathbf{A})^{-1}(\hat{\mathbf{f}} + x_0) .$$

Finally, we obtain the solution $\mathbf{x}(t)$ by taking the inverse Laplace transform:

$$\mathbf{x} = \mathscr{L}^{-1}\{\hat{\mathbf{x}}\} = \mathscr{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}(\hat{\mathbf{f}} + x_0)\} . \tag{8}$$

In applying the matrix Laplace transform method, it is straightforward (but possibly tedious) to compute $(s\mathbf{I} - \mathbf{A})^{-1}$, but the computation of the inverse transform may require some of the special techniques (such as partial fractions) discussed in Chapter 7.

(a) In the above procedure, we used the property that $\mathscr{L}\{\mathbf{A}\mathbf{x}\} = \mathbf{A}\mathscr{L}\{\mathbf{x}\}$ for any constant $n \times n$ matrix $\mathbf{A}$. Show that this property follows from the linearity of the transform in the scalar case.

(b) Use the matrix Laplace transform method to solve the following initial value problems:

(i) $\mathbf{x}'(t) = \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix} \mathbf{x}(t) , \qquad \mathbf{x}(0) = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$

(ii) $\mathbf{x}'(t) = \begin{bmatrix} 3 & -2 \\ 4 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \sin t \\ -\cos t \end{bmatrix}, \qquad \mathbf{x}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$

(c) By comparing the Laplace transform solution formula (8) with the matrix exponential solution formula given in Section 9.8 [relation (16), page 551] for the homogeneous case $\mathbf{f}(t) \equiv \mathbf{0}$ and $t_0 = 0$, derive the Laplace transform formula for the matrix exponential

$$e^{\mathbf{A}t} = \mathscr{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\}(t) . \tag{9}$$

(d) What is $e^{\mathbf{A}t}$ for the coefficient matrices in part (b) above?

(e) Use (9) to rework Problems 1, 2, 7, and 8 in Exercises 9.8, page 551.

C Undamped Second-Order Systems

We have seen that the coupled mass–spring system depicted in Figure 9.5, page 536, is governed by equations (10) of Section 9.6, which we reproduce here:

$$m_1 x_1'' = -k_1 x_1 + k_2(x_2 - x_1) ,$$

$$m_2 x_2'' = -k_2(x_2 - x_1) - k_3 x_2 .$$

This system was rewritten in normal form as equation (11) in Section 9.6; however, there is some advantage in expressing it as a second-order system in the form

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}'' = -\begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 + k_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} .$$

This structure,

$$\textbf{(10)} \qquad \mathbf{M}\mathbf{x}'' = -\mathbf{K}\mathbf{x} ,$$

with a diagonal *mass matrix* **M** and a symmetric *stiffness matrix* **K**, is typical for most undamped vibrating systems. Our experience with other mass–spring systems (Section 5.6) suggests that we seek solutions to (10) of the form

$$\textbf{(11)} \qquad \mathbf{x} = (\cos \omega t)\,\mathbf{v} \quad \text{or} \quad \mathbf{x} = (\sin \omega t)\,\mathbf{v} ,$$

where $\mathbf{v}$ is a constant vector and ω is a positive constant.

(a) Show that the system (10) has a nontrivial solution of the form (11) if and only if ω and $\mathbf{v}$ satisfy the "generalized eigenvalue problem" $\mathbf{K}\mathbf{v} = \omega^2 \mathbf{M}\mathbf{v}$.

(b) By employing the inverse of the mass matrix, one can rewrite (10) as

$$\mathbf{x}'' = -\mathbf{M}^{-1}\mathbf{K}\mathbf{x} =: \mathbf{B}\mathbf{x} .$$

Show that $-\omega^2$ must be an eigenvalue of **B** if (11) is a nontrivial solution.

(c) If **B** is an $n \times n$ constant matrix, then $\mathbf{x}'' = \mathbf{B}\mathbf{x}$ can be written as a system of $2n$ first-order equations in normal form. Thus, a general solution can be formed from $2n$ linearly independent solutions. Use the observation in part (b) to find a general solution to the following second-order systems:

(i) $\mathbf{x}'' = \begin{bmatrix} 0 & 1 \\ -4 & -5 \end{bmatrix} \mathbf{x}$.

(ii) $\mathbf{x}'' = \begin{bmatrix} -2 & 2 \\ 2 & -5 \end{bmatrix} \mathbf{x}$.

(iii) $\mathbf{x}'' = \begin{bmatrix} -5 & 4 & -4 \\ -1 & 0 & -1 \\ 1 & -2 & -1 \end{bmatrix} \mathbf{x}$.

(iv) $\mathbf{x}'' = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} \mathbf{x}$.

CHAPTER 10

Partial Differential Equations

10.1 Introduction: A Model for Heat Flow

Develop a model for the flow of heat through a thin, insulated wire whose ends are kept at a constant temperature of 0°C and whose initial temperature distribution is to be specified.

Suppose the wire is placed along the x-axis with $x = 0$ at the left end of the wire and $x = L$ at the right end (see Figure 10.1). If we let u denote the temperature of the wire, then u depends on the time t and on the position x within the wire. (We will assume the wire is thin and hence u is constant throughout a cross section of the wire corresponding to a fixed value of x.) Because the wire is insulated, we assume no heat enters or leaves through the sides of the wire.

To develop a model for heat flow through the thin wire, let's consider the small volume element V of wire between the two cross-sectional planes A and B that are perpendicular to the x-axis, with plane A located at x and plane B located at $x + \Delta x$ (see Figure 10.1).

The temperature on plane A at time t is $u(x, t)$ and on plane B is $u(x + \Delta x, t)$. We will need the following principles of physics that describe heat flow.†

1. **Heat Conduction:** The rate of heat flow (the amount of heat per unit time flowing through a unit of cross-sectional area at A) is proportional to $\partial u/\partial x$, the temperature gradient at A (Fick's law). The proportionality constant k is called the **thermal conductivity** of the material. In general, the thermal conductivity can vary from point to point: $k = k(x)$.
2. **Direction of Heat Flow:** The direction of heat flow is always from points of higher temperature to points of lower temperature.

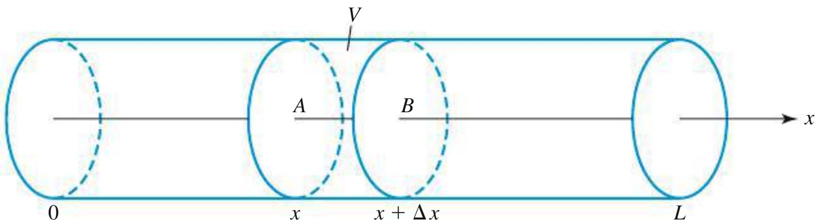


Figure 10.1 Heat flow through a thin piece of wire

†For a discussion of heat transfer, see *University Physics with Modern Physics*™, 13th ed., by H. D. Young and R. A. Freedman (Addison-Wesley, Reading, Mass., 2011).

3. **Specific Heat Capacity:** The amount of heat necessary to raise the temperature of an object of mass m by an amount Δu is $cm\,\Delta u$, where the constant c is the **specific heat capacity** of the material. The specific heat capacity, like the thermal conductivity, can vary with position: $c = c(x)$.

If we let H represent the amount of heat flowing from left to right through the surface A during an interval of time Δt, then the formula for heat conduction becomes

$$H(x) = -k(x)\,a\,\Delta t\frac{\partial u}{\partial x}(x, t)\ ,$$

where a is the cross-sectional area of the wire. The negative sign follows from the second principle—if $\partial u/\partial x$ is positive, then heat flows from right to left (hotter to colder).

Similarly, the amount of heat flowing from left to right across plane B during an interval of time Δt is

$$H(x + \Delta x) = -k(x + \Delta x)\,a\,\Delta t\frac{\partial u}{\partial x}(x + \Delta x, t)\ .$$

The net change in the heat ΔE in volume V is the amount entering at end A minus the amount leaving at end B, plus any heat generated by sources (such as electric currents, chemical reactions, heaters, etc.). The latter is modeled by a term $Q(x, t)\Delta x a\,\Delta t$, where Q is the energy rate (power) density. Therefore,

$$\begin{aligned} \text{(1)} \qquad \Delta E &= H(x) - H(x + \Delta x) + Q(x, t)\Delta x a\,\Delta t \\ &= a\,\Delta t\left[k(x + \Delta x)\frac{\partial u}{\partial x}(x + \Delta x, t) - k(x)\frac{\partial u}{\partial x}(x, t)\right] + Q(x, t)\Delta x\,a\,\Delta t\,. \end{aligned}$$

Now by the third principle, the net change is given by $\Delta E = cm\,\Delta u$, where Δu is the change in temperature and c is the specific heat capacity. If we assume that the change in temperature in the volume V is essentially equal to the change in temperature at x—that is, $\Delta u = u(x, t + \Delta t) - u(x, t)$, and that the mass of the volume V of wire is $a\rho\,\Delta x$, where $\rho = \rho(x)$ is the density of the wire, then

$$\text{(2)} \qquad \Delta E = c(x)a\rho(x)\,\Delta x[u(x, t + \Delta t) - u(x, t)]\,.$$

Equating the two expressions for ΔE given in equations (1) and (2) yields

$$\begin{aligned} a\,\Delta t\left[k(x + \Delta x)\frac{\partial u}{\partial x}(x + \Delta x, t) - k(x)\frac{\partial u}{\partial x}(x, t)\right] + Q(x, t)\,\Delta x\,a\,\Delta t \\ = c(x)a\rho(x)\,\Delta x[u(x, t + \Delta t) - u(x, t)]\,. \end{aligned}$$

Now dividing both sides by $a\ \Delta x\ \Delta t$ and then taking the limits as Δx and Δt approach zero, we obtain

$$\text{(3)} \qquad \frac{\partial}{\partial x}\left[k(x)\frac{\partial u}{\partial x}(x, t)\right] + Q(x, t) = c(x)\rho(x)\frac{\partial u}{\partial t}(x, t)\ .$$

If the physical parameters k, c, and ρ are uniform along the length of the wire, then (3) reduces to the **one-dimensional heat flow equation**

$$\text{(4)} \qquad \frac{\partial u}{\partial t}(x, t) = \beta\frac{\partial^2 u}{\partial x^2}(x, t) + P(x, t)\ ,$$

where the positive constant $\beta := k/(c\rho)$ is the **diffusivity** of the material and $P(x, t) := Q(x, t)/(c\rho)$.

Equation (4) governs the flow of heat in the wire. We have two other constraints in our original problem. First, we are keeping the ends of the wire at 0°C. Thus, we require that

$$u(0, t) = u(L, t) = 0 \tag{5}$$

for all t. These are called **boundary conditions.** Second, we must be given the initial temperature distribution $f(x)$. That is, we require

$$u(x, 0) = f(x), \qquad 0 < x < L. \tag{6}$$

Equation (6) is referred to as the **initial condition** on u.

Combining equations (4), (5), and (6), we have the following mathematical model for the heat flow in a uniform wire without internal sources $(P = 0)$ whose ends are kept at the constant temperature 0°C:

$$\frac{\partial u}{\partial t}(x, t) = \beta \frac{\partial^2 u}{\partial x^2}(x, t), \qquad 0 < x < L, \qquad t > 0, \tag{7}$$

$$u(0, t) = u(L, t) = 0, \qquad t > 0, \tag{8}$$

$$u(x, 0) = f(x), \qquad 0 < x < L. \tag{9}$$

This model is an example of an **initial-boundary value problem.** Intuitively we expect that equations (7)–(9) completely and unambiguously specify the temperature in the wire. Once we have found a function $u(x, t)$ that meets all three of these conditions, we can be assured that u is the temperature. (Theorem 7 in Section 10.5, page 602, will bear this out.)

In higher dimensions, the heat flow equation (or just **heat equation**) is adjusted to account for the additional heat flow contribution along the other axes by the simple modification

$$\frac{\partial u}{\partial t} = \beta \Delta u + P(x, y, z, t),$$

where Δu, known as the **Laplacian** in this context,† is defined in two and three dimensions, respectively, as

$$\Delta u := \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad \text{and} \quad \Delta u := \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

When the temperature reaches a steady state—that is, when u does not depend on time, and there are no sources—then $\partial u / \partial t = 0$ and the temperature satisfies **Laplace's equation**

$$\Delta u = 0.$$

One classical technique for solving the initial-boundary value problem for the heat equation (7)–(9) is the method of *separation of variables,* which effectively allows us to replace the partial differential equation by ordinary differential equations. This technique is discussed in the next section. In using separation of variables, one is often required to express a given function as a trigonometric series. Such series are called *Fourier series;* their properties are discussed in Sections 10.3 and 10.4. We devote the remaining three sections to the three basic partial differential equations that arise most commonly in applications: the heat equation, the wave equation, and Laplace's equation.

Many computer-based algorithms for solving partial differential equations have been developed. These are based on finite differences, finite elements, variational principles, and projection methods that include the method of moments and its wavelet-based implementations. Like the Runge–Kutta or Euler methods for ordinary differential equations, such techniques are more universally applicable than the analytic procedures of separation of variables, but their accuracy is often difficult to assess. Indeed, it is customary practice to gauge any newly proposed numerical procedure by comparing its predictions with those of separation of variables.

†Regrettably, it is the same notation as we just used for $u(x, t + \Delta t) - u(x, t)$. Some authors prefer the symbol ∇^2 for the Laplacian.

10.2 Method of Separation of Variables

The method of separation of variables is a classical technique that is effective in solving several types of partial differential equations. The idea is roughly the following. We think of a solution, say $u(x, t)$, to a partial differential equation as being a linear combination of simple component functions $u_n(x, t)$, $n = 0, 1, 2, \ldots$, which also satisfy the equation and certain boundary conditions. (This is a reasonable assumption provided the partial differential equation and the boundary conditions are linear.) To determine a component solution, $u_n(x, t)$, we assume it can be written with its variables separated; that is, as

$$u_n(x, t) = X_n(x)T_n(t) .$$

Substituting this form for a solution into the partial differential equation and using the boundary conditions leads, in many circumstances, to two *ordinary* differential equations for the unknown functions $X_n(x)$ and $T_n(t)$. In this way we have reduced the problem of solving a partial differential equation to the more familiar problem of solving a differential equation that involves only one variable. In this section we will illustrate this technique for the heat equation and the wave equation.

In the previous section we derived the following initial-boundary value problem as a mathematical model for the sourceless heat flow in a uniform wire whose ends are kept at the constant temperature zero:

$$\textbf{(1)} \qquad \frac{\partial u}{\partial t}(x, t) = \beta \frac{\partial^2 u}{\partial x^2}(x, t) , \qquad 0 < x < L , \qquad t > 0 ,$$

$$\textbf{(2)} \qquad u(0, t) = u(L, t) = 0 , \qquad t > 0 ,$$

$$\textbf{(3)} \qquad u(x, 0) = f(x) , \qquad 0 < x < L .$$

To solve this problem by the method of separation of variables, we begin by addressing equation (1). We propose that it has solutions of the form

$$u(x, t) = X(x)T(t) ,$$

where X is a function of x alone and T is a function of t alone. To determine X and T, we first compute the partial derivatives of u to obtain

$$\frac{\partial u}{\partial t} = X(x)T'(t) \quad \text{and} \quad \frac{\partial^2 u}{\partial x^2} = X''(x)T(t) .$$

Substituting these expressions into (1) gives

$$X(x)T'(t) = \beta X''(x)T(t) ,$$

and separating variables yields

$$\textbf{(4)} \qquad \frac{T'(t)}{\beta T(t)} = \frac{X''(x)}{X(x)} .$$

We now observe that the functions on the left-hand side of (4) depend only on t, while those on the right-hand side depend only on x. If we fix t and vary x, the ratio on the right cannot change; it must be *constant*. Most authors define this constant with a minus sign, so we adhere to the convention:

$$\frac{X''(x)}{X(x)} = -\lambda \quad \text{and} \quad \frac{T'(t)}{\beta T(t)} = -\lambda ,$$

or

(5) $\quad X''(x) = -\lambda X(x) \quad$ and $\quad T'(t) = -\lambda\beta T(t)$.

Consequently, for separable solutions, we have reduced the problem of solving the partial differential equation (1) to solving the two *ordinary* differential equations in (5).

Next we consider the boundary conditions in (2). Since $u(x, t) = X(x)T(t)$, these conditions are

$$X(0)T(t) = 0 \quad \text{and} \quad X(L)T(t) = 0\,, \qquad t > 0\,.$$

Hence, either $T(t) = 0$ for all $t > 0$, which implies that $u(x, t) \equiv 0$, or

(6) $\quad X(0) = X(L) = 0$.

Ignoring the trivial solution $u(x, t) \equiv 0$, we combine the boundary conditions in (6) with the differential equation for X in (5) and obtain the *boundary value problem*

(7) $\quad X''(x) + \lambda X(x) = 0\,, \qquad X(0) = X(L) = 0\,,$

where λ can be any constant.

Notice that the function $X(x) \equiv 0$ is a solution of (7) for every λ. Depending on the choice of λ, this may be the *only* solution to the boundary value problem (7). Thus, if we seek a nontrivial solution $u(x, t) = X(x)T(t)$ to (1)–(2), we must first determine those values of λ for which the boundary value problem (7) has nontrivial solutions. These solutions are called the **eigenfunctions** of the problem; the **eigenvalues** are the special values of λ. If we write the first equation in (7) as $-D^2[X] = \lambda X$ and compare with the matrix eigenvector equation $\mathbf{Au} = r\mathbf{u}$ (Section 9.5, page 523), the designation of λ as an eigenvalue (of $-D^2$) becomes clear.

To solve the (constant-coefficient) equation in (7), we try $X(x) = e^{rx}$, derive the auxiliary equation $r^2 + \lambda = 0$, and consider three cases.

Case 1. $\lambda < 0$. In this case, the roots of the auxiliary equation are $\pm\sqrt{-\lambda}$, so a general solution to the differential equation in (7) is

$$X(x) = C_1 e^{\sqrt{-\lambda}x} + C_2 e^{-\sqrt{-\lambda}x}\,.$$

To determine C_1 and C_2, we appeal to the boundary conditions:

$$X(0) = C_1 + C_2 = 0\,,$$

$$X(L) = C_1 e^{\sqrt{-\lambda}L} + C_2 e^{-\sqrt{-\lambda}L} = 0\,.$$

From the first equation, we see that $C_2 = -C_1$. The second equation can then be written as $C_1(e^{\sqrt{-\lambda}L} - e^{-\sqrt{-\lambda}L}) = 0$ or $C_1(e^{2\sqrt{-\lambda}L} - 1) = 0$. Since $-\lambda > 0$, it follows that $(e^{2\sqrt{-\lambda}L} - 1) > 0$. Therefore C_1, and hence C_2, is zero. Consequently, there is *no* nontrivial solution to (7) for $\lambda < 0$.

Case 2. $\lambda = 0$. Here $r = 0$ is a repeated root to the auxiliary equation, and a general solution to the differential equation is

$$X(x) = C_1 + C_2 x\,.$$

The boundary conditions in (7) yield $C_1 = 0$ and $C_1 + C_2 L = 0$, which imply that $C_1 = C_2 = 0$. Thus, for $\lambda = 0$, there is *no* nontrivial solution to (7).

Case 3. $\lambda > 0$. In this case the roots of the auxiliary equation are $\pm i\sqrt{\lambda}$. Thus a general solution to $X'' + \lambda X = 0$ is

(8) $\quad X(x) = C_1 \cos \sqrt{\lambda}x + C_2 \sin \sqrt{\lambda}x$.

This time the boundary conditions $X(0) = X(L) = 0$ give the system

$$C_1 = 0\,,$$
$$C_1 \cos\sqrt{\lambda}L + C_2 \sin\sqrt{\lambda}L = 0\,.$$

Because $C_1 = 0$, the system reduces to solving $C_2 \sin\sqrt{\lambda}L = 0$. Hence, either $\sin\sqrt{\lambda}L = 0$ or $C_2 = 0$. Now $\sin\sqrt{\lambda}\mathrm{L} = 0$ only when $\sqrt{\lambda}L = n\pi$, where n is an integer. Therefore, (7) has a nontrivial solution $(C_2 \neq 0)$ when $\sqrt{\lambda}L = n\pi$ or $\lambda = (n\pi/L)^2$, $n = 1, 2, 3, \ldots$ (we exclude $n = 0$, since it makes $\lambda = 0$). Furthermore, the nontrivial solutions (eigenfunctions) X_n corresponding to the eigenvalue $\lambda = (n\pi/L)^2$ are given by [cf. (8)]

$$X_n(x) = a_n \sin\left(\frac{n\pi x}{L}\right), \tag{9}$$

where the a_n's are arbitrary nonzero constants.

The eigenvalues and eigenfunctions for this example have many features that are common to all of the separation of variables solutions that we will study. Take note of these features of Figure 10.2.

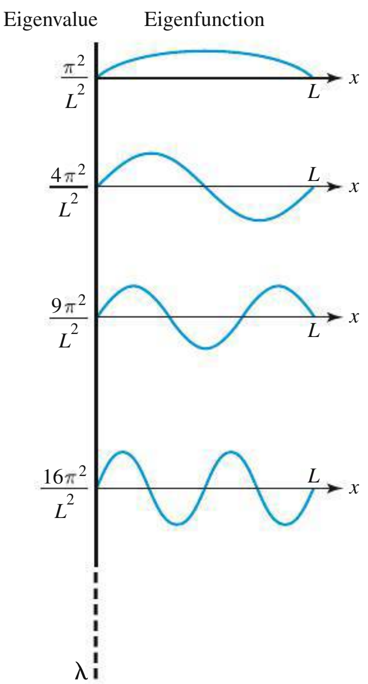


Figure 10.2 Eigenfunctions

SEPARATION OF VARIABLES:
EIGENFUNCTION PROPERTIES

(*i*) The eigenfunctions are solutions to a second-order ordinary differential equation (7) containing a parameter λ called the eigenvalue.

(*ii*) Each eigenfunction satisfies a homogeneous boundary condition at each end (7).

(*iii*) The only values of λ that admit nontrivial solutions, i.e., the eigenvalues, form an infinite set accumulating at ∞.

(*iv*) The eigenfunctions oscillate faster as λ increases; the first contains no interior zeros, and each subsequent eigenfunction contains one more zero than its predecessor.

(*v*) If any two distinct eigenfunctions are multiplied together, the resulting function is zero on the average; specifically,

$$\sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L} = \frac{1}{2}\left[\cos\frac{(m-n)\pi x}{L} - \cos\frac{(m+n)\pi x}{L}\right]$$

and both cosines integrate to 0 over $[0, L]$.

Having determined that $\lambda = (n\pi/L)^2$ for any positive integer n, let's consider the second equation in (5):

$$T'(t) + \beta\left(\frac{n\pi}{L}\right)^2 T(t) = 0\,.$$

For each $n = 1, 2, 3, \ldots$, a general solution to this linear first-order equation is

$$T_n(t) = b_n e^{-\beta(n\pi/L)^2 t}\,.$$

(Note that the time factor has none of the eigenfunction properties.) Combining this with equation (9) we obtain, for each $n = 1, 2, 3, \ldots$, the functions

(10) $$u_n(x, t) := X_n(x)T_n(t) = a_n \sin(n\pi x/L)\, b_n e^{-\beta(n\pi/L)^2 t}$$
$$= c_n e^{-\beta(n\pi/L)^2 t} \sin(n\pi x/L)\,,$$

where c_n is also an arbitrary constant.

It is easy to see that each $u_n(x, t)$ is a solution to (1)–(2). A simple computation also shows that if u_n and u_m are solutions to (1)–(2), then so is any linear combination $au_n + bu_m$. (This is a consequence of the facts that the operator $\mathcal{L} := \partial/\partial t - \beta\partial^2/\partial x^2$ is a *linear* operator and the boundary conditions in (2) are *linear homogeneous*.)

This enables us to solve the following example.

Example 1 Find the solution to the heat flow problem

(11) $$\frac{\partial u}{\partial t} = 7\frac{\partial^2 u}{\partial x^2}\,, \qquad 0 < x < \pi\,, \qquad t > 0\,,$$

(12) $$u(0, t) = u(\pi, t) = 0\,, \qquad t > 0\,,$$

(13) $$u(x, 0) = 3\sin 2x - 6\sin 5x\,, \qquad 0 < x < \pi\,.$$

Solution Comparing equation (11) with (1), we see that $\beta = 7$ and $L = \pi$. Hence, we need only find a combination of terms like (10) that satisfies the initial condition (13):

$$u(x, 0) = \sum c_n e^0 \sin nx = 3\sin 2x - 6\sin 5x.$$

For these data the task is simple; $c_2 = 3$ and $c_5 = -6$. The solution to the heat flow problem (11)–(13) is

$$u(x, t) = c_2 e^{-\beta(2\pi/L)^2 t} \sin(2\pi x/L) + c_5 e^{-\beta(5\pi/L)^2 t} \sin(5\pi x/L)$$
$$= 3e^{-28t} \sin 2x - 6e^{-175t} \sin 5x\,. \;\blacklozenge$$

What would we do if the initial condition (13) had been an *arbitrary* function, rather than a simple combination of a few of the eigenfunctions we found? Possibly the most beautiful property of the eigenfunctions—one that we did not list on page 565—is *completeness*; virtually any function f likely to arise in applications can be expressed as a convergent series of eigenfunctions! For the sines we have been working with, the **Fourier sine series** looks like

(14) $$f(x) = \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi x}{L}\right) \quad \text{for } 0 < x < L\,.$$

This enables the complete solution to the generic problem given by (1)–(3):

(15) $$\boldsymbol{u(x, t) = \sum_{n=1}^{\infty} c_n e^{-\beta(n\pi/L)^2 t} \sin\left(\frac{n\pi x}{L}\right)}\,,$$

provided this expansion and its first two derivatives converge.

The convergence of the sine series (among others), as well as the procedure for finding the coefficients c_n, will be discussed in the next two sections (and generalized in Chapter 11). For now, let us turn to the mathematical description of a vibrating string, another situation in which the separation of variables approach applies. This concerns the transverse vibrations of a string stretched between two points, such as a guitar string or piano wire. The goal is to find a function $u(x, t)$ that gives the displacement (deflection) of the string at any point

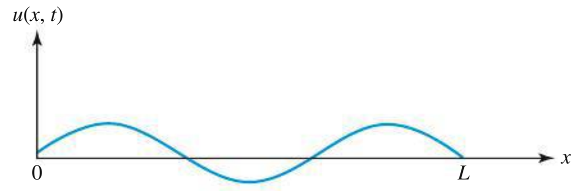


Figure 10.3 Displacement of string at time t

x $(0 \le x \le L)$ and any time $t \ge 0$ (see Figure 10.3). In developing the mathematical model, we assume that the string is perfectly flexible and has constant linear density, the tension on the string is constant, gravity is negligible, and no other forces are acting on the string. Under these conditions and the additional assumption that the displacements $u(x, t)$ are small in comparison to the length of the string, it turns out that the motion of the string is governed by the following initial-boundary value problem.[†]

$$\textbf{(16)} \qquad \frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \qquad t > 0,$$

$$\textbf{(17)} \qquad u(0, t) = u(L, t) = 0, \qquad t \ge 0,$$

$$\textbf{(18)} \qquad u(x, 0) = f(x), \qquad 0 \le x \le L,$$

$$\textbf{(19)} \qquad \frac{\partial u}{\partial t}(x, 0) = g(x), \qquad 0 \le x \le L.$$

Equation (16) essentially states Newton's third law, $F = ma$, for the string. The second (time) derivative on the left is the acceleration, and the second (spatial) derivative on the right arises because the restoring force is produced by the string's *curvature*. The constant α^2 is strictly positive and equals the ratio of the tension to the (linear) density of the string. The physical significance of α, which has units of velocity, will be revealed in Section 10.6. The boundary conditions in (17) reflect the fact that the string is held fixed at the two endpoints $x = 0$ and $x = L$. Equations (18) and (19), respectively, specify the initial displacement of the string and the initial velocity of each point on the string. Recall that these are the typical initial data for all mechanical systems. For the initial and boundary conditions to be consistent, we assume $f(0) = f(L) = 0$ and $g(0) = g(L) = 0$.

Let's apply the method of separation of variables to the initial-boundary value problem for the vibrating string (16)–(19). Thus, we begin by assuming equation (16) has a solution of the form

$$u(x, t) = X(x)T(t),$$

where X is a function of x alone and T is a function of t alone. Differentiating u, we obtain

$$\frac{\partial^2 u}{\partial t^2} = X(x)T''(t), \qquad \frac{\partial^2 u}{\partial x^2} = X''(x)T(t).$$

Substituting these expressions into (16), we have

$$X(x)T''(t) = \alpha^2 X''(x)T(t),$$

[†]For a derivation of this mathematical model, see *Partial Differential Equations: Sources and Solutions*, by Arthur D. Snider (Dover Publications, N.Y., 2006).

and separating variables gives

$$\frac{T''(t)}{\alpha^2 T(t)} = \frac{X''(x)}{X(x)} .$$

Just as before, these ratios must equal some constant $-\lambda$:

$$\textbf{(20)} \qquad \frac{X''(x)}{X(x)} = -\lambda \qquad \text{and} \qquad \frac{T''(t)}{\alpha^2 T(t)} = -\lambda .$$

Furthermore, with $u(x, t) = X(x)T(t)$, the boundary conditions in (17) give

$$X(0)T(t) = 0 , \qquad X(L)T(t) = 0 , \qquad t \geq 0 .$$

In order for these equations to hold for all $t \geq 0$, either $T(t) \equiv 0$, which implies that $u(x, t) \equiv 0$, or

$$X(0) = X(L) = 0 .$$

Ignoring the trivial solution, we combine these boundary conditions with the differential equation for X in (20) and obtain the boundary value problem

$$\textbf{(21)} \qquad X''(x) + \lambda X(x) = 0 , \qquad X(0) = X(L) = 0 ,$$

where λ can be any constant.

This is the same boundary value problem that we encountered earlier while solving the heat equation. There we found that the suitable values for λ are

$$\lambda = \left(\frac{n\pi}{L}\right)^2 , \qquad n = 1, 2, 3, \ldots ,$$

with corresponding eigenfunctions (nontrivial solutions)

$$\textbf{(22)} \qquad X_n(x) = c_n \sin\left(\frac{n\pi x}{L}\right) ,$$

where the c_n's are arbitrary nonzero constants. Recall Figure 10.2, page 565.

Having determined that $\lambda = (n\pi/L)^2$ for some positive integer n, let's consider the second equation in (20) for such λ:

$$T''(t) + \frac{\alpha^2 n^2 \pi^2}{L^2} T(t) = 0 .$$

For each $n = 1, 2, 3, \ldots$, a general solution is

$$T_n(t) = c_{n,1} \cos\frac{n\pi\alpha}{L} t + c_{n,2} \sin\frac{n\pi\alpha}{L} t .$$

Combining this with equation (22), we obtain, for each $n = 1, 2, 3, \ldots$, the function

$$u_n(x, t) = X_n(x)T_n(t) = \left(c_n \sin\frac{n\pi x}{L}\right)\left(c_{n,1} \cos\frac{n\pi\alpha}{L} t + c_{n,2} \sin\frac{n\pi\alpha}{L} t\right) ,$$

or, reassembling the constants,

$$\textbf{(23)} \qquad u_n(x, t) = \left(a_n \cos\frac{n\pi\alpha}{L} t + b_n \sin\frac{n\pi\alpha}{L} t\right) \sin\frac{n\pi x}{L} .$$

Using the fact that linear combinations of solutions to (16)–(17) are again solutions, we consider a superposition of the functions in (23):

$$\textbf{(24)}\qquad u(x,t) = \sum_{n=1}^{\infty}\left[a_n \cos\frac{n\pi\alpha}{L}t + b_n \sin\frac{n\pi\alpha}{L}t\right]\sin\frac{n\pi x}{L}.$$

For a solution of the form (24), substitution into the initial conditions (18)–(19) gives

$$\textbf{(25)}\qquad u(x,0) = \sum_{n=1}^{\infty} a_n \sin\frac{n\pi x}{L} = f(x)\,, \qquad 0 \le x \le L\,,$$

$$\textbf{(26)}\qquad \frac{\partial u}{\partial t}(x,0) = \sum_{n=1}^{\infty} \frac{n\pi\alpha}{L} b_n \sin\frac{n\pi x}{L} = g(x)\,, \qquad 0 \le x \le L\,.$$

We have now reduced the vibrating string problem (16)–(19) to the problem of determining the Fourier sine series expansions for $f(x)$ and $g(x)$:

$$\textbf{(27)}\qquad f(x) = \sum_{n=1}^{\infty} a_n \sin\frac{n\pi x}{L}\,, \qquad g(x) = \sum_{n=1}^{\infty} B_n \sin\frac{n\pi x}{L}\,,$$

where $B_n = (n\pi\alpha/L)b_n$. If we choose the a_n's and b_n's so that the equations in (25) and (26) hold, then the expansion for $u(x,t)$ in (24) is a **formal solution** to the vibrating string problem (16)–(19). If this expansion is finite, or converges to a function with continuous second partial derivatives, then the formal solution is an actual (genuine) solution.

Example 2 Find the solution to the vibrating string problem

$$\textbf{(28)}\qquad \frac{\partial^2 u}{\partial t^2} = 4\frac{\partial^2 u}{\partial x^2}\,, \qquad 0 < x < \pi\,, \qquad t > 0\,,$$

$$\textbf{(29)}\qquad u(0,t) = u(\pi,t) = 0\,, \qquad t \ge 0\,,$$

$$\textbf{(30)}\qquad u(x,0) = \sin 3x - 4\sin 10x\,, \qquad 0 \le x \le \pi\,,$$

$$\textbf{(31)}\qquad \frac{\partial u}{\partial t}(x,0) = 2\sin 4x + \sin 6x\,, \qquad 0 \le x \le \pi\,.$$

Solution Comparing equation (28) with equation (16), we see that $\alpha = 2$ and $L = \pi$. Hence, we need only determine the values of the coefficients a_n and b_n in formula (24). The a_n's are chosen so that equation (25) holds; that is,

$$u(x,0) = \sin 3x - 4\sin 10x = \sum_{n=1}^{\infty} a_n \sin nx\,.$$

Equating coefficients of like terms, we see that

$$a_3 = 1\,, \qquad a_{10} = -4\,,$$

and the remaining a_n's are zero. Similarly, referring to equation (26), we must choose the b_n's so that

$$\frac{\partial u}{\partial t}(x,0) = 2\sin 4x + \sin 6x = \sum_{n=1}^{\infty} n2b_n \sin nx\,.$$

Comparing coefficients, we find

$$2 = (4)(2)b_4 \quad \text{or} \quad b_4 = \frac{1}{4}\,,$$

$$1 = (6)(2)b_6 \quad \text{or} \quad b_6 = \frac{1}{12}\,,$$

and the remaining b_n's are zero. Hence, from formula (24), the solution to the vibrating string problem (28)–(31) is

$$u(x, t) = \cos 6t \sin 3x + \frac{1}{4}\sin 8t \sin 4x + \frac{1}{12}\sin 12t \sin 6x - 4\cos 20t \sin 10x\,. \tag{32}$$ ◆

In later sections the method of separation of variables is used to study a wide variety of problems for the heat, wave, and Laplace's equations. However, to use the method effectively, one must be able to compute trigonometric series (or, more generally, eigenfunction expansions) such as the Fourier sine series that we encountered here. These expansions are discussed in the next two sections.

10.2 EXERCISES

In Problems 1–8, determine all the solutions, if any, to the given boundary value problem by first finding a general solution to the differential equation.

1. $y'' - y = 0\,; \quad 0 < x < 1\,,$
$y(0) = 0\,, \quad y(1) = -4$

2. $y'' - 6y' + 5y = 0\,; \quad 0 < x < 2\,,$
$y(0) = 1\,, \quad y(2) = 1$

3. $y'' + 4y = 0\,; \quad 0 < x < \pi\,,$
$y(0) = 0\,, \quad y'(\pi) = 0$

4. $y'' + 9y = 0\,; \quad 0 < x < \pi\,,$
$y(0) = 0\,, \quad y'(\pi) = -6$

5. $y'' - y = 1 - 2x\,; \quad 0 < x < 1\,,$
$y(0) = 0\,, \quad y(1) = 1 + e$

6. $y'' + y = 0\,; \quad 0 < x < 2\pi\,,$
$y(0) = 0\,, \quad y(2\pi) = 1$

7. $y'' + y = 0\,; \quad 0 < x < 2\pi\,,$
$y(0) = 1\,, \quad y(2\pi) = 1$

8. $y'' - 2y' + y = 0\,; \quad -1 < x < 1\,,$
$y(-1) = 0\,, \quad y(1) = 2$

In Problems 9–14, find the values of λ (eigenvalues) for which the given problem has a nontrivial solution. Also determine the corresponding nontrivial solutions (eigenfunctions).

9. $y'' + \lambda y = 0\,; \quad 0 < x < \pi\,,$
$y(0) = 0\,, \quad y'(\pi) = 0$

10. $y'' + \lambda y = 0\,; \quad 0 < x < \pi\,,$
$y'(0) = 0\,, \quad y(\pi) = 0$

11. $y'' + \lambda y = 0\,; \quad 0 < x < 2\pi\,,$
$y(0) = y(2\pi)\,, \quad y'(0) = y'(2\pi)$

12. $y'' + \lambda y = 0\,; \quad 0 < x < \pi/2\,,$
$y'(0) = 0\,, \quad y'(\pi/2) = 0$

13. $y'' + \lambda y = 0\,; \quad 0 < x < \pi\,,$
$y(0) - y'(0) = 0\,, \quad y(\pi) = 0$

14. $y'' - 2y' + \lambda y = 0\,; \quad 0 < x < \pi\,,$
$y(0) = 0\,, \quad y(\pi) = 0$

In Problems 15–18, solve the heat flow problem (1)–(3) with $\beta = 3$, $L = \pi$, and the given function $f(x)$.

15. $f(x) = \sin x - 6\sin 4x$

16. $f(x) = \sin 3x + 5\sin 7x - 2\sin 13x$

17. $f(x) = \sin x - 7\sin 3x + \sin 5x$

18. $f(x) = \sin 4x + 3\sin 6x - \sin 10x$

In Problems 19–22, solve the vibrating string problem (16)–(19) with $\alpha = 3$, $L = \pi$, and the given initial functions $f(x)$ and $g(x)$.

19. $f(x) = 3\sin 2x + 12\sin 13x\,, \quad g(x) \equiv 0$

20. $f(x) \equiv 0\,,$
$g(x) = -2\sin 3x + 9\sin 7x - \sin 10x$

21. $f(x) = 6\sin 2x + 2\sin 6x\,,$
$g(x) = 11\sin 9x - 14\sin 15x$

22. $f(x) = \sin x - \sin 2x + \sin 3x\,,$
$g(x) = 6\sin 3x - 7\sin 5x$

23. Find the formal solution to the heat flow problem (1)–(3) with $\beta = 2$ and $L = 1$ if

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{n^2}\sin n\pi x\,.$$

24. Find the formal solution to the vibrating string problem (16)–(19) with $\alpha = 4$, $L = \pi$, and

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{n^2}\sin nx\,,$$

$$g(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}\sin nx\,.$$

25. By considering the behavior of the solutions of the equation

$$T'(t) = -\lambda\beta T(t)\,, \quad t > 0\,,$$

give an argument that is based on physical grounds to rule out the case where $\lambda < 0$ in equation (5).

26. Verify that $u_n(x, t)$ given in equation (10) satisfies equation (1) and the boundary conditions in (2) by substituting $u_n(x, t)$ directly into the equations involved.

In Problems 27–30, a partial differential equation (PDE) is given along with the form of a solution having separated variables. Show that such a solution must satisfy the indicated set of ordinary differential equations.

27. $\dfrac{\partial^2 u}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial u}{\partial r} + \dfrac{1}{r^2}\dfrac{\partial^2 u}{\partial \theta^2} = 0$

with $u(r, \theta) = R(r)\Theta(\theta)$ yields

$$r^2 R''(r) + rR'(r) - \lambda R(r) = 0 ,$$
$$\Theta''(\theta) + \lambda\Theta(\theta) = 0 ,$$

where λ is a constant.

28. $\dfrac{\partial^2 u}{\partial t^2} + \dfrac{\partial u}{\partial t} + u = \alpha^2 \dfrac{\partial^2 u}{\partial x^2}$

with $u(x, t) = X(x)T(t)$ yields

$$X''(x) + \lambda X(x) = 0 ,$$
$$T''(t) + T'(t) + (1 + \lambda\alpha^2)T(t) = 0 ,$$

where λ is a constant.

29. $\dfrac{\partial u}{\partial t} = \beta\left\{\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2}\right\}$

with $u(x, y, t) = X(x)Y(y)T(t)$ yields

$$T'(t) + \beta\lambda T(t) = 0 ,$$
$$X''(x) + \mu X(x) = 0 ,$$
$$Y''(y) + (\lambda - \mu)Y(y) = 0 ,$$

where λ, μ are constants.

30. $\dfrac{\partial^2 u}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial u}{\partial r} + \dfrac{1}{r^2}\dfrac{\partial^2 u}{\partial \theta^2} + \dfrac{\partial^2 u}{\partial z^2} = 0$

with $u(r, \theta, z) = R(r)\Theta(\theta)Z(z)$ yields

$$\Theta''(\theta) + \mu\Theta(\theta) = 0 ,$$
$$Z''(z) + \lambda Z(z) = 0 ,$$
$$r^2 R''(r) + rR'(r) - (r^2\lambda + \mu)R(r) = 0 ,$$

where μ, λ are constants.

31. For the PDE in Problem 27, assume that the following boundary conditions are imposed:

$$u(r, 0) = u(r, \pi) = 0 ,$$
$$u(r, \theta) \text{ remains bounded as } r \to 0^+ .$$

Show that a nontrivial solution of the form $u(r, \theta) = R(r)\Theta(\theta)$ must satisfy the boundary conditions

$$\Theta(0) = \Theta(\pi) = 0 ,$$
$$R(r) \text{ remains bounded as } r \to 0^+ .$$

32. For the PDE in Problem 29, assume that the following boundary conditions are imposed:

$$u(0, y, t) = u(a, y, t) = 0 ; \quad 0 \le y \le b , \quad t \ge 0 ,$$
$$\frac{\partial u}{\partial y}(x, 0, t) = \frac{\partial u}{\partial y}(x, b, t) = 0 ; \quad 0 \le x \le a , \quad t \ge 0 .$$

Show that a nontrivial solution of the form $u(x, y, t) = X(x)Y(y)T(t)$ must satisfy the boundary conditions

$$X(0) = X(a) = 0 ,$$
$$Y'(0) = Y'(b) = 0 .$$

33. When the temperature in a wire reaches a steady state, that is, when u depends only on x, then $u(x)$ satisfies Laplace's equation $\partial^2 u/\partial x^2 = 0$.

(a) Find the steady-state solution when the ends of the wire are kept at a constant temperature of 50°C, that is, when $u(0) = u(L) = 50$.

(b) Find the steady-state solution when one end of the wire is kept at 10°C, while the other is kept at 40°C, that is, when $u(0) = 10$ and $u(L) = 40$.

10.3 Fourier Series

While analyzing heat flow and vibrating strings in the previous section, we encountered the problem of expressing a function in a trigonometric series [compare equations (12) and (25) in Section 10.2]. In the next two sections, we discuss the theory of Fourier series, which deals with trigonometric series expansions. First, however, we review some function properties that are particularly relevant to this study: piecewise continuity, periodicity, and even and odd symmetry.

In Section 7.2 we defined a **piecewise continuous** function on $[a, b]$ as a function f that is continuous at every point in $[a, b]$, except possibly for a finite number of points at which f has a *jump discontinuity*. Such functions are necessarily integrable over any finite interval on which they are piecewise continuous.

Recall also that a function is **periodic of period** T if $f(x+T) = f(x)$ for all x in the domain of f. The smallest positive value of T is called the **fundamental period**. The trigonometric functions $\sin x$ and $\cos x$ are examples of periodic functions with fundamental period 2π and $\tan x$ is periodic with fundamental period π. A constant function is a periodic function with arbitrary period T.

Two symmetry properties of functions will be useful in the study of Fourier series. A function f that satisfies $f(-x) = f(x)$ for all x in the domain of f has a graph that is symmetric with respect to the y-axis [see Figure 10.4(a)]. We say that such a function is **even**. A function f that satisfies $f(-x) = -f(x)$ for all x in the domain of f has a graph that is symmetric with respect to the origin [see Figure 10.4(b)]. It is said to be an **odd** function. The functions $1, x^2, x^4, \ldots$ are examples of even functions, while the functions $x, x^3, x^5, \ldots$ are odd. The trigonometric functions $\sin x$ and $\tan x$ are odd functions and $\cos x$ is an even function.

Example 1 Determine whether the given function is even, odd, or neither.

(a) $f(x) = \sqrt{1+x^2}$ **(b)** $g(x) = x^{1/3} - \sin x$ **(c)** $h(x) = e^x$

Solution

(a) Since $f(-x) = \sqrt{1+(-x)^2} = \sqrt{1+x^2} = f(x)$, then $f(x)$ is an even function.

(b) Because $g(-x) = (-x)^{1/3} - \sin(-x) = -x^{1/3} + \sin x = -(x^{1/3} - \sin x) = -g(x)$, it follows that $g(x)$ is an odd function.

(c) Here $h(-x) = e^{-x}$. Since $e^{-x} = e^x$ only when $x = 0$ and e^{-x} is never equal to $-e^x$, then $h(x)$ is neither an even nor an odd function. ◆

Knowing that a function is even or odd can be useful in evaluating definite integrals. The following result, illustrated in Figure 10.4, is a straightforward consequence of the definition of the definite integral.

Properties of Symmetric Functions

Theorem 1. If f is an even piecewise continuous function on $[-a, a]$, then

(1) $$\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx\,.$$

If f is an odd piecewise continuous function on $[-a, a]$, then

(2) $$\int_{-a}^{a} f(x)\,dx = 0\,.$$

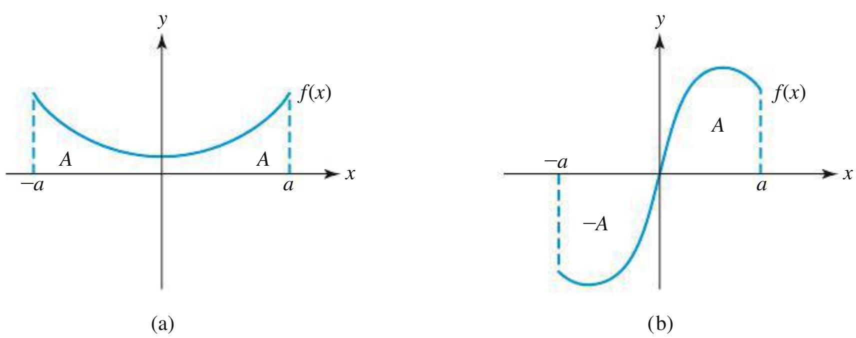


Figure 10.4 (a) Even function $\int_{-a}^{a} f = A + A = 2\int_0^a f$ (b) Odd function $\int_{-a}^{a} f = A - A = 0$

The next example deals with certain integrals that are crucial in Fourier series.

Example 2 Evaluate the following integrals when m and n are nonnegative integers:

(a) $\displaystyle\int_{-L}^{L} \sin\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\,dx$. **(b)** $\displaystyle\int_{-L}^{L} \sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L}\,dx$.

(c) $\displaystyle\int_{-L}^{L} \cos\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\,dx$.

Solution The even and odd functions occurring in the integrands are sketched in Figure 10.5.

The given integrals are easily evaluated by invoking the trigonometric formula for products of sines and cosines:

$$\sin\frac{m\pi x}{L}\cos\frac{n\pi x}{L} = \frac{1}{2}\sin\frac{(m-n)\pi x}{L} + \frac{1}{2}\sin\frac{(m+n)\pi x}{L},$$
$$\sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L} = \frac{1}{2}\cos\frac{(m-n)\pi x}{L} - \frac{1}{2}\cos\frac{(m+n)\pi x}{L},$$
$$\cos\frac{m\pi x}{L}\cos\frac{n\pi x}{L} = \frac{1}{2}\cos\frac{(m-n)\pi x}{L} + \frac{1}{2}\cos\frac{(m+n)\pi x}{L}.$$

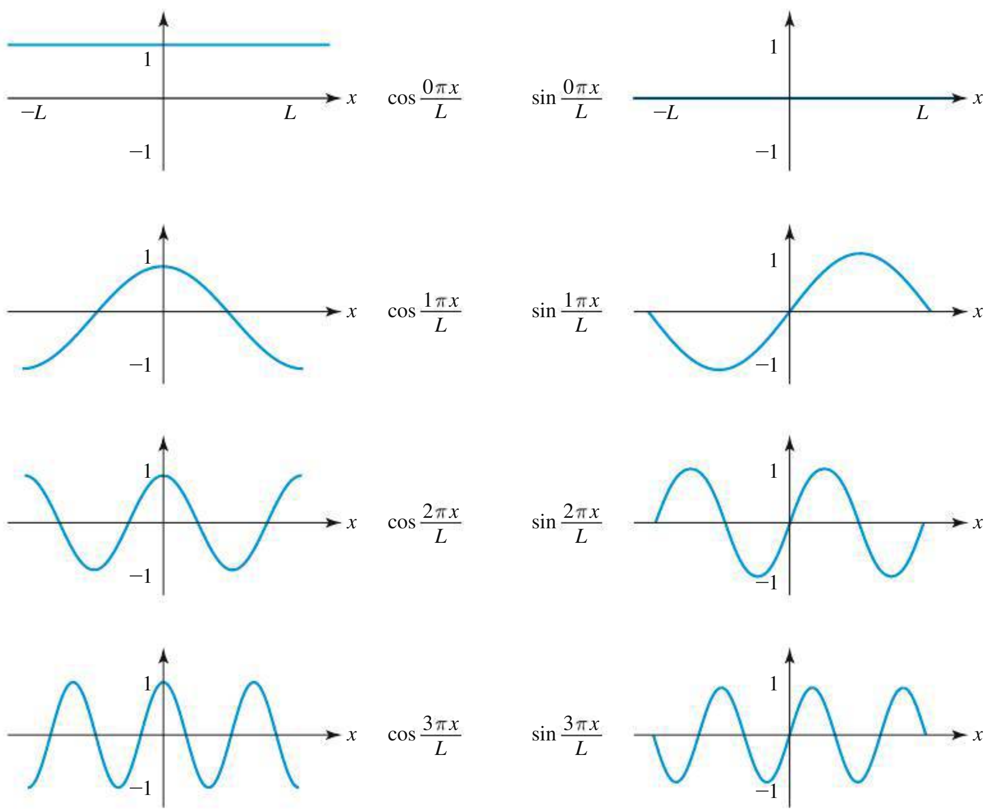


Figure 10.5 The sinusoids

Thus if $m \neq n$, each of the integrals calls for the (signed) area under an oscillating sinusoid over a whole number of periods, and is therefore zero. In fact, the only way to avoid these "oscillators" is to take $m = n$; whence $\cos[(m-n)\pi x/L] = \cos 0 = 1$ and subtends an area of $2L$, while $\sin[(m-n)\pi x/L] = \sin 0 = 0$ subtending zero area (again). But note that if m and n are *both* zero, then $\cos[(m+n)\pi x/L] = \cos 0$ also subtends area $2L$. Restoring the factors of $1/2$, we summarize with

$$\int_{-L}^{L} \sin\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\,dx = 0\,, \tag{3}$$

$$\int_{-L}^{L} \sin\frac{m\pi x}{L}\sin\frac{n\pi x}{L}\,dx = \begin{cases} 0\,, & m \neq n\,,\\ L\,, & m = n\,,\end{cases} \tag{4}$$

$$\int_{-L}^{L} \cos\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\,dx = \begin{cases} 0\,, & m \neq n\,,\\ L\,, & m = n \neq 0\,,\\ 2L\,, & m = n = 0\,.\end{cases} \quad \blacklozenge \tag{5}$$

Equations (3)–(5) express an **orthogonality condition**† satisfied by the set of trigonometric functions $\{1 = \cos 0x,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots\}$, where $L = \pi$. We will say more about this later in this section.

It is easy to verify that if each of the functions $f_1, \ldots, f_n$ is periodic of period T, then so is any linear combination

$$c_1 f_1(x) + \cdots + c_n f_n(x)\,.$$

For example, the sum $7 + 3\cos\pi x - 8\sin\pi x + 4\cos 2\pi x - 6\sin 2\pi x$ has period 2, since each term has period 2. Furthermore, if the infinite series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos\frac{n\pi x}{L} + b_n\sin\frac{n\pi x}{L}\right)$$

consisting of $2L$-periodic functions converges for all x, then the function to which it converges will be periodic of period $2L$.

Just as we can associate a Taylor series with a function that has derivatives of all orders at a fixed point, we can identify a particular trigonometric series with a piecewise continuous function. To illustrate this, let's assume that $f(x)$ has the series expansion‡

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left\{a_n\cos\frac{n\pi x}{L} + b_n\sin\frac{n\pi x}{L}\right\}, \tag{6}$$

where the a_n's and b_n's are constants. (Necessarily, f has period $2L$.)

To determine the coefficients $a_0, a_1, b_1, a_2, b_2, \ldots$, we proceed as follows. Let's integrate $f(x)$ from $-L$ to L, assuming that we can integrate term by term:

$$\int_{-L}^{L} f(x)\,dx = \int_{-L}^{L}\frac{a_0}{2}\,dx + \sum_{n=1}^{\infty} a_n\int_{-L}^{L}\cos\frac{n\pi x}{L}\,dx + \sum_{n=1}^{\infty} b_n\int_{-L}^{L}\sin\frac{n\pi x}{L}\,dx\,.$$

†This nomenclature is suggested by the fact that the formulas for the Riemann sums approximating the integrals (3)–(5) look like dot products (Section 9.1, page 496) of higher-dimensional vectors. In most calculus texts, it is shown that the dot product of orthogonal vectors in two dimensions is zero.

‡The choice of constant $a_0/2$ instead of just a_0 will be motivated shortly.

The (signed) area under the "oscillators" is zero. Hence,

$$\int_{-L}^{L} f(x)\,dx = \int_{-L}^{L} \frac{a_0}{2}\,dx = a_0 L\,,$$

and so

$$a_0 = \frac{1}{L}\int_{-L}^{L} f(x)\,dx\,.$$

(Notice that $a_0/2$ is the average value of f over one period $2L$.) Next, to find the coefficient a_m when $m \geq 1$, we multiply (6) by $\cos(m\pi x/L)$ and integrate:

$$\begin{aligned}\int_{-L}^{L} f(x)\cos\frac{m\pi x}{L}\,dx &= \frac{a_0}{2}\int_{-L}^{L}\cos\frac{m\pi x}{L}\,dx + \sum_{n=1}^{\infty} a_n \int_{-L}^{L}\cos\frac{n\pi x}{L}\cos\frac{m\pi x}{L}\,dx \\ &\quad + \sum_{n=1}^{\infty} b_n \int_{-L}^{L}\sin\frac{n\pi x}{L}\cos\frac{m\pi x}{L}\,dx\,.\end{aligned} \tag{7}$$

The orthogonality conditions (3)–(5) render the integrals on the right-hand side quite immediately. We have already observed that

$$\int_{-L}^{L}\cos\frac{m\pi x}{L}\,dx = 0\,, \qquad m \geq 1\,,$$

and, by formulas (3) and (5), we have

$$\int_{-L}^{L}\sin\frac{n\pi x}{L}\cos\frac{m\pi x}{L}\,dx = 0\,,$$

$$\int_{-L}^{L}\cos\frac{n\pi x}{L}\cos\frac{m\pi x}{L}\,dx = \begin{cases} 0\,, & n \neq m\,, \\ L\,, & n = m\,. \end{cases}$$

Hence, in (7) we see that only one term on the right-hand side survives the integration:

$$\int_{-L}^{L} f(x)\cos\frac{m\pi x}{L}\,dx = a_m L\,.$$

Thus, we have a formula for the coefficient a_m:

$$a_m = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{m\pi x}{L}\,dx\,.$$

Similarly, multiplying (6) by $\sin(m\pi x/L)$ and integrating yields

$$\int_{-L}^{L} f(x)\sin\frac{m\pi x}{L}\,dx = b_m L$$

so that the formula for b_m is

$$b_m = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{m\pi x}{L}\,dx\,.$$

Motivated by the above computations, we now make the following definition.

Fourier Series

Definition 1. Let f be a piecewise continuous function on the interval $[-L, L]$. The **Fourier series**[†] of f is the trigonometric series

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \left\{ a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right\}, \tag{8}$$

where the a_n's and b_n's are given by the formulas[‡]

$$a_n = \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi x}{L}\, dx\,, \qquad n = 0, 1, 2, \ldots, \tag{9}$$

$$b_n = \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi x}{L}\, dx\,, \qquad n = 1, 2, 3, \ldots. \tag{10}$$

Formulas (9) and (10) are called the **Euler-Fourier formulas.** We use the symbol $\sim$ in (8) to remind us that this series is associated with $f(x)$ but may not converge to $f(x)$. We will return to the question of convergence later in this section. Let's first consider a few examples of Fourier series.

Example 3 Compute the Fourier series for

$$f(x) = \begin{cases} 0\,, & -\pi < x < 0\,, \\ x\,, & 0 < x < \pi\,. \end{cases}$$

Solution Here $L = \pi$. Using formulas (9) and (10), we have

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, dx = \frac{1}{\pi} \int_{0}^{\pi} x\, dx = \left. \frac{x^2}{2\pi} \right|_0^{\pi} = \frac{\pi}{2}\,,$$

$$\begin{aligned} a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\, dx = \frac{1}{\pi} \int_{0}^{\pi} x \cos nx\, dx \\ &= \frac{1}{\pi n^2} \int_{0}^{\pi n} u \cos u\, du = \left. \frac{1}{\pi n^2} [\cos u + u \sin u] \right|_0^{\pi n} \\ &= \frac{1}{\pi n^2} (\cos n\pi - 1) = \frac{1}{\pi n^2} [(-1)^n - 1]\,, \qquad n = 1, 2, 3, \ldots\,, \end{aligned}$$

$$\begin{aligned} b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx\, dx = \frac{1}{\pi} \int_{0}^{\pi} x \sin nx\, dx \\ &= \frac{1}{\pi n^2} \int_{0}^{\pi n} u \sin u\, du = \left. \frac{1}{\pi n^2} [\sin u - u \cos u] \right|_0^{\pi n} \\ &= \frac{-\cos n\pi}{n} = \frac{(-1)^{n+1}}{n}\,, \qquad n = 1, 2, 3, \ldots\,. \end{aligned}$$

[†] *Historical Footnote:* Joseph B. J. Fourier (1768–1830) developed his series for solving heat flow problems. Lagrange expressed doubts about the validity of the representation, but Dirichlet devised conditions that ensured its convergence. Note that the notational choice of $a_0/2$ in equation (6) made it unnecessary to insert an extra formula for a_0 in (9).

[‡] Notice that $f(x)$ need not be defined for every x in $[-L, L]$; we need only that the integrals in (9) and (10) exist.

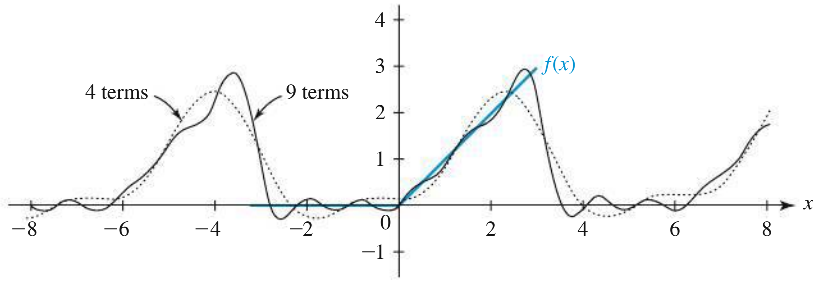


Figure 10.6 Partial sums of Fourier series in Example 3

Therefore,

$$f(x) \sim \frac{\pi}{4} + \sum_{n=1}^{\infty} \left\{ \frac{1}{\pi n^2}[(-1)^n - 1]\cos nx + \frac{(-1)^{n+1}}{n}\sin nx \right\} \tag{11}$$

$$= \frac{\pi}{4} - \frac{2}{\pi}\left\{\cos x + \frac{1}{9}\cos 3x + \frac{1}{25}\cos 5x + \cdots\right\}$$

$$+ \left\{\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x + \cdots\right\}.$$

Some partial sums of this series are displayed in Figure 10.6. ◆

Example 4 Compute the Fourier series for

$$f(x) = \begin{cases} -1, & -\pi < x < 0, \\ 1, & 0 < x < \pi. \end{cases}$$

Solution Again, $L = \pi$. Notice that f is an odd function. Since the product of an odd function and an even function is odd (see Problem 7), $f(x)\cos nx$ is also an odd function. Thus,

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx = 0, \qquad n = 0, 1, 2, \ldots.$$

Furthermore, $f(x)\sin nx$ is the product of two odd functions and therefore is even, so

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = \frac{2}{\pi}\int_0^{\pi} \sin nx\,dx$$

$$= \frac{2}{\pi}\left[\frac{-\cos nx}{n}\right]\Bigg|_0^{\pi} = \frac{2}{\pi}\left[\frac{1}{n} - \frac{(-1)^n}{n}\right], \qquad n = 1, 2, 3, \ldots,$$

$$= \begin{cases} 0, & n \text{ even}, \\ \dfrac{4}{\pi n}, & n \text{ odd}. \end{cases}$$

Thus,

$$f(x) \sim \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{[1 - (-1)^n]}{n}\sin nx = \frac{4}{\pi}\left[\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \cdots\right]. \tag{12}$$

Some partial sums of (12) are sketched in Figure 10.7 on page 578. ◆

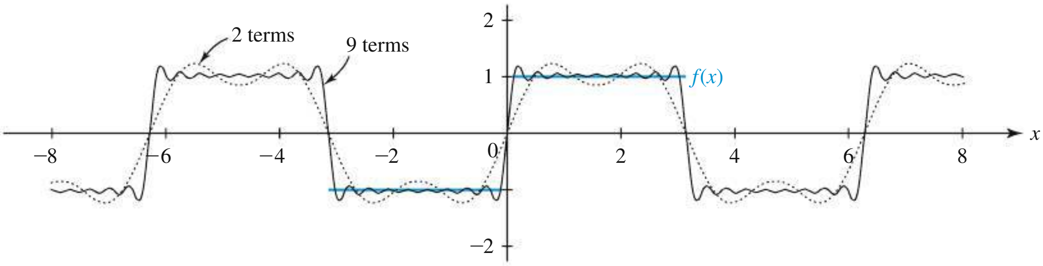


Figure 10.7 Partial sums of Fourier series in Example 4

In Example 4 the odd function f has a Fourier series consisting only of sine functions. It is easy to see that, in general, if f is any odd function, then its Fourier series consists only of sine terms.

Example 5 Compute the Fourier series for $f(x) = |x|$, $-1 < x < 1$.

Solution Here $L = 1$. Since f is an even function, $f(x)\sin(n\pi x)$ is an odd function. Therefore,

$$b_n = \int_{-1}^{1} f(x)\sin(n\pi x)\,dx = 0\,, \qquad n = 1, 2, 3, \ldots\,.$$

Since $f(x)\cos(n\pi x)$ is an even function, we have

$$a_0 = \int_{-1}^{1} f(x)\,dx = 2\int_0^1 x\,dx = x^2\Big|_0^1 = 1\,,$$

$$a_n = \int_{-1}^{1} f(x)\cos(n\pi x)\,dx = 2\int_0^1 x\cos(n\pi x)\,dx$$

$$= \frac{2}{\pi^2 n^2}\int_0^{\pi n} u\cos u\,du = \frac{2}{\pi^2 n^2}[\cos u + u\sin u]\Big|_0^{\pi n} = \frac{2}{\pi^2 n^2}(\cos n\pi - 1)$$

$$= \frac{2}{\pi^2 n^2}[(-1)^n - 1]\,, \qquad n = 1, 2, 3, \ldots\,.$$

Therefore,

$$f(x) \sim \frac{1}{2} + \sum_{n=1}^{\infty} \frac{2}{\pi^2 n^2}[(-1)^n - 1]\cos(n\pi x) \tag{13}$$

$$= \frac{1}{2} - \frac{4}{\pi^2}\left\{\cos(\pi x) + \frac{1}{9}\cos(3\pi x) + \frac{1}{25}\cos(5\pi x) + \cdots\right\}.$$

Partial sums for (13) are displayed in Figure 10.8. ◆

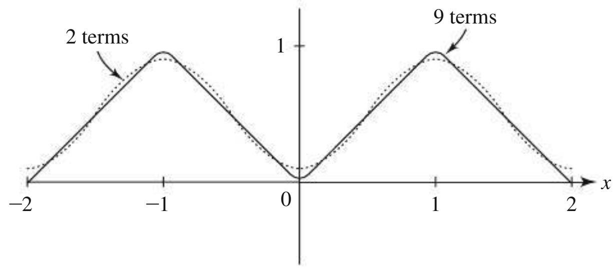


Figure 10.8 Partial sums of Fourier series in Example 5

Notice that the even function f of Example 5 has a Fourier series consisting only of cosine functions and the constant function $1 = \cos(0\pi x)$. In general, if f is an even function, then its Fourier series consists only of cosine functions [including $\cos(0\pi x)$].

Orthogonal Expansions

Fourier series are examples of orthogonal expansions.[†] A set of functions $\{f_n(x)\}_{n=1}^{\infty}$ is said to be an **orthogonal system** or just **orthogonal** with respect to the nonnegative weight function $w(x)$ on the interval $[a, b]$ if

(14) $$\int_a^b f_m(x) f_n(x) w(x)\, dx = 0\,, \qquad \text{whenever} \quad m \neq n\,.$$

As we have seen, the set of trigonometric functions

(15) $$\{1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots\}$$

is orthogonal on $[-\pi, \pi]$ with respect to the weight function $w(x) \equiv 1$. If we define the **norm** of f as

(16) $$\|f\| := \left[\int_a^b f^2(x)\, w(x)\, dx\right]^{1/2},$$

then we say that a set of functions $\{f_n(x)\}_{n=1}^{\infty}$ (or $\{f_n(x)\}_{n=1}^{N}$) is an **orthonormal system with respect to** $w(x)$ if (14) holds and also $\|f_n\| = 1$ for each n. Equivalently, we say the set is an orthonormal system if

(17) $$\int_a^b f_m(x) f_n(x) w(x)\, dx = \begin{cases} 0\,, & m \neq n\,, \\ 1\,, & m = n\,. \end{cases}$$

We can always obtain an orthonormal system from an orthogonal system just by dividing each function by its norm. In particular, since

$$\int_{-\pi}^{\pi} \cos^2 nx\, dx = \int_{-\pi}^{\pi} \sin^2 nx\, dx = \pi\,, \qquad n = 1, 2, 3, \ldots$$

and

$$\int_{-\pi}^{\pi} 1\, dx = 2\pi\,,$$

then the orthogonal system (15) gives rise on $[-\pi, \pi]$ to the orthonormal system

$$\{(2\pi)^{-1/2},\ \pi^{-1/2}\cos x,\ \pi^{-1/2}\sin x,\ \pi^{-1/2}\cos 2x,\ \pi^{-1/2}\sin 2x,\ \ldots\}\,.$$

If $\{f_n(x)\}_{n=1}^{\infty}$ is an orthogonal system with respect to $w(x)$ on $[a, b]$, we might ask if we can expand a function $f(x)$ in terms of these functions; that is, can we express f in the form

(18) $$f(x) = c_1 f_1(x) + c_2 f_2(x) + c_3 f_3(x) + \cdots$$

for a suitable choice of constants $c_1, c_2, \ldots$? Such an expansion is called an **orthogonal expansion,** or a **generalized Fourier series.**

[†]Orthogonality is also discussed in Section 8.8 on page 474.

To determine the constants in (18), we can proceed as we did in deriving Euler's formulas for the coefficients of a Fourier series, this time using the orthogonality of the system. Presuming the representation (18) is valid, we multiply by $f_m(x)w(x)$ and integrate to obtain

(19)
$$\int_a^b f(x)f_m(x)\,w(x)\,dx = c_1\int_a^b f_1(x)f_m(x)\,w(x)\,dx + c_2\int_a^b f_2(x)f_m(x)\,w(x)\,dx + \cdots$$
$$= \sum_{n=1}^{\infty} c_n \int_a^b f_n(x)f_m(x)\,w(x)\,dx\,.$$

(Here we have also assumed that we can integrate term by term.) Because the system is orthogonal with respect to $w(x)$, every integral on the right-hand side of (19) is zero except when $n = m$. Solving for c_m gives

(20)
$$c_m = \frac{\displaystyle\int_a^b f(x)f_m(x)\,w(x)\,dx}{\displaystyle\int_a^b f_m^2(x)\,w(x)\,dx} = \frac{\displaystyle\int_c^b f(x)f_m(x)\,w(x)\,dx}{\|f_m\|^2}\,, \qquad n = 1, 2, 3, \ldots\,.$$

The derivation of the formula for c_m was only *formal,* since the question of the convergence of the expansion in (18) was not answered. If the series $\sum_{n=1}^{\infty} c_n f_n(x)$ converges *uniformly* to $f(x)$ on $[a, b]$, then each step can be justified, and indeed, the coefficients are given by formula (20). The notion of uniform convergence is discussed in the next subsection and in Section 13.2.†

Convergence of Fourier Series

Let's turn to the question of the convergence of a Fourier series. For Example 5 it is possible to use a comparison or limit comparison test to show that the series is absolutely dominated by a p-series of the form $\sum_{n=1}^{\infty} 1/n^2$, which converges. However, this is much harder to do in Example 4, since the terms go to zero like $1/n$. Matters can be even worse, since there exist Fourier series that *diverge.*‡ We state two theorems that deal with the convergence of a Fourier series and two dealing with the properties of termwise differentiation and integration. For proofs of these results, see *Partial Differential Equations of Mathematical Physics,* 2nd ed., by Tyn Myint-U (Elsevier North Holland, Inc., New York, 1983), Chapter 5; *Advanced Calculus with Applications,* by N. J. DeLillo (Macmillan, New York, 1982), Chapter 9; or an advanced text on the theory of Fourier series.

Before proceeding, we need a notation for the left- and right-hand limits of a function. Let

$$f(x^+) := \lim_{h\to 0^+} f(x+h) \quad \text{and} \quad f(x^-) := \lim_{h\to 0^+} f(x-h)\,.$$

We now present the fundamental pointwise convergence theorem for Fourier series. While reading it, keep in mind Figures 10.6-10.8 .

---

†All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems,* 7th ed.

‡In fact, there are trigonometric series that converge but are not Fourier series; an example is $\displaystyle\sum_{n=1}^{\infty} \frac{\sin nx}{\ln(n+1)}$.

Pointwise Convergence of Fourier Series

Theorem 2. If f and f' are piecewise continuous on $[-L, L]$, then for any x in $(-L, L)$

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left\{ a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L} \right\} = \frac{1}{2}[f(x^+) + f(x^-)] , \tag{21}$$

where the a_n's and b_n's are given by the Euler-Fourier formulas (9) and (10). For $x = \pm L$, the series converges to $\frac{1}{2}[f(-L^+) + f(L^-)]$.[†]

In other words, when f and f' are piecewise continuous on $[-L, L]$, the Fourier series converges to $f(x)$ whenever f is continuous at x and converges to the average of the left- and right-hand limits at points where f is discontinuous.

Observe that the left-hand side of (21) is periodic of period $2L$. This means that if we extend $f(x)$ from the interval $(-L, L)$ to the entire real line using $2L$-periodicity, then equation (21) holds for all x for the $2L$-periodic extension of $f(x)$.

Example 6 To which function does the Fourier series for

$$f(x) = \begin{cases} -1 , & -\pi < x < 0 , \\ 1 , & 0 < x < \pi , \end{cases}$$

converge?

Solution In Example 4 we found that the Fourier series for $f(x)$ is given by (12), and in Figure 10.7, we sketched the graphs of two of its partial sums. Now $f(x)$ and $f'(x)$ are piecewise continuous in $[-\pi, \pi]$. Moreover, f is continuous except at $x = 0$. Thus, by Theorem 2, the Fourier series of f in (12) converges to the 2π-periodic function $g(x)$, where $g(x) = f(x) = -1$ for $-\pi < x < 0$, $g(x) = f(x) = 1$ for $0 < x < \pi$, $g(0) = [f(0^+) + f(0^-)]/2 = 0$, and at $\pm\pi$ we have $g(\pm\pi) = [f(-\pi^+) + f(\pi^-)]/2 = (-1 + 1)/2 = 0$. The graph of $g(x)$ is given in Figure 10.9. ◆

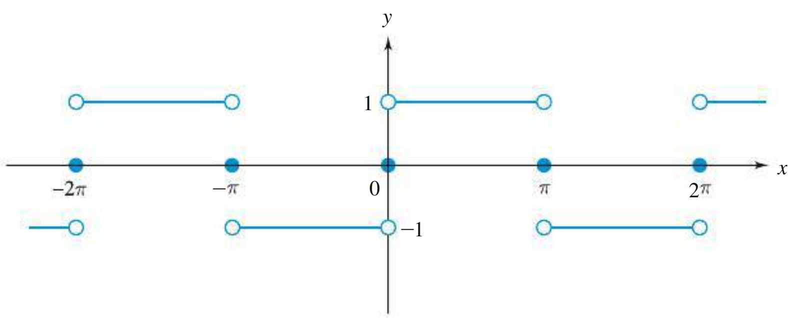


Figure 10.9 The limit function of the Fourier series for $f(x) = \begin{cases} -1 , & -\pi < x < 0 , \\ 1 , & 0 < x < \pi \end{cases}$

[†]From formula (21), we see that it doesn't matter how we define $f(x)$ at its points of discontinuity, since only the left-and right-hand limits are involved. The derivative $f'(x)$, of course, is undefined at such points.

When f is a $2L$-periodic function that is continuous on $(-\infty, \infty)$ and has a piecewise continuous derivative, its Fourier series not only converges at each point—it converges **uniformly on** $(-\infty, \infty)$. This means that for any prescribed tolerance $\varepsilon > 0$, the graph of the partial sum

$$s_N(x) := \frac{a_0}{2} + \sum_{n=1}^{N} \left\{ a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L} \right\}$$

will, for all N sufficiently large, lie in an ε-corridor about the graph of f on $(-\infty, \infty)$ (see Figure 10.10). The property of uniform convergence of Fourier series is particularly helpful when one needs to verify that a formal solution to a partial differential equation is an actual (genuine) solution.

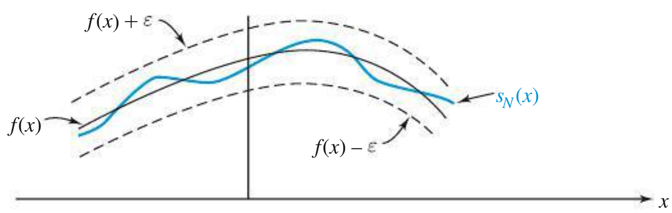


Figure 10.10 An ε-corridor about f

Uniform Convergence of Fourier Series

Theorem 3. Let f be a continuous function on $(-\infty, \infty)$ and periodic of period $2L$. If f' is piecewise continuous on $[-L, L]$, then the Fourier series for f converges uniformly to f on $[-L, L]$ and hence on any interval. That is, for each $\varepsilon > 0$, there exists an integer N_0 (that depends on ε) such that

$$\left| f(x) - \left[\frac{a_0}{2} + \sum_{n=1}^{N} \left\{ a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L} \right\} \right] \right| < \varepsilon ,$$

for all $N \geq N_0$, and all $x \in (-\infty, \infty)$.

In Example 5 we obtained the Fourier series expansion given in (13) for $f(x) = |x|$, $-1 < x < 1$. Since $g(x)$, the periodic extension of $f(x)$ (see Figure 10.11 on page 583) is continuous on $(-\infty, \infty)$ and

$$f'(x) = \begin{cases} -1\,, & -1 < x < 0\,, \\ 1\,, & 0 < x < 1\,, \end{cases}$$

is piecewise continuous on $[-1, 1]$, the Fourier series expansion (13) converges uniformly to $|x|$ on $[-1, 1]$. Compare Figure 10.8 on page 578.

The term-by-term differentiation of a Fourier series is not always permissible. For example, the Fourier series for $f(x) = x$, $-\pi < x < \pi$ (see Problem 9), is

$$f(x) \sim 2 \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\sin nx}{n}\,, \qquad (22)$$

which converges for all x, whereas its derived series

$$2\sum_{n=1}^{\infty}(-1)^{n+1}\cos nx$$

diverges for every x. The following theorem gives sufficient conditions for using termwise differentiation.

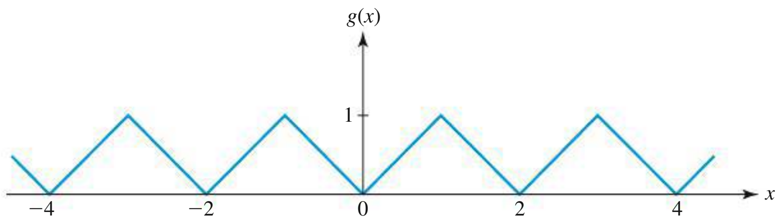


Figure 10.11 Periodic extension of $f(x) = |x|, -1 < x < 1$

Differentiation of Fourier Series

Theorem 4. Let $f(x)$ be continuous on $(-\infty, \infty)$ and $2L$-periodic. Let $f'(x)$ and $f''(x)$ be piecewise continuous on $[-L, L]$. Then, the Fourier series of $f'(x)$ can be obtained from the Fourier series for $f(x)$ by termwise differentiation. In particular, if

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left\{a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}\right\},$$

then

$$f'(x) \sim \sum_{n=1}^{\infty}\frac{\pi n}{L}\left\{-a_n \sin\frac{n\pi x}{L} + b_n \cos\frac{n\pi x}{L}\right\}.$$

Notice that Theorem 4 does not apply to the example function $f(x)$ and its Fourier series expansion shown in (22), since the 2π-periodic extension of this $f(x)$ fails to be continuous on $(-\infty, \infty)$.

Termwise integration of a Fourier series is permissible under much weaker conditions.

Integration of Fourier Series

Theorem 5. Let $f(x)$ be piecewise continuous on $[-L, L]$ with Fourier series

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty}\left\{a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}\right\}.$$

Then, for any x in $[-L, L]$,

$$\int_{-L}^{x} f(t)\,dt = \int_{-L}^{x}\frac{a_0}{2}dt + \sum_{n=1}^{\infty}\int_{-L}^{x}\left\{a_n \cos\frac{n\pi t}{L} + b_n \sin\frac{n\pi t}{L}\right\}dt\,.$$

A final note on the convergence issue: The question as to under what conditions the Fourier series converges has led to a tremendous amount of beautiful mathematics. As a practical matter, however, time and economics will often dictate that you can afford to compute only a few terms of an eigenfunction series for your particular partial differential equation. Convergence fades into the background, and you need to know how to "do the best with what you've got." So it is extremely gratifying to know that even if you're going to use only part of an eigenfunction expansion, the Fourier coefficients (9)–(10) are still the best choice; Problem 37 demonstrates that every partial sum of a Fourier series outperforms any comparable superposition of the trigometric functions, in terms of mean-square approximation.

10.3 EXERCISES

In Problems 1–6, determine whether the given function is even, odd, or neither.

1. $f(x) = x^3 + \sin 2x$

2. $f(x) = \sin^2 x$

3. $f(x) = (1 - x^2)^{-1/2}$

4. $f(x) = \sin(x + 1)$

5. $f(x) = e^{-x}\cos 3x$

6. $f(x) = x^{1/5}\cos x^2$

7. Prove the following properties:

(a) If f and g are even functions, then so is the product fg.

(b) If f and g are odd functions, then fg is an even function.

(c) If f is an even function and g is an odd function, then fg is an odd function.

8. Verify formula (5). [*Hint:* Use the identity $2\cos A\cos B = \cos(A + B) + \cos(A - B)$.]

In Problems 9–16, compute the Fourier series for the given function f on the specified interval. Use a computer or graphing calculator to plot a few partial sums of the Fourier series.

9. $f(x) = x\,, \quad -\pi < x < \pi$

10. $f(x) = |x|\,, \quad -\pi < x < \pi$

11. $f(x) = \begin{cases} 1\,, & -2 < x < 0\,, \\ x\,, & 0 < x < 2 \end{cases}$

12. $f(x) = \begin{cases} 0\,, & -\pi < x < 0\,, \\ x^2\,, & 0 < x < \pi \end{cases}$

13. $f(x) = x^2\,, \quad -1 < x < 1$

14. $f(x) = \begin{cases} x\,, & 0 < x < \pi\,, \\ x + \pi\,, & -\pi < x < 0 \end{cases}$

15. $f(x) = e^x\,, \quad -\pi < x < \pi$

16. $f(x) = \begin{cases} 0\,, & -\pi < x < -\pi/2\,, \\ -1\,, & -\pi/2 < x < 0\,, \\ 1\,, & 0 < x < \pi/2\,, \\ 0\,, & \pi/2 < x < \pi \end{cases}$

In Problems 17–24, determine the function to which the Fourier series for $f(x)$, given in the indicated problem, converges.

17. Problem 9

18. Problem 10

19. Problem 11

20. Problem 12

21. Problem 13

22. Problem 14

23. Problem 15

24. Problem 16

25. Find the functions represented by the series obtained by the termwise integration of the given series from $-\pi$ to x.

(a) $2\displaystyle\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin nx \sim x\,, \quad -\pi < x < \pi$

(b) $\dfrac{4}{\pi}\displaystyle\sum_{n=0}^{\infty}\frac{\sin(2n+1)x}{(2n+1)} \sim f(x)\,,$

$$f(x) = \begin{cases} -1\,, & -\pi < x < 0\,, \\ 1\,, & 0 < x < \pi \end{cases}$$

26. Show that the set of functions

$$\left\{\cos\frac{\pi}{2}x,\ \sin\frac{\pi}{2}x,\ \cos\frac{3\pi}{2}x,\ \sin\frac{3\pi}{2}x,\ \ldots,\right.$$
$$\left.\cos\frac{(2n-1)\pi}{2}x,\ \sin\frac{(2n-1)\pi}{2}x,\ \ldots\right\}$$

is an orthonormal system on $[-1, 1]$ with respect to the weight function $w(x) \equiv 1$.

27. Find the orthogonal expansion (generalized Fourier series) for

$$f(x) = \begin{cases} 0\,, & -1 < x < 0\,, \\ 1\,, & 0 < x < 1\,, \end{cases}$$

in terms of the orthonormal system of Problem 26.

28. **(a)** Show that the function $f(x) = x^2$ has the Fourier series, on $-\pi < x < \pi$,

$$f(x) \sim \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos nx\,.$$

(b) Use the result of part (a) and Theorem 2 to show that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12} .$$

(c) Use the result of part (a) and Theorem 2 to show that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} .$$

29. In Section 8.8, it was shown that the Legendre polynomials $P_n(x)$ are orthogonal on the interval $[-1, 1]$ with respect to the weight function $w(x) \equiv 1$. Using the fact that the first three Legendre polynomials are

$$P_0(x) \equiv 1 ,\ P_1(x) = x ,\ P_2(x) = (3/2)x^2 - (1/2) ,$$

find the first three coefficients in the expansion

$$f(x) = c_0P_0(x) + c_1P_1(x) + c_2P_2(x) + \cdots ,$$

where $f(x)$ is the function

$$f(x) := \begin{cases} -1 , & -1 < x < 0 , \\ 1 , & 0 < x < 1 . \end{cases}$$

30. As in Problem 29, find the first three coefficients in the expansion

$$f(x) = c_0P_0(x) + c_1P_1(x) + c_2P_2(x) + \cdots ,$$

when $f(x) = |x|, -1 < x < 1$.

31. The Hermite polynomials $H_n(x)$ are orthogonal on the interval $(-\infty, \infty)$ with respect to the weight function $W(x) = e^{-x^2}$. Verify this fact for the first three Hermite polynomials:

$$H_0(x) \equiv 1 , \quad H_1(x) = 2x , \quad H_2(x) = 4x^2 - 2 .$$

32. The Chebyshev (Tchebichef) polynomials $T_n(x)$ are orthogonal on the interval $[-1, 1]$ with respect to the weight function $w(x) = (1 - x^2)^{-1/2}$. Verify this fact for the first three Chebyshev polynomials:

$$T_0(x) \equiv 1 , \quad T_1(x) = x , \quad T_2(x) = 2x^2 - 1 .$$

33. Let $\{f_n(x)\}$ be an orthogonal set of functions on the interval $[a, b]$ with respect to the weight function $w(x)$. Show that they satisfy the **Pythagorean property**

$$\|f_m + f_n\|^2 = \|f_m\|^2 + \|f_n\|^2$$

if $m \neq n$.

34. Norm. The norm of a function $\|f\|$ is like the length of a vector in R^n. In particular, show that the norm defined in (16) satisfies the following properties associated with length (assume f and g are continuous and $w(x) > 0$ on $[a, b]$):

(a) $\|f\| \geq 0$, and $\|f\| = 0$ if and only if $f \equiv 0$.

(b) $\|cf\| = |c|\ \|f\|$, where c is any real number.

(c) $\|f + g\| \leq \|f\| + \|g\|$.

35. Inner Product. The integral in the orthogonality condition (14) is like the dot product of two vectors in R^n. In particular, show that the **inner product** of two functions defined by

$$\text{(23)} \qquad \langle f, g \rangle := \int_a^b f(x)g(x)w(x)\,dx ,$$

where $w(x)$ is a positive weight function, satisfies the following properties associated with the dot product (assume f, g, and h are continuous on $[a, b]$):

(a) $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$.

(b) $\langle cf, h \rangle = c\langle f, h \rangle$, where c is any real number.

(c) $\langle f, g \rangle = \langle g, f \rangle$.

36. Complex Form of the Fourier Series.

(a) Using the Euler formula $e^{i\theta} = \cos\theta + i\sin\theta$, $i = \sqrt{-1}$, prove that

$$\cos nx = \frac{e^{inx} + e^{-inx}}{2} \quad \text{and} \quad \sin nx = \frac{e^{inx} - e^{-inx}}{2i} .$$

(b) Show that the Fourier series

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \{a_n \cos nx + b_n \sin nx\}$$
$$= c_0 + \sum_{n=1}^{\infty} \{c_n e^{inx} + c_{-n} e^{-inx}\} ,$$

where

$$c_0 = \frac{a_0}{2}, \quad c_n = \frac{a_n - ib_n}{2}, \quad c_{-n} = \frac{a_n + ib_n}{2} .$$

(c) Finally, use the results of part (b) to show that

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx} , \quad -\pi < x < \pi ,$$

where

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\, e^{-inx}\,dx .$$

37. Least-Squares Approximation Property. The Nth partial sum of the Fourier series

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \{a_n \cos nx + b_n \sin nx\}$$

gives the best mean-square approximation of f by a trigonometric polynomial. To prove this, let $F_N(x)$ denote an arbitrary trigonometric polynomial of degree N:

$$F_N(x) = \frac{\alpha_0}{2} + \sum_{n=1}^{N} \{\alpha_n \cos nx + \beta_n \sin nx\} ,$$

and define

$$E := \int_{-\pi}^{\pi} [f(x) - F_N(x)]^2\,dx ,$$

which is the *total square error.* Expanding the integrand, we get

$$E = \int_{-\pi}^{\pi} f^2(x)\, dx - 2\int_{-\pi}^{\pi} f(x)F_N(x)\, dx + \int_{-\pi}^{\pi} F_N^2(x)\, dx\,.$$

(a) Use the orthogonality of the functions $\{1, \cos x, \sin x, \cos 2x, \dots\}$ to show that

$$\int_{-\pi}^{\pi} F_N^2(x)\, dx = \pi\left(\frac{\alpha_0^2}{2} + \alpha_1^2 + \cdots + \alpha_N^2 + \beta_1^2 + \cdots + \beta_N^2\right)$$

and

$$\int_{-\pi}^{\pi} f(x)F_N(x)\, dx = \pi\left(\frac{\alpha_0 a_0}{2} + \alpha_1 a_1 + \cdots + \alpha_N a_N + \beta_1 b_1 + \cdots + \beta_N b_N\right).$$

(b) Let E^* be the error when we approximate f by the Nth partial sum of its Fourier series, that is, when we choose $\alpha_n = a_n$ and $\beta_n = b_n$. Show that

$$E^* = \int_{-\pi}^{\pi} f^2(x) dx - \pi\left(\frac{a_0^2}{2} + a_1^2 + \cdots + a_N^2 + b_1^2 + \cdots + b_N^2\right).$$

(c) Using the results of parts (a) and (b), show that $E - E^* \ge 0$, that is, $E \ge E^*$, by proving that

$$E - E^* = \pi\left\{\frac{(\alpha_0 - a_0)^2}{2} + (\alpha_1 - a_1)^2 + \cdots + (\alpha_N - a_N)^2 + (\beta_1 - b_1)^2 + \cdots + (\beta_N - b_N)^2\right\}.$$

Hence, the Nth partial sum of the Fourier series gives the least total square error, since $E \ge E^*$.

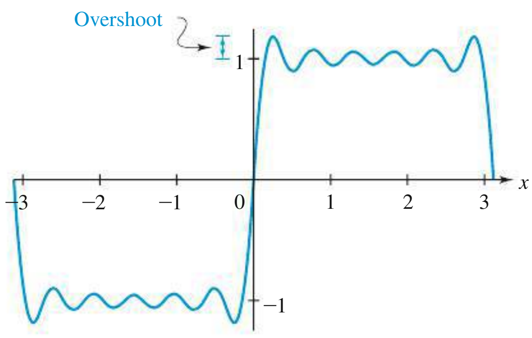
38. Bessel's Inequality. Use the fact that E^*, defined in part (b) of Problem 37, is nonnegative to prove **Bessel's inequality**

$$\frac{a_0^2}{2} + \sum_{n=1}^{\infty} \{a_n^2 + b_n^2\} \le \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\, dx\,. \tag{24}$$

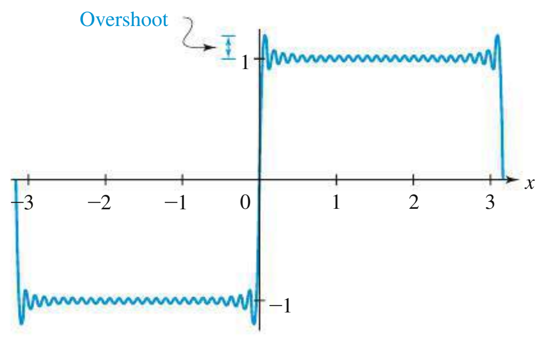
(If f is piecewise continuous on $[-\pi, \pi]$, then we have equality in (24). This result is called **Parseval's identity**.)

39. Gibbs Phenomenon.[†] Josiah Willard Gibbs, who was awarded the first American doctorate in engineering (Yale, 1863), observed that near points of discontinuity of f, the partial sums of the Fourier series for f may overshoot by approximately 9% of the jump, regardless of the number of terms. This is illustrated in Figure 10.12 for the function

$$f(x) = \begin{cases} -1, & -\pi < x < 0, \\ 1, & 0 < x < \pi, \end{cases}$$



(a) Graph of $f_{11}(x)$



(b) Graph of $f_{51}(x)$

Figure 10.12 Gibbs phenomenon for partial sums of Fourier series

[†]*Historical Footnote:* Actually, H. Wilbraham discovered this phenomenon some 50 years earlier than Gibbs did. It is more appropriately called the *Gibbs–Wilbraham* phenomenon.

whose Fourier series has the partial sums

$$f_{2n-1}(x) = \frac{4}{\pi}\left[\sin x + \frac{1}{3}\sin 3x + \cdots + \frac{\sin(2n-1)x}{(2n-1)}\right].$$

To verify this for $f(x)$, proceed as follows:

(a) Show that

$$\pi(\sin x)f'_{2n-1}(x) = 4\sin x[\cos x + \cos 3x + \cdots + \cos(2n-1)x] = 2\sin 2nx.$$

(b) Infer from part (a) and the figure that the maximum occurs at $x = \pi/(2n)$ and has the value

$$f_{2n-1}\left(\frac{\pi}{2n}\right) = \frac{4}{\pi}\left[\sin\frac{\pi}{2n} + \frac{1}{3}\sin\frac{3\pi}{2n} + \cdots + \frac{1}{2n-1}\sin\frac{(2n-1)\pi}{2n}\right].$$

(c) Show that if one approximates

$$\int_0^\pi \frac{\sin x}{x}dx$$

using the partition $x_k := (2k-1)(\pi/2n)$, $k = 1, 2, \ldots, n$, $\Delta x_k = \pi/n$ and choosing the midpoint of each interval as the place to evaluate the integrand, then

$$\int_0^\pi \frac{\sin x}{x}dx \approx \frac{\sin(\pi/2n)}{\pi/2n}\frac{\pi}{n} + \cdots + \frac{\sin[(2n-1)\pi/2n]}{(2n-1)\pi/2n}\frac{\pi}{n} = \frac{\pi}{2}f_{2n-1}\left(\frac{\pi}{2n}\right).$$

(d) Use the result of part (c) to show that the overshoot satisfies

$$\lim_{n\to\infty} f_{2n-1}\left(\frac{\pi}{2n}\right) = \frac{2}{\pi}\int_0^\pi \frac{\sin x}{x}dx.$$

(e) Using the result of part (d) and a numerical integration algorithm (such as Simpson's rule, Appendix C) for the **sine integral function**

$$Si(z) := \int_0^z \frac{\sin x}{x}dx,$$

show that $\lim_{n\to\infty} f_{2n-1}(\pi/(2n)) \approx 1.18$. Thus, the approximations overshoot the true value of $f(0^+) = 1$ by 0.18, or 9% of the jump from $f(0^-)$ to $f(0^+)$.

10.4 Fourier Cosine and Sine Series

A typical problem encountered in using separation of variables to solve a partial differential equation is the problem of representing a function defined on some finite interval by a trigonometric series consisting of only sine functions or only cosine functions. For example, in Section 10.2, equation (25), page 569, we needed to express the initial values $u(x, 0) = f(x)$, $0 < x < L$, of the solution to the initial-boundary value problem associated with the problem of a vibrating string as a trigonometric series of the form

$$f(x) = \sum_{n=1}^{\infty} a_n \sin\left(\frac{n\pi x}{L}\right). \tag{1}$$

Recalling that the Fourier series for an odd function defined on $[-L, L]$ consists entirely of sine terms, we might try to achieve (1) by artificially extending the function $f(x)$, $0 < x < L$, to the interval $(-L, L)$ in such a way that the extended function is odd. This is accomplished by defining the function

$$f_o(x) := \begin{cases} f(x), & 0 < x < L, \\ -f(-x), & -L < x < 0, \end{cases}$$

and extending $f_o(x)$ to all x using $2L$-periodicity.† Since $f_o(x)$ is an odd function, it has a Fourier series consisting entirely of sine terms. Moreover, $f_o(x)$ is an extension of $f(x)$,

†Strictly speaking, we have extended $f_o(x)$ to all x other than the integer multiples of L. Continuity considerations often suggest appropriate values for the extended functions at some of these points, as well. Figure 10.13 on page 589 illustrates this.

since $f_o(x) = f(x)$ on $(0, L)$. This extension is called the **odd $2L$-periodic extension** of $f(x)$. The resulting Fourier series expansion is called a half-range expansion for $f(x)$, since it represents the function $f(x)$ on $(0, L)$, which is half of the interval $(-L, L)$ where it represents $f_o(x)$.

In a similar fashion, we can define the **even $2L$-periodic extension** of $f(x)$ as the function

$$f_e(x) := \begin{cases} f(x), & 0 < x < L, \\ f(-x), & -L < x < 0, \end{cases}$$

with $f_e(x + 2L) = f_e(x)$.

To illustrate the various extensions, let's consider the function $f(x) = x, 0 < x < \pi$. If we extend $f(x)$ to the interval $(-\pi, \pi)$ using π-periodicity, then the extension $\tilde{f}$ is given by

$$\tilde{f}(x) = \begin{cases} x, & 0 < x < \pi, \\ x + \pi, & -\pi < x < 0, \end{cases}$$

with $\tilde{f}(x + 2\pi) = \tilde{f}(x)$. In Problem 14 of Exercises 10.3, the Fourier series for $\tilde{f}(x)$ was found to be

$$\tilde{f}(x) \sim \frac{\pi}{2} - \sum_{n=1}^{\infty} \frac{1}{n} \sin 2nx,$$

which consists of both odd functions (the sine terms) and even functions (the constant term), since the π-periodic extension $\tilde{f}(x)$ is neither an even nor an odd function. The odd 2π-periodic extension of $f(x)$ is just $f_o(x) = x, -\pi < x < \pi$, which has the Fourier series expansion

$$f_o(x) \sim 2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin nx \tag{2}$$

(see Problem 9 in Exercises 10.3). Because $f_o(x) = f(x)$ on the interval $(0, \pi)$, the expansion in (2) is a half-range expansion for $f(x)$. The even 2π-periodic extension of $f(x)$ is the function $f_e(x) = |x|, -\pi < x < \pi$, which has the Fourier series expansion

$$f_e(x) = \frac{\pi}{2} - \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \cos(2n+1)x \tag{3}$$

(see Problem 10 in Exercises 10.3).

The preceding three extensions, the π-periodic function $\tilde{f}(x)$, the odd 2π-periodic function $f_o(x)$, and the even 2π-periodic function $f_e(x)$, are natural extensions of $f(x)$. There are many other ways of extending $f(x)$. For example, the function

$$g(x) = \begin{cases} x, & 0 < x < \pi, \\ 0, & -\pi < x < 0, \end{cases} \qquad g(x + 2\pi) = g(x),$$

which we studied in Example 3 of Section 10.3, is also an extension of $f(x)$. However, its Fourier series contains both sine and cosine terms and hence is not as useful as previous extensions. The graphs of these extensions of $f(x)$ are given in Figure 10.13 on page 589.

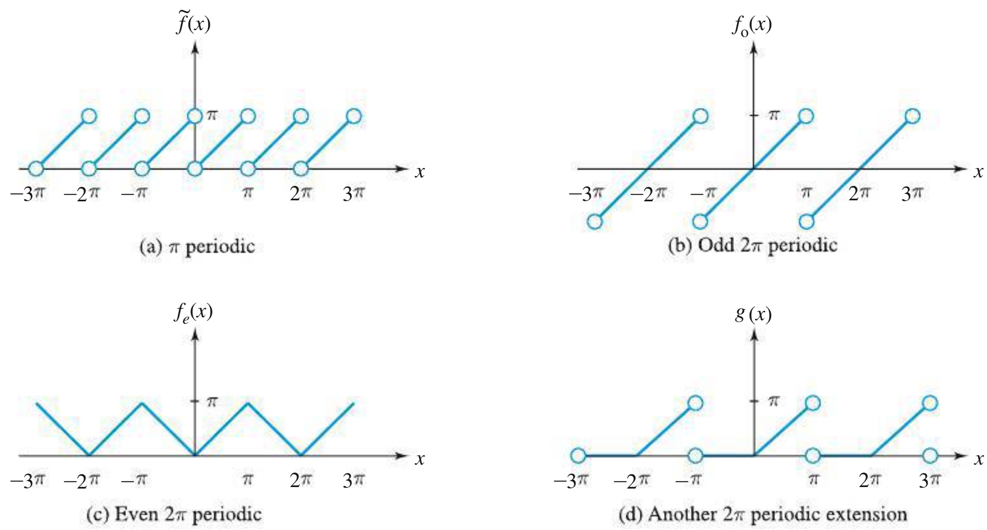


Figure 10.13 Extensions of $f(x) = x, 0 < x < \pi$

The Fourier series expansions for $f_o(x)$ and $f_e(x)$ given in (2) and (3) represent $f(x)$ on the interval $(0, \pi)$ $\big($actually, they equal $f(x)$ on $(0, \pi)\big)$. This motivates the following definitions.

Fourier Cosine and Sine Series

Definition 2. Let $f(x)$ be piecewise continuous on the interval $[0, L]$. The **Fourier cosine series** of $f(x)$ on $[0, L]$ is

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{L}, \tag{4}$$

where

$$a_n = \frac{2}{L}\int_0^L f(x) \cos\frac{n\pi x}{L}\, dx, \qquad n = 0, 1, \ldots. \tag{5}$$

The **Fourier sine series** of $f(x)$ on $[0, L]$ is

$$\sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{L}, \tag{6}$$

where

$$b_n = \frac{2}{L}\int_0^L f(x) \sin\frac{n\pi x}{L}\, dx, \qquad n = 1, 2, \ldots. \tag{7}$$

The trigonometric series in (4) is just the Fourier series for $f_e(x)$, the even $2L$-periodic extension of $f(x)$, and that in (6) is the Fourier series for $f_o(x)$, the odd $2L$-periodic extension of $f(x)$. These are called **half-range expansions** for $f(x)$.

Example 1 Compute the Fourier sine series for

$$f(x) = \begin{cases} x\,, & 0 \le x \le \pi/2\,, \\ \pi - x\,, & \pi/2 \le x \le \pi\,. \end{cases}$$

Solution Using formula (7) with $L = \pi$, we find

$$\begin{aligned} b_n &= \frac{2}{\pi}\int_0^{\pi} f(x)\sin nx\,dx = \frac{2}{\pi}\int_0^{\pi/2} x\sin nx\,dx + \frac{2}{\pi}\int_{\pi/2}^{\pi}(\pi - x)\sin nx\,dx \\ &= \frac{2}{\pi n^2}\int_0^{\pi n/2} u\sin u\,du + 2\int_{\pi/2}^{\pi}\sin nx\,dx - \frac{2}{\pi n^2}\int_{\pi n/2}^{\pi n} u\sin u\,du \\ &= \frac{2}{\pi n^2}[\sin u - u\cos u]\Big|_0^{\pi n/2} - \frac{2}{n}\left[\cos \pi n - \cos\frac{n\pi}{2}\right] \\ &\quad - \frac{2}{\pi n^2}[\sin u - u\cos u]\Big|_{\pi n/2}^{\pi n} \\ &= \frac{4}{\pi n^2}\sin\frac{n\pi}{2} = \begin{cases} 0\,, & n \text{ even}\,, \\ \dfrac{4(-1)^{(n-1)/2}}{\pi n^2}\,, & n \text{ odd}\,. \end{cases} \end{aligned}$$

So on letting $n = 2k + 1$, we find the Fourier sine series for $f(x)$ to be

$$\textbf{(8)} \qquad \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^2}\sin(2k+1)x = \frac{4}{\pi}\left\{\sin x - \frac{1}{9}\sin 3x + \frac{1}{25}\sin 5x + \cdots\right\}. \ \blacklozenge$$

Since, in Example 1, the function $f(x)$ is continuous and $f'(x)$ is piecewise continuous on $(0, \pi)$, it follows from Theorem 2 on pointwise convergence of Fourier series that

$$f(x) = \frac{4}{\pi}\left\{\sin x - \frac{1}{9}\sin 3x + \frac{1}{25}\sin 5x - \frac{1}{49}\sin 7x + \cdots\right\}$$

for all x in $[0, \pi]$.

Let's return to the problem of heat flow in one dimension.

Example 2 Find the solution to the heat flow problem

$$\textbf{(9)} \qquad \frac{\partial u}{\partial t} = 2\frac{\partial^2 u}{\partial x^2}\,, \qquad 0 < x < \pi\,, \qquad t > 0\,,$$

$$\textbf{(10)} \qquad u(0, t) = u(\pi, t) = 0\,, \qquad t > 0\,,$$

$$\textbf{(11)} \qquad u(x, 0) = \begin{cases} x\,, & 0 < x \le \pi/2\,, \\ \pi - x\,, & \pi/2 \le x < \pi\,. \end{cases}$$

Solution Comparing equation (9) with equation (1) in Section 10.2 (page 563), we see that $\beta = 2$ and $L = \pi$. Hence, we need only represent $u(x, 0) = f(x)$ in a Fourier sine series of the form

$$\sum_{n=1}^{\infty} c_n \sin nx\,.$$

In Example 1 we obtained this expansion and showed that

$$c_n = \frac{4}{\pi n^2}\sin\frac{n\pi}{2} = \begin{cases} 0\,, & n \text{ even}\,, \\ \dfrac{4(-1)^{(n-1)/2}}{\pi n^2}\,, & n \text{ odd}\,. \end{cases}$$

Hence, from equation (15) on page 566, the solution to the heat flow problem (9)–(11) is

$$\begin{aligned} u(x,t) &= \sum_{n=1}^{\infty} c_n e^{-2n^2 t}\sin nx \\ &= \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^2}e^{-2(2k+1)^2 t}\sin(2k+1)x \\ &= \frac{4}{\pi}\left\{e^{-2t}\sin x - \frac{1}{9}e^{-18t}\sin 3x + \frac{1}{25}e^{-50t}\sin 5x + \cdots\right\}. \end{aligned}$$

A sketch of a partial sum for $u(x,t)$ is displayed in Figure 10.14. ◆

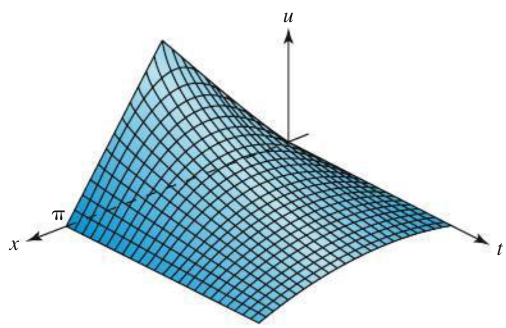


Figure 10.14 Partial sum for $u(x,t)$ in Example 2

10.4 EXERCISES

In Problems 1–4, determine (a) the π-periodic extension $\tilde{f}$, (b) the odd 2π-periodic extension f_o, and (c) the even 2π-periodic extension f_e for the given function f and sketch their graphs.

1. $f(x) = x^2\,, \quad 0 < x < \pi$

2. $f(x) = \sin 2x\,, \quad 0 < x < \pi$

3. $f(x) = \begin{cases} 0\,, & 0 < x < \pi/2\,, \\ 1\,, & \pi/2 < x < \pi \end{cases}$

4. $f(x) = \pi - x\,, \quad 0 < x < \pi$

In Problems 5–10, compute the Fourier sine series for the given function.

5. $f(x) = -1\,, \quad 0 < x < 1$

6. $f(x) = \cos x\,, \quad 0 < x < \pi$

7. $f(x) = x^2\,, \quad 0 < x < \pi$

8. $f(x) = \pi - x\,, \quad 0 < x < \pi$

9. $f(x) = x - x^2\,, \quad 0 < x < 1$

10. $f(x) = e^x\,, \quad 0 < x < 1$

In Problems 11–16, compute the Fourier cosine series for the given function.

11. $f(x) = \pi - x\,, \quad 0 < x < \pi$

12. $f(x) = 1 + x\,, \quad 0 < x < \pi$

13. $f(x) = e^x\,, \quad 0 < x < 1$

14. $f(x) = e^{-x}\,, \quad 0 < x < 1$

15. $f(x) = \sin x\,, \quad 0 < x < \pi$

16. $f(x) = x - x^2\,, \quad 0 < x < 1$

In Problems 17–19, for the given $f(x)$, find the solution to the heat flow problem.

$$\frac{\partial u}{\partial t} = 5\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < \pi, \quad t > 0,$$
$$u(0,t) = u(\pi,t) = 0, \quad t > 0,$$
$$u(x,0) = f(x), \quad 0 < x < \pi,$$

17. $f(x) = 1 - \cos 2x$

18. $f(x) = x(\pi - x)$

19. $f(x) = \begin{cases} -x, & 0 < x \le \pi/2, \\ x - \pi, & \pi/2 \le x < \pi \end{cases}$

10.5 The Heat Equation

In Section 10.1 we developed a model for heat flow in an insulated uniform wire whose ends are kept at the constant temperature 0°C. In particular, we found that the temperature $u(x, t)$ in the wire is governed by the initial-boundary value problem

$$\text{(1)} \qquad \frac{\partial u}{\partial t} = \beta\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < L, \quad t > 0,$$

$$\text{(2)} \qquad u(0,t) = u(L,t) = 0, \quad t > 0,$$

$$\text{(3)} \qquad u(x,0) = f(x), \quad 0 < x < L$$

[see equations (7)–(9) in Section 10.1, page 562]. Here equation (2) specifies that the temperature at the ends of the wire is zero, whereas equation (3) specifies the initial temperature distribution.

In Section 10.2 we also derived a formal solution to (1)–(3) using separation of variables. There we found the solution to (1)–(3) to have the form

$$\text{(4)} \qquad u(x,t) = \sum_{n=1}^{\infty} c_n e^{-\beta(n\pi/L)^2 t} \sin\frac{n\pi x}{L},$$

where the c_n's are the coefficients in the Fourier sine series for $f(x)$:

$$\text{(5)} \qquad f(x) = \sum_{n=1}^{\infty} c_n \sin\frac{n\pi x}{L}.$$

In other words, solving (1)–(3) reduces to computing the Fourier sine series for the initial value function $f(x)$.

In this section we discuss heat flow problems where the ends of the wire are insulated or kept at a constant, but nonzero, temperature. (The latter involves nonhomogeneous boundary conditions.) We will also discuss the problem in which a heat source is adding heat to the wire. (This results in a nonhomogeneous partial differential equation.) The problem of heat flow in a rectangular plate is also discussed and leads to the topic of double Fourier series. We conclude this section with a discussion of the existence and uniqueness of solutions to the heat flow problem.

In the model of heat flow in a uniform wire, let's replace the assumption that the ends of the wire are kept at a constant temperature zero and instead assume that the ends of the wire are *insulated*—that is, no heat flows out (or in) at the ends of the wire. It follows from the principle of heat conduction (see Section 10.1) that the temperature gradient must be zero at these endpoints, that is,

$$\frac{\partial u}{\partial x}(0,t) = \frac{\partial u}{\partial x}(L,t) = 0, \quad t > 0.$$

In the next example we obtain the formal solution to the heat flow problem with these boundary conditions.

Example 1 Find a formal solution to the heat flow problem governed by the initial-boundary value problem

(6) $$\frac{\partial u}{\partial t} = \beta \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \qquad t > 0,$$

(7) $$\frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(L, t) = 0, \quad t > 0,$$

(8) $$u(x, 0) = f(x), \quad 0 < x < L.$$

Solution Using the method of separation of variables, we first assume that

$$u(x, t) = X(x)T(t).$$

Substituting into equation (6) and separating variables, as was done in Section 10.2 [compare equation (5) on page 564], we get the two equations

(9) $$X''(x) + \lambda X(x) = 0,$$

(10) $$T'(t) + \beta\lambda T(t) = 0,$$

where λ is some constant. The boundary conditions in (7) become

$$X'(0)T(t) = 0 \quad \text{and} \quad X'(L)T(t) = 0.$$

For these equations to hold for all $t > 0$, either $T(t) \equiv 0$, which implies that $u(x, t) \equiv 0$, or

(11) $$X'(0) = X'(L) = 0.$$

Combining the boundary conditions in (11) with equation (9) gives the boundary value problem

(12) $$X''(x) + \lambda X(x) = 0; \qquad X'(0) = X'(L) = 0,$$

where λ can be any constant.

To solve for the nontrivial solutions to (12), we propose $X(x) = e^{rx}$ and form the auxiliary equation $r^2 + \lambda = 0$. When $\lambda < 0$, arguments similar to those used in Section 10.2 show that there are no nontrivial solutions to (12).

When $\lambda = 0$, the auxiliary equation has the repeated root 0 and a general solution to the differential equation is

$$X(x) = A + Bx.$$

The boundary conditions in (12) reduce to $B = 0$ with A arbitrary. Thus, for $\lambda = 0$, the nontrivial solutions to (12) are of the form

$$X(x) = c_0,$$

where $c_0 = A$ is an arbitrary nonzero constant.

When $\lambda > 0$, the auxiliary equation has the roots $r = \pm i\sqrt{\lambda}$. Thus, a general solution to the differential equation in (12) is

$$X(x) = C_1 \cos\sqrt{\lambda}x + C_2 \sin\sqrt{\lambda}x.$$

The boundary conditions in (12) lead to the system

$$\begin{aligned} \sqrt{\lambda}C_2 &= 0, \\ -\sqrt{\lambda}C_1 \sin\sqrt{\lambda}L + \sqrt{\lambda}C_2 \cos\sqrt{\lambda}L &= 0. \end{aligned}$$

Hence, $C_2 = 0$ and the system reduces to solving $C_1 \sin\sqrt{\lambda}L = 0$. Since $\sin\sqrt{\lambda}L = 0$ only when $\sqrt{\lambda}L = n\pi$, where n is an integer, we obtain a nontrivial solution only when $\sqrt{\lambda} = n\pi/L$ or $\lambda = (n\pi/L)^2$, $n = 1, 2, 3, \ldots$. Furthermore, the nontrivial solutions (eigenfunctions) X_n corresponding to the eigenvalue $\lambda = (n\pi/L)^2$ are given by

$$X_n(x) = c_n \cos\frac{n\pi x}{L}, \tag{13}$$

where the c_n's are arbitrary nonzero constants. In fact, formula (13) also holds for $n = 0$, since $\lambda = 0$ has the eigenfunctions $X_0(x) = c_0$.

Note that these eigenvalues and eigenfunctions (see Figure 10.15) have the same properties as those on page 565 (which we now recognize as the components of the sine series): homogeneous boundary conditions, eigenvalues clustering at infinity, increasingly oscillatory, orthogonal.

Having determined that $\lambda = (n\pi/L)^2$, $n = 0, 1, 2, \ldots$, let's consider equation (10) for such λ:

$$T'(t) + \beta(n\pi/L)^2 T(t) = 0 .$$

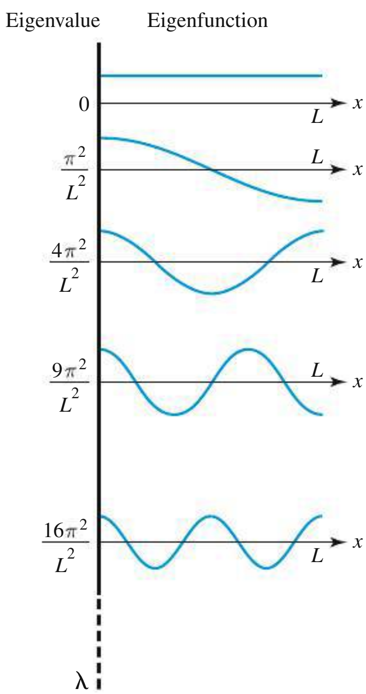


Figure 10.15 Cosine eigenfunctions

For $n = 0, 1, 2, \ldots$, the general solution is

$$T_n(t) = b_n e^{-\beta(n\pi/L)^2 t},$$

where the b_n's are arbitrary constants. Combining this with equation (13), we obtain the functions

$$u_n(x, t) = X_n(x)T_n(t) = \left[c_n \cos\frac{n\pi x}{L}\right]\left[b_n e^{-\beta(n\pi/L)^2 t}\right],$$

$$u_n(x, t) = a_n e^{-\beta(n\pi/L)^2 t} \cos\frac{n\pi x}{L},$$

where $a_n = b_n c_n$ is, again, an arbitrary constant.

If we take an infinite series of these functions, we obtain

$$u(x, t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n e^{-\beta(n\pi/L)^2 t} \cos\frac{n\pi x}{L}, \tag{14}$$

which will be a solution to (6)–(7) provided the series has the proper convergence behavior. Notice that in (14) we have altered the constant term and written it as $a_0/2$, thus producing the standard form for cosine expansions.

Assuming a solution to (6)–(7) is given by the series in (14) and substituting into the initial condition (8), we get

$$u(x, 0) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{L} = f(x), \quad 0 < x < L . \tag{15}$$

This means that if we choose the a_n's as the coefficients in the Fourier cosine series for f,

$$a_n = \frac{2}{L}\int_0^L f(x)\cos\frac{n\pi x}{L}dx\,, \quad n = 0, 1, 2, \ldots\,,$$

then $u(x, t)$ given in (14) will be a **formal solution** to the heat flow problem (6)–(8). Again, if this expansion converges to a continuous function with continuous second partial derivatives, then the formal solution is an actual solution. ◆

Note that the individual terms in the series (4) and (14) are genuine solutions to the heat equation and the associated boundary conditions (but not the initial condition, of course). These solutions are called the **modes** of the system. Observe that the time factor $e^{-\beta(n\pi/L)^2 t}$ decays faster for the more oscillatory (higher n) modes. This makes sense physically; a temperature pattern of many alternating hot and cold strips arranged closely together will equilibrate faster than a more uniform pattern. See Figure 10.16.

A good way of expressing how separation of variables "works" is as follows: The initial temperature profile $f(x)$ is decomposed into modes through Fourier analysis—equations (5) or (15). Then each mode is matched with a time decay factor—equations (4) or (14)—and it evolves in time accordingly.

When the ends of the wire are kept at 0°C or when the ends are insulated, the boundary conditions are said to be **homogeneous.** But when the ends of the wire are kept at constant temperatures different from zero, that is,

$$u(0, t) = U_1 \quad \text{and} \quad u(L, t) = U_2\,, \qquad t > 0\,, \tag{16}$$

then the boundary conditions are called **nonhomogeneous.**

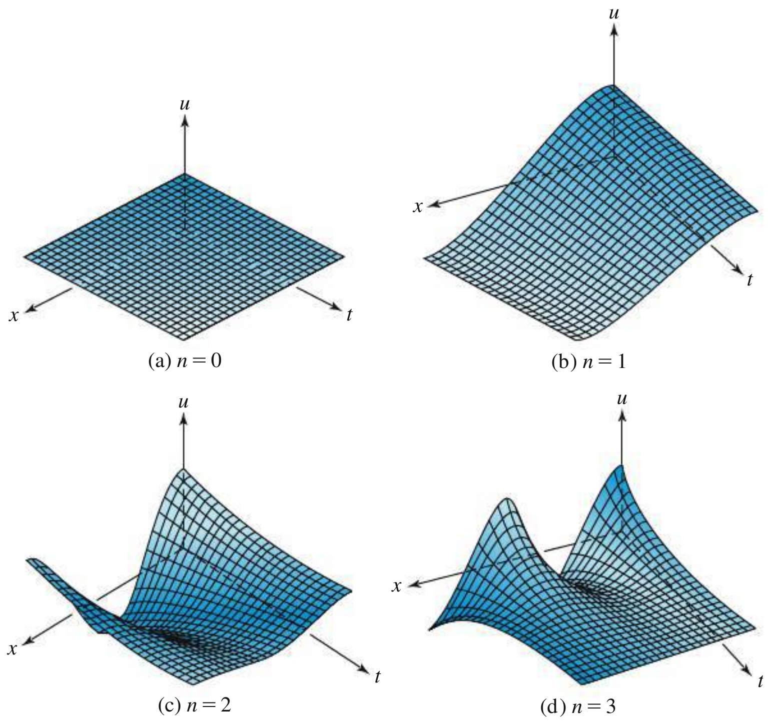


Figure 10.16 Modes for equation (14)

From our experience with vibration problems in Section 4.9 we expect that the solution to the heat flow problem with nonhomogeneous boundary conditions will consist of a **steady-state solution** $v(x)$ that satisfies the nonhomogeneous boundary conditions in (16) plus a **transient solution** $w(x, t)$. That is,

$$\boldsymbol{u(x,t) = v(x) + w(x,t)} \,,$$

where $w(x, t)$ and its partial derivatives tend to zero as $t \to \infty$. The function $w(x, t)$ will then satisfy homogeneous boundary conditions, as illustrated in the next example.

Example 2 Find a formal solution to the heat flow problem governed by the initial-boundary value problem

$$\frac{\partial u}{\partial t} = \beta \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \qquad t > 0, \tag{17}$$

$$u(0, t) = U_1, \quad u(L, t) = U_2, \qquad t > 0, \tag{18}$$

$$u(x, 0) = f(x), \quad 0 < x < L. \tag{19}$$

Solution Let's assume that the solution $u(x, t)$ consists of a steady-state solution $v(x)$ and a transient solution $w(x, t)$—that is,

$$u(x, t) = v(x) + w(x, t). \tag{20}$$

Substituting for $u(x, t)$ in equations (17)–(19) leads to

$$\frac{\partial u}{\partial t} = \frac{\partial w}{\partial t} = \beta v''(x) + \beta \frac{\partial^2 w}{\partial x^2}, \quad 0 < x < L, \qquad t > 0, \tag{21}$$

$$v(0) + w(0, t) = U_1, \quad v(L) + w(L, t) = U_2, \qquad t > 0, \tag{22}$$

$$v(x) + w(x, 0) = f(x), \quad 0 < x < L. \tag{23}$$

If we allow $t \to \infty$ in (21)–(22), assuming that $w(x, t)$ is a transient solution, we obtain the steady-state boundary value problem

$$v''(x) = 0, \quad 0 < x < L,$$
$$v(0) = U_1, \quad v(L) = U_2.$$

Solving for v, we obtain $v(x) = Ax + B$, and choosing A and B so that the boundary conditions are satisfied yields

$$v(x) = U_1 + \frac{(U_2 - U_1)x}{L} \tag{24}$$

as the steady-state solution.

With this choice for $v(x)$, the initial-boundary value problem (21)–(23) reduces to the following initial-boundary value problem for $w(x, t)$:

$$\frac{\partial w}{\partial t} = \beta \frac{\partial^2 w}{\partial x^2}, \quad 0 < x < L, \qquad t > 0, \tag{25}$$

$$w(0, t) = w(L, t) = 0, \qquad t > 0, \tag{26}$$

$$w(x, 0) = f(x) - U_1 - \frac{(U_2 - U_1)x}{L}, \qquad 0 < x < L. \tag{27}$$

Recall that a formal solution to (25)–(27) is given by equation (4). Hence,

$$w(x,t) = \sum_{n=1}^{\infty} c_n e^{-\beta(n\pi/L)^2 t} \sin\frac{n\pi x}{L},$$

where the c_n's are the coefficients of the Fourier sine series expansion

$$f(x) - U_1 - \frac{(U_2 - U_1)x}{L} = \sum_{n=1}^{\infty} c_n \sin\frac{n\pi x}{L}.$$

Therefore, the formal solution to (17)–(19) is

$$u(x,t) = U_1 + \frac{(U_2 - U_1)x}{L} + \sum_{n=1}^{\infty} c_n e^{-\beta(n\pi/L)^2 t} \sin\frac{n\pi x}{L}, \tag{28}$$

with

$$c_n = \frac{2}{L}\int_0^L \left[f(x) - U_1 - \frac{(U_2 - U_1)x}{L}\right] \sin\frac{n\pi x}{L}\,dx\,. \quad \blacklozenge$$

In the next example, we consider the heat flow problem when a heat source is present but is independent of time. (Recall the derivation in Section 10.1.)

Example 3 Find a formal solution to the heat flow problem governed by the initial-boundary value problem

$$\frac{\partial u}{\partial t} = \beta\frac{\partial^2 u}{\partial x^2} + P(x)\,, \qquad 0 < x < L\,, \qquad t > 0\,, \tag{29}$$

$$u(0,t) = U_1\,, \qquad u(L,t) = U_2\,, \qquad t > 0\,, \tag{30}$$

$$u(x,0) = f(x)\,, \qquad 0 < x < L\,. \tag{31}$$

Solution We begin by assuming that the solution consists of a steady-state solution $v(x)$ and a transient solution $w(x,t)$, namely,

$$u(x,t) = v(x) + w(x,t)\,,$$

where $w(x,t)$ and its partial derivatives tend to zero as $t \to \infty$. Substituting for $u(x,t)$ in (29)–(31) yields

$$\frac{\partial u}{\partial t} = \frac{\partial w}{\partial t} = \beta v''(x) + \beta\frac{\partial^2 w}{\partial x^2} + P(x)\,, \qquad 0 < x < L\,, \qquad t > 0\,, \tag{32}$$

$$v(0) + w(0,t) = U_1\,, \qquad v(L) + w(L,t) = U_2\,, \qquad t > 0\,, \tag{33}$$

$$v(x) + w(x,0) = f(x)\,, \qquad 0 < x < L\,. \tag{34}$$

Letting $t \to \infty$ in (32)–(33), we obtain the steady-state boundary value problem

$$v''(x) = -\frac{1}{\beta}P(x)\,, \qquad 0 < x < L\,,$$

$$v(0) = U_1\,, \qquad v(L) = U_2\,.$$

The solution to this boundary value problem can be obtained by two integrations using the boundary conditions to determine the constants of integration. The reader can verify that the solution is given by the formula

$$v(x) = \left[U_2 - U_1 + \int_0^L \left(\int_0^z \frac{1}{\beta}P(s)\,ds\right)dz\right]\frac{x}{L} + U_1 - \int_0^x \left(\int_0^z \frac{1}{\beta}P(s)\,ds\right)dz\,. \tag{35}$$

With this choice for $v(x)$, we find that the initial-boundary value problem (32)–(34) reduces to the following initial-boundary value problem for $w(x, t)$:

$$\frac{\partial w}{\partial t} = \beta\frac{\partial^2 w}{\partial x^2}\,, \qquad 0 < x < L\,, \qquad t > 0\,, \tag{36}$$

$$w(0, t) = w(L, t) = 0\,, \qquad t > 0\,, \tag{37}$$

$$w(x, 0) = f(x) - v(x)\,, \qquad 0 < x < L\,, \tag{38}$$

where $v(x)$ is given by formula (35). As before, the solution to this initial-boundary value problem is

$$w(x, t) = \sum_{n=1}^{\infty} c_n e^{-\beta(n\pi/L)^2 t}\sin\frac{n\pi x}{L}\,, \tag{39}$$

where the c_n's are determined from the Fourier sine series expansion of $f(x) - v(x)$:

$$f(x) - v(x) = \sum_{n=1}^{\infty} c_n \sin\frac{n\pi x}{L}\,. \tag{40}$$

Thus the formal solution to (29)–(31) is given by

$$u(x, t) = v(x) + w(x, t)\,,$$

where $v(x)$ is given in (35) and $w(x, t)$ is prescribed by (39)–(40). ◆

The method of separation of variables is also applicable to problems in higher dimensions. For example, consider the problem of heat flow in a rectangular plate with sides $x = 0$, $x = L$, $y = 0$, and $y = W$. If the two sides $y = 0$, $y = W$ are kept at a constant temperature of 0°C and the two sides $x = 0$, $x = L$ are perfectly insulated, then heat flow is governed by the initial-boundary value problem in the following example (see Figure 10.17).

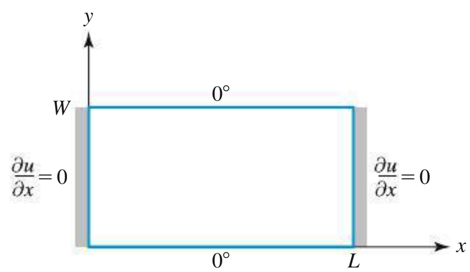


Figure 10.17 Plate with insulated sides

Example 4 Find a formal solution $u(x, y, t)$ to the initial-boundary value problem

(41) $$\frac{\partial u}{\partial t} = \beta\left\{\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right\}, \qquad 0 < x < L, \qquad 0 < y < W, \qquad t > 0,$$

(42) $$\frac{\partial u}{\partial x}(0, y, t) = \frac{\partial u}{\partial x}(L, y, t) = 0, \qquad 0 < y < W, \qquad t > 0,$$

(43) $$u(x, 0, t) = u(x, W, t) = 0, \qquad 0 < x < L, \qquad t > 0,$$

(44) $$u(x, y, 0) = f(x, y), \qquad 0 < x < L, \qquad 0 < y < W.$$

Solution If we assume a solution of the form $u(x, y, t) = V(x, y)T(t)$, then equation (41) separates into the two equations

(45) $$T'(t) + \beta\lambda T(t) = 0,$$

(46) $$\frac{\partial^2 V}{\partial x^2}(x, y) + \frac{\partial^2 V}{\partial y^2}(x, y) + \lambda V(x, y) = 0,$$

where λ can be any constant. To solve equation (46), we again use separation of variables. Here we assume $V(x, y) = X(x)Y(y)$. This allows us to separate equation (46) into the two equations

(47) $$X''(x) + \mu X(x) = 0,$$

(48) $$Y''(y) + (\lambda - \mu)Y(y) = 0,$$

where μ can be any constant (see Problem 29 in Exercises 10.2 on page 571).

To solve for $X(x)$, we observe that the boundary conditions in (42), in terms of the separated variables, become

$$X'(0)Y(y)T(t) = X'(L)Y(y)T(t) = 0; \qquad 0 < y < W, \qquad t > 0.$$

Hence, in order to get a nontrivial solution, we must have

(49) $$X'(0) = X'(L) = 0.$$

The boundary value problem for X given in equations (47) and (49) was solved in Example 1 [compare equations (12) and (13)]. Here $\mu = (m\pi/L)^2$, $m = 0, 1, 2, \ldots$, and

$$X_m(x) = c_m \cos\frac{m\pi x}{L},$$

where the c_m's are arbitrary. (We forgo the "$c_0/2$" convention here.)

To solve for $Y(y)$, we first observe that the boundary conditions in (43) become

(50) $$Y(0) = Y(W) = 0.$$

Next, substituting $\mu = (m\pi/L)^2$ into equation (48) yields

$$Y''(y) + \left(\lambda - (m\pi/L)^2\right)Y(y) = 0,$$

which we can rewrite as

(51) $$Y''(y) + EY(y) = 0,$$

where $E = \lambda - (m\pi/L)^2$. The boundary value problem for Y consisting of (50)–(51) has also been solved before. In Section 10.2 [compare equations (7) and (9), pages 564 and 565] we showed that $E = (n\pi/W)^2$, $n = 1, 2, 3, \ldots$, and the nontrivial solutions are given by

$$Y_n(y) = a_n \sin\frac{n\pi y}{W},$$

where the a_n's are arbitrary.

Since $\lambda = E + (m\pi/L)^2$, we have

$$\lambda = (n\pi/W)^2 + (m\pi/L)^2 ; \qquad m = 0, 1, 2, \dots , \qquad n = 1, 2, 3, \dots .$$

Substituting λ into equation (45), we can solve for $T(t)$ and obtain

$$T_{mn}(t) = b_{mn} e^{-(m^2/L^2+n^2/W^2)\beta\pi^2 t} .$$

Substituting in for X_m, Y_n, and T_{mn}, we get

$$u_{mn}(x, y, t) = \left(c_m \cos\frac{m\pi x}{L}\right)\left(a_n \sin\frac{n\pi y}{W}\right)\left(b_{mn} e^{-(m^2/L^2+n^2/W^2)\beta\pi^2 t}\right) ,$$

$$u_{mn}(x, y, t) = a_{mn} e^{-(m^2/L^2+n^2/W^2)\beta\pi^2 t} \cos\frac{m\pi x}{L} \sin\frac{n\pi y}{W} ,$$

where $a_{mn} \coloneqq a_n b_{mn} c_m$ $(m = 0, 1, 2, \dots, n = 1, 2, 3, \ \dots)$ are arbitrary constants.

If we now take a doubly infinite series of such functions, then we obtain the formal series

(52) $$u(x, y, t) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} a_{mn} e^{-(m^2/L^2+n^2/W^2)\beta\pi^2 t} \cos\frac{m\pi x}{L} \sin\frac{n\pi y}{W} .$$

We are now ready to apply the initial conditions (44). Setting $t = 0$, we obtain

(53) $$u(x, y, 0) = f(x, y) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} a_{mn} \cos\frac{m\pi x}{L} \sin\frac{n\pi y}{W} .$$

This is a **double Fourier series.**[†] The formulas for the coefficients a_{mn} are obtained by exploiting the orthogonality conditions twice. Presuming (53) is valid and permits term-by-term integration, we multiply each side by $\cos(p\pi x/L)\sin(q\pi y/W)$ and integrate over x and y:

$$\int_0^L \int_0^W f(x, y) \cos\frac{p\pi x}{L} \sin\frac{q\pi y}{W} dy\, dx$$
$$= \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} a_{mn} \int_0^L \int_0^W \cos\frac{m\pi x}{L} \sin\frac{n\pi y}{W} \cos\frac{p\pi x}{L} \sin\frac{q\pi y}{W} dy\, dx .$$

According to the orthogonality conditions, each integral on the right is zero, except when $m = p$ and $n = q$; thus,

$$\int_0^L \int_0^W f(x, y) \cos\frac{p\pi x}{L} \sin\frac{q\pi y}{W} dy\, dx$$
$$= a_{pq} \int_0^L \cos^2\frac{p\pi x}{L} dx \int_0^W \sin^2\frac{q\pi y}{W} dy = \begin{cases} \dfrac{LW}{4} a_{pq} , & p \neq 0 , \\ \dfrac{LW}{2} a_{pq} , & p = 0 . \end{cases}$$

Hence,

(54) $$a_{0q} = \frac{2}{LW} \int_0^L \int_0^W f(x, y) \sin\frac{q\pi y}{W} dy\, dx , \qquad q = 1, 2, 3, \dots ,$$

[†]For a discussion of double Fourier series, see *Partial Differential Equations of Mathematical Physics,* 2nd ed. by Tyn Myint-U (Elsevier North Holland, New York, 1983), Sec. 5.14.

and for $p \geq 1$, $q \geq 1$,

$$a_{pq} = \frac{4}{LW} \int_0^L \int_0^W f(x, y) \cos \frac{p\pi x}{L} \sin \frac{q\pi y}{W} \, dy \, dx \,. \tag{55}$$

Finally, the solution to the initial-boundary value problem (41)–(44) is given by equation (52), where the coefficients are prescribed by equations (54) and (55). ◆

We derived formulas (54) and (55) under the assumption that the double series (53) truly converged (uniformly). How can we justify the assumption that a double Fourier series converges? A rough argument goes as follows. If we *fix* y in $f(x, y)$, then $f(x, y)$ presumably has a convergent cosine series in x: $f(x, y) = \sum_{m=0}^{\infty} d_m \cos(m\pi x/L)$. But the coefficients d_m depend on the value of the y that we fixed: $d_m = d_m(y)$. So presumably each *function* $d_m(y)$ has a convergent sine series in y: $d_m(y) = \sum_{n=1}^{\infty} a_{mn} \sin(n\pi y/W)$. Assembling all this, we get $f(x, y) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} a_{mn} \cos(m\pi x/L) \sin(n\pi y/W)$.

Existence and Uniqueness of Solutions

In the examples that we have studied in this section and in Section 10.2, we were able to obtain formal solutions in the sense that we could express the solution in terms of a series expansion consisting of exponentials, sines, and cosines. To prove that these series converge to actual solutions requires results on the convergence of Fourier series and results from real analysis on uniform convergence. We will not go into these details here, but refer the reader to Section 6.5 of the text by Tyn Myint-U (see footnote on page 600) for a proof of the existence of a solution to the heat flow problem discussed in Sections 10.1 and 10.2. (A proof of uniqueness is also given there.)

As might be expected, by using Fourier series and the method of separation of variables one can also obtain "solutions" when the initial data are discontinuous, since the formal solutions require only the existence of a convergent Fourier series. This allows one to study idealized problems in which the initial conditions do not agree with the boundary conditions or the initial conditions involve a jump discontinuity. For example, we may assume that initially one half of the wire is at one temperature, whereas the other half is at a different temperature, that is,

$$f(x) = \begin{cases} U_1 \,, & 0 < x < L/2 \,, \\ U_2 \,, & L/2 < x < L \,. \end{cases}$$

The formal solution that we obtain will make sense for $0 < x < L$, $t > 0$, but near the points of discontinuity $x = 0$, $L/2$, and L, we have to expect behavior like that exhibited in Figure 10.7 of Section 10.3, page 578.

The question of the uniqueness of the solution to the heat flow problem can be answered in various ways. One is tempted to argue that the method of separation of variables yields *formulas* for the solutions and therefore a unique solution. However, this does *not* exclude the possibility of solutions existing that cannot be obtained by the method of separation of variables.

From physical considerations, we know that the *peak* temperature along a wire does not increase spontaneously if there are no heat sources to drive it up. Indeed, if hot objects could draw heat from colder objects without external intervention, we would have violations of the second law of thermodynamics. Thus no point on an unheated wire will ever reach a higher temperature than the initial peak temperature. Such statements are called **maximum principles** and can be proved mathematically for the heat equation and Laplace's equation. One such result is the following.†

Maximum Principle for the Heat Equation

Theorem 6. Let $u(x, t)$ be a continuously differentiable function that satisfies the heat equation

(56) $$\frac{\partial u}{\partial t} = \beta \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \qquad t > 0,$$

and the boundary conditions

(57) $$u(0, t) = u(L, t) = 0, \qquad t \geq 0.$$

Then $u(x, t)$ attains its maximum value at $t = 0$, for some x in $[0, L]$—that is,

$$\max_{\substack{t \geq 0 \\ 0 \leq x \leq L}} u(x, t) = \max_{0 \leq x \leq L} u(x, 0).$$

We can use the maximum principle to show that the heat flow problem has a unique solution.

Uniqueness of Solution

Theorem 7. The initial-boundary value problem

(58) $$\frac{\partial u}{\partial t} = \beta \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \qquad t > 0,$$

(59) $$u(0, t) = u(L, t) = 0, \qquad t \geq 0,$$

(60) $$u(x, 0) = f(x), \qquad 0 < x < L,$$

has at most one continuously differentiable solution.

Proof. Assume $u(x, t)$ and $v(x, t)$ are continuously differentiable functions that satisfy the initial-boundary value problem (58)–(60). Let $w = u - v$. Now w is a continuously differentiable solution to the boundary value problem (56)–(57). By the maximum principle, w must attain its maximum at $t = 0$, and since

$$w(x, 0) = u(x, 0) - v(x, 0) = f(x) - f(x) = 0,$$

we have $w(x, t) \leq 0$. Hence, $u(x, t) \leq v(x, t)$ for all $0 \leq x \leq L, t \geq 0$. A similar argument using $\hat{w} = v - u$ yields $v(x, t) \leq u(x, t)$. Therefore we have $u(x, t) = v(x, t)$ for all $0 \leq x \leq L$, $t \geq 0$. Thus, there is at most one continuously differentiable solution to the problem (58)–(60). ◆

†For a discussion of maximum principles and their applications, see *Maximum Principles in Differential Equations,* by M. H. Protter and H. F. Weinberger (Springer-Verlag, New York, 1984).

10.5 EXERCISES

In Problems 1–10, find a formal solution to the given initial-boundary value problem.

1. $\dfrac{\partial u}{\partial t} = 5\dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < 1\,, \quad t > 0\,,$
$u(0, t) = u(1, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = (1 - x)x^2\,, \quad 0 < x < 1$

2. $\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$u(0, t) = u(\pi, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = x^2\,, \quad 0 < x < \pi$

3. $\dfrac{\partial u}{\partial t} = 3\dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$\dfrac{\partial u}{\partial x}(0, t) = \dfrac{\partial u}{\partial x}(\pi, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = x\,, \quad 0 < x < \pi$

4. $\dfrac{\partial u}{\partial t} = 2\dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < 1\,, \quad t > 0\,,$
$\dfrac{\partial u}{\partial x}(0, t) = \dfrac{\partial u}{\partial x}(1, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = x(1 - x)\,, \quad 0 < x < 1$

5. $\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$\dfrac{\partial u}{\partial x}(0, t) = \dfrac{\partial u}{\partial x}(\pi, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = e^x\,, \quad 0 < x < \pi$

6. $\dfrac{\partial u}{\partial t} = 7\dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$\dfrac{\partial u}{\partial x}(0, t) = \dfrac{\partial u}{\partial x}(\pi, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = 1 - \sin x\,, \quad 0 < x < \pi$

7. $\dfrac{\partial u}{\partial t} = 2\dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$u(0, t) = 5\,, \quad u(\pi, t) = 10\,, \quad t > 0\,,$
$u(x, 0) = \sin 3x - \sin 5x\,, \quad 0 < x < \pi$

8. $\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$u(0, t) = 0\,, \quad u(\pi, t) = 3\pi\,, \quad t > 0\,,$
$u(x, 0) = 0\,, \quad 0 < x < \pi$

9. $\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2} + e^{-x}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$u(0, t) = u(\pi, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = \sin 2x\,, \quad 0 < x < \pi$

10. $\dfrac{\partial u}{\partial t} = 3\dfrac{\partial^2 u}{\partial x^2} + x\,, \quad 0 < x < \pi\,, \quad t > 0\,,$
$u(0, t) = u(\pi, t) = 0\,, \quad t > 0\,,$
$u(x, 0) = \sin x\,, \quad 0 < x < \pi$

11. Find a formal solution to the initial-boundary value problem

$$\frac{\partial u}{\partial t} = 4\frac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$$
$$\frac{\partial u}{\partial x}(0, t) = 0\,, \quad u(\pi, t) = 0\,, \quad t > 0\,,$$
$$u(x, 0) = f(x)\,, \quad 0 < x < \pi\,.$$

12. Find a formal solution to the initial-boundary value problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}\,, \quad 0 < x < \pi\,, \quad t > 0\,,$$
$$u(0, t) = 0\,, \quad u(\pi, t) + \frac{\partial u}{\partial x}(\pi, t) = 0\,, \quad t > 0\,,$$
$$u(x, 0) = f(x)\,, \quad 0 < x < \pi\,.$$

13. Find a formal solution to the initial-boundary value problem

$$\frac{\partial u}{\partial t} = 2\frac{\partial^2 u}{\partial x^2} + 4x\,, \quad 0 < x < \pi\,, \quad t > 0\,,$$
$$u(0, t) = u(\pi, t) = 0\,, \quad t > 0\,,$$
$$u(x, 0) = \sin x\,, \quad 0 < x < \pi\,.$$

14. Find a formal solution to the initial-boundary value problem

$$\frac{\partial u}{\partial t} = 3\frac{\partial^2 u}{\partial x^2} + 5\,, \quad 0 < x < \pi\,, \quad t > 0\,,$$
$$u(0, t) = u(\pi, t) = 1\,, \quad t > 0\,,$$
$$u(x, 0) = 1\,, \quad 0 < x < \pi\,.$$

In Problems 15–18, find a formal solution to the initial-boundary value problem.

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\,, \quad 0 < x < \pi\,, \quad 0 < y < \pi\,, \quad t > 0\,,$$
$$\frac{\partial u}{\partial x}(0, y, t) = \frac{\partial u}{\partial x}(\pi, y, t) = 0\,, \quad 0 < y < \pi\,, \quad t > 0\,,$$
$$u(x, 0, t) = u(x, \pi, t) = 0\,, \quad 0 < x < \pi\,, \quad t > 0\,,$$
$$u(x, y, 0) = f(x, y)\,, \quad 0 < x < \pi\,, \quad 0 < y < \pi\,,$$

for the given function $f(x, y)$.

15. $f(x, y) = \cos 6x \sin 4y - 3 \cos x \sin 11y$

16. $f(x, y) = \cos x \sin y + 4 \cos 2x \sin y - 3 \cos 3x \sin 4y$

17. $f(x, y) = y$

18. $f(x, y) = x \sin y$

19. Chemical Diffusion. Chemical diffusion through a thin layer is governed by the equation

$$\frac{\partial C}{\partial t} = k\frac{\partial^2 C}{\partial x^2} - LC ,$$

where $C(x, t)$ is the concentration in moles/cm^3, the diffusivity k is a positive constant with units cm^2/sec, and $L > 0$ is a consumption rate with units sec^{-1}. Assume the boundary conditions are

$$C(0, t) = C(a, t) = 0 , \qquad t > 0 ,$$

and the initial concentration is given by

$$C(x, 0) = f(x) , \qquad 0 < x < a .$$

Use the method of separation of variables to solve formally for the concentration $C(x, t)$. What happens to the concentration as $t \to +\infty$?

10.6 The Wave Equation

In Section 10.2 we presented a model for the motion of a vibrating string. If $u(x, t)$ represents the displacement (deflection) of the string and the ends of the string are held fixed, then the motion of the string is governed by the initial-boundary value problem

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2\frac{\partial^2 u}{\partial x^2} , \qquad 0 < x < L , \qquad t > 0 , \tag{1}$$

$$u(0, t) = u(L, t) = 0 , \qquad t > 0 , \tag{2}$$

$$u(x, 0) = f(x) , \qquad 0 < x < L , \tag{3}$$

$$\frac{\partial u}{\partial t}(x, 0) = g(x) , \qquad 0 < x < L . \tag{4}$$

Equation (1) is called the **wave equation.**

The constant α^2 appearing in (1) is strictly positive and depends on the linear density and tension of the string. The boundary conditions in (2) reflect the fact that the string is held fixed at the two endpoints $x = 0$ and $x = L$.

Equations (3) and (4) specify, respectively, the initial displacement and the initial velocity of each point on the string. For the initial and boundary conditions to be consistent, we assume $f(0) = f(L) = 0$ and $g(0) = g(L) = 0$.

Using the method of separation of variables, we found in Section 10.2 that a formal solution to (1)–(4) is given by [compare equations (24)–(26) on page 569]

$$u(x, t) = \sum_{n=1}^{\infty}\left[a_n \cos\frac{n\pi\alpha}{L}t + b_n \sin\frac{n\pi\alpha}{L}t\right]\sin\frac{n\pi x}{L} , \tag{5}$$

where the a_n's and b_n's are determined from the Fourier sine series

$$f(x) = \sum_{n=1}^{\infty} a_n \sin\frac{n\pi x}{L} , \tag{6}$$

$$g(x) = \sum_{n=1}^{\infty} b_n\left(\frac{n\pi\alpha}{L}\right)\sin\frac{n\pi x}{L} . \tag{7}$$

See Figure 10.18 on page 605 for a sketch of partial sums for (5).

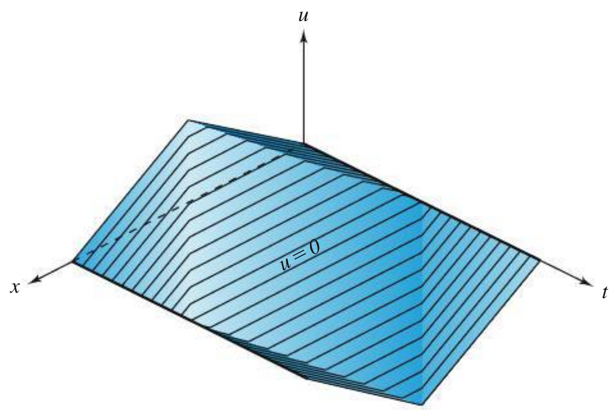


Figure 10.18 Partial sums for equation (5) with $f(x) = \begin{cases} x, & 0 < x \le \pi/2, \\ \pi - x, & \pi/2 \le x < \pi, \end{cases}$ and $g(x) = 0$

Each term (or mode) in expansion (5) can be viewed as a **standing wave** (a wave that vibrates in place without lateral motion along the string). For example, the first term,

$$\left(a_1 \cos\frac{\pi\alpha}{L}t + b_1 \sin\frac{\pi\alpha}{L}t\right) \sin\frac{\pi x}{L},$$

consists of a sinusoidal shape function $\sin(\pi x/L)$ multiplied by a time-varying amplitude. The second term is also a sinusoid $\sin(2\pi x/L)$ with a time-varying amplitude. In the latter case, there is a *node* in the middle at $x = L/2$ that never moves. For the nth term, we have a sinusoid $\sin(n\pi x/L)$ with a time-varying amplitude and $(n-1)$ nodes. This is illustrated in Figure 10.19. Thus, separation of variables decomposes the initial data into sinusoids or modes [equations (6)–(7)], assigns each mode a frequency at which to vibrate [equation (5)], and represents the solution as the superposition of infinitely many standing waves. Modes with more nodes vibrate at higher frequencies. By choosing the initial configuration of the string to have the same shape as one of the individual terms in the solution, we can activate only that mode.

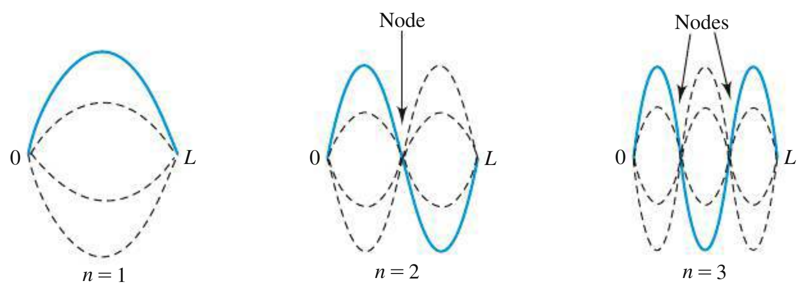


Figure 10.19 Standing waves. Time-varying amplitudes are shown by dashed curves

The different modes of vibration of a guitar string are distinguishable to the human ear by the frequency of the sound they generate; different frequencies are discerned as different "pitches." The *fundamental* is the frequency of the lowest mode, and integer multiples of the fundamental frequency are called "harmonics." A cello and a trombone sound different even if they are playing the same fundamental, say F$^{\#}$, because of the difference in the intensities of the harmonics they generate.

Many engineering devices operate better in some modes than in others. For example, certain modes propagate down a waveguide or an optical fiber with less attenuation than the others. In such cases, engineers design the systems to suppress the undesirable modes, either by shaping the initial excitation or by introducing devices (like resistor cards in a waveguide) which damp out certain modes preferentially.

The fundamental mode on a guitar string can be suppressed by holding the finger lightly in contact with the midpoint, thereby creating a stationary point or node there. According to equation (5), the pitch of the note then doubles, thus producing an "octave." This style of fingering, called "playing harmonics," is used frequently in musical performance.

An important characteristic of an engineering device is the set of angular frequencies supported by its eigenmodes—its "eigenfrequencies." According to equation (5), the eigenfrequencies of the vibrating string are the harmonics $\omega_n = n\pi\alpha/L$, $n = 1, 2, 3, \ldots$. These blend together pleasantly to the human ear, and we enjoy melodies played on string instruments. The eigenfrequencies of a drum are not harmonics (see Problem 21). Therefore, drums are used for rhythm, not for melody.

As we have seen in the preceding section, the method of separation of variables can be used to solve problems with nonhomogeneous boundary conditions and nonhomogeneous equations where the forcing term is time independent. In the next example, we will consider a problem with a time-dependent forcing term $h(x, t)$.

Example 1 For given functions f, g, and h, find a formal solution to the initial-boundary value problem

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2} + h(x,t)\,, \quad 0 < x < L\,, \qquad t > 0\,, \tag{8}$$

$$u(0,t) = u(L,t) = 0\,, \quad t > 0\,, \tag{9}$$

$$u(x,0) = f(x)\,, \quad 0 < x < L\,, \tag{10}$$

$$\frac{\partial u}{\partial t}(x,0) = g(x)\,, \quad 0 < x < L\,. \tag{11}$$

Solution The boundary conditions in (9) certainly require that the solution be zero for $x = 0$ and $x = L$. Motivated by the fact that the solution to the corresponding homogeneous system (1)–(4) consists of a superposition of standing waves, let's try to find a solution to (8)–(11) of the form

$$u(x,t) = \sum_{n=1}^{\infty} u_n(t) \sin\frac{n\pi x}{L}\,, \tag{12}$$

where the $u_n(t)$'s are functions of t to be determined.

For each fixed t, we can compute a Fourier sine series for $h(x, t)$. If we assume that the series is convergent to $h(x, t)$, then

$$h(x,t) = \sum_{n=1}^{\infty} h_n(t) \sin\frac{n\pi x}{L}\,, \tag{13}$$

where the coefficient $h_n(t)$ is given by [recall equation (7) on page 589]

$$h_n(t) = \frac{2}{L}\int_0^L h(x,t)\sin\frac{n\pi x}{L}dx\,, \qquad n = 1, 2, \ldots\,.$$

If the series in (13) has the proper convergence properties, then we can substitute (12) and (13) into equation (8) and obtain

$$\sum_{n=1}^{\infty}\left[u_n''(t) + \left(\frac{n\pi\alpha}{L}\right)^2 u_n(t)\right]\sin\frac{n\pi x}{L} = \sum_{n=1}^{\infty} h_n(t)\sin\frac{n\pi x}{L}\,.$$

Equating the coefficients in each series (why?), we have

$$u_n''(t) + \left(\frac{n\pi\alpha}{L}\right)^2 u_n(t) = h_n(t)\ .$$

This is a nonhomogeneous, constant-coefficient equation that can be solved using variation of parameters. You should verify that

$$u_n(t) = a_n \cos\frac{n\pi\alpha}{L}t + b_n \sin\frac{n\pi\alpha}{L}t + \frac{L}{n\pi\alpha}\int_0^t h_n(s)\sin\left[\frac{n\pi\alpha}{L}(t-s)\right]ds$$

[compare Problem 20 of Exercises 4.6]. Hence, with this choice of $u_n(t)$, the series in (12) is a formal solution to the partial differential equation (8).

Since[†]

$$u_n(0) = a_n \quad \text{and} \quad u_n'(0) = b_n\left(\frac{n\pi\alpha}{L}\right),$$

substituting (12) into the initial conditions (10)–(11) yields

(14) $$u(x,0) = f(x) = \sum_{n=1}^{\infty} a_n \sin\frac{n\pi x}{L},$$

(15) $$\frac{\partial u}{\partial t}(x,0) = g(x) = \sum_{n=1}^{\infty} b_n\left(\frac{n\pi\alpha}{L}\right)\sin\frac{n\pi x}{L}.$$

Thus, if we choose the a_n's and b_n's so that equations (14) and (15) are satisfied, a formal solution to (8)–(11) is given by

(16) $$u(x,t) = \sum_{n=1}^{\infty}\left\{a_n \cos\frac{n\pi\alpha}{L}t + b_n \sin\frac{n\pi\alpha}{L}t + \frac{L}{n\pi\alpha}\int_0^t h_n(s)\sin\left[\frac{n\pi\alpha}{L}(t-s)\right]ds\right\}\sin\frac{n\pi x}{L}.$$ ◆

The method of separation of variables can also be used to solve initial-boundary value problems for the wave equation in higher dimensions. For example, a vibrating rectangular membrane of length L and width W (see Figure 10.20 on page 608) is governed by the following initial-boundary value problem for $u(x, y, t)$:

(17) $$\frac{\partial^2 u}{\partial t^2} = \alpha^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \quad 0 < x < L, \quad 0 < y < W, \quad t > 0,$$

(18) $$u(0,y,t) = u(L,y,t) = 0, \quad 0 < y < W, \quad t > 0,$$

(19) $$u(x,0,t) = u(x,W,t) = 0, \quad 0 < x < L, \quad t > 0,$$

(20) $$u(x,y,0) = f(x,y), \quad 0 < x < L, \quad 0 < y < W,$$

(21) $$\frac{\partial u}{\partial t}(x,y,0) = g(x,y), \quad 0 < x < L, \quad 0 < y < W.$$

[†]To compute $u_n'(0)$, we use the fact that $\dfrac{d}{dt}\displaystyle\int_0^t G(s,t)\,ds = G(t,t) + \int_0^t \frac{\partial G}{\partial t}(s,t)\,ds.$

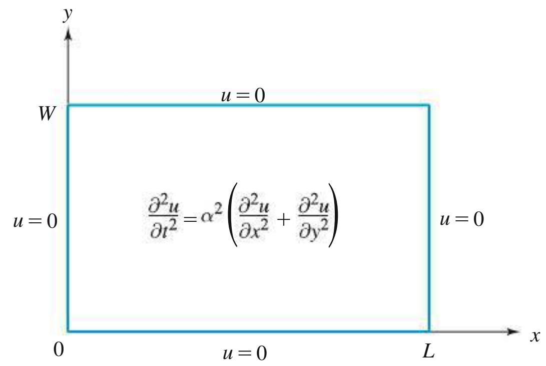


Figure 10.20 Vibrating membrane

Using an argument similar to the one given for the problem of heat flow in a rectangular plate (Example 4 in Section 10.5), we find that the initial-boundary value problem (17)–(21) has a formal solution

(22)
$$u(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \left\{ a_{mn} \cos\left(\sqrt{\frac{m^2}{L^2} + \frac{n^2}{W^2}}\, \alpha\pi t\right) + b_{mn} \sin\left(\sqrt{\frac{m^2}{L^2} + \frac{n^2}{W^2}}\, \alpha\pi t\right) \right\} \sin\frac{m\pi x}{L} \sin\frac{n\pi y}{W} ,$$

where the constants a_{mn} and b_{mn} are determined from the double Fourier series

$$f(x, y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} \sin\frac{m\pi x}{L} \sin\frac{n\pi y}{W} ,$$

$$g(x, y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \alpha\pi \sqrt{\frac{m^2}{L^2} + \frac{n^2}{W^2}}\, b_{mn} \sin\frac{m\pi x}{L} \sin\frac{n\pi y}{W} .$$

In particular,

(23)
$$a_{mn} = \frac{4}{LW} \int_0^L \int_0^W f(x, y) \sin\frac{m\pi x}{L} \sin\frac{n\pi y}{W}\, dy\, dx ,$$

(24)
$$b_{mn} = \frac{4}{LW\pi\alpha\sqrt{\frac{m^2}{L^2} + \frac{n^2}{W^2}}} \int_0^L \int_0^W g(x, y) \sin\frac{m\pi x}{L} \sin\frac{n\pi y}{W}\, dy\, dx .$$

We leave the derivation of this solution as an exercise (see Problem 19).

We mentioned earlier that the solution to the vibrating string problem (1)–(4) consisted of a superposition of standing waves. There are also "traveling waves" associated with the wave equation. Traveling waves arise naturally out of d'Alembert's solution to the wave equation for an "infinite" string.

To obtain d'Alembert's solution to the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2} ,$$

we use the change of variables

$$\psi = x + \alpha t\,, \qquad \eta = x - \alpha t\,.$$

If u has continuous second partial derivatives, then $\partial u/\partial x = \partial u/\partial\psi + \partial u/\partial\eta$ and $\partial u/\partial t = \alpha(\partial u/\partial\psi - \partial u/\partial\eta)$, from which we obtain

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial \psi^2} + 2\frac{\partial^2 u}{\partial\psi\partial\eta} + \frac{\partial^2 u}{\partial\eta^2}\,,$$

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2\left\{\frac{\partial^2 u}{\partial \psi^2} - 2\frac{\partial^2 u}{\partial\psi\partial\eta} + \frac{\partial^2 u}{\partial\eta^2}\right\}.$$

Substituting these expressions into the wave equation and simplifying yields

$$\frac{\partial^2 u}{\partial\psi\,\partial\eta} = 0\,.$$

We can solve this equation directly by first integrating with respect to ψ to obtain

$$\frac{\partial u}{\partial\eta} = b(\eta)\,,$$

where $b(\eta)$ is an arbitrary function of η, and then integrating with respect to η to find

$$u(\psi, \eta) = A(\psi) + B(\eta)\,,$$

where $A(\psi)$ and $B(\eta) = \int b(\eta)\,d\eta$ are arbitrary functions. Substituting the original variables x and t gives **d'Alembert's solution**

(25) $$\boldsymbol{u(x, t) = A(x + \alpha t) + B(x - \alpha t)\,.}$$

It is easy to check by direct substitution that $u(x, t)$, defined by formula (25), is indeed a solution to the wave equation, provided A and B are twice-differentiable functions.

Example 2 Using d'Alembert's formula (25), find a solution to the initial value problem

(26) $$\frac{\partial^2 u}{\partial t^2} = \alpha^2\frac{\partial^2 u}{\partial x^2} \qquad -\infty < x < \infty\,, \qquad t > 0\,,$$

(27) $$u(x, 0) = f(x) \qquad -\infty < x < \infty\,,$$

(28) $$\frac{\partial u}{\partial t}(x, 0) = g(x) \qquad -\infty < x < \infty\,.$$

Solution A solution to (26) is given by formula (25), so we need only choose the functions A and B so that the initial conditions (27)–(28) are satisfied. For this we need

(29) $$u(x, 0) = A(x) + B(x) = f(x)\,,$$

(30) $$\frac{\partial u}{\partial t}(x, 0) = \alpha A'(x) - \alpha B'(x) = g(x)\,.$$

Integrating equation (30) from x_0 to x (x_0 arbitrary) and dividing by α gives

(31) $$A(x) - B(x) = \frac{1}{\alpha}\int_{x_0}^{x} g(s)\,ds + C\,,$$

where C is also arbitrary. Solving the system (29) and (31), we obtain

$$A(x) = \frac{1}{2}f(x) + \frac{1}{2\alpha}\int_{x_0}^{x} g(s)\,ds + \frac{C}{2},$$

$$B(x) = \frac{1}{2}f(x) - \frac{1}{2\alpha}\int_{x_0}^{x} g(s)\,ds - \frac{C}{2}.$$

Using these functions in formula (25) gives

$$u(x,t) = \frac{1}{2}[f(x+\alpha t) + f(x-\alpha t)] + \frac{1}{2\alpha}\left[\int_{x_0}^{x+\alpha t} g(s)ds - \int_{x_0}^{x-\alpha t} g(s)ds\right],$$

which simplifies to

$$u(x,t) = \frac{1}{2}[f(x+\alpha t) + f(x-\alpha t)] + \frac{1}{2\alpha}\int_{x-\alpha t}^{x+\alpha t} g(s)\,ds\,. \quad ◆ \tag{32}$$

Example 3 Find the solution to the initial value problem

$$\frac{\partial^2 u}{\partial t^2} = 4\frac{\partial^2 u}{\partial x^2}, \qquad -\infty < x < \infty\,, \qquad t > 0\,, \tag{33}$$

$$u(x,0) = \sin x\,, \qquad -\infty < x < \infty\,, \tag{34}$$

$$\frac{\partial u}{\partial t}(x,0) = 1\,, \qquad -\infty < x < \infty\,. \tag{35}$$

Solution This is just a special case of the preceding example where $\alpha = 2, f(x) = \sin x$, and $g(x) = 1$. Substituting into (32), we obtain the solution

$$\begin{aligned} u(x,t) &= \frac{1}{2}[\sin(x+2t) + \sin(x-2t)] + \frac{1}{4}\int_{x-2t}^{x+2t} ds \\ &= \sin x \cos 2t + t\,. \quad ◆ \end{aligned} \tag{36}$$

We now use d'Alembert's formula to show that the solution to the "infinite" string problem consists of traveling waves.

Let $h(x)$ be a function defined on $(-\infty, \infty)$. The function $h(x+a)$, where $a > 0$, is a translation of the function $h(x)$ in the sense that its "shape" is the same as $h(x)$, but its position has been shifted to the left by an amount a. This is illustrated in Figure 10.21 for a function $h(x)$ whose graph consists of a triangular "bump." If we let $t \geq 0$ be a parameter (say, time), then the functions $h(x+\alpha t)$ represent a family of functions with the same shape but shifted farther and farther to the left as $t \to \infty$ (Figure 10.22 on page 611). We say that $h(x+\alpha t)$ is a **traveling wave** moving to the left with speed α. In a similar fashion, $h(x-\alpha t)$ is a traveling wave moving to the right with speed α.

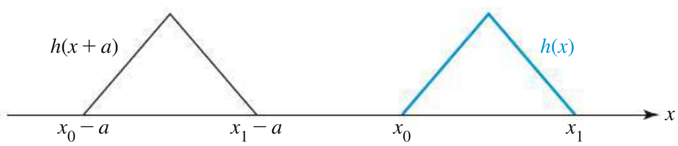


Figure 10.21 Graphs of $h(x)$ and $h(x+a)$

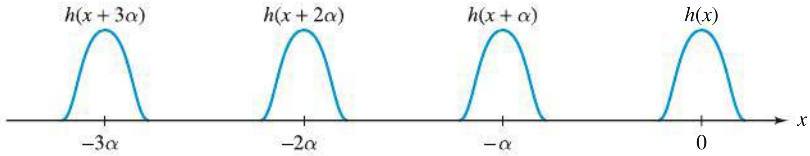


Figure 10.22 Traveling wave

If we refer to formula (25), we find that the solution to $\partial^2 u/\partial t^2 = \alpha^2 \partial^2 u/\partial x^2$ consists of traveling waves $A(x + \alpha t)$ moving to the left with speed α and $B(x - \alpha t)$ moving to the right at the same speed.

In the special case when the initial velocity $g(x) \equiv 0$, we have

$$u(x, t) = \frac{1}{2}[f(x + \alpha t) + f(x - \alpha t)] .$$

Hence, $u(x, t)$ is the sum of the traveling waves

$$\frac{1}{2}f(x + \alpha t) \quad \text{and} \quad \frac{1}{2}f(x - \alpha t) .$$

These waves are initially superimposed, since

$$u(x, 0) = \frac{1}{2}f(x) + \frac{1}{2}f(x) = f(x) .$$

As t increases, the two waves move away from each other with speed 2α. This is illustrated in Figure 10.23 for a triangular wave.

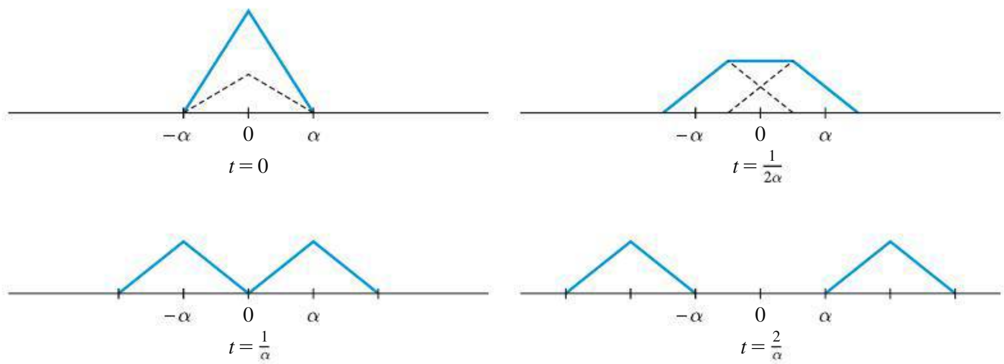


Figure 10.23 Decomposition of initial displacement into traveling waves

Example 4 Express the standing wave $\cos\left(\frac{\pi\alpha}{L}t\right)\sin\left(\frac{\pi x}{L}\right)$ as a superposition of traveling waves.

Solution By a familiar trigonometric identity,

$$\begin{aligned}\cos\left(\frac{\pi\alpha}{L}t\right)\sin\left(\frac{\pi x}{L}\right) &= \frac{1}{2}\sin\frac{\pi}{L}(x - \alpha t) + \frac{1}{2}\sin\frac{\pi}{L}(x + \alpha t) \\ &= h(x - \alpha t) + h(x + \alpha t) ,\end{aligned}$$

where

$$h(x) = \frac{1}{2}\sin\frac{\pi x}{L} .$$

This standing wave results from adding a wave traveling to the right to the same waveform traveling to the left. ◆

Existence and Uniqueness of Solutions

In Example 1 the method of separation of variables was used to derive a formal solution to the given initial-boundary value problem. To show that these series converge to an actual solution requires results from real analysis, just as was the case for the formal solutions to the heat equation in Section 10.5. In Examples 2 and 3, we can establish the validity of d'Alembert's solution by direct substitution into the initial value problem, assuming sufficient differentiability of the initial functions. We leave it as an exercise for you to show that if f has a continuous second derivative and g has a continuous first derivative, then d'Alembert's solution is a true solution (see Problem 12).

The question of the uniqueness of the solution to the initial-boundary value problem (1)–(4) can be answered using an **energy argument.**

Uniqueness of the Solution to the Vibrating String Problem

Theorem 8. The initial-boundary value problem

(37) $$\frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \qquad t > 0,$$

(38) $$u(0, t) = u(L, t) = 0, \qquad t > 0,$$

(39) $$u(x, 0) = f(x), \qquad 0 < x < L,$$

(40) $$\frac{\partial u}{\partial t}(x, 0) = g(x), \qquad 0 < x < L,$$

has at most one twice continuously differentiable solution.

Proof. Assume both $u(x, t)$ and $v(x, t)$ are twice continuously differentiable solutions to (37)–(40) and let $w(x, t) := u(x, t) - v(x, t)$. It is easy to check that $w(x, t)$ satisfies the initial-boundary value problem (37)–(40) with zero initial data; that is, for $0 \le x \le L$,

(41) $$w(x, 0) = 0 \quad \text{and} \quad \frac{\partial w}{\partial t}(x, 0) = 0.$$

We now show that $w(x, t) \equiv 0$ for $0 \le x \le L$, $t \ge 0$.

If $w(x, t)$ is the displacement of the vibrating string at location x for time t, then with the appropriate units, the total energy $E(t)$ of the vibrating string at time t equals the integral

(42) $$E(t) := \frac{1}{2} \int_0^L \left[\alpha^2 \left(\frac{\partial w}{\partial x} \right)^2 + \left(\frac{\partial w}{\partial t} \right)^2 \right] dx.$$

(The first term in the integrand relates to the stretching of the string at location x and represents the potential energy. The second term is the square of the velocity of the vibrating string at x and represents the kinetic energy.)

We now consider the derivative of $E(t)$:

$$\frac{dE}{dt} = \frac{d}{dt} \left\{ \frac{1}{2} \int_0^L \left[\alpha^2 \left(\frac{\partial w}{\partial x} \right)^2 + \left(\frac{\partial w}{\partial t} \right)^2 \right] dx \right\}.$$

Since w has continuous second partial derivatives (because u and v do), we can interchange the order of integration and differentiation. This gives

$$\frac{dE}{dt} = \int_0^L \left[\alpha^2 \frac{\partial w}{\partial x}\frac{\partial^2 w}{\partial t \partial x} + \frac{\partial w}{\partial t}\frac{\partial^2 w}{\partial t^2}\right] dx \,. \tag{43}$$

Again the continuity of the second partials of w guarantees that the mixed partials are equal; that is,

$$\frac{\partial^2 w}{\partial t \partial x} = \frac{\partial^2 w}{\partial x \partial t} .$$

Combining this fact with integration by parts, we obtain

$$\begin{aligned} \int_0^L \alpha^2 \frac{\partial w}{\partial x}\frac{\partial^2 w}{\partial t \partial x} dx &= \int_0^L \alpha^2 \frac{\partial w}{\partial x}\frac{\partial^2 w}{\partial x \partial t} dx \\ &= \alpha^2 \frac{\partial w}{\partial x}(L,t)\frac{\partial w}{\partial t}(L,t) - \alpha^2 \frac{\partial w}{\partial x}(0,t)\frac{\partial w}{\partial t}(0,t) \\ &\quad - \int_0^L \alpha^2 \frac{\partial^2 w}{\partial x^2}\frac{\partial w}{\partial t} dx \,. \end{aligned} \tag{44}$$

The boundary conditions $w(0,t) = w(L,t) = 0$, $t \geq 0$, imply that $(\partial w/\partial t)(0,t) = (\partial w/\partial t)(L,t) = 0$, $t \geq 0$. This reduces equation (44) to

$$\int_0^L \alpha^2 \frac{\partial w}{\partial x}\frac{\partial^2 w}{\partial t \partial x} dx = -\int_0^L \alpha^2 \frac{\partial w}{\partial t}\frac{\partial^2 w}{\partial x^2} dx \,.$$

Substituting this in for the first integrand in (43), we find

$$\frac{dE}{dt} = \int_0^L \frac{\partial w}{\partial t}\left[\frac{\partial^2 w}{\partial t^2} - \alpha^2 \frac{\partial^2 w}{\partial x^2}\right] dx \,.$$

Since w satisfies equation (37), the integrand is zero for all x. Thus, $dE/dt = 0$, and so $E(t) \equiv C$, where C is a constant. This means that the total energy is conserved within the vibrating string.

The first boundary condition in (41) states that $w(x,0) = 0$ for $0 \leq x \leq L$. Hence, $(\partial w/\partial x)(x,0) = 0$ for $0 < x < L$. Combining this with the second boundary condition in (41), we find that when $t = 0$, the integrand in (42) is zero for $0 < x < L$. Therefore, $E(0) = 0$. Since $E(t) \equiv C$, we must have $C = 0$. Hence,

$$E(t) = \frac{1}{2}\int_0^L \left[\alpha^2 \left(\frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial w}{\partial t}\right)^2\right] dx \equiv 0 \,. \tag{45}$$

That is, the total energy of this system is zero.

Because the integrand in (45) is nonnegative and continuous and the integral is zero, the integrand must be zero for $0 \leq x \leq L$. Moreover, the integrand is the sum of two squares and so each term must be zero. Hence,

$$\frac{\partial w}{\partial x}(x,t) = 0 \quad \text{and} \quad \frac{\partial w}{\partial t}(x,t) = 0$$

for all $0 \leq x \leq L$, $t \geq 0$. Thus $w(x,t) = K$, where K is a constant. Physically, this says that there is no motion in the string.

Finally, since w is constant and w is zero when $t = 0$, then $w(x, t) \equiv 0$. Consequently, $u(x, t) = v(x, t)$ and the initial-boundary value problem has at most one solution. ◆

10.6 EXERCISES

In Problems 1–4, find a formal solution to the vibrating string problem governed by the given initial-boundary value problem.

1. $$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < 1, \quad t > 0,$$
$$u(0, t) = u(1, t) = 0, \quad t > 0,$$
$$u(x, 0) = x(1 - x), \quad 0 < x < 1,$$
$$\frac{\partial u}{\partial t}(x, 0) = \sin 7\pi x, \quad 0 < x < 1$$

2. $$\frac{\partial^2 u}{\partial t^2} = 16\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < \pi, \quad t > 0,$$
$$u(0, t) = u(\pi, t) = 0, \quad t > 0,$$
$$u(x, 0) = \sin^2 x, \quad 0 < x < \pi,$$
$$\frac{\partial u}{\partial t}(x, 0) = 1 - \cos x, \quad 0 < x < \pi$$

3. $$\frac{\partial^2 u}{\partial t^2} = 4\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < \pi, \quad t > 0,$$
$$u(0, t) = u(\pi, t) = 0, \quad t > 0,$$
$$u(x, 0) = x^2(\pi - x), \quad 0 < x < \pi,$$
$$\frac{\partial u}{\partial t}(x, 0) = 0, \quad 0 < x < \pi$$

4. $$\frac{\partial^2 u}{\partial t^2} = 9\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < \pi, \quad t > 0,$$
$$u(0, t) = u(\pi, t) = 0, \quad t > 0,$$
$$u(x, 0) = \sin 4x + 7\sin 5x, \quad 0 < x < \pi,$$
$$\frac{\partial u}{\partial t}(x, 0) = \begin{cases} x, & 0 < x < \pi/2, \\ \pi - x, & \pi/2 < x < \pi \end{cases}$$

5. The Plucked String. A vibrating string is governed by the initial-boundary value problem (1)–(4). If the string is lifted to a height h_0 at $x = a$ and released, then the initial conditions are

$$f(x) = \begin{cases} h_0 x/a, & 0 < x \le a, \\ h_0(L - x)/(L - a), & a < x < L, \end{cases}$$

and $g(x) \equiv 0$. Find a formal solution.

6. A vibrating string is governed by the initial-boundary value problem (1)–(4). The initial conditions are given by $f(x) \equiv 0$ and

$$g(x) = \begin{cases} v_0 x/a, & 0 < x \le a, \\ v_0(L - x)/(L - a) & a < x < L, \end{cases}$$

where v_0 is a constant. Find a formal solution.

In Problems 7 and 8, find a formal solution to the vibrating string problem governed by the given nonhomogeneous initial-boundary value problem.

7. $$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + tx, \quad 0 < x < \pi, \quad t > 0,$$
$$u(0, t) = u(\pi, t) = 0, \quad t > 0,$$
$$u(x, 0) = \sin x, \quad 0 < x < \pi,$$
$$\frac{\partial u}{\partial t}(x, 0) = 5\sin 2x - 3\sin 5x, \quad 0 < x < \pi$$

8. $$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + x\sin t, \quad 0 < x < \pi, \quad t > 0,$$
$$u(0, t) = u(\pi, t) = 0, \quad t > 0,$$
$$u(x, 0) = 0, \quad 0 < x < \pi,$$
$$\frac{\partial u}{\partial t}(x, 0) = 0, \quad 0 < x < \pi$$

9. If one end of a string is held fixed while the other is free, then the motion of the string is governed by the initial-boundary value problem

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < L, \quad t > 0,$$
$$u(0, t) = 0 \quad \text{and} \quad \frac{\partial u}{\partial x}(L, t) = 0, \quad t > 0,$$
$$u(x, 0) = f(x), \quad 0 < x < L,$$
$$\frac{\partial u}{\partial t}(x, 0) = g(x), \quad 0 < x < L.$$

Derive a formula for a formal solution.

10. Derive a formula for the solution to the following initial-boundary value problem involving nonhomogeneous boundary conditions

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \qquad t > 0,$$
$$u(0, t) = U_1, \qquad u(L, t) = U_2, \qquad t > 0,$$
$$u(x, 0) = f(x), \qquad 0 < x < L,$$
$$\frac{\partial u}{\partial t}(x, 0) = g(x), \qquad 0 < x < L,$$

where U_1 and U_2 are constants.

11. The Telegraph Problem.[†] Use the method of separation of variables to derive a formal solution to the telegraph problem

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} + u = \alpha^2 \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < L, \quad t > 0,$$
$$u(0, t) = u(L, t) = 0, \qquad t > 0,$$
$$u(x, 0) = f(x), \qquad 0 < x < L,$$
$$\frac{\partial u}{\partial t}(x, 0) = 0, \qquad 0 < x < L.$$

12. Verify d'Alembert's solution (32) to the initial value problem (26)–(28) when $f(x)$ has a continuous second derivative and $g(x)$ has a continuous first derivative by substituting it directly into the equations.

In Problems 13–18, find the solution to the initial value problem.

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2}, \qquad -\infty < x < \infty, \qquad t > 0,$$
$$u(x, 0) = f(x), \qquad -\infty < x < \infty,$$
$$\frac{\partial u}{\partial t}(x, 0) = g(x), \qquad -\infty < x < \infty,$$

for the given functions $f(x)$ and $g(x)$.

13. $f(x) \equiv 0, \quad g(x) = \cos x$

14. $f(x) = x^2, \quad g(x) \equiv 0$

15. $f(x) = x, \quad g(x) = x$

16. $f(x) = \sin 3x, \quad g(x) \equiv 1$

17. $f(x) = e^{-x^2}, \quad g(x) = \sin x$

18. $f(x) = \cos 2x, \quad g(x) = 1 - x$

19. Derive the formal solution given in equations (22)–(24) to the vibrating membrane problem governed by the initial-boundary value problem (17)–(21).

20. Long Water Waves. The motion of long water waves in a channel of constant depth is governed by the **linearized Korteweg and de Vries (KdV) equation**

$$u_t + \alpha u_x + \beta u_{xxx} = 0, \tag{46}$$

where $u(x, t)$ is the displacement of the water from its equilibrium depth at location x and at time t, and α and β are positive constants.

(a) Show that equation (46) has a solution of the form

$$u(x, t) = V(z), \quad z = kx - w(k)t, \tag{47}$$

where k is a fixed constant and $w(k)$ is a function of k, provided V satisfies

$$-w\frac{dV}{dz} + \alpha k \frac{dV}{dz} + \beta k^3 \frac{d^3V}{dz^3} = 0. \tag{48}$$

These solutions, defined by (47), are called **uniform waves.**

(b) Physically, we are interested only in solutions $V(z)$ that are bounded and nonconstant on the infinite interval $(-\infty, \infty)$. Show that such solutions exist only if $\alpha k - w > 0$.

(c) Let $\lambda^2 = (\alpha k - w)/(\beta k^3)$. Show that the solutions from part (b) can be expressed in the form

$$V(x, t) = A \sin[\lambda k x - (\alpha \lambda k - \beta \lambda^3 k^3)t + B],$$

where A and B are arbitrary constants. [*Hint:* Solve for w in terms of λ and k and use (47).]

(d) Since both λ and k can be chosen arbitrarily and they always appear together as the product λk, we can set $\lambda = 1$ without loss of generality. Hence, we have

$$V(x, t) = A \sin[kx - (\alpha k - \beta k^3)t + B]$$

as a uniform wave solution to (46). The defining relation $w(k) = \alpha k - \beta k^3$ is called the **dispersion relation,** the ratio $w(k)/k = \alpha - \beta k^2$ is called the **phase velocity,** and the derivative $dw/dk = \alpha - 3\beta k^2$ is called the **group velocity.** When the group velocity is not constant, the waves are called **dispersive.** Show that the standard wave equation $u_{tt} = \alpha^2 u_{xx}$ has only nondispersive waves.

21. Vibrating Drum. A vibrating circular membrane of unit radius whose edges are held fixed in a plane and

[†]For a discussion of the telegraph problem, see Project E at the end of this chapter.

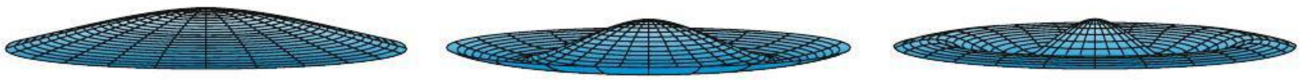
whose displacement $u(r, t)$ depends only on the radial distance r from the center and on the time t is governed by the initial-boundary value problem.

$$\frac{\partial^2 u}{\partial t^2} = \alpha^2\left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r}\right),$$
$$0 < r < 1, \qquad t > 0,$$
$$u(1, t) = 0, \qquad t > 0,$$
$$u(r, t) \quad \text{remains finite as } r \to 0^+ .$$

Show that there is a family of solutions of the form

$$u_n(r, t) = [a_n \cos(k_n\alpha t) + b_n \sin(k_n\alpha t)]J_0(k_n r),$$

where J_0 is the Bessel function of the first kind of order zero (see page 478) and $0 < k_1 < k_2 < \cdots < k_n < \cdots$ are the positive zeros of J_0. See Figure 10.24.



(a) (b) (c)

Figure 10.24 Mode shapes for vibrating drum. (a) $J_0(2.405r)$, (b) $J_0(5.520r)$, (c) $J_0(8.654r)$

10.7 Laplace's Equation

In Section 10.1 we showed how Laplace's equation,

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

arises in the study of steady-state or time-independent solutions to the heat equation. Because these solutions do not depend on time, initial conditions are irrelevant and only boundary conditions are specified. Other applications include the static displacement $u(x, y)$ of a stretched membrane fastened in space along the boundary of a region (here u must satisfy Laplace's equation inside the region); the electrostatic and gravitational potentials in certain force fields (here u must satisfy Laplace's equation in any region that is free of electrical charges or mass); and, in fluid mechanics for an idealized fluid, the stream function $u(x, y)$ whose level curves (stream lines), $u(x, y) =$ constant, represent the path of particles in the fluid (again u satisfies Laplace's equation in the flow region).

There are two basic types of boundary conditions that are usually associated with Laplace's equation: **Dirichlet boundary conditions,** where the solution $u(x, y)$ to Laplace's equation in a domain D is required to satisfy

$$u(x, y) = f(x, y) \qquad \text{on } \partial D,$$

with $f(x, y)$ a specified function defined on the boundary ∂D of D; and **Neumann boundary conditions,** where the directional derivative $\partial u/\partial n$ along the outward normal to the boundary is required to satisfy

$$\frac{\partial u}{\partial n}(x, y) = g(x, y) \qquad \text{on } \partial D,$$

with $g(x, y)$ a specified function defined on ∂D.[†] We say that the boundary conditions are **mixed** if the solution is required to satisfy $u(x, y) = f(x, y)$ on part of the boundary and $(\partial u/\partial n)(x, y) = g(x, y)$ on the remaining portion of the boundary.

In this section we use the method of separation of variables to find solutions to Laplace's equation with various boundary conditions for rectangular, circular, and cylindrical domains. We also discuss the existence and uniqueness of such solutions.

Example 1 Find a solution to the following mixed boundary value problem for a rectangle (see Figure 10.25):

(1) $$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0\,, \qquad 0 < x < a\,, \qquad 0 < y < b\,,$$

(2) $$\frac{\partial u}{\partial x}(0, y) = \frac{\partial u}{\partial x}(a, y) = 0\,, \qquad 0 \le y \le b\,,$$

(3) $$u(x, b) = 0\,, \qquad 0 \le x \le a\,,$$

(4) $$u(x, 0) = f(x)\,, \qquad 0 \le x \le a\,.$$

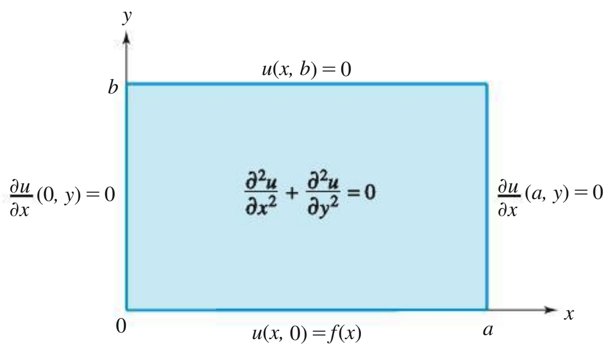


Figure 10.25 Mixed boundary value problem

Solution Separating variables, we first let $u(x, y) = X(x)Y(y)$. Substituting into equation (1), we have

$$X''(x)Y(y) + X(x)Y''(y) = 0\,,$$

which separates into

$$\frac{X''(x)}{X(x)} = -\frac{Y''(y)}{Y(y)} = -\lambda\,,$$

where λ is some constant. This leads to the two ordinary differential equations

(5) $$X''(x) + \lambda X(x) = 0\,,$$

(6) $$Y''(y) - \lambda Y(y) = 0\,.$$

From the boundary condition (2), we observe that

(7) $$X'(0) = X'(a) = 0\,.$$

[†]These generalize the boundary conditions (5) in Section 10.1 and (42) in Section 10.5.

We have encountered the eigenvalue problem in (5) and (7) previously (see Example 1 in Section 10.5). The eigenvalues are $\lambda = \lambda_n = (n\pi/a)^2$, $n = 0, 1, 2, \ldots$, with corresponding solutions

(8) $$X_n(x) = a_n \cos\left(\frac{n\pi x}{a}\right),$$

where the a_n's are arbitrary constants.

Setting $\lambda = \lambda_n = (n\pi/a)^2$ in equation (6) and solving for Y gives[†]

(9) $$Y_0(y) = A_0 + B_0 y\,,$$
$$Y_n(y) = A_n \cosh\left(\frac{n\pi y}{a}\right) + B_n \sinh\left(\frac{n\pi y}{a}\right), \qquad n = 1, 2, \ldots$$

Now the boundary condition $u(x, b) = 0$ in (3) will be satisfied if $Y(b) = 0$. Setting $y = b$ in (9), we see that we want $A_0 = -bB_0$ and

$$A_n = -B_n \tanh\left(\frac{n\pi b}{a}\right), \quad \text{or equivalently,}$$

$$Y_n(y) = C_n \sinh\left[\frac{n\pi}{a}(y - b)\right], \qquad n = 1, 2, \ldots$$

(see Problem 18). Thus we find that there are solutions to (1)–(3) of the form

$$u_0(x, y) = X_0(x)Y_0(y) = a_0B_0(y - b) = E_0(y - b)\,,$$
$$u_n(x, y) = X_n(x)Y_n(y) = a_n \cos\left(\frac{n\pi x}{a}\right) C_n \sinh\left[\frac{n\pi}{a}(y - b)\right]$$
$$= E_n \cos\left(\frac{n\pi x}{a}\right) \sinh\left[\frac{n\pi}{a}(y - b)\right], \qquad n = 1, 2, \ldots\,,$$

where the E_n's are constants. In fact, by the superposition principle,

(10) $$u(x, y) = E_0(y - b) + \sum_{n=1}^{\infty} E_n \cos\left(\frac{n\pi x}{a}\right) \sinh\left[\frac{n\pi}{a}(y - b)\right]$$

is a formal solution to (1)–(3).

Applying the remaining nonhomogeneous boundary condition in (4), we have

$$u(x, 0) = f(x) = -E_0 b + \sum_{n=1}^{\infty} E_n \sinh\left(-\frac{n\pi b}{a}\right) \cos\left(\frac{n\pi x}{a}\right).$$

This is a Fourier cosine series for $f(x)$ and hence the coefficients are given by the formulas

(11) $$E_0 = \frac{1}{(-ba)} \int_0^a f(x)\,dx\,,$$
$$E_n = \frac{2}{a \sinh\left(-\frac{n\pi b}{a}\right)} \int_0^a f(x) \cos\left(\frac{n\pi x}{a}\right) dx\,, \qquad n = 1, 2, \ldots\,.$$

Thus a formal solution is given by (10) with the constants E_n given by (11). See Figure 10.26 on page 619 for a sketch of partial sums for (10). ◆

[†]We usually express $Y_n(y) = a_n e^{n\pi y/a} + b_n e^{-n\pi y/a}$. However, computation is simplified in this case by using the hyperbolic functions $\cosh z = (e^z + e^{-z})/2$ and $\sinh z = (e^z - e^{-z})/2$.

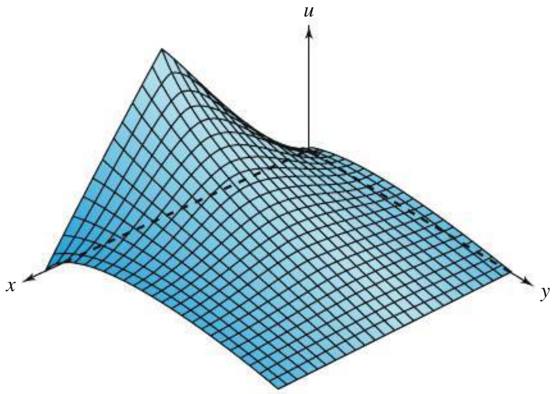


Figure 10.26 Partial sums for Example 1 with $f(x) = \begin{cases} x, & 0 < x \le \pi/2, \\ \pi - x, & \pi/2 \le x < \pi \end{cases}$

In Example 1 the boundary conditions were homogeneous on three sides of the rectangle and nonhomogeneous on the fourth side, $\{(x, y): y = 0, 0 \le x \le a\}$. It is important to note that the method used in Example 1 can also be used to solve problems for which the boundary conditions are nonhomogeneous on all sides. This is accomplished by solving four separate boundary value problems in which three sides have homogeneous boundary conditions and only one side is nonhomogeneous. The solution is then obtained by summing these four solutions (see Problem 5).

For problems involving circular domains, it is usually more convenient to use polar coordinates. In rectangular coordinates the Laplacian has the form

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} .$$

In polar coordinates (r, θ), we let

$$x = r\cos\theta , \qquad y = r\sin\theta$$

so that

$$r = \sqrt{x^2 + y^2} , \qquad \tan\theta = y/x .$$

With patience and a little care in applying the chain rule, one can show that the Laplacian in polar coordinates is

$$\textbf{(12)} \qquad \Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}$$

(see Problem 6). In the next example we obtain a solution to the **Dirichlet problem** in a disk of radius a.

Example 2 A circular metal disk of radius a has its top and bottom insulated. The edge $(r = a)$ of the disk is kept at a specified temperature that depends on its location (varies with θ). The steady-state temperature inside the disk satisfies Laplace's equation. Determine the temperature distribution $u(r, \theta)$ inside the disk by finding the solution to the following Dirichlet boundary value problem, depicted in Figure 10.27 on page 620:

$$\textbf{(13)} \qquad \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0 , \qquad 0 \le r < a , \qquad -\pi \le \theta \le \pi ,$$

$$\textbf{(14)} \qquad u(a, \theta) = f(\theta) , \qquad -\pi \le \theta \le \pi .$$

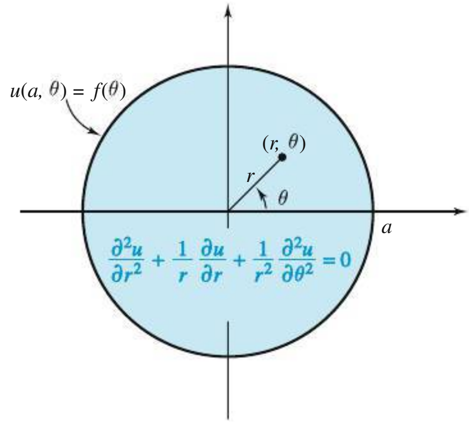


Figure 10.27 Steady-state temperature distribution in a disk

Solution To use the method of separation of variables, we first set

$$u(r, \theta) = R(r)\Theta(\theta) ,$$

where $0 \le r < a$ and $-\pi \le \theta \le \pi$. Substituting into (13) and separating variables give

$$\frac{r^2R''(r) + rR'(r)}{R(r)} = -\frac{\Theta''(\theta)}{\Theta(\theta)} = \lambda ,$$

where λ is any constant. This leads to the two ordinary differential equations

(15) $$r^2R''(r) + rR'(r) - \lambda R(r) = 0 ,$$

(16) $$\Theta''(\theta) + \lambda\Theta(\theta) = 0 .$$

For $u(r, \theta)$ to be continuous in the disk $0 \le r < a$, we need $\Theta(\theta)$ to be 2π-periodic; in particular, we require

(17) $$\Theta(-\pi) = \Theta(\pi) \quad \text{and} \quad \Theta'(-\pi) = \Theta'(\pi) .$$

Therefore, we seek nontrivial solutions to the eigenvalue problem (16)–(17).

When $\lambda < 0$, the general solution to (16) is the sum of two exponentials. Hence we have only trivial 2π-periodic solutions.

When $\lambda = 0$, we find $\Theta(\theta) = A\theta + B$ to be the solution to (16). This linear function is periodic only when $A = 0$, that is, $\Theta_0(\theta) = B$ is the only 2π-periodic solution corresponding to $\lambda = 0$.

When $\lambda > 0$, the general solution to (16) is

$$\Theta(\theta) = A \cos \sqrt{\lambda}\theta + B \sin \sqrt{\lambda}\theta .$$

Here we get a nontrivial 2π-periodic solution only when $\sqrt{\lambda} = n$, $n = 1, 2, \ldots$. [You can check this using (17).] Hence, we obtain the nontrivial 2π-periodic solutions

(18) $$\Theta_n(\theta) = A_n \cos n\theta + B_n \sin n\theta$$

corresponding to $\sqrt{\lambda} = n$, $n = 1, 2, \ldots$. Notice that the eigenfunctions are the terms of the Fourier series discussed in Section 10.3 (recall Figure 10.5, page 573).

Now for $\lambda = n^2$, $n = 0, 1, 2, \ldots$, equation (15) is the Cauchy–Euler equation

$$r^2R''(r) + rR'(r) - n^2R(r) = 0 \tag{19}$$

(see Section 4.7, page 192). When $n = 0$, the general solution is

$$R_0(r) = C + D \ln r .$$

Since $\ln r \to -\infty$ as $r \to 0^+$, this solution is unbounded near $r = 0$ when $D \neq 0$. Therefore, we must choose $D = 0$ if $u(r, \theta)$ is to be continuous at $r = 0$. We now have $R_0(r) = C$ and so $u_0(r, \theta) = R_0(r)\Theta(\theta) = CB$, which for convenience we write in the form

$$u_0(r, \theta) = \frac{A_0}{2}, \tag{20}$$

where A_0 is an arbitrary constant.

When $\lambda = n^2$, $n = 1, 2, \ldots$, you should verify that equation (19) has the general solution

$$R_n(r) = C_n r^n + D_n r^{-n} .$$

Since $r^{-n} \to \infty$ as $r \to 0^+$, we must set $D_n = 0$ in order for $u(r, \theta)$ to be bounded at $r = 0$. Thus,

$$R_n(r) = C_n r^n .$$

Now for each $n = 1, 2, \ldots$, we have the solutions

$$u_n(r, \theta) = R_n(r)\Theta_n(\theta) = C_n r^n(A_n \cos n\theta + B_n \sin n\theta) , \tag{21}$$

and by forming an infinite series from the solutions in (20) and (21), we get the following formal solution to (13):

$$u(r, \theta) = \frac{A_0}{2} + \sum_{n=1}^{\infty} C_n r^n(A_n \cos n\theta + B_n \sin n\theta) .$$

It is more convenient to write this series in the equivalent form

$$u(r, \theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(\frac{r}{a}\right)^n (a_n \cos n\theta + b_n \sin n\theta) , \tag{22}$$

where the a_n's and b_n's are constants. These constants can be determined from the boundary condition; indeed, with $r = a$ in (22), condition (14) becomes

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta) .$$

Hence, we recognize that a_n, b_n are Fourier coefficients for $f(\theta)$. Thus,

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \cos n\theta \, d\theta , \qquad n = 0, 1, \ldots , \tag{23}$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \sin n\theta \, d\theta , \qquad n = 1, 2, \ldots . \tag{24}$$

To summarize, if a_n and b_n are defined by formulas (23) and (24), then $u(r, \theta)$ given in (22) is a formal solution to the Dirichlet problem (13)–(14). ◆

The procedure in Example 2 can also be used to study the **Neumann problem** in a disk:

$$\Delta u = 0\,, \qquad 0 \le r < a\,, \qquad -\pi \le \theta \le \pi\,, \tag{25}$$

$$\frac{\partial u}{\partial r}(a, \theta) = f(\theta)\,, \qquad -\pi \le \theta \le \pi\,. \tag{26}$$

For this problem there is no unique solution, since if u is a solution, then the function u plus a constant is also a solution. Moreover, f must satisfy the **consistency condition**

$$\int_{-\pi}^{\pi} f(\theta)\, d\theta = 0\,. \tag{27}$$

If we interpret the solution $u(r, \theta)$ of equation (25) as the steady-state temperature distribution inside a circular disk that does not contain either a heat source or heat sink, then equation (26) specifies the flow of heat across the boundary of the disk. Here the consistency condition (27) is simply the requirement that the net flow of heat across the boundary is zero. (If we keep pumping heat in, the temperature won't reach equilibrium!) We leave the solution of the Neumann problem and the derivation of the consistency condition for the exercises.

The technique used in Example 2 also applies to annular domains, $\{(r, \theta): 0 < a < r < b\}$, and to exterior domains, $\{(r, \theta): a < r\}$. We also leave these applications as exercises.

Laplace's equation in cylindrical coordinates arises in the study of steady-state temperature distributions in a solid cylinder and in determining the electric potential inside a cylinder. In cylindrical coordinates,

$$x = r\cos\theta\,, \qquad y = r\sin\theta\,, \qquad z = z\,,$$

Laplace's equation becomes

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = 0\,. \tag{28}$$

The Dirichlet problem for the cylinder $\{(r, \theta, z): 0 \le r \le a, 0 \le z \le b\}$ has the boundary conditions

$$u(a, \theta, z) = f(\theta, z)\,, \qquad -\pi \le \theta \le \pi\,, \qquad 0 \le z \le b\,, \tag{29}$$

$$u(r, \theta, 0) = g(r, \theta)\,, \qquad 0 \le r \le a\,, \qquad -\pi \le \theta \le \pi\,, \tag{30}$$

$$u(r, \theta, b) = h(r, \theta)\,, \qquad 0 \le r \le a\,, \qquad -\pi \le \theta \le \pi\,. \tag{31}$$

To solve the Dirichlet boundary value problem (28)–(31), we first solve the three boundary value problems corresponding to: (i) $g \equiv 0$ and $h \equiv 0$; (ii) $f \equiv 0$ and $h \equiv 0$; and (iii) $f \equiv 0$ and $g \equiv 0$. Then by the superposition principle, the solution to (28)–(31) will be the sum of these three solutions. This is the same method that was discussed in dealing with Dirichlet problems on rectangular domains. (See the remarks following Example 1.) In the next example we solve the Dirichlet problem when $g \equiv 0$ and $h \equiv 0$.

Example 3 The base $(z = 0)$ and the top $(z = b)$ of a charge-free cylinder are grounded and therefore are at zero potential. The potential on the lateral surface $(r = a)$ of the cylinder is given by $u(a, \theta, z) = f(\theta, z)$ where $f(\theta, 0) = f(\theta, b) = 0$. Inside the cylinder, the potential $u(r, \theta, z)$

satisfies Laplace's equation. Determine the potential u inside the cylinder by finding a solution to the Dirichlet boundary value problem

(32) $$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}+\frac{\partial^2 u}{\partial z^2}=0\,, \quad 0\le r<a\,, \quad -\pi\le\theta\le\pi\,, \quad 0<z<b\,,$$

(33) $$u(a,\theta,z)=f(\theta,z)\,, \quad -\pi\le\theta\le\pi\,, \quad 0\le z\le b\,,$$

(34) $$u(r,\theta,0)=u(r,\theta,b)=0\,, \quad 0\le r<a\,, \quad -\pi\le\theta\le\pi\,.$$

Solution Using the method of separation of variables, we first assume that

$$u(r,\theta,z)=R(r)\Theta(\theta)Z(z)\,.$$

Substituting into equation (32) and separating out the Z's, we find

$$\frac{R''(r)+(1/r)R'(r)}{R(r)}+\frac{1}{r^2}\frac{\Theta''(\theta)}{\Theta(\theta)}=-\frac{Z''(z)}{Z(z)}=\lambda\,,$$

where λ can be any constant. Separating further the R's and Θ's gives

$$\frac{r^2R''(r)+rR'(r)}{R(r)}-r^2\lambda=-\frac{\Theta''(\theta)}{\Theta(\theta)}=\mu\,,$$

where μ can also be any constant. We now have the three ordinary differential equations:

(35) $$r^2R''(r)+rR'(r)-(r^2\lambda+\mu)R(r)=0\,,$$

(36) $$\Theta''(\theta)+\mu\Theta(\theta)=0\,,$$

(37) $$Z''(z)+\lambda Z(z)=0\,.$$

For u to be continuous in the cylinder, $\Theta(\theta)$ must be 2π-periodic. Thus, let's begin with the eigenvalue problem

$$\Theta''(\theta)+\mu\Theta(\theta)=0\,, \qquad -\pi<\theta<\pi\,,$$

$$\Theta(-\pi)=\Theta(\pi) \quad \text{and} \quad \Theta'(-\pi)=\Theta'(\pi)\,.$$

In Example 2 we showed that this problem has nontrivial solutions for $\mu=n^2$, $n=0,1,2,\ldots$, that are given by the Fourier series components

(38) $$\Theta_n(\theta)=A_n\cos n\theta+B_n\sin n\theta\,,$$

where the A_n's and B_n's are arbitrary constants.

The boundary conditions in (34) imply that $Z(0)=Z(b)=0$. Therefore, Z must satisfy the eigenvalue problem

$$Z''(z)+\lambda Z(z)=0\,, \quad 0<z<b\,, \quad Z(0)=Z(b)=0\,.$$

We have seen this eigenvalue problem several times before. Nontrivial solutions exist for $\lambda=(m\pi/b)^2$, $m=1,2,3,\ldots$ and are given by the sine series components

(39) $$Z_m(z)=C_m\sin\left(\frac{m\pi z}{b}\right),$$

where the C_m's are arbitrary constants.

Substituting for μ and λ in equation (35) gives

$$r^2R''(r)+rR'(r)-\left(r^2\left(\frac{m\pi}{b}\right)^2+n^2\right)R(r)=0\,, \quad 0\le r<a\,.$$

You should verify that the change of variables $s = (m\pi r/b)$ transforms this equation into the **modified Bessel's equation of order n** (page 483)[†]

$$s^2R''(s) + sR'(s) - (s^2 + n^2)R(s) = 0\,, \qquad 0 \le s < \frac{m\pi a}{b}\,. \tag{40}$$

The modified Bessel's equation of order n has two linearly independent solutions—the **modified Bessel function of the first kind:**

$$I_n(s) = \sum_{k=0}^{\infty} \frac{(s/2)^{2k+n}}{k!\,\Gamma(k+n+1)}\,,$$

which remains bounded near zero, and the **modified Bessel function of the second kind:**

$$K_n(s) = \lim_{\nu \to n} \frac{\pi}{2} \frac{I_{-\nu}(s) - I_{\nu}(s)}{\sin \nu\pi}\,,$$

which becomes unbounded as $s \to 0$. (Recall that Γ is the gamma function discussed in Section 7.6.) A general solution to (40) has the form $CK_n + DI_n$, where C and D are constants. Since u must remain bounded near $s = 0$, we must take $C = 0$. Thus, the desired solutions to (40) have the form

$$R_{mn}(r) = D_{mn} I_n\left(\frac{m\pi r}{b}\right); \qquad n = 0, 1, \ldots\,, \qquad m = 1, 2, \ldots\,, \tag{41}$$

where the D_{mn}'s are arbitrary constants.

If we multiply the functions in (38), (39), and (41) and then sum over m and n, we obtain the following series solution to (32) and (34):

$$\begin{aligned} u(r, \theta, z) = {} & \sum_{m=1}^{\infty} a_{m0} I_0\left(\frac{m\pi r}{b}\right) \sin\left(\frac{m\pi z}{b}\right) \\ & + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (a_{mn} \cos n\theta + b_{mn} \sin n\theta)\, I_n\left(\frac{m\pi r}{b}\right) \sin\left(\frac{m\pi z}{b}\right), \end{aligned} \tag{42}$$

where the a_{mn}'s and b_{mn}'s are arbitrary constants.

The constants in (42) can be obtained by imposing the boundary condition (33). Setting $r = a$ and rearranging terms, we have

$$\begin{aligned} f(\theta, z) = {} & \sum_{m=1}^{\infty} a_{m0} I_0\left(\frac{m\pi a}{b}\right) \sin\left(\frac{m\pi z}{b}\right) \\ & + \sum_{n=1}^{\infty} \left[\sum_{m=1}^{\infty} a_{mn} I_n\left(\frac{m\pi a}{b}\right) \sin\left(\frac{m\pi z}{b}\right)\right] \cos n\theta \\ & + \sum_{n=1}^{\infty} \left[\sum_{m=1}^{\infty} b_{mn} I_n\left(\frac{m\pi a}{b}\right) \sin\left(\frac{m\pi z}{b}\right)\right] \sin n\theta\,. \end{aligned} \tag{43}$$

[†]The modified Bessel's equation of order n arises in many applications and has been studied extensively. We refer the reader to the text *Special Functions* by E. D. Rainville (Chelsea Publishing, New York, 1972), for details about its solution.

Treating this double Fourier series like the one in Section 10.5 (page 600), we find

$$a_{m0} = \int_0^b \int_{-\pi}^{\pi} f(\theta, z) \sin\left(\frac{m\pi z}{b}\right) d\theta\, dz \Big/ \pi b I_0\left(\frac{m\pi a}{b}\right), \tag{44}$$

$$a_{mn} = 2\int_0^b \int_{-\pi}^{\pi} f(\theta, z) \sin\left(\frac{m\pi z}{b}\right) \cos n\theta\, d\theta\, dz \Big/ \pi b I_n\left(\frac{m\pi a}{b}\right), \qquad n \geq 1, \tag{45}$$

$$b_{mn} = 2\int_0^b \int_{-\pi}^{\pi} f(\theta, z) \sin\left(\frac{m\pi z}{b}\right) \sin n\theta\, d\theta\, dz \Big/ \pi b I_n\left(\frac{m\pi a}{b}\right), \quad n \geq 1. \tag{46}$$

Consequently, a formal solution to (32)–(34) is given by equation (42) with the constants a_{mn} and b_{mn} determined by equations (44)–(46). ◆

Existence and Uniqueness of Solutions

The existence of solutions to the boundary value problems for Laplace's equation can be established by studying the convergence of the formal solutions that we obtained using the method of separation of variables.

To answer the question of the uniqueness of the solution to a Dirichlet boundary value problem for Laplace's equation, recall that Laplace's equation arises in the search for steady-state solutions to the heat equation. Just as there are maximum principles for the heat equation, there are also maximum principles for Laplace's equation. We state one such result here.†

Maximum Principle for Laplace's Equation

Theorem 9. Let $u(x, y)$ be a solution to Laplace's equation in a bounded domain D with $u(x, y)$ continuous in $\overline{D}$, the closure of D. ($\overline{D} = D \cup \partial D$, where ∂D denotes the boundary of D.) Then $u(x, y)$ attains its maximum value on ∂D.

The uniqueness of the solution to the Dirichlet boundary value problem follows from the maximum principle. We state this result in the next theorem but leave its proof as an exercise (see Problem 19).

Uniqueness of Solution

Theorem 10. Let D be a bounded domain. If there is a continuous solution to the Dirichlet boundary value problem

$$\Delta u(x, y) = 0 \quad \text{in } D,$$
$$u(x, y) = f(x, y) \quad \text{on } \partial D,$$

then the solution is unique.

†A proof can be found in Section 8.2 of the text, *Partial Differential Equations of Mathematical Physics,* 2nd ed., by Tyn Myint-U (Elsevier North Holland, New York, 1983).

Solutions to Laplace's equation in two variables are called **harmonic** functions. These functions arise naturally in the study of analytic functions of a single complex variable. Moreover, complex analysis provides many useful results about harmonic functions. For a discussion of this interaction, we refer the reader to an introductory text on complex analysis such as *Fundamentals of Complex Analysis,* 3rd ed., by E. B. Saff and A. D. Snider (Prentice Hall, Englewood Cliffs, N.J., 2003).

10.7 EXERCISES

In Problems 1–5, find a formal solution to the given boundary value problem.

1. $$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0\,, \quad 0<x<\pi\,, \quad 0<y<1\,,$$
$$\frac{\partial u}{\partial x}(0,y)=\frac{\partial u}{\partial x}(\pi,y)=0\,, \quad 0\le y\le 1\,,$$
$$u(x,0)=4\cos 6x+\cos 7x\,, \quad 0\le x\le\pi\,,$$
$$u(x,1)=0\,, \quad 0\le x\le\pi$$

2. $$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0\,, \quad 0<x<\pi\,, \quad 0<y<\pi\,,$$
$$\frac{\partial u}{\partial x}(0,y)=\frac{\partial u}{\partial x}(\pi,y)=0\,, \quad 0\le y\le \pi\,,$$
$$u(x,0)=\cos x-2\cos 4x\,, \quad 0\le x\le\pi\,,$$
$$u(x,\pi)=0\,, \quad 0\le x\le\pi$$

3. $$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0\,, \quad 0<x<\pi\,, \quad 0<y<\pi\,,$$
$$u(0,y)=u(\pi,y)=0\,, \quad 0\le y\le\pi\,,$$
$$u(x,0)=f(x)\,, \quad 0\le x\le\pi\,,$$
$$u(x,\pi)=0\,, \quad 0\le x\le\pi$$

4. $$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0\,, \quad 0<x<\pi\,, \quad 0<y<\pi\,,$$
$$u(0,y)=u(\pi,y)=0\,, \quad 0\le y\le\pi\,,$$
$$u(x,0)=\sin x+\sin 4x\,, \quad 0\le x\le\pi\,,$$
$$u(x,\pi)=0\,, \quad 0\le x\le\pi$$

5. $$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0\,, \quad 0<x<\pi\,, \quad 0<y<1\,,$$
$$\frac{\partial u}{\partial x}(0,y)=\frac{\partial u}{\partial x}(\pi,y)=0\,, \quad 0\le y\le 1\,,$$
$$u(x,0)=\cos x-\cos 3x\,, \quad 0\le x\le\pi\,,$$
$$u(x,1)=\cos 2x\,, \quad 0\le x\le\pi$$

6. Derive the polar coordinate form of the Laplacian given in equation (12).

In Problems 7 and 8, find a solution to the Dirichlet boundary value problem for a disk:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,, \quad 0\le r<2\,, \quad -\pi\le\theta\le\pi\,,$$
$$u(2,\theta)=f(\theta)\,, \quad -\pi\le\theta\le\pi$$

for the given function $f(\theta)$.

7. $f(\theta)=|\theta|, \quad -\pi\le\theta\le\pi$

8. $f(\theta)=\cos^2\theta\,, \quad -\pi\le\theta\le\pi$

9. Find a solution to the Neumann boundary value problem for a disk:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,, \quad 0\le r<a\,, \quad -\pi\le\theta\le\pi\,,$$
$$\frac{\partial u}{\partial r}(a,\theta)=f(\theta)\,, \quad -\pi\le\theta\le\pi\,.$$

10. A solution to the Neumann problem (25)–(26) must also satisfy the consistency condition in (27). To show this, use **Green's second formula**

$$\iint_D (v\Delta u-u\Delta v)\,dx\,dy=\int_{\partial D}\left(v\frac{\partial u}{\partial n}-u\frac{\partial v}{\partial n}\right)ds\,,$$

where $\partial/\partial n$ is the outward normal derivative and ds is the differential of arc length. [*Hint:* Take $v\equiv 1$ and observe that $\partial u/\partial n=\partial u/\partial r$.]

11. Find a solution to the following Dirichlet problem for an annulus:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,, \quad 1<r<2\,, \quad -\pi\le\theta\le\pi\,,$$
$$u(1,\theta)=\sin 4\theta-\cos\theta\,, \quad -\pi\le\theta\le\pi\,,$$
$$u(2,\theta)=\sin\theta\,, \quad -\pi\le\theta\le\pi\,.$$

12. Find a solution to the following Dirichlet problem for an annulus:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,,\qquad 1<r<3\,,\qquad -\pi\le\theta\le\pi\,,$$
$$u(1,\theta)=0\,,\qquad -\pi\le\theta\le\pi\,,$$
$$u(3,\theta)=\cos 3\theta+\sin 5\theta\,,\quad -\pi\le\theta\le\pi\,.$$

13. Find a solution to the following Dirichlet problem for an exterior domain:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,,\qquad 1<r\,,\qquad -\pi\le\theta\le\pi\,,$$
$$u(1,\theta)=f(\theta)\qquad -\pi\le\theta\le\pi\,,$$
$$u(r,\theta)\quad \text{remains bounded as } r\to\infty\,.$$

14. Find a solution to the following Neumann problem for an exterior domain:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,,\qquad 1<r\,,\qquad -\pi\le\theta\le\pi\,,$$
$$\frac{\partial u}{\partial r}(1,\theta)=f(\theta)\,,\qquad -\pi\le\theta\le\pi\,,$$
$$u(r,\theta)\quad \text{remains bounded as } r\to\infty\,.$$

15. Find a solution to the following Dirichlet problem for a half disk:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,,\qquad 0<r<1\,,\qquad 0<\theta<\pi\,,$$
$$u(r,0)=0\,,\qquad 0\le r\le 1\,,$$
$$u(r,\pi)=0\,,\qquad 0\le r\le 1\,,$$
$$u(1,\theta)=\sin 3\theta\,,\qquad 0\le\theta\le\pi\,,$$
$$u(0,\theta)\quad \text{bounded.}$$

16. Find a solution to the following Dirichlet problem for a half annulus:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,,\qquad \pi<r<2\pi\,,\qquad 0<\theta<\pi\,,$$
$$u(r,0)=\sin r\,,\qquad \pi\le r\le 2\pi\,,$$
$$u(r,\pi)=0\,,\qquad \pi\le r\le 2\pi\,,$$
$$u(\pi,\theta)=u(2\pi,\theta)=0\,,\qquad 0\le\theta\le\pi\,.$$

17. Find a solution to the mixed boundary value problem

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0\,,\quad 1<r<3\,,\qquad -\pi\le\theta\le\pi\,,$$
$$u(1,\theta)=f(\theta)\,,\qquad -\pi\le\theta\le\pi\,,$$
$$\frac{\partial u}{\partial r}(3,\theta)=g(\theta)\,,\qquad -\pi\le\theta\le\pi\,.$$

18. Show that

$$-B_n\tanh\left(\frac{n\pi b}{a}\right)\cosh\left(\frac{n\pi y}{a}\right)+B_n\sinh\left(\frac{n\pi y}{a}\right)$$
$$=C_n\sinh\left[\frac{n\pi}{a}(y-b)\right],$$

where

$$C_n=B_n\Big/\cosh\left(\frac{n\pi b}{a}\right).$$

19. Prove Theorem 10 on the uniqueness of the solution to the Dirichlet problem.

20. **Stability.** Use the maximum principle to prove the following theorem on the continuous dependence of the solution on the boundary conditions:

Theorem. Let f_1 and f_2 be continuous functions on ∂D, where D is a bounded domain. For $i=1$ and 2, let u_i be the solution to the Dirichlet problem

$$\Delta u=0\,,\quad \text{in } D\,,$$
$$u=f_i\,,\quad \text{on } \partial D\,.$$

If the boundary values satisfy

$$|f_1(x,y)-f_2(x,y)|\le\varepsilon\qquad \text{for all } (x,y) \text{ on } \partial D\,,$$

where $\epsilon>0$ is some constant, then

$$|u_1(x,y)-u_2(x,y)|\le\varepsilon\qquad \text{for all } (x,y) \text{ in } D\,.$$

21. For the Dirichlet problem described in Example 3, let $a=b=\pi$ and assume the potential on the lateral side $(r=\pi)$ of the cylinder is $f(\theta,z)=\sin z$. Use equations (44)–(46) to compute the solution given by equation (42).

22. **Invariance of Laplace's Equation.** A complex-valued function $f(z)$ of the complex variable $z=x+iy$ can be written in the form $f(z)=u(x,y)+iv(x,y)$, where u and v are real-valued functions. If $f(z)$ is analytic in a planar region D, then its real and imaginary parts satisfy the Cauchy–Riemann equations in D; that is, in D

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}\,,\qquad \frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}\,.$$

Let f be analytic and one-to-one in D and assume its inverse f^{-1} is analytic in D', where D' is the image of D under f. Then

$$\frac{\partial x}{\partial u}=\frac{\partial y}{\partial v}\,,\qquad \frac{\partial x}{\partial v}=-\frac{\partial y}{\partial u}$$

are just the Cauchy–Riemann equations for f^{-1}. Show that if $\phi(x,y)$ satisfies Laplace's equation for (x,y) in D, then $\psi(u,v):=\phi\big(x(u,v),y(u,v)\big)$ satisfies Laplace's equation for (u,v) in D'.

23. Fluid Flow Around a Corner. The stream lines that describe the fluid flow around a corner (see Figure 10.28) are given by $\phi(x, y) = k$, where k is a constant and ϕ, the stream function, satisfies the boundary value problem

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0\,, \qquad x > 0\,, \qquad y > 0\,,$$
$$\phi(x, 0) = 0\,, \qquad 0 \le x\,,$$
$$\phi(0, y) = 0\,, \qquad 0 \le y\,.$$

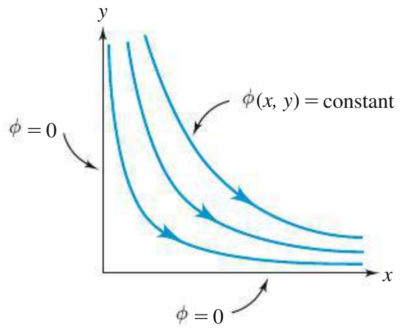


Figure 10.28 Flow around a corner

(a) Using the results of Problem 22, show that this problem can be reduced to finding the flow above a flat plate (see Figure 10.29). That is, show that the problem reduces to finding the solution to

$$\frac{\partial^2 \psi}{\partial u^2} + \frac{\partial^2 \psi}{\partial v^2} = 0\,, \qquad v > 0\,, \qquad -\infty < u < \infty\,,$$
$$\psi(u, 0) = 0\,, \qquad -\infty < u < \infty\,,$$

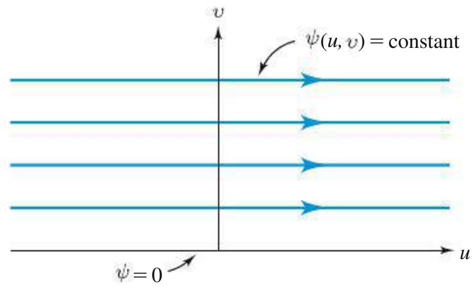


Figure 10.29 Flow above a flat surface

where ψ and ϕ are related as follows: $\psi(u, v) = \phi\big(x(u, v), y(u, v)\big)$ with the mapping between (u, v) and (x, y) given by the analytic function $f(z) = z^2$.

(b) Verify that a nonconstant solution to the problem in part (a) is given by $\psi(u, v) = v$.

(c) Using the result of part (b), find a stream function $\phi(x, y)$ for the original problem.

24. Unbounded Domain. Using separation of variables, find a solution of Laplace's equation in the infinite rectangle $0 < x < \pi, 0 < y < \infty$ that is zero on the sides $x = 0$ and $x = \pi$, approaches zero as $y \to \infty$, and equals $f(x)$ for $y = 0$.

Chapter 10 Summary

Separation of Variables. A classical technique that is effective in solving boundary value problems for partial differential equations is the method of separation of variables. Briefly, the idea is to assume first that there exists a solution that can be written with the variables separated: e.g., $u(x, t) = X(x)T(t)$. Substituting $u = XT$ into the partial differential equation and then imposing the boundary conditions leads to the problem of finding the eigenvalues and eigenfunctions for a boundary value problem for an ordinary differential equation. Solving for the eigenvalues and eigenfunctions, one eventually obtains solutions $u_n(x, t) = X_n(x)T_n(t)$, $n = 1, 2, 3, \ldots$ that solve the partial differential equation and the homogeneous boundary conditions. Taking infinite linear combinations of the $u_n(x, t)$ yields complete solutions of the form

$$u(x, t) = \sum_{n=1}^{\infty} c_n u_n(x, t)\,.$$

The coefficients c_n are then obtained using the initial conditions or nonhomogeneous conditions.

Fourier Series. Let $f(x)$ be a piecewise continuous function on the interval $[-L, L]$. The Fourier series of f is the trigonometric series

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \left\{ a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right\},$$

where the a_n's and b_n's are given by **Euler's formulas:**

$$a_n = \frac{1}{L} \int_{-L}^{L} f(x) \cos\left(\frac{n\pi x}{L}\right) dx, \qquad n = 0, 1, 2, \ldots,$$

$$b_n = \frac{1}{L} \int_{-L}^{L} f(x) \sin\left(\frac{n\pi x}{L}\right) dx, \qquad n = 1, 2, 3, \ldots.$$

Let f be piecewise continuous on the interval $[0, L]$. The **Fourier cosine series** of $f(x)$ on $[0, L]$ is

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right),$$

where

$$a_n = \frac{2}{L} \int_{0}^{L} f(x) \cos\left(\frac{n\pi x}{L}\right) dx.$$

The **Fourier sine series** of $f(x)$ on $[0, L]$ is

$$f(x) \sim \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right),$$

where

$$b_n = \frac{2}{L} \int_{0}^{L} f(x) \sin\left(\frac{n\pi x}{L}\right) dx.$$

Fourier series and the method of separation of variables are used to solve boundary value problems and initial-boundary value problems for the three classical equations:

Heat equation	$\dfrac{\partial u}{\partial t} = \beta \dfrac{\partial^2 u}{\partial x^2}.$
Wave equation	$\dfrac{\partial^2 u}{\partial t^2} = \alpha^2 \dfrac{\partial^2 u}{\partial x^2}.$
Laplace's equation	$\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 0.$

For the heat equation, there is a unique solution $u(x, t)$ when the boundary values at the ends of a conducting wire are specified along with the initial temperature distribution. The wave equation yields the displacement $u(x, t)$ of a vibrating string when the ends are held fixed and the initial displacement and initial velocity at each point of the string are given. In such a case, separation of variables yields a solution that is the sum of standing waves. For an infinite string with specified initial displacement and velocity, d'Alembert's solution yields traveling waves. Laplace's equation arises in the study of steady-state solutions to the heat and wave equations.

TECHNICAL WRITING EXERCISES FOR CHAPTER 10

1. The method of separation of variables is an important technique in solving initial-boundary value problems and boundary value problems for *linear* partial differential equations. Explain where the linearity of the differential equation plays a crucial role in the method of separation of variables.

2. In applying the method of separation of variables, we have encountered a variety of *special functions,* such as sines, cosines, Bessel functions, and modified Bessel functions. Describe three or four examples of partial differential equations that involve other special functions, such as Legendre polynomials, Hermite polynomials, and Laguerre polynomials. (Some exploring in the library may be needed; start with the table on page 483.)

3. A constant-coefficient second-order partial differential equation of the form

$$a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} = 0$$

can be classified using the discriminant $D := b^2 - 4ac$. In particular, the equation is called

hyperbolic if $D > 0$, and

elliptic if $D < 0$.

Verify that the wave equation is hyperbolic and Laplace's equation is elliptic. It can be shown that such hyperbolic (elliptic) equations can be transformed by a linear change of variables into the wave (Laplace's) equation. Based on your knowledge of the latter equations, describe which types of problems (initial value, boundary value, etc.) are appropriate for hyperbolic equations and elliptic equations.

Projects for Chapter 10

A Steady-State Temperature Distribution in a Circular Cylinder

When the temperature u inside a circular cylinder reaches a steady state, it satisfies Laplace's equation $\Delta u = 0$. If the temperature on the lateral surface $(r = a)$ is kept at zero, the temperature on the top $(z = b)$ is kept at zero, and the temperature on the bottom $(z = 0)$ is given by $u(r, \theta, 0) = f(r, \theta)$, then the steady-state temperature satisfies the boundary value problem

$$\begin{aligned}
&\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = 0\,, \quad 0 \le r < a\,, \quad -\pi \le \theta \le \pi\,, \quad 0 < z < b\,,\\
&u(a, \theta, z) = 0\,, \quad -\pi \le \theta \le \pi\,, \quad 0 \le z \le b\,,\\
&u(r, \theta, b) = 0\,, \quad 0 \le r < a\,, \quad -\pi \le \theta \le \pi\,,\\
&u(r, \theta, 0) = f(r, \theta)\,, \quad 0 \le r < a\,, \quad -\pi \le \theta \le \pi\,,
\end{aligned}$$

where $f(a, \theta) = 0$ for $-\pi \le \theta \le \pi$ (see Figure 10.30). To find a solution to this boundary value problem, proceed as follows:

(a) Let $u(r, \theta, z) = R(r)\Theta(\theta)Z(z)$. Show that R, Θ, and Z must satisfy the three ordinary differential equations

$$\begin{aligned}
r^2R'' + rR' - (r^2\lambda + \mu)R &= 0\,,\\
\Theta'' + \mu\Theta &= 0\,,\\
Z'' + \lambda Z &= 0\,.
\end{aligned}$$

(b) Show that $\Theta(\theta)$ has the form

$$\Theta(\theta) = A\cos n\theta + B\sin n\theta$$

for $\mu = n^2$, $n = 0, 1, 2, \ldots$.

(c) Show that $Z(z)$ has the form

$$Z(z) = C\sinh[\beta(b - z)]$$

for $\lambda = -\beta^2$, where $\beta > 0$.

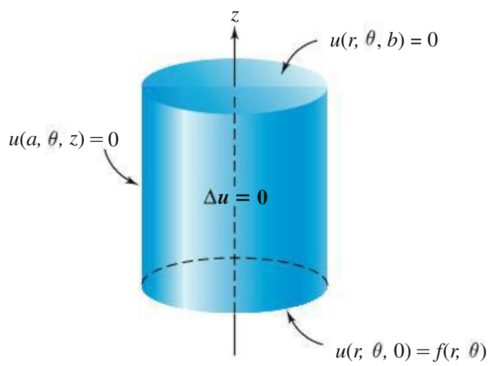


Figure 10.30 Dirichlet problem for cylinder

(d) Show that $R(r)$ has the form, for each n,

$$R_n(r) = DJ_n(\beta r) ,$$

where J_n is the Bessel function of the first kind.

(e) Show that the boundary conditions require $R(a) = 0$, and so

$$J_n(\beta a) = 0 .$$

Hence, for each n, if $0 < \alpha_{n1} < \alpha_{n2} < \cdots < \alpha_{nm} < \cdots$ are the zeros of J_n, then

$$\beta_{nm} = \alpha_{nm}/a .$$

Moreover,

$$R_n(r) = DJ_n(\alpha_{nm}r/a) .$$

Some typical eigenfunctions are displayed in Figure 10.31.

(f) Use the preceding results to show that $u(r, \theta, z)$ has the form

$$u(r, \theta, z) = \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} J_n\left(\frac{\alpha_{nm}r}{a}\right)(a_{nm}\cos n\theta + b_{nm}\sin n\theta)\sinh\left(\frac{\alpha_{nm}(b-z)}{a}\right),$$

where a_{nm} and b_{nm} are constants.

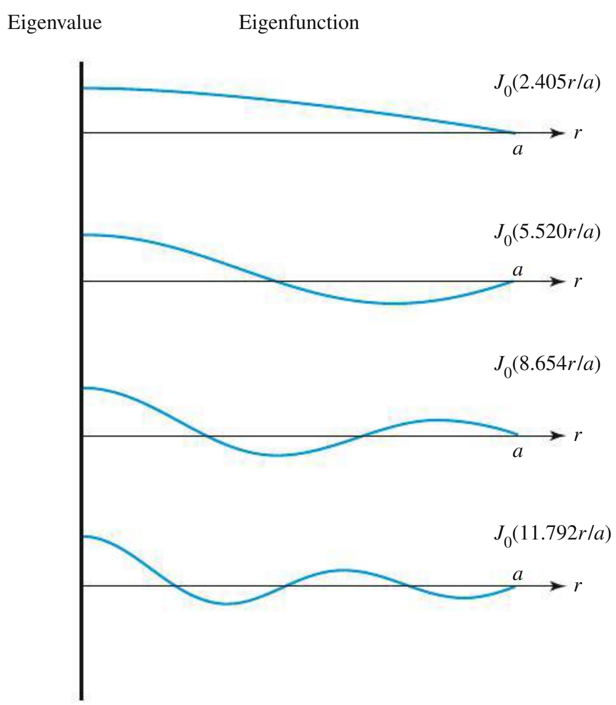


Figure 10.31 Bessel eigenfunctions. See also Figure 10.24, page 616.

(g) Use the final boundary condition and a double orthogonal expansion involving Bessel functions and trigonometric functions to derive the formulas

$$a_{0m} = \frac{1}{\pi a^2 \sinh(\alpha_{0m} b/a)[J_1(\alpha_{0m})]^2} \int_0^a \int_0^{2\pi} f(r,\theta) J_0\left(\frac{\alpha_{0m} r}{a}\right) r\,dr\,d\theta ,$$

for $m = 1, 2, 3, \ldots$, and for $n, m = 1, 2, 3, \ldots$,

$$a_{nm} = \frac{2}{\pi a^2 \sinh(\alpha_{nm} b/a)[J_{n+1}(\alpha_{nm})]^2} \int_0^a \int_0^{2\pi} f(r,\theta) J_n\left(\frac{\alpha_{nm} r}{a}\right) \cos(n\theta)\, r\,dr\,d\theta ,$$

$$b_{nm} = \frac{2}{\pi a^2 \sinh(\alpha_{nm} b/a)[J_{n+1}(\alpha_{nm})]^2} \int_0^a \int_0^{2\pi} f(r,\theta) J_n\left(\frac{\alpha_{nm} r}{a}\right) \sin(n\theta)\, r\,dr\,d\theta .$$

(The orthogonality relations for Bessel function expansions are described by equations (21), (22) in Section 11.7, page 689.)*

B Laplace Transform Solution of the Wave Equation

Laplace transforms can be used to solve certain partial differential equations. To illustrate this technique, consider the initial-boundary value problem

(1) $$\frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < \infty, \qquad t > 0,$$

(2) $$u(0,t) = h(t), \qquad t > 0,$$

(3) $$u(x,0) = 0, \qquad 0 < x < \infty,$$

(4) $$\frac{\partial u}{\partial t}(x,0) = 0, \qquad 0 < x < \infty,$$

(5) $$\lim_{x\to\infty} u(x,t) = 0, \qquad t \geq 0.$$

This problem arises in studying a semi-infinite string that is initially horizontal and at rest and where one end is being moved vertically. Let $u(x,t)$ be the solution to (1)–(5). For each x, let

$$U(x,s) \coloneqq \mathscr{L}\{u(x,t)\}(x,s) = \int_0^\infty e^{-st} u(x,t)\,dt .$$

(a) Using the fact that

$$\mathscr{L}\left\{\frac{\partial^2 u}{\partial x^2}\right\} = \frac{\partial^2}{\partial x^2}\mathscr{L}\{u\},$$

show that $U(x,s)$ satisfies the equation

(6) $$s^2 U(x,s) = \alpha^2 \frac{\partial^2 U}{\partial x^2}, \qquad 0 < x < \infty .$$

(b) Show that the general solution to (6) is

$$U(x,s) = A(s)e^{-sx/\alpha} + B(s)e^{sx/\alpha},$$

where $A(s)$ and $B(s)$ are arbitrary functions of s.

[†]All references to Chapters 11–13 refer to the expanded text, *Fundamentals of Differential Equations and Boundary Value Problems*, 7th ed.

(c) Since $u(x, t) \to 0$ as $x \to \infty$ for all $0 \le t < \infty$, we have $U(x, s) \to 0$ as $x \to \infty$. Use this fact to show that the $B(s)$ in part (b) must be zero.

(d) Using equation (2), show that

$$A(s) = H(s) = \mathscr{L}\{h\}(s) ,$$

where $A(s)$ is given in part (b).

(e) Use the results of parts (b), (c), and (d) to obtain a formal solution to (1)–(5).

C Green's Function

Let Ω be a region in the xy-plane having a smooth boundary $\partial\Omega$. Associated with Ω is a **Green's function** $G(x, y; \xi, \eta)$ defined for pairs of distinct points (x, y), (ξ, η) in Ω. The function $G(x, y; \xi, \eta)$ has the following property.

Let $\Delta u := \partial^2 u/\partial x^2 + \partial^2 u/\partial y^2$ denote the Laplacian operator on Ω; let $h(x, y)$ be a given continuous function on Ω; and let $f(x, y)$ be a given continuous function on $\partial\Omega$. Then, a continuous solution to the Dirichlet boundary value problem

$$\Delta u(x, y) = h(x, y) , \qquad \text{in } \Omega ,$$
$$u(x, y) = f(x, y) , \qquad \text{on } \partial\Omega ,$$

is given by

$$\text{(7)} \qquad u(x, y) = \iint_{\Omega} G(x, y; \xi, \eta)\, h(\xi, \eta)\, d\xi\, d\eta + \int_{\partial\Omega} f(\xi, \eta) \frac{\partial G}{\partial n}(x, y; \xi, \eta)\, d\sigma(\xi, \eta) ,$$

where n is the outward normal to the boundary $\partial\Omega$ of Ω and the second integral is the line integral, with respect to arc length, around the boundary of Ω with the interior of Ω on the left as the boundary is traversed. In (7) we assume that $\partial\Omega$ is sufficiently smooth so that the integrands and integrals exist.

When Ω is the upper half-plane, the Green's function is[†]

$$G(x, y; \xi, \eta) = \frac{1}{4\pi} \ln\left[\frac{(x-\xi)^2 + (y-\eta)^2}{(x-\xi)^2 + (y+\eta)^2}\right] .$$

(a) Using (7), show that a solution to

$$\text{(8)} \qquad \Delta u(x, y) = h(x, y) , \qquad -\infty < x < \infty , \qquad 0 < y ,$$
$$\text{(9)} \qquad u(x, 0) = f(x) , \qquad -\infty < x < \infty ,$$

is given by

$$u(x, y) = \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{f(\xi)}{(x-\xi)^2 + y^2}\, d\xi + \frac{1}{4\pi} \int_0^{\infty} \int_{-\infty}^{\infty} \ln\left[\frac{(x-\xi)^2 + (y-\eta)^2}{(x-\xi)^2 + (y+\eta)^2}\right] h(\xi, \eta)\, d\xi\, d\eta .$$

[†]Techniques for determining Green's functions can be found in texts on partial differential equations such as *Partial Differential Equations of Mathematical Physics,* 2nd ed., by Tyn Myint-U (Elsevier North Holland, New York, 1983), Chapter 10.

(b) Use the result of part (a) to determine a solution to (8)–(9) when $h \equiv 0$ and $f \equiv 1$.

(c) When Ω is the interior of the unit circle $x^2 + y^2 = 1$, the Green's function, in polar coordinates (r, θ) and (ρ, ϕ), is given by

$$G(r, \theta; \rho, \phi) = \frac{1}{4\pi} \ln\left[r^2 + \rho^2 - 2r\rho\cos(\phi - \theta)\right] - \frac{1}{4\pi} \ln\left[r^2 + \rho^{-2} - 2r\rho^{-1}\cos(\phi - \theta)\right] - \frac{1}{4\pi} \ln \rho .$$

Using (7), show that the solution to

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0 , \qquad 0 \le r < 1 , \qquad 0 \le \theta \le 2\pi ,$$

$$u(1, \theta) = f(\theta) , \qquad 0 \le \theta \le 2\pi ,$$

is given by

$$u(r, \theta) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - r^2}{1 + r^2 - 2r\cos(\phi - \theta)} f(\phi)\, d\phi .$$

This is known as **Poisson's integral formula.**

(d) Use Poisson's integral formula to derive the following mean value property for solutions to Laplace's equation.

Mean Value Property

Theorem 11. If $u(x, y)$ satisfies $\Delta u = 0$ in a bounded domain Ω in $\boldsymbol{R}^2$ and (x_0, y_0) lies in Ω, then,

(10) $$u(x_0, y_0) = \frac{1}{2\pi} \int_0^{2\pi} u(x_0 + r\cos\theta, y_0 + r\sin\theta)\, d\theta$$

for all $r > 0$ for which the disk $\overline{B}(x_0, y_0; r) = \{(x, y) \colon (x - x_0)^2 + (y - y_0)^2 \le r^2\}$ lies entirely in Ω.

[*Hint:* Use a change of variables that maps the disk $\overline{B}(x_0, y_0; r)$ to the unit disk $\overline{B}(0, 0; 1)$.]

D Numerical Method for $\Delta u = f$ on a Rectangle

Let R denote the open rectangle

$$R = \{(x, y) \colon a < x < b, \ \ c < y < d\}$$

and ∂R be its boundary. Here we describe a numerical technique for solving the generalized Dirichlet problem

(11) $$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y) , \qquad \text{for } (x, y) \text{ in } R ,$$

$$u(x, y) = g(x, y) , \qquad \text{for } (x, y) \text{ on } \partial R .$$

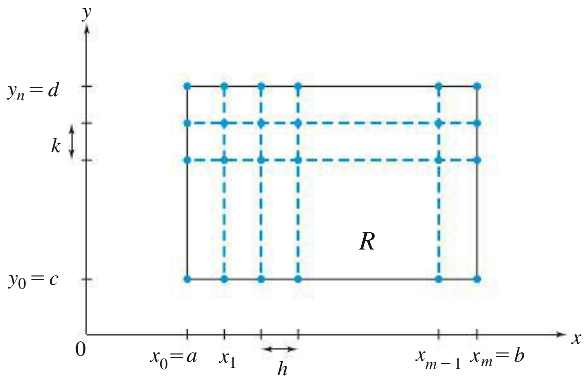


Figure 10.32 Rectangular grid

The method is similar to the finite-difference technique discussed in Chapter 4. We begin by selecting positive integers m and n and step sizes h and k so that $b - a = hm$ and $d - c = kn$. The interval $[a, b]$ is now partitioned into m equal subintervals and $[c, d]$ into n equal subintervals. The partition points are

$$x_i = a + ih\,, \qquad 0 \le i \le m\,,$$
$$y_j = c + jk\,, \qquad 0 \le j \le n$$

(see Figure 10.32). The (dashed) lines $x = x_i$ and $y = y_j$ are called **grid lines** and their intersections (x_i, y_j) are the **mesh points** of the partition. Our goal is to obtain approximations to the solution $u(x, y)$ of problem (11) at each *interior* mesh point, i.e., at (x_i, y_j) where $1 \le i \le m - 1$, $1 \le j \le n - 1$. [Of course, from (11), we are given the values of $u(x, y)$ at the boundary mesh points, e.g., $u(x_0, y_j) = g(x_0, y_j)$, $0 \le j \le n$.]

The next step is to approximate the partial derivatives $\partial^2 u/\partial x^2$ and $\partial^2 u/\partial y^2$ using the *centered-difference approximations*

(12) $$\frac{\partial^2 u}{\partial x^2}(x_i, y_j) \approx \frac{1}{h^2}[u(x_{i+1}, y_j) - 2u(x_i, y_j) + u(x_{i-1}, y_j)] \qquad (\text{for } 1 \le i \le m - 1)\,,$$

and

(13) $$\frac{\partial^2 u}{\partial y^2}(x_i, y_j) \approx \frac{1}{k^2}[u(x_i, y_{j+1}) - 2u(x_i, y_j) + u(x_i, y_{j-1})] \qquad (\text{for } 1 \le i \le n - 1)\,.$$

These approximations are based on the generic formula

$$y''(x) = [y(x + h) - 2y(x) + y(x - h)]/h^2 + [\text{terms involving } h^2, h^3, \ldots]\,,$$

which follows from the Taylor series expansion for $y(x)$.

(a) Show that substituting the approximations (12) and (13) into the Dirichlet problem (11) yields the following system:

(14) $$2\left[\left(\frac{h}{k}\right)^2 + 1\right]u_{i,j} - (u_{i+1,j} + u_{i-1,j}) - \left(\frac{h}{k}\right)^2(u_{i,j+1} + u_{i,j-1}) = -h^2 f(x_i, y_j)$$

for $i = 1, 2, \ldots, m - 1$, and $j = 1, 2, \ldots, n - 1$;

$$u_{0,j} = g(a, y_j)\,, \qquad u_{m,j} = g(b, y_j)\,, \qquad j = 0, 1, \ldots, n\,,$$
$$u_{i,0} = g(x_i, c)\,, \qquad u_{i,n} = g(x_i, d)\,, \qquad i = 1, 2, \ldots, m - 1\,,$$

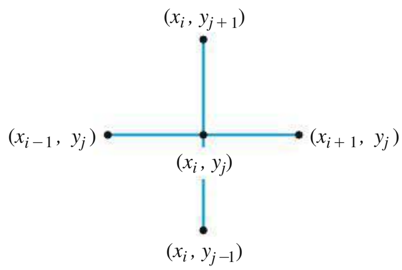


Figure 10.33 Approximate solution $u_{i,j}$ at (x_i, y_j) is obtained from values at the ends of the cross

where $u_{i,j}$ approximates $u(x_i, y_j)$. Notice that each equation in (14) involves approximations to the solution that appear in a cross centered at a mesh point (see Figure 10.33).

(b) Show that the system in part (a) is a linear system of $(m-1)(n-1)$ unknowns in $(m-1)(n-1)$ equations.

(c) For Laplace's equation where $f(x, y) \equiv 0$, show that when $h = k$, equation (14) yields

$$u_{i,j} = \frac{1}{4}(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}) .$$

Compare this averaging formula with the mean value property of Theorem 11 (Project C).

(d) For the square plate

$$R = \{(x, y): 0 \le x \le 0.4, 0 \le y \le 0.4\} ,$$

the boundary is maintained at the following temperatures:

$$u(0, y) = 0 , \qquad u(0.4, y) = 10y , \qquad 0 \le y \le 0.4 .$$
$$u(x, 0) = 0 , \qquad u(x, 0.4) = 10x , \qquad 0 \le x \le 0.4 .$$

Using the system in part (a) with $f(x, y) \equiv 0$, $m = n = 4$, and $h = k = 0.1$, find approximations to the steady-state temperatures at the mesh points of the plate. [*Hint:* It is helpful to label these grid points with a single index, say $p_1, p_2, \ldots, p_q$, choosing the ordering in a book-reading sequence.]

E The Telegrapher's Equation and the Cable Equation

Electrical engineers often employ transmission lines to guide electromagnetic waves from a source to a device. The transmission line may take any of several forms, such as depicted in Figure 10.34(a) on page 638. They have common characteristics: uniformity in one direction (z) and uniform capacitance, inductance, longitudinal resistance, and transversal resistance *per unit length* in the z direction (if they are operated at low frequency). Thus any short length dz of the line can be modeled as an *RLC* circuit as depicted in Figure 10.34(b).

In the figure, the circuit parameters are interpreted *per unit length*; thus the incremental inductance is $L\,dz$, the incremental capacitance is $C\,dz$, the (very low) incremental longitudinal

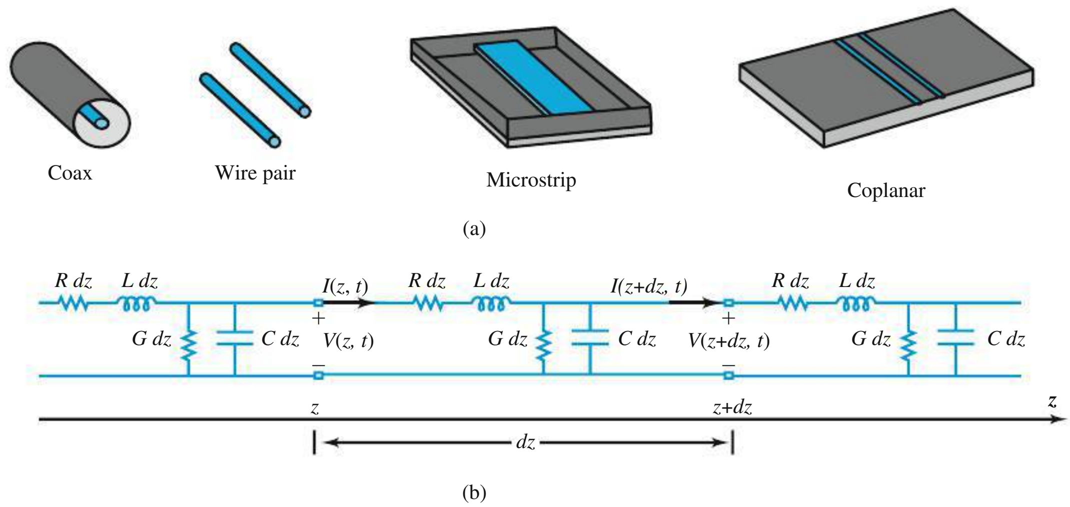


Figure 10.34 Transmission lines (a) common forms (b) circuit model

resistance is $R\,dz$, and the (very low) incremental transversal conductance (reciprocal of resistance) is $G\,dz$.

(a) Apply the circuit analysis of Section 5.7 to the length dz of the transmission line to derive the equations

$$V(z+dz,t) = V(z,t) - (R\,dz)I(z,t) - (L\,dz)\frac{\partial I(z,t)}{\partial t}, \tag{15}$$

$$I(z+dz,t) = I(z,t) - (G\,dz)V(z,t) - (C\,dz)\frac{\partial V(z,t)}{\partial t}, \tag{16}$$

(b) Divide these equations by dz and take limits to obtain the system

$$\frac{\partial V(z,t)}{\partial z} = -RI(z,t) - L\frac{\partial I(z,t)}{\partial t}, \tag{17}$$
$$\frac{\partial I(z,t)}{\partial z} = -GV(z,t) - C\frac{\partial V(z,t)}{\partial t}.$$

(c) Eliminate $I(z,t)$ from the system (17) to derive

$$\frac{\partial^2 V(z,t)}{\partial z^2} = LC\frac{\partial^2 V(z,t)}{\partial t^2} + [RC+LG]\frac{\partial V(z,t)}{\partial t} + RGV(z,t). \tag{18}$$

[*Hint:* Differentiate the first equation with respect to z and the second with respect to t, and observe the equality of the mixed partials of $I(z,t)$.]

Equation (18) is known as the *telegrapher's equation* for the transmission line; (17) is also known as the *system of telegrapher's equations.*

(d) Make the substitution $V(z,t) = T(t)v(z,t)$ in the telegrapher's equation. For what choices of $T(t)$ is the term proportional to $\partial v/\partial t$ eliminated? For what choices of $T(t)$ is the term proportional to v eliminated?

(e) Show that the telegrapher's equation for a *lossless line,* where the longitudinal resistance R and the transversal conductance G of the line are zero, reduces to the wave equation of Section 10.6. What is the speed of the voltage waves on a lossless line?

(f) Naturally engineers are interested in oscillatory waves on transmission lines, and thanks to Euler's formula (Section 4.3, page 166), solutions to the system (17) can be constructed from the real and imaginary parts of solutions of the form $V(z,t) = V(z)e^{i\omega t}$ and $I(z,t) = I(z)e^{i\omega t}$. (Elaboration of this procedure is given in Project B, Chapter 7, page 418.) Derive a first-order matrix system of the form

$$\textbf{(19)} \qquad \begin{bmatrix} V(z) \\ I(z) \end{bmatrix}' = \mathbf{A}\begin{bmatrix} V(z) \\ I(z) \end{bmatrix}$$

for such solutions, where **A** is a constant matrix.

(g) For the lossless case $(R = G = 0)$, construct the matrix exponential $e^{\mathbf{A}t}$ for the matrix in (19) by direct calculation of the series expansion in equation (2) of Section 9.8 (page 545). Express a general solution to the oscillatory transmission line system (19) for the lossless line.

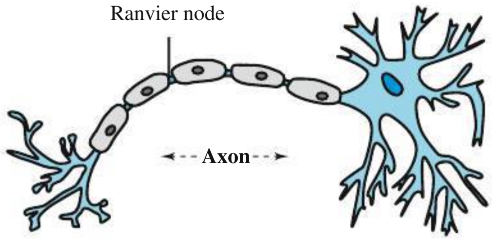


Figure 10.35 Axon model

(h) In the neurons of an invertebrate's central nervous system, electric signals are propagated between *Ranvier nodes* along *axons*, which are fibers containing an axoplasm liquid and insulated by a myelin sheaf; see Figure 10.35. Their electrical properties are quite similar to those of the coaxial cable in Figure 10.34 on page 638. However, the inductance is negligible. Therefore the voltages in a myelinated axon propagate in accordance with the *cable equation,* which is the transmission line equation (18) with L omitted:

$$\textbf{(20)} \qquad \frac{\partial^2 V(z,t)}{\partial z^2} = RC\frac{\partial V(z,t)}{\partial t} + RGV(z,t)\ .$$

(i) Use the strategy of part (d) to find $T(t)$ such that the substitution $V(z,t) = T(t)v(z,t)$ results in an analog to the heat equation (Section 10.5)

$$\textbf{(21)} \qquad \frac{\partial^2 v(z,t)}{\partial z^2} = RC\frac{\partial v(z,t)}{\partial t}\ .$$

(Therefore propagation along an axon proceeds not like a wavefront, but rather like a warm front.)

Appendices

APPENDIX A
Review of Integration Techniques

In solving differential equations, we cannot overemphasize the importance of integration techniques. Indeed, we sometimes refer to the problem of solving a differential equation by asking, "How can we integrate this equation?"

This section is devoted to reviewing three of the standard techniques that will be useful for integrating many of the functions encountered in this text. Of course, one can always search in a table of integrals (a brief table appears on the inside back cover of this text), but familiarity with tricks of integration is a worthwhile investment of time for several reasons: (i) we may not find the particular integral we need in the table; (ii) using one of the techniques may be quicker; and (iii) we might be curious to know how the various integrals in the table were obtained.

So what techniques can we use for the following integrals?

(a) $\int x\sqrt{2x+1}\,dx$ **(b)** $\int \frac{x^2}{\sqrt{9-x^2}}\,dx$ **(c)** $\int x \ln x \; dx$

(d) $\int \frac{7x^2+10x-1}{x^3+3x^2-x-3}\,dx$ **(e)** $\int e^{x^2}dx$

For **(a)** we can use *algebraic substitution*, while **(b)** can be handled with *trigonometric substitution*. For **(c)** we have available *integration by parts*, and the integral in **(d)** will yield to *partial fractions*. But try as we may, no trick is sufficient to give **(e)** in a finite number of terms. By this we mean that if we have available the elementary functions (rational functions, trigonometric functions, inverse trigonometric functions, exponential functions, and logarithmic functions) and permit use of the standard operations—addition, subtraction, multiplication, division, taking roots, and composite functions—a finite number of times, we cannot find a formula for the antiderivative of e^{x^2}. A similar conundrum arises with integration of $\sqrt{1+\cos^2 x}$. So we can add a fourth reason for honing our skills with integration techniques: They will help us to recognize those integrals that *can* be expressed in a finite number of terms.

In this context, we are speaking of integration in closed form—analytically—as opposed to numerical integration. Thus integration is the inverse of differentiation; it requires finding an anti*derivative*. As a warmup, then, we begin this appendix by jogging your recollection of rules for *derivatives*. We hope you will find our approach a little different and refreshing.

Differentiation

Of course you remember that the derivative of a constant equals zero, and the derivative of a sum equals the sum of the derivatives. So let's jump in with the basic power rule.

Derivative of a power function $\frac{d}{dx}x^p = px^{p-1}$.

This holds for *any* constant exponent p, be it integer, fraction, or negative.

Examples: $\frac{d}{dx}x^{100} = 100x^{99}$, $\frac{d}{dx}x^{7/2} = \frac{7}{2}x^{5/2}$, $\frac{d}{dx}x^{1/3} = \frac{1}{3}x^{-2/3}$, $\frac{d}{dx}x^{-e} = -ex^{-e-1}$

Derivative of the exponential $\frac{d}{dx}e^x = e^x$.

This is the essential feature of the exponential function.

Derivative of trig functions $\frac{d}{dx}\sin x = \cos x$, $\frac{d}{dx}\cos x = -\sin x$.

Remember which one gets the minus sign.

Derivative of a product $\frac{d}{dx}(fg) = f\frac{d}{dx}g + g\frac{d}{dx}f$.

Example: $\frac{d}{dx}(x^2 \sin x) = x^2 \cos x + (\sin x)(2x)$

(*Bonus*: If we rewrite this rule as $f\frac{d}{dx}g = \frac{d}{dx}(fg) - g\frac{d}{dx}f$, we begin to see how to build an antiderivative (fg) for $f\frac{d}{dx}g$, by compensating with $-g\frac{d}{dx}f$. This is the essence of *integration by parts*.)

Derivative of a quotient $\frac{d}{dx}\left(\frac{f}{g}\right) = \frac{g\frac{d}{dx}f - f\frac{d}{dx}g}{g^2}$.

(If you forget this, you can just apply the product rule to fg^{-1})

Examples:

$$\frac{d}{dx}\tan x = \frac{d}{dx}\frac{\sin x}{\cos x} = \frac{(\cos x)(\cos x) - (\sin x)(-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x}$$
$$= \frac{1}{\cos^2 x} = \sec^2 x = 1 + \tan^2 x$$

(*Bonus*: if you reverse the roles, so $\tan x = y$ and $x = \arctan y$, and invert, you get a new rule: $\frac{d}{dy}\arctan y = \frac{1}{1+y^2}$.)

$$\frac{d}{dx}\sec x = \frac{d}{dx}\frac{1}{\cos x} = \frac{(\cos x)(0) - 1(-\sin x)}{\cos^2 x} = \frac{\sin x}{\cos^2 x} = (\sec x)(\tan x)$$

Chain rule $\frac{d}{dx}f(u) = \left[\frac{d}{du}f(u)\right]\left[\frac{du}{dx}\right]$. $\left(\text{Think } \frac{df(u)}{\cancel{du}}\frac{\cancel{du}}{dx}.\right)$

Examples: $$\frac{d}{dx}(7x^2 + x - 2)^{-1/4} = \left[-\frac{1}{4}(7x^2 + x - 2)^{-5/4}\right]\left[\frac{d}{dx}(7x^2 + x - 2)\right]$$
$$= -\frac{1}{4}\frac{14x+1}{(7x^2+x-2)^{5/4}}.$$
$$\frac{d}{dx}e^{x^2} = \left[e^{(x^2)}\right]\left[\frac{d}{dx}(x^2)\right] = 2xe^{x^2}.$$

Now we begin our review of integration with the *method of substitution* (also called *change of variable*), which is derived from the chain rule.

Method of Substitution

From the formula $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$ we obtain

$$\int f'\big(g(x)\big)g'(x)dx = f\big(g(x)\big) + C\,.$$

Thus if we can recognize an integrand as being of the form $f'(g(x))g'(x)$, then integration is immediate via the substitution $u = g(x)$. Here are some simple illustrations:

- $\displaystyle\int \sin^3 2x \cos 2x\, dx$

  Since $d(\sin 2x) = 2\cos 2x\, dx$, the substitution $u = \sin 2x$ reduces the integration to that of $u^3/2\, du$.

- $\displaystyle\int x^3 e^{5x^4}\, dx$

  Observe that x^3 is essentially (except for a constant factor) the derivative of $5x^4$, so we set $u = 5x^4$ and integrate $e^u/20\, du$.

- $\displaystyle\int \frac{6\ln(x+5)}{x+5}\,dx$

  Recognizing that $1/(x+5)$ is the derivative of $\ln(x+5)$, we put $u = \ln(x+5)$ and integrate $6u\, du$.

- $\displaystyle\int \tan^3 x \sec x\, dx$

  Here it is convenient to use the identity $\tan^2 x = \sec^2 x - 1$ and write the integrand as $(\sec^2 x - 1)\sec x \tan x$. Recalling that $(\sec x)' = \sec x \tan x$, the substitution $u = \sec x$ reduces the problem to integrating $(u^2 - 1)\, du$.

When a portion of the integrand involves some fractional power of an expression, an algebraic substitution can sometimes simplify the integration process. The technique is amply illu-strated in the following example.

Example 1 Find

$$(1) \qquad I_1 := \int x\sqrt{2x+1}\, dx\,.$$

Solution The difficulty here arises from the square root factor, so let's try to rid ourselves of this problem term by making the substitution $u = \sqrt{2x+1}$. Here are the details:

$$(2) \qquad u^2 = 2x + 1, \qquad x = \frac{u^2 - 1}{2}, \qquad dx = u\, du\,.$$

If we use these expressions in (1), we find that in terms of the new variable u,

$$\textbf{(3)}\qquad \int x\sqrt{2x+1}\,dx = \int\left(\frac{u^2-1}{2}\right)(u)\,(u\,du) = \frac{1}{2}\int (u^4-u^2)\,du$$
$$= \frac{1}{2}\left(\frac{u^5}{5}-\frac{u^3}{3}\right)+C$$
$$= \frac{u^3}{30}(3u^2-5)+C\,.$$

Returning to the original variable x, via $u = \sqrt{2x+1}$, we deduce that

$$I_1 = \frac{(2x+1)^{3/2}}{30}[3(2x+1)-5]+C = \frac{(2x+1)^{3/2}}{15}(3x-1)+C\,. \quad \blacklozenge$$

We remark that the substitution $u = \sqrt{2x+1}$ is not the only substitution that will allow us to compute the integral in (1). Another substitution that works just as well is $u = 2x+1$.

In a similar manner we can compute the integrals

$$\int \frac{x^3}{(1+x^2)^{3/2}}\,dx \qquad \text{and} \qquad \int \frac{e^{x^{1/3}+1}}{x^{2/3}}\,dx$$

by making the substitution $u = 1+x^2$ in the first and $u = x^{1/3}+1$ in the second.

The integration of certain algebraic expressions can sometimes be simplified by utilizing *trigonometric substitutions*. The effectiveness of these substitutions in handling quadratic expressions under a radical derives from the familiar identities

$$1-\sin^2\theta = \cos^2\theta\,, \quad 1+\tan^2\theta = \sec^2\theta, \quad \sec^2\theta - 1 = \tan^2\theta\,.$$

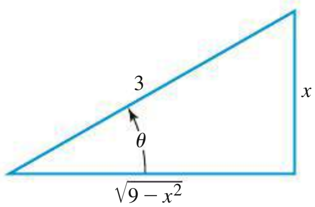
Accordingly, if the integrand involves:

$$\begin{array}{rll} \textbf{(I)} & \sqrt{a^2-x^2}, & \text{set } x = a\sin\theta, \\ \textbf{(II)} & \sqrt{a^2+x^2}, & \text{set } x = a\tan\theta, \\ \textbf{(III)} & \sqrt{x^2-a^2}, & \text{set } x = a\sec\theta. \end{array}$$

These substitutions are easy to remember if we associate with each of the cases a right triangle whose sides are a, x, and a suitable radical.† The labeling of the triangle depends on the particular case at hand. For example, in (II) it is clear that the hypotenuse must be $\sqrt{a^2+x^2}$, while in (I) the hypotenuse must be a, and in (III) the hypotenuse is x.

Example 2 Find the indefinite integral $I_2 := \displaystyle\int \frac{x^2\,dx}{\sqrt{9-x^2}}$.

Solution This integral is of type (I) with $a = 3$. We first construct a triangle as shown in Figure A.1.



$$x = 3\sin\theta,$$
$$\sqrt{9-x^2} = 3\cos\theta,$$
$$dx = 3\cos\theta\,d\theta.$$

Figure A.1 Triangle for Example 2

†The radical sign may not be visible. For example, if the integrand is $1/(a^2+x^2)^5$, this comes under case (II), because $1/(a^2+x^2)^5 = 1/(\sqrt{a^2+x^2})^{10}$.

From this triangle we see immediately that

$$I_2 = \int \frac{x^2\,dx}{\sqrt{9-x^2}} = \int \frac{(3\sin\theta)^2\,3\cos\theta\,d\theta}{3\cos\theta} = 9\int \sin^2\theta\,d\theta\,.$$

To integrate $\sin^2\theta$ we use the trigonometric identity

$$\sin^2\theta = \frac{1-\cos 2\theta}{2}$$

and obtain

$$I_2 = \frac{9}{2}\int (1-\cos 2\theta)\,d\theta = \frac{9}{2}\theta - \frac{9}{4}\sin 2\theta + C$$

$$= \frac{9}{2}\theta - \frac{9}{2}\sin\theta\cos\theta + C\,.$$

Returning to the original variable x, we can derive the expressions for θ, $\sin\theta$, and $\cos\theta$ directly from the triangle in Figure A.1. We find that

$$\theta = \arcsin\frac{x}{3}, \quad \sin\theta = \frac{x}{3}, \quad \cos\theta = \frac{\sqrt{9-x^2}}{3}.$$

Hence

$$I_2 = \frac{9}{2}\arcsin\frac{x}{3} - \frac{9}{2}\frac{x}{3}\frac{\sqrt{9-x^2}}{3} + C = \frac{9}{2}\arcsin\frac{x}{3} - \frac{1}{2}x\sqrt{9-x^2} + C\,. \quad ◆$$

Integration by Parts

The method of integration by parts is of both practical and theoretical importance. It is based on the product rule for differentiation which we now write in the format

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx} = uv' + vu'\,.$$

Integrating both sides of this equation and transposing, we obtain

$$\int uv'\,dx = uv - \int vu'\,dx\,. \tag{4}$$

Since $dv = v'dx$ and $du = u'dx$, equation (4) can be put in the more compact form

$$\int u\,dv = uv - \int v\,du\,. \tag{5}$$

Equation (5) is the basic equation for integration by parts. The idea is that the integrand on the left side may be very complicated, but if we select the part for u and the part for dv quite carefully, the integral on the right side may be much easier. (Some experimentation may be necessary to find the right selection.)

Example 3 Find the indefinite integral $I_3 := \int x \ln x \, dx$.

Solution There are two natural ways of selecting u and dv so that the product is $x \ln x \, dx$:

First possibility:	let $u = x$,	$dv = \ln x \, dx$.
Second possibility:	let $u = \ln x$,	$dv = x \, dx$.

With the first selection we cannot easily find v, the integral of $\ln x$, so we come to a dead end. In the second case we can find v. We arrange the work as follows. Let

$$u = \ln x, \quad dv = x \, dx .$$

Then

$$du = \frac{1}{x} dx, \quad v = \frac{x^2}{2} .$$

Making these substitutions in (5), we have for the left side

$$\int u \, dv = \int (\ln x)(x \, dx) = \int x \ln x \, dx .$$

and for the right side

$$uv - \int v \, du = \frac{x^2}{2} \ln x - \int \frac{x^2}{2} \frac{1}{x} dx = \frac{x^2}{2} \ln x - \frac{1}{2} \int x \, dx . \tag{6}$$

Therefore, by (5),

$$I_3 = \int x \ln x \, dx = \frac{x^2}{2} \ln x - \frac{1}{4} x^2 + C . \ \blacklozenge \tag{7}$$

Note that we used $v = x^2/2$ rather than $v = x^2/2 + C$. It is clear that equation (5) is correct for every value of C, and in most cases it is simpler to set $C = 0$. The arbitrary constant that we need can be added in the last step, as we have done in going from equation (6) to (7).

It is sometimes necessary to apply integration by parts more than once in order to evaluate a particular integral, such as

$$I_4 := \int x^2 e^x \, dx .$$

In the first application, we set $u = x^2$, $dv = e^x \, dx$ to deduce that

$$I_4 = x^2 e^x - 2 \int x e^x \, dx \tag{8}$$

and to evaluate the last integral in (8) we perform a second application of integration by parts with $u = x$, $dv = e^x \, dx$.

There are some integrals that may appear at first not to be split into two parts, but we need to keep in mind that "1" can be considered as a factor of any expression. For example, to compute the integrals

$$I_5 := \int \arctan(3x) \, dx \qquad \text{and} \qquad I_6 := \int \ln x \, dx .$$

an effective splitting for evaluating I_5 is $u = \arctan(3x)$, $dv = 1 \, dx$, while for I_6 an appropriate choice is $u = \ln x$, $dv = 1 \, dx$. The reader should verify that integration by parts then reduces these integrals to routine problems.

Partial Fractions

Any rational function, that is, any function of the form $R(x) = P(x)/Q(x)$, where $P(x)$ and $Q(x)$ are polynomials, can be integrated in finite terms. The first step in the process is to compare the degree of $P(x)$ with that of $Q(x)$. If deg $P \geq$ deg Q, then division gives

$$\frac{P(x)}{Q(x)} = s(x) + \frac{p(x)}{Q(x)}, \tag{9}$$

where $s(x)$ and $p(x)$ are polynomials and deg $p <$ deg Q. Since the integration of the polynomial $s(x)$ is a trivial matter, we need only focus on the integration of the remainder term $p(x)/Q(x)$.

The idea of the method of partial fractions is to express $p(x)/Q(x)$ as the sum of simpler fractions that are easy to integrate. For example, knowing that

$$R(x) = \frac{7x^2 + 10x - 1}{x^3 + 3x^2 - x - 3} \tag{10}$$

can be written as

$$\frac{2}{x-1} + \frac{1}{x+1} + \frac{4}{x+3}$$

leads immediately to the evaluation

$$\int R(x)\,dx = 2\ln|x-1| + \ln|x+1| + 4\ln|x+3| + C\,.$$

The key to finding such a simple fraction representation is to first factor the denominator polynomial $Q(x)$. Indeed, in (10) the denominator is just the product $(x-1)(x+1)(x+3)$. In general, we recall that any polynomial $Q(x)$ with real coefficients can be factored into a product of linear and irreducible quadratic factors with real coefficients. There are four cases that may arise in the factorization: **(i)** nonrepeated linear factors; **(ii)** repeated linear factors; **(iii)** nonrepeated quadratic factors; and **(iv)** repeated quadratic factors.

In the case when a nonrepeated linear factor $(x-a)$ occurs in the factorization of $Q(x)$, we associate the partial fraction $A/(x-a)$, where A is a constant to be determined. If a linear factor $(x-a)$ is repeated m times in the factorization of $Q(x)$, we associate the sum of m fractions

$$\sum_{k=1}^{m} \frac{A_k}{(x-a)^k},$$

where the m constants A_k are to be determined. In the case of a nonrepeated quadratic factor $ax^2 + bx + c$ (where $b^2 - 4ac < 0$) we associate the single fraction

$$\frac{Ax + B}{ax^2 + bx + c},$$

where we now have two constants in the numerator to be determined, while if this quadratic factor is repeated m times, then we associate the sum

$$\sum_{k=1}^{m} \frac{A_k x + B_k}{(ax^2 + bx + c)^k},$$

which involves the determination of $2m$ constants A_k and B_k.

We illustrate the approach in the following example. Further discussion of partial fractions appears in Section 7.4, page 370.

Example 4 Find the indefinite integral $I_7 = \displaystyle\int \frac{3x^3 + 3x^2 + 3x + 2}{x^3(x+1)}\,dx$.

Solution Since the factor x is repeated three times in the denominator, while $(x+1)$ is a nonrepeated factor, our partial fraction decomposition must be

$$\frac{3x^3 + 3x^2 + 3x + 2}{x^3(x+1)} = \frac{A}{x+1} + \frac{B}{x} + \frac{C}{x^2} + \frac{D}{x^3}, \tag{11}$$

where A, B, C, and D are unknowns to be determined. Multiplying both sides of (11) by $x^3(x+1)$, we have

$$3x^3 + 3x^2 + 3x + 2 = Ax^3 + Bx^2(x+1) + Cx(x+1) + D(x+1)\,. \tag{12}$$

In order for (12) to be true for all x, the corresponding coefficients must be the same:

$$\begin{aligned}
&\text{The coefficients of } x^3 \text{ yield} \qquad & A + B &= 3. \\
&\text{The coefficients of } x^2 \text{ yield} \qquad & B + C &= 3. \\
&\text{The coefficients of } x^1 \text{ yield} \qquad & C + D &= 3. \\
&\text{The coefficients of } x^0 \text{ yield} \qquad & D &= 2.
\end{aligned}$$

Solving this system of four linear equations in four unknowns gives $A = 1$, $B = 2$, $C = 1$, and $D = 2$. Hence

$$I_7 = \int\left(\frac{1}{x+1} + \frac{2}{x} + \frac{1}{x^2} + \frac{2}{x^3}\right)dx = \ln\left|(x+1)x^2\right| - \frac{1}{x} - \frac{1}{x^2} + C\,. \ ◆$$

We remark that the coefficients A and D in the representation (11) can be determined directly. Indeed, if we set $x = 0$ in (12), we obtain $2 = D$, and if we set $x = -1$ in (12), we obtain $-3 + 3 - 3 + 2 = A(-1)$ or $A = 1$. But B and C cannot be obtained so readily. If we are fortunate to have a denominator with only nonrepeated linear factors, then this simple substitution method will work to quickly determine all the unknown coefficients.

In dealing with the integration of rational functions that have a quadratic factor in the denominator, the familiar formula

$$\int \frac{1}{1+x^2}\,dx = \arctan x + C$$

plays a central role. For example, by completing the square and making a simple change of variables, we find

$$\int \frac{3}{x^2 + 4x + 5}\,dx = \int \frac{3}{(x+2)^2 + 1}\,dx = 3\arctan(x+2) + C\,.$$

A EXERCISES

In Problems 1–26, find the indicated indefinite integral.

1. $\displaystyle\int x^3(1+x^2)^{3/2}\,dx$

2. $\displaystyle\int x\cos(2x)\,dx$

3. $\displaystyle\int 4t^4 e^{6t^5}\,dt$

4. $\displaystyle\int x^3 e^{3x}\,dx$

5. $\displaystyle\int \cos^2(\pi\theta)\,d\theta$

6. $\displaystyle\int \frac{x^{1/3}}{x^{1/3}+1}\,dx$

7. $\displaystyle\int \frac{t+1}{t^2+4}\,dt$

8. $\displaystyle\int 10\sec^2(5x)\,dx$

9. $\displaystyle\int \frac{1}{\sqrt{1-2x^2}}\,dx$

10. $\displaystyle\int \frac{1}{2x^2+4x+4}\,dx$

11. $\displaystyle\int \frac{4}{(x-1)(x-2)(x-3)}\,dx$

12. $\displaystyle\int \sec(2x)\tan(2x)\;dx$

13. $\displaystyle\int \frac{\sqrt{9x^2-1}}{x}\,dx$

14. $\displaystyle\int \frac{2x+41}{x^2+5x-14}\,dx$

15. $\displaystyle\int \frac{x^3+2x^2+8}{x^2(x^2+4)}\,dx$

16. $\displaystyle\int \frac{x^3+4x+1}{x^3+3x^2+4x+2}\,dx$

17. $\displaystyle\int \frac{1}{x\ln x}\,dx$

18. $\displaystyle\int \frac{x^2}{x^2-1}\,dx$

19. $\displaystyle\int y\sinh y\;dy$

20. $\displaystyle\int \sec(3\theta+1)\;d\theta$

21. $\displaystyle\int \frac{1}{(x^2+4)^{3/2}}\,dx$

22. $\displaystyle\int 36x^5\exp(2x^3)\;dx$

23. $\displaystyle\int t\ln(t+3)\;dt$

24. $\displaystyle\int \sin\sqrt{x}\,dx$

25. $\displaystyle\int \frac{\sin t}{1+\cos^2 t}\,dt$

26. $\displaystyle\int \frac{e^{3y}}{1+e^{6y}}\,dy$

In Problems 27–34, use an appropriate trigonometric identity to help find the indicated integral.

27. $\displaystyle\int \sin^{2/5}x\cos^3 x\,dx$

28. $\displaystyle\int \tan^3 x\sqrt{\sec x}\,dx$

29. $\displaystyle\int \cos(3x)\cos(7x)\;dx$

30. $\displaystyle\int \cot^3\theta\;d\theta$

31. $\displaystyle\int \tan^4\theta\sec^6\theta\;d\theta$

32. $\displaystyle\int \cos^{1/3}x\sin^3 x\,dx$

33. $\displaystyle\int \sin^2(3x)\cos^2(3x)\;dx$

34. $\displaystyle\int \cot^2 x\csc^4 x\,dx$

APPENDIX B

Newton's Method

To solve an equation $g(x) = 0$, we must find the point or points where the graph of $y = g(x)$ meets the x-axis. One procedure for approximating a solution is **Newton's method.**

To motivate Newton's method geometrically, we let $\tilde{x}$ be a root of $g(x) = 0$ and let x_1 be our guess at the value of $\tilde{x}$. If $g(x_1) = 0$, we are done. If $g(x_1) \neq 0$, then we are off by some amount that we call dy (see Figure B.1). Then

$$\frac{dy}{dx} = g'(x_1) ,$$

and so

$$dx = \frac{dy}{g'(x_1)} . \tag{1}$$

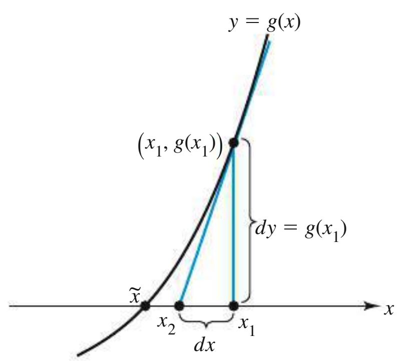


Figure B.1 Tangent line approximation of root

Now $dx = x_1 - x_2$ or $x_2 = x_1 - dx$, where x_2 is at the point where the tangent line through $(x_1, g(x_1))$ intersects the x-axis (Figure B.1). Using equation (1) and the fact that $dy = g(x_1)$, we obtain

$$x_2 = x_1 - \frac{dy}{g'(x_1)} = x_1 - \frac{g(x_1)}{g'(x_1)},$$

which we use as the next approximation to the root $\tilde{x}$.

Repeating this process with x_2 in place of x_1, we obtain the next approximation x_3 to the root $\tilde{x}$. In general, we find the next approximation x_{n+1} by the formula

$$\textbf{(2)} \qquad x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}, \quad n = 1, 2, \ldots .$$

The process is illustrated in Figure B.2.

If the initial guess x_1 is sufficiently close to a root $\tilde{x}$, then the sequence of iterations $\{x_n\}_{n=1}^{\infty}$ usually converges to the root $\tilde{x}$. However, if we make a bad guess for x_1, then the process may lead away from $\tilde{x}$.

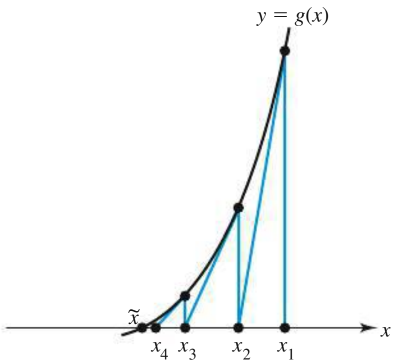


Figure B.2 Sequence of iterations converging to root

Example 1 Find a root to four decimal places of the equation

$$\textbf{(3)} \qquad x^3 + 2x - 4 = 0 .$$

Solution Setting $g(x) = x^3 + 2x - 4$, we find that $g'(x) = 3x^2 + 2$ is positive for all x. Hence, g is increasing and has at most one zero. Furthermore, since $g(1) = -1$ and $g(2) = 8$, this zero must lie between 1 and 2. Thus, we begin the procedure with the initial guess $x_1 = 1.5$. For $g(x) = x^3 + 2x - 4$, equation (2) becomes

$$\textbf{(4)} \qquad x_{n+1} = x_n - \frac{x_n^3 + 2x_n - 4}{3x_n^2 + 2}, \quad n = 1, 2, \ldots .$$

With $x_1 = 1.5$, equation (4) gives

$$x_2 = 1.5 - \frac{(1.5)^3 + 2(1.5) - 4}{3(1.5)^2 + 2} = 1.5 - \frac{2.375}{8.75} \approx 1.22857 .$$

Using x_2 to compute x_3 and so on, we find

$$x_3 = 1.18085\,,$$

$$x_4 = 1.17951\,,$$

$$x_5 = 1.17951\,,$$

where we have rounded off the computations to five decimal places. Since x_4 and x_5 agree to four decimal places and we are uncertain of the fifth decimal place because of roundoff, we surmise that the root $\tilde{x}$ of (3) agrees with 1.1795 to four decimal places. Indeed,

$$g(1.1795) = -0.00005\ldots \quad \text{and} \quad g(1.1796) = 0.00056\ldots\,,$$

so $1.1795 < \tilde{x} < 1.1796$. Consequently, $\tilde{x} = 1.1795\ldots$. ◆

Observe that Newton's method transforms the problem of finding a root to the equation $g(x) = 0$ into the problem of finding a fixed point for the function $h(x) = x - g(x)/g'(x)$; that is, finding a number x such that $x = h(x)$. [See equation (2).]

Several theorems give conditions that guarantee that the sequence of iterations $\{x_n\}_{n=1}^{\infty}$ defined by (2) will converge to a zero of $g(x)$. We mention one such result.

Convergence of Newton's Method

Theorem 1. Suppose a zero $\tilde{x}$ of $g(x)$ lies in the interval (a, b) and that in this interval

$$g'(x) > 0 \quad \text{and} \quad g''(x) > 0\,.$$

If we select x_1 so that $\tilde{x} < x_1 < b$, then the sequence of iterations defined by (2) will decrease to $\tilde{x}$.

We do not give a proof of this theorem, but refer the reader to an introductory numerical analysis text such as *Numerical Analysis,* 10th ed., by R. Burden, J. Faires, and A. Burden (Cengage Learning, 2016).

APPENDIX C

Simpson's Rule

A useful procedure for approximating the value of a definite integral is **Simpson's rule.**

Let the interval $[a, b]$ be divided into $2n$ equal parts and let $x_0, x_1, \ldots, x_{2n}$ be the points of the partition, that is,

$$x_k := a + kh \ , \quad k = 0, 1, \ldots, 2n \ ,$$

where $h := (b - a)/(2n)$. If

$$y_k := f(x_k) \ , \quad k = 0, 1, \ldots, 2n \ ,$$

then the Simpson's rule approximation I_S for the value of the definite integral

$$\int_a^b f(x)\,dx$$

is given by

$$\textbf{(1)}\qquad I_S = \frac{h}{3}\left[y_0 + 4y_1 + 2y_2 + 4y_3 + \cdots + 2y_{2n-2} + 4y_{2n-1} + y_{2n}\right]$$

$$= \frac{h}{3}\sum_{k=1}^{n}\left(y_{2k-2} + 4y_{2k-1} + y_{2k}\right).$$

If

$$E := \int_a^b f(x)\,dx - I_S$$

is the error that results from using Simpson's rule to approximate the value of the definite integral, then

$$\textbf{(2)}\qquad |E| \le \frac{(b-a)}{180}h^4 M\,,$$

where $M := \max\left|f^{(4)}(x)\right|$ for all x in $[a, b]$.

Example 1 Use Simpson's rule with $n = 4$ to approximate the value of the definite integral

$$\textbf{(3)}\qquad \int_0^1 \frac{1}{1+x^2}\,dx\,.$$

Solution Here $h = 1/8$, $x_k = k/8$, $k = 0, 1, \ldots, 8$, and

$$y_k = \left(1 + x_k^2\right)^{-1} = \frac{1}{1 + \dfrac{k^2}{64}} = \frac{64}{64 + k^2}\,.$$

By Simpson's rule (1), we find

$$I_S = \frac{\left(\frac{1}{8}\right)}{3}\left[1 + 4\left(\frac{64}{64+1}\right) + 2\left(\frac{64}{64+4}\right) + 4\left(\frac{64}{64+9}\right)\right.$$
$$+ 2\left(\frac{64}{64+16}\right) + 4\left(\frac{64}{64+25}\right) + 2\left(\frac{64}{64+36}\right)$$
$$\left. + 4\left(\frac{64}{64+49}\right) + \left(\frac{64}{64+64}\right)\right] = 0.7854\,.$$

Hence, the value of the definite integral in (3) is approximately $I_S = 0.7854$. ◆

For a more detailed discussion of Simpson's rule, we refer the reader to a numerical analysis book such as *Numerical Analysis,* 10th ed., by R. Burden, J. Faires, and A. Burden (Cengage Learning, 2016).

APPENDIX D Cramer's Rule

When a system of n linear equations in n unknowns has a unique solution, determinants can be used to obtain a formula for the unknowns. This procedure is called **Cramer's rule.** When n is small, these formulas provide a simple procedure for solving the system.

Suppose that for a system of n linear equations in n unknowns,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 , \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 , \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \quad &\quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n , \end{aligned} \tag{1}$$

the coefficient matrix

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

has a nonzero determinant. Then Cramer's rule gives the solutions

$$x_i = \frac{\det \mathbf{A}_i}{\det \mathbf{A}} , \qquad i = 1, 2, \ldots, n , \tag{3}$$

where $\mathbf{A}_i$ is the matrix obtained from $\mathbf{A}$ by replacing the ith column of $\mathbf{A}$ by the column vector

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

consisting of the constants on the right-hand side of system (1).

Example 1 Use Cramer's rule to solve the system

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 0 , \\ 2x_1 + x_2 + x_3 &= 9 , \\ x_1 - x_2 - 2x_3 &= 1 . \end{aligned}$$

Solution We first compute the determinant of the coefficient matrix:

$$\det \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -1 & -2 \end{bmatrix} = 12 .$$

Using formula (3), we find

$$x_1 = \frac{1}{12} \det \begin{bmatrix} 0 & 2 & -1 \\ 9 & 1 & 1 \\ 1 & -1 & -2 \end{bmatrix} = \frac{48}{12} = 4 ,$$

$$x_2 = \frac{1}{12} \det \begin{bmatrix} 1 & 0 & -1 \\ 2 & 9 & 1 \\ 1 & 1 & -2 \end{bmatrix} = \frac{-12}{12} = -1 ,$$

$$x_3 = \frac{1}{12} \det \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 9 \\ 1 & -1 & 1 \end{bmatrix} = \frac{24}{12} = 2 . \ \blacklozenge$$

For a more detailed discussion of Cramer's rule, we refer the reader to *Fundamentals of Matrix Analysis with Applications*, by Edward Barry Saff and Arthur David Snider (John Wiley & Sons, Hoboken, New Jersey, 2016).

APPENDIX E Method of Least Squares

The **method of least squares** is a procedure for fitting a straight line to a set of measured data. Consider the *scatter diagram* in Figure E.1, consisting of an x, y-plot of some data points $\{(x_i, y_i): i = 1, 2, \ldots, N\}$. It is desired to construct the straight line $y = \alpha + \beta x$ that best fits the data points, in the sense that the sum of the squares of the vertical deviations from the points to the line is minimized.

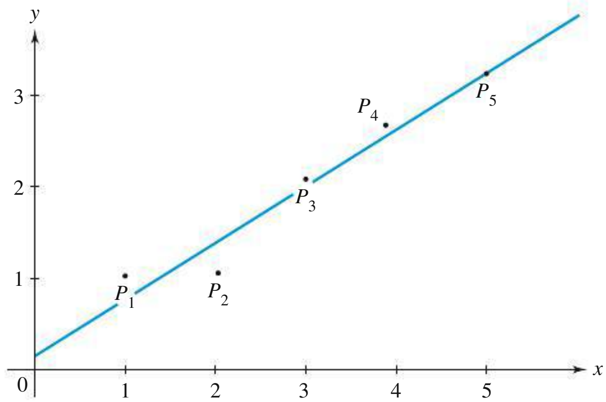


Figure E.1 Scatter diagram and least-squares linear fit

If the y-axis intercept α and the slope β of the line are known, then the y-value on the line corresponding to the measured x-value x_i is given by $\alpha + \beta x_i$. The corresponding measured y-value, y_i, thus deviates from the line by $[y_i - (\alpha + \beta x_i)]$, and the total sum of squares of deviations is

$$S := \sum_{i=1}^{N} [y_i - (\alpha + \beta x_i)]^2 .$$

The interesting feature of the function S is that the symbols x_i and y_i are constants, while α and β are the (unknown) variables. The values of α and β that minimize S will force its partial derivatives to be zero:

$$0 = \frac{\partial S}{\partial \alpha} = \sum_{i=1}^{N} 2[y_i - (\alpha + \beta x_i)](-1) ,$$

$$0 = \frac{\partial S}{\partial \beta} = \sum_{i=1}^{N} 2[y_i - (\alpha + \beta x_i)](-x_i) .$$

Displaying these conditions in a form that emphasizes the roles of α and β, we rewrite them after a little algebra as

$$\alpha N + \beta \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i ,$$

$$\alpha \sum_{i=1}^{N} x_i + \beta \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i$$

and obtain the formulas for the optimal values of intercept and slope:

$$\alpha = \frac{\left(\sum_{i=1}^{N} x_i^2\right)\left(\sum_{i=1}^{N} y_i\right) - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} x_i y_i\right)}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2},$$

$$\beta = \frac{N\sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}.$$

Example 1 Find the least-squares linear fit to the data points $P_1(1, 1)$, $P_2(2, 1)$, $P_3(3, 2)$, $P_4(4, 2.5)$, and $P_5(5, 3.1)$, which are plotted in Figure E.1.

Solution Arranging the data as in Table E.1 yields

$$\alpha = \frac{55(9.6) - 15(34.5)}{5(55) - (15)^2} = \frac{10.5}{50} = 0.21\,,$$

$$\beta = \frac{5(34.5) - 15(9.6)}{5(55) - (15)^2} = \frac{28.5}{50} = 0.57\,.$$

TABLE E.1

i	x_i	y_i	$x_i y_i$	x_i^2
1	1	1	1	1
2	2	1	2	4
3	3	2	6	9
4	4	2.5	10	16
5	5	3.1	15.5	25
Sums	15	9.6	34.5	55

Thus, $y = 0.21 + 0.57x$ is the equation for the best-fitting line, which is graphed in Figure E.1. ◆

APPENDIX F Runge–Kutta Procedure for n Equations

Below are program outlines for the classical fourth-order Runge–Kutta algorithm for systems discussed in Section 5.3. The first of these, called the "subroutine," provides approximations to the solution functions of a system of n first-order ordinary differential equations over an interval. The second outline on page A-17 utilizes the method of halving the step size in order to obtain approximations (to within a prescribed tolerance) to the solution functions at a single given point.

Classical Fourth-Order Runge–Kutta Subroutine (n Equations)

Purpose To approximate the solution to the initial value problem

$$x_i' = f_i(t, x_1, \ldots, x_n)\ ;$$
$$x_i(t_0) = a_i\ , \qquad i = 1, 2, \ldots, n$$

for $t_0 \le t \le c$.

INPUT $n, t_0, a_1, \ldots, a_n, c, N$ (number of steps), PRNTR ($= 1$ to print table).

Step 1 Set step size

$$h = (c - t_0)/N\ , \quad t = t_0\ , \quad x_1 = a_1\ , \ \ldots\ , \quad x_n = a_n\ .$$

Step 2 For $j = 1$ to N do Steps 3–5.

Step 3 Set

$$k_{i,1} = hf_i(t, x_1, \ldots, x_n)\ ,$$
$$i = 1, \ldots, n\ ;$$
$$k_{i,2} = hf_i\left(t + \frac{h}{2}, x_1 + \frac{1}{2}k_{1,1}, \ldots, x_n + \frac{1}{2}k_{n,1}\right),$$
$$i = 1, \ldots, n\ ;$$
$$k_{i,3} = hf_i\left(t + \frac{h}{2}, x_1 + \frac{1}{2}k_{1,2}, \ldots, x_n + \frac{1}{2}k_{n,2}\right),$$
$$i = 1, \ldots, n\ ;$$
$$k_{i,4} = hf_i(t + h, x_1 + k_{1,3}, \ \ldots, x_n + k_{n,3})\ ,$$
$$i = 1, \ldots, n\ .$$

Step 4 Set

$$t = t + h\ ;$$
$$x_i = x_i + \frac{1}{6}(k_{i,1} + 2k_{i,2} + 2k_{i,3} + k_{i,4})\ , \quad i = 1, \ldots, n\ .$$

Step 5 If PRNTR $= 1$, print $t, x_1, x_2, \ldots, x_n$.

Classical Fourth-Order Runge–Kutta Algorithm with Tolerance (n Equations)

Purpose To approximate the solution to the initial value problem

$$x_i' = f_i(t, x_1, \ldots, x_n)\ ;$$
$$x_i(t_0) = a_i\,, \quad i = 1, 2, \ldots, n$$

at $t = c$, with tolerance ε.

INPUT $n, t_0, a_1, \ldots, a_n, c$
ε (tolerance)
M (maximum number of iterations)

Step 1 Set $z_i = a_i$, $\quad i = 1, 2, \ldots, n$; set PRNTR $= 0$.
Step 2 For $m = 0$ to M do Steps 3–7 (or, to save time, start with $m > 0$).
Step 3 Set $N = 2^m$.
Step 4 Call FOURTH-ORDER RUNGE–KUTTA SUBROUTINE (n EQUATIONS).
Step 5 Print $h, x_1, x_2, \ldots, x_n$.
Step 6 If $|z_i - x_i| < \varepsilon$ for $i = 1, \ldots, n$, go to Step 10.
Step 7 Set $z_i = x_i\,, \quad i = 1, \ldots, n$.
Step 8 Print "$x_i(c)$ is approximately"; x_i (for $i = 1, \ldots, n$); "but may not be within the tolerance"; ε.
Step 9 Go to Step 11.
Step 10 Print "$x_i(c)$ is approximately"; x_i (for $i = 1, \ldots, n$); "with tolerance"; ε.
Step 11 STOP.
OUTPUT Approximations of the solution to the initial value problem at $t = c$ using 2^m steps.

APPENDIX G Software for Analyzing Differential Equations

In this section we shall list some commercial software and some freeware that we have found to be useful in the various aspects of analyzing systems of differential equations numerically: direction field plotting, Euler and Runge–Kutta codes for solutions, and phase plane experimentation. These items are available at the time of printing this edition, but the accessibility of freeware, in particular, is notoriously tenuous. Each of the web sites mentioned was successfully accessed on July 6, 2016.

Some ***commercial products*** that are helpful:

MATLAB® (Mathworks, 1 Apple Hill Drive, Natick, MA 01760-2098 USA) (http://www.mathworks.com/products/matlab/):

"**The MATLAB ODE Suite**" at
http://www.mathworks.com/videos/solving-odes-in-matlab-9-the-matlab-ode-suite-117653.html?s_tid=srchtitle
"**Choose an ODE Solver**" at
http://www.mathworks.com/help/matlab/math/choose-an-ode-solver.html
"**Vector Fields**" at
http://www.mathworks.com/help/matlab/vector-fields.html .

MATHEMATICA® (Wolfram Research, 100 Trade Center Dr, Champaign, IL 61820 USA)
(https://www.wolfram.com/mathematica/):

"**How to Solve a Differential Equation**" at
https://reference.wolfram.com/language/howto/SolveADifferentialEquation.html
"**How to Plot a Direction Field**" at
https://reference.wolfram.com/language/howto/PlotAVectorField.html .

MAPLE® (Maplesoft, 615 Kumpf Drive, Waterloo, ON N2V 1K8 CAN)
(http://www.maplesoft.com/).

"**dsolve**" at
https://www.maplesoft.com/support/help/maple/view.aspx?path=dsolve
"**dfieldplot**" at
https://www.maplesoft.com/support/help/maple/view.aspx?path=DEtools%2Fdfieldplot .

PTC Mathcad® (Parametric Technology Corporation, 140 Kendrick Street, Needham, MA 02494 USA)
(http://www.ptc.com/engineering-math-software/mathcad).

"**Differential Equations worksheet**" in
PTC Mathcad Worksheet Library - Applied Math.

Some ***freeware sources*** that are helpful:

Casio Computer Co. Ltd® (6-2, Hon-machi 1-chome, Shibuya-ku, Tokyo 151-8543, Japan)
(http://world.casio.com/)

"**Euler's Method Calculator**" at
http://keisan.casio.com/exec/system/1392171850
"**Runge–Kutta method (2nd-order) Calculator**" at
http://keisan.casio.com/exec/system/1392171606
"**Runge–Kutta method (4th-order) Calculator**" at
http://keisan.casio.com/exec/system/1222997077

The codes *dfield* and *pplane* are graciously provided by our differential equations colleague John C. Polking of the Department of Mathematics at Rice University. He holds a copyright and they are not in the public domain; however, they are being made available free for use in educational institutions.

"**dfield and pplane: The Java Versions**" (Polking, J. C. and Castellanos, J.) at
http://math.rice.edu/~dfield/dfpp.html

The Dynamic Web Tools site provides codes dedicated to the FitzHugh-Nagumo equation:

"**FitzHugh-Nagumo System of Differential Equations**" (Martin, M.) at
http://math.jccc.edu:8180/webMathematica/JSP/mmartin/fitznag.jsp

The MIT Interactive Mathematics Site provides the following dedicated codes:

"**Vector Fields**" (Miller, H., MIT Mathlets) at
http://mathlets.org/mathlets/vector-fields/
"**Linear Phase Portraits: Matrix Entry**" (Miller, H., MIT Mathlets) at
http://mathlets.org/mathlets/linear-phase-portraits-matrix-entry/

Answers to Odd-Numbered Problems

CHAPTER 1

Exercises 1.1, page 5

1. ODE, 2nd-order, ind. var. t, dep. var. x, linear
3. ODE, 1st-order, ind. var. x, dep. var. y, nonlinear
5. ODE, 1st-order, ind. var. x, dep. var. y, nonlinear
7. ODE, 1st-order, ind. var. t, dep. var. p, nonlinear
9. ODE, 2nd-order, ind. var. x, dep. var. y, linear
11. PDE, 2nd-order, ind. var. t, r, dep. var. N
13. $dp/dt = kp$, where k is the proportionality constant
15. $dT/dt = k(M - T)$, where k is the proportionality constant
17. Kevin wins by $6\sqrt{3} - 4\sqrt{6} \approx 0.594$ sec.

Exercises 1.2, page 13

3. Yes **5.** No **7.** Yes **9.** No **11.** Yes
13. Yes **19.** The left-hand side is always ≥ 4.
21. **(a)** $1/3, -3$ **(b)** $1 \pm \sqrt{6}$ **23.** Yes **25.** Yes
27. No **29.** **(b)** Yes **31.** **(a)** No
(c) $y = 2x$ and $y = -2x$

Exercises 1.3, page 21

1. **(b)**

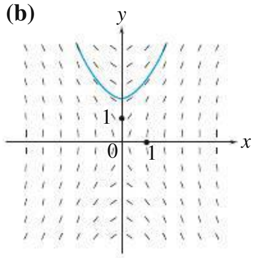


Figure B.1
Solution to Problem 1(b)

(c)

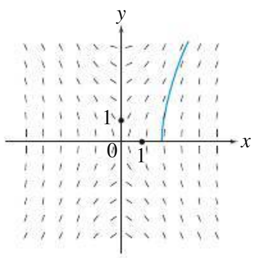


Figure B.2
Solution to Problem 1(c)

(d) The solutions in parts (b) and (c) both become infinite and have the line $y = 2x$ as an asymptote as $x \to \infty$. As $x \to -\infty$, the solution in part (b) becomes infinite and has $y = -2x$ as an asymptote, but the solution in part (c) does not even exist for x negative.

3. All solutions have limiting value 8 as $t \to +\infty$.

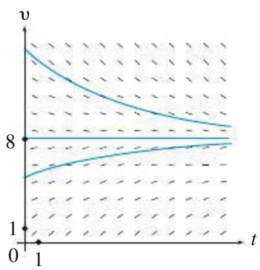


Figure B.3 Solutions to Problem 3 satisfying $v(0) = 5$, $v(0) = 8$, and $v(0) = 15$

5. **(a)**

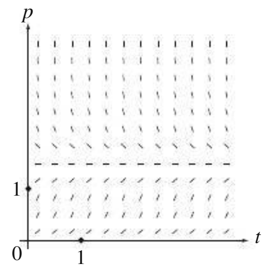


Figure B.4 Direction field for Problem 5(a)

(b) $\lim_{t\to\infty} p(t) = 3/2$ **(c)** $\lim_{t\to\infty} p(t) = 3/2$
(d) No

7. **(a)**

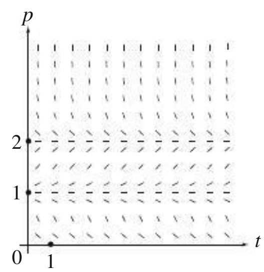


Figure B.5 Direction field for Problem 7(a)

(b) 2 **(c)** 2 **(d)** 0 **(e)** No

9. **(d)** It increases and asymptotically approaches the line $y = x - 1$.
(f) and **(g)**

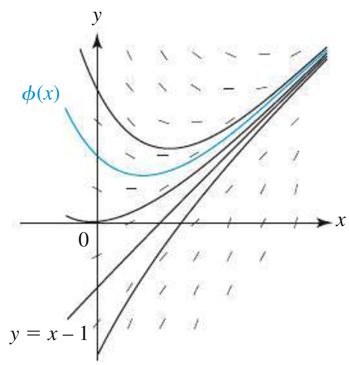


Figure B.6 Direction field and sketch of $\phi(x)$ for Problems 9 (f), (g)

11.

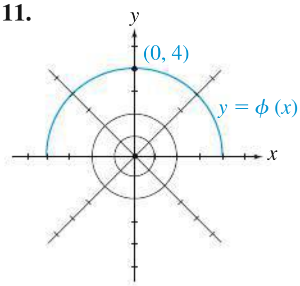


Figure B.7 Solution to Problem 11

13.

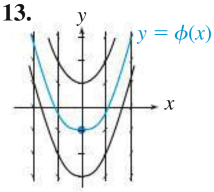


Figure B.8 Solution to Problem 13

15.

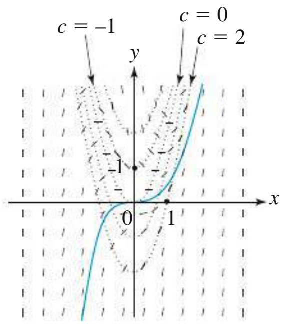


Figure B.9 Solution to Problem 15

17. It approaches 3.

Exercises 1.4, page 28

1. (rounded to three decimal places)

x_n	0.1	0.2	0.3	0.4	0.5
y_n	4.000	3.998	3.992	3.985	3.975

3. (rounded to three decimal places)

x_n	0.1	0.2	0.3	0.4	0.5
y_n	1.100	1.220	1.362	1.528	1.721

5. (rounded to three decimal places)

x_n	1.1	1.2	1.3	1.4	1.5
y_n	0.100	0.209	0.325	0.444	0.564

7. (rounded to three decimal places)

N	h	y_N
1	π	3.142
2	$\pi/2$	1.571
4	$\pi/4$	1.207
8	$\pi/8$	1.148

9.

x_n	y_n
1.1	−0.9
1.2	−0.81654
1.3	−0.74572
1.4	−0.68480
1.5	−0.63176
1.6	−0.58511
1.7	−0.54371
1.8	−0.50669
1.9	−0.47335
2.0	−0.44314

11. $x(1)$ stabilizes at 1.56 ± 0.01 for h below 2^{-10}; $[\tan(1) = 1.5574\ldots]$ $x(t_0) = 1$ for some t_0 in $(0.78, 0.80)$; $[\arctan(1) = 0.785\ldots]$.

15. **(a)** $T(30) = 311.7$ **(b)** $T(60) = 298.2$

Review Problems, page 29

1. ind. var. t, dep. var. x, nonlinear

3. ind. var. y, dep. var. x, linear

5. ind. var. t, dep. var. v, linear

7. True **9.** False **11.** True **13.** True

15. $y_n = \left(1 - \dfrac{1}{n}\right)^n$, $y(1) = e^{-1} \approx 0.3679$

17. **(a)** 500 **(b)** No **(c)** 500

CHAPTER 2

Exercises 2.2, page 46

1. No **3.** Yes **5.** Yes

7. $y = \pm\sqrt[4]{\ln(x^4) + C}$

9. $e^{2x}(2x-1) + 4e^{-t}(t+1) = C$

11. $v = \pm\dfrac{\sqrt{1 - Cx^{-8/3}}}{2}$

13. $\dfrac{y}{\sqrt{1+y^2}} - x^3 = C$

15. $y = \pm\sqrt{C\exp[\exp(-x^2)] - 1}$

17. $y = \tan\left(-\ln|\cos x| + \dfrac{\pi}{3}\right)$

19. $y = (\sin x + 1)^2 - 1$

21. $y^2/2 + \ln y = \sin\theta - \theta\cos\theta + 1/2 - \pi$

23. $y = \arctan(t^2 + 1)$

25. $y = 4e^{x^3/3} - 1$

27. **(a)** $y(x) = \displaystyle\int_0^x e^{t^2}\,dt$

(b) $y(x) = \left(1 + 3\displaystyle\int_0^x e^{t^2}\,dt\right)^{1/3}$

(c) $y(x) = \tan\left(\displaystyle\int_0^x \sqrt{1 + \sin t}\,dt + \pi/4\right)$

(d) $y(0.5) \approx 1.381$

29. **(d)** $\partial f/\partial y$ is not continuous at $(0, 0)$.

31. **(a)** $x^2 + y^{-2} = C$

(b) $\dfrac{1}{\sqrt{1-x^2}}$; $\dfrac{1}{\sqrt{4-x^2}}$; $\dfrac{1}{\sqrt{1/4 - x^2}}$

(c) $-1 < x < 1$; $-2 < x < 2$; $-\dfrac{1}{2} < x < \dfrac{1}{2}$

(d) $|x| < \dfrac{1}{a}$ is domain.

33. 28.1 kg

35. **(a)** 82.2 min **(b)** 31.8 min

(c) Never attains desired temperature

37. **(a)** \$1105.17 **(b)** 27.73 years
(c) \$4427.59

39. A wins.

Exercises 2.3, page 54

1. Linear **3.** Both **5.** Neither

7. $y = (1/2)e^{3x} + Ce^{x}$

9. $r = \sin\theta + C\cos\theta$

11. $y = -t - 2 + Ce^{t}$

13. $x = y^3 + Cy^{-2}$

15. $y = 1 + C(x^2+1)^{-1/2}$

17. $y = xe^{x} - x$

19. $x = \dfrac{t^3}{6}\ln t - \dfrac{t^3}{36} + \dfrac{1}{2t} - \dfrac{17}{36t^3}$

21. $y = x^2\cos x - \pi^2\cos x$

23. $y(t) = (38/3)e^{-5t} - (8/3)e^{-20t}$

25. **(b)** $y(3) \approx 0.183$

27. **(b)** 0.9960 **(c)** 0.9486 , 0.9729

29. $x = e^{4y}/2 + Ce^{2y}$

31. **(a)** $y = x - 1 + Ce^{-x}$
(b) $y = x - 1 + 2e^{-x}$
(c) $y = x/3 - 1/9 + Ce^{-3x}$
(d) $y = \begin{cases} x - 1 + 2e^{-x}, & \text{if } 0 \le x \le 2 \\ x/3 - 1/9 + (4e^6/9 + 2e^4)e^{-3x}, & \text{if } 2 < x \end{cases}$
(e)

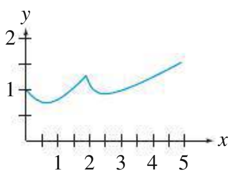


Figure B.10 Solution to Problem 31(e)

33. **(a)** $y = x$ is only solution in a neighborhood of $x = 0$.
(b) $y = -3x + Cx^2$ satisfies $y(0) = 0$ for any C.

35. **(a)** 0.0281 kg/L **(b)** 0.0598 kg/L

37. $$x(t) = \frac{1}{2} - \frac{2\cos(\pi t/12)}{4 + (\pi/12)^2} - \frac{(\pi/12)\sin(\pi t/12)}{4 + (\pi/12)^2} + \left(\frac{19}{2} + \frac{2}{4 + (\pi/12)^2}\right)e^{-2t}$$

39. 71.8°F at noon; 26.9°F at 5 P.M.

Exercises 2.4, page 64

1. Linear with y as dep. var.

3. Separable

5. Separable, also linear with x as dep. var.

7. Exact

9. $y = (C - 3x)/(x^2 - 1)$

11. $e^{x}\sin y - x^3 + \sqrt[3]{y} = C$

13. $y = \dfrac{(t-1)e^{t} + C}{1 + e^{t}}$

15. $r = (C - e^{\theta})\sec\theta$

17. Not exact

19. $x^2 - y^2 + \arctan(xy) = C$

21. $\ln x + x^2y^2 - \sin y = \pi^2$

23. $y = -2/(te^{t} + 2)$

25. $\sin x - x\cos x = \ln y + 1/y + \pi - 1$ (equation is separable, not exact)

27. **(a)** $-\ln|y| + f(x)$
(b) $\cos x \sin y - y^2/2 + f(x)$, where f is a function of x only

29. **(c)** $y = x^2/(C - x)$ **(d)** Yes, $y \equiv 0$

33. **(a)** $x = cy^2$; $x \equiv 0$; $y \equiv 0$
(b) $x^2 + 4y^2 = c$
(c) $2y^2\ln y - y^2 + 2x^2 = c$
(d) $2x^2 + y^2 = c$

Exercises 2.5, page 69

1. Integrating factor depending on x alone

3. Exact

5. Linear with x as dep. var., has integrating factor depending on x alone

7. $\mu = y^{-4}$; $x^2y^{-3} - y^{-1} = C$ and $y \equiv 0$

9. $\mu = x^{-2}$; $y = x^4/3 - x\ln|x| + Cx$ and $x \equiv 0$

11. $\mu = y^{-2}$; $x^2y^{-1} + x = C$ and $y \equiv 0$

13. $\mu = xy$; $x^2y^3 - 2x^3y^2 = C$

15. **(a)** $\mu(z) = \exp\left[\int H(z)\,dz\right]$; $z = xy$
(b) $\ln(x^3y) + 2xy = 2$

17. **(a)** $\dfrac{\partial N/\partial x - \partial M/\partial y}{x^2M - 2xyN}$ depends only on x^2y .
(b) $\exp(x^2y)[x^2 + 2xy] = C$

19. **(b)** $\mu = e^{y}$; $x = y - 1 + Ce^{-y}$

Exercises 2.6, page 76

1. Homogeneous and Bernoulli **3.** Bernoulli

5. Homogeneous and Bernoulli

7. $y' = G(ax + by)$

9. $y = -x/(\ln|x| + C)$ and $y \equiv 0$

11. $y = x/(\ln|x| + C)$ and $y \equiv 0$, $x \equiv 0$

13. $\sqrt{1 + x^2/t^2} = \ln|t| + C$

15. $(x^2 - 4y^2)^3x^2 = C$

17. $y = (x + C)^2/4 - x$ and $y = -x$

19. $y = x + (6 + 4Ce^{2x})/(1 + Ce^{2x})$ and $y = x + 4$

21. $y = 2/(Cx - x^3)$ and $y \equiv 0$

23. $y = 5x^2/(x^5 + C)$ and $y \equiv 0$

25. $x^{-2} = 2t^2\ln|t| + Ct^2$ and $x \equiv 0$

27. $r = \theta^2/(C - \theta)$ and $r \equiv 0$

29. $2\arctan\left(\dfrac{y-3}{x+2}\right) - \ln\left[(x+2)^2 + (y-3)^2\right] = C$.

31. $(2x + 2y - 3)^3 = C(2x + y - 2)^2$

33. $x^2 + t^2 - Ct = 0$ and $t \equiv 0$

35. $y = -3/(x^4 + Cx)$ and $y \equiv 0$

37. $y = (\theta/4)(\ln|\theta| + C)^2$ and $\theta \equiv 0$

39. $\sin(x + y) + \cos(x + y) = Ce^{x-y}$

41. $(x - y + 2)^2 = Ce^{2x} + 1$

45. $(y - 4x)^2(y + x)^3 = C$

47. **(a)** $v' + [2Pu + Q]v = -P$
(b) $y = x + 5x/(C - x^5)$

Review Problems, page 79

1. $e^x + ye^{-y} = C$
3. $x^2y - x^3 + y^{-2} = C$
5. $y + x\sin(xy) = C$
7. $t = C\exp(1/(7y^7))$
9. $(x^2 + 4y^2)^3x^2 = C$
11. $\tan(t - x) + t = C$
13. $y = -(x^2/2)\cos(2x) + (x/4)\sin(2x) + Cx$
15. $y = 2x + 3 - (x + C)^2/4$
17. $y = 2/(1 + Ce^{2\theta})$ and $y \equiv 0$
19. $y^2 = x^2 + Cx^3$ and $x \equiv 0$
21. $xy - x^2 - x + y^2/2 - 4y = C$
23. $y^2 + 2xy - x^2 = C$
25. $x^2y^{-2} - 2xy^{-1} - 4xy^{-2} = C$ and $y \equiv 0$
27. $$[(y-4)^2 - 3(x-3)^2]\left|[\sqrt{3}(x-3) + (y-4)]\right.$$ $$\left./[\sqrt{3}(x-3) - (y-4)]\right|^{1/\sqrt{3}} = C$$
29. $x^4y^3 - 3x^3y^2 + x^4y^2 = C$ **31.** $y = -x^3/2 + 7x/2$
33. $x = -t - 2 + 3e^{-t}$ **35.** $y = -2x\sqrt{2x^2 - 1}$
37. $$\ln[(y-2)^2 + 2(x-1)^2] + \sqrt{2}\arctan\left[\frac{y-2}{\sqrt{2}(x-1)}\right] = \ln 2$$
39. $y = \sqrt{(19x^4 - 1)/2}$
41. $y(t) = e^{-t}\int_2^t \frac{e^r}{1 + r^2}\,dr + 3e^{-(t-2)}, \quad y(3) \approx 1.1883\ldots$

CHAPTER 3

Exercises 3.2, page 100

1. $5 - 4.5e^{-2t/25}$ kg; 5.07 min
3. $(0.4)(100 - t) - (3.9 \times 10^{-7})(100 - t)^4$L; 19.96 min
5. 0.0097%; 73.24 h **7.** 20 min later; $1/e$ times as salty
9. 110,868 **13.** 5970; 6000 **15.** 6572; 6693
17. (a)

Year	$\frac{1}{p}\frac{dp}{dt}$	Logistic (Least Squares)
1790	0.0351	3.29
1800	0.0363	4.51
1810	0.0331	6.16
1820	0.0335	8.41
1830	0.0326	11.45
1840	0.0359	15.53
1850	0.0356	20.97
1860	0.0267	28.14
1870	0.0260	37.45
1880	0.0255	49.31
1890	0.0210	64.07
1900	0.0210	81.89
1910	0.0150	102.65
1920	0.0162	125.87
1930	0.0073	150.64
1940	0.0145	175.80
1950	0.0185	200.10
1960	0.0134	222.47
1970	0.0114	242.16
1980	0.0098	258.81
1990	0.0132	272.44
2000	0.0097	283.28
2010		291.73

(b) $p_1 = 316.920$, $A = 0.00010050$
(c) $p_0 = 3.28780$ (using all data, including 2010)
(d) See table in part (a).
19. $(1/2)\ln(15) \approx 1.354$ yr; 14 million tons per yr
21. 1 hr; 2 h **23.** 11.7% **25.** 31,606 yr
27. e^{-2t} kg of Hh, $2e^{-t} - 2e^{-2t}$ kg of It, and $1 - 2e^{-t} + e^{-2t}$ kg of Bu

Exercises 3.3, page 107

1. 20.7 min **3.** 22.6 min **5.** 9:08 A.M.
7. 28.3°C; 32.5°C; 1:16 P.M.
9. 16.3°C; 19.1°C; 31.7°C; 28.9°C **11.** 39.5 min
13. 148.6°F
15. $T - M = C(T + M)\exp[2\arctan(T/M) - 4M^3kt]$; for T near M, $M^4 - T^4 \approx 4M^3(M - T)$, and so $dT/dt \approx k_1(M - T)$, where $k_1 = 4M^3k$

Exercises 3.4, page 115

1. $(0.981)t + (0.0981)e^{-10t} - 0.0981$ m; 1019 sec
3. 18.6 sec **5.** $4.91t + 22.55 - 22.55e^{-2t}$ m; 97.3 sec
7. 241 sec
9. $95.65t + 956.5e^{-t/10} - 956.5$ m; 13.2 sec
11. $e^{bv}|bv - mg|^{mg} = e^{v_0b}|bv_0 - mg|^{mg}e^{-b^2x/m}$
13. 2.69 sec; 101.19 m **15.** $(\omega_0 - T/k)e^{-kt/I} + T/k$
17. 300 sec
19. $2636e^{-t/20} + 131.8t - 2636$ m; 1.768 sec
21. $5e^{-2t}/2 + 6t - 5/2$; 6 m/sec **23.** Sailboat B
25. (e) 11.18 km/sec **(f)** 2.38 km/sec

Exercises 3.5, page 121

1. $I = \{1.44e^{-100t} + \cos 120t + 1.2\sin 120t\}/2.44$;
$E_L = (-7.2e^{-100t} - 6\sin 120t + 7.2\cos 120t)/2.44$
3. $-(\ln .4) \times 10^{-10} \approx 9.2 \times 10^{-11}$ sec
5. $VI = IRI = I^2R$; $VI = L\frac{dI}{dt}I = \frac{d}{dt}\frac{LI^2}{2}$;
$$VI = E_C\frac{dCE_C}{dt} = \frac{d}{dt}\frac{CE_C^2}{2}$$
7. $-(10\ln .1)/3 \approx 7.68$ sec

Exercises 3.6, page 129

3.

h	"e"
1	3
0.1	2.72055
0.01	2.71830
0.001	2.71828
0.0001	2.71828

7.

x_n	y_n
1.1	0.10450
1.2	0.21668
1.3	0.33382
1.4	0.45300
1.5	0.57135

9.

x_n	y_n
0.2	0.61784
0.4	1.23864
0.6	1.73653
0.8	1.98111
1.0	1.99705
1.2	1.88461
1.4	1.72447
1.6	1.56184
1.8	1.41732
2.0	1.29779

11. $\phi(1) \approx x(1; 2^{-3}) = 1.25494$
13. $\phi(1) \approx y(1; 2^{-3}) = 0.71698$
15. $x = 1.27$
17.

x_n	$y_n(h = 0.2)$	$y_n(h = 0.1)$	$y_n(h = 0.025)$
0.1		−1	0.06250
0.2	−3	1	0.00391
0.3		−1	0.00024
0.4	9	1	0.00002
0.5		−1	0.00000
0.6	−27	1	0.00000
0.7		−1	0.00000
0.8	81	1	0.00000
0.9		−1	0.00000
1.0	−243	1	0.00000

We conclude that step size can dramatically affect convergence.

19.

	T_n		
Time	$K = 0.2$	$K = 0.4$	$K = 0.6$
Midnight	65.0000	65.0000	65.0000
4 A.M.	69.1639	68.5644	68.1300
8 A.M.	71.4836	72.6669	73.6678
Noon	72.9089	75.1605	76.9783
4 P.M.	72.0714	73.5977	74.7853
8 P.M.	69.8095	69.5425	69.2831
Midnight	68.3852	67.0500	65.9740

Exercises 3.7, page 139

1. $y_{n+1} = y_n + h\cos(x_n + y_n) - \dfrac{h^2}{2}\sin(x_n + y_n)[1 + \cos(x_n + y_n)]$

3. $y_{n+1} = y_n + h(x_n - y_n) + \dfrac{h^2}{2}(1 - x_n + y_n) - \dfrac{h^3}{6}(1 - x_n + y_n) + \dfrac{h^4}{24}(1 - x_n + y_n)$

5. Order 2, $\phi(1) \approx 1.3725$; order 4, $\phi(1) \approx 1.3679$
7. -11.7679 **9.** 1.36789 **11.** $x = 1.41$
13. $x = 0.50$
15.

x_n	y_n
0.5	0.21462
1.0	0.13890
1.5	−0.02668
2.0	−0.81879
2.5	−1.69491
3.0	−2.99510

19. $v(3) \approx 0.24193$ with $h = 0.0625$
21. $z(1) \approx 2.87083$ with $h = 0.03125$

CHAPTER 4

Exercises 4.1, page 156

3. Both approach zero. **5.** 0
7. $y(t) = -(30/61)\cos 3t - (25/61)\sin 3t$
9. $y(t) = -2\cos 2t + (3/2)\sin 2t$

Exercises 4.2, page 164

1. $c_1e^{t/2} + c_2e^{-4t}$
3. $c_1e^{-3t} + c_2e^{-2t}$
5. $c_1e^{-4t} + c_2te^{-4t}$
7. $c_1e^{t/2} + c_2e^{-2t/3}$
9. $c_1e^{t/2} + c_2te^{t/2}$
11. $c_1e^{-5t/2} + c_2te^{-5t/2}$
13. $3e^{-4t}$ **15.** $\frac{4}{3}e^{t} - \frac{1}{3}e^{3t}$
17. $\left(\frac{7t}{3} + 2\right)e^{3t}$. **19.** $e^{-t} - 2te^{-t}$
21. **(a)** $ar + b = 0$ **(b)** $ce^{-bt/a}$
23. $ce^{-4t/5}$ **25.** $ce^{13t/6}$
27. Lin. dep. **29.** Lin. indep. **31.** Lin. dep.
33. If $c_1 \neq 0$, then $y_1 = -(c_2/c_1)y_2$.
35. **(a)** Lin. indep. **(b)** Lin. dep. **(c)** Lin. indep. **(d)** Lin. dep.
37. $c_1e^{t} + c_2e^{(-1-\sqrt{5})t} + c_3e^{(-1+\sqrt{5})t}$
39. $c_1e^{-2t} + c_2te^{-2t} + c_3e^{2t}$
41. $c_1e^{-3t} + c_2e^{-2t} + c_3e^{2t}$
43. $3 + e^{t} - 2e^{-t}$
45. **(a)** $c_1e^{r_1t} + c_2e^{r_2t} + c_3e^{r_3t}$ (where $r_1 = -4.832$, $r_2 = -1.869$, and $r_3 = 0.701$)
(b) $c_1e^{r_1t} + c_2e^{-r_1t} + c_3e^{r_2t} + c_4e^{-r_2t}$ (where $r_1 = 1.176$, $r_2 = 1.902$)
(c) $c_1e^{-t} + c_2e^{t} + c_3e^{-2t} + c_4e^{2t} + c_5e^{3t}$

Exercises 4.3, page 172

1. $c_1\cos 3t + c_2\sin 3t$
3. $c_1e^{3t}\cos t + c_2e^{3t}\sin t$
5. $c_1e^{-2t}\cos\sqrt{2}t + c_2e^{-2t}\sin\sqrt{2}t$
7. $c_1e^{-t/2}\cos(\sqrt{5}t/2) + c_2e^{-t/2}\sin(\sqrt{5}t/2)$
9. $c_1e^{t} + c_2e^{7t}$

11. $c_1e^{-5t} + c_2te^{-5t}$
13. $c_1e^t\cos 5t + c_2e^t\sin 5t$
15. $c_1e^{(3+\sqrt{53})t/2} + c_2e^{(3-\sqrt{53})t/2}$.
17. $c_1e^{t/2}\cos(3\sqrt{3}t/2) + c_2e^{t/2}\sin(3\sqrt{3}t/2)$
19. $c_1e^t + c_2e^{-t}\cos 2t + c_3e^{-t}\sin 2t$
21. $2e^{-t}\cos t + 3e^{-t}\sin t$
23. $(\sqrt{2}/4)[e^{(2+\sqrt{2})t} - e^{(2-\sqrt{2})t}]$
25. $e^t\sin t - e^t\cos t$
27. $e^{2t} - \sqrt{2}\,e^t\sin\sqrt{2}t$
29. **(a)** $c_1e^{-t} + c_2e^t\cos\sqrt{2}t + c_3e^t\sin\sqrt{2}t$
(b) $c_1e^{2t} + c_2e^{-2t}\cos 3t + c_3e^{-2t}\sin 3t$
(c) $c_1\cos 2t + c_2\sin 2t + c_3\cos 3t + c_4\sin 3t$
31. **(a)** Oscillatory **(b)** Tends to zero
(c) Tends to $-\infty$ **(d)** Tends to $-\infty$
(e) Tends to $+\infty$
33. **(a)** $y(t) = 0.3e^{-3t}\cos 4t + 0.2e^{-3t}\sin 4t$
(b) $[\pi + \arctan(-1.5)]/4$
(c) $2/\pi$
(d) Decreases the frequency of oscillation, introduces the factor e^{-3t}, causing the solution to decay to zero
35. $b \geq 2\sqrt{Ik}$
37. **(a)** $c_1\cos t + c_2\sin t + c_3t\cos t + c_4t\sin t$
(b) $(c_1 + c_2t)e^{-t}\cos(\sqrt{3}t) + (c_3 + c_4t)e^{-t}\sin(\sqrt{3}t)$

Exercises 4.4, page 180

1. No **3.** Yes **5.** No **7.** Yes
9. $y_p \equiv -3$
11. $y_p(x) = [(\ln 2)^2 + 1]^{-1}2^x$
13. $\cos 3t$
15. $xe^x/2 + 3e^x/4$
17. $-2t\cos 2t$
19. $\left(\frac{t}{13} + \frac{8}{169}\right)te^{-3t}$ **21.** $t^3e^{2t}/6$
23. $-\frac{1}{21}\theta^3 - \frac{1}{49}\theta^2 - \frac{2}{343}\theta$
25. $e^{2t}(\cos 3t + 6\sin 3t)$
27. $(A_3t^4 + A_2t^3 + A_1t^2 + A_0t)\cos 3t$
$+ (B_3t^4 + B_2t^3 + B_1t^2 + B_0t)\sin 3t$
29. $e^t(A_4t^4 + A_3t^3 + A_2t^2 + A_1t + A_0)$
31. $(A_3t^4 + A_2t^3 + A_1t^2 + A_0t)e^{-t}\cos t$
$+ (B_3t^4 + B_2t^3 + B_1t^2 + B_0t)e^{-t}\sin t$
33. $(1/5)\cos t + (2/5)\sin t$
35. $\left(\frac{1}{10}t^2 - \frac{4}{25}t\right)e^t$

Exercises 4.5, page 185

1. **(a)** $5\cos t$ **(b)** $\cos t - e^{2t}$ **(c)** $4\cos t + 6e^{2t}$
3. $y = -t + c_1e^t + c_2e^{-t}$
5. $\theta = t - 1 + c_1e^{2t} + c_2e^{-t}$
7. $y = \tan x + c_1e^{\sqrt{2}x} + c_2e^{-\sqrt{2}x}$
9. Yes **11.** No **13.** Yes **15.** No
17. $y(t) = -t^2 + \frac{4t}{3} + \frac{1}{9} + c_1e^{3t} + c_2e^{-t}$.
19. $y = (\cos x - \sin x)e^x/2 + c_1e^x + c_2e^{2x}$
21. $y = (1/2)\theta e^{-\theta}\sin\theta + (c_1\cos\theta + c_2\sin\theta)e^{-\theta}$
23. $y = e^t - 1$
25. $z = e^{-x} - \cos x + \sin x$
27. $y = -(3/10)\cos x - (1/10)\sin x - (1/20)\cos 2x$
$+(3/20)\sin 2x$
29. $y = -(1/2)\sin\theta - (1/3)e^{2\theta} + (3/4)e^\theta + (7/12)e^{-\theta}$
31. $y_p = (A_1t + A_0)t\cos t + (B_1t + B_0)t\sin t + C\cdot 10^t$
33. $x_p(t) = (A\cos t + B\sin t)e^t + C_2t^2 + C_1t + C_0$
$+ D_1\cos 3t + D_2\sin 3t + E_1\cos t + E_2\sin t$
35. $y_p = (A_1t + A_0)\cos 3t + (B_1t + B_0)\sin 3t + Ce^{5t}$
37. $y_p = t^2 + 3t - 1$
39. $y_p = (t/10 - 4/25)te^t - 1/2$
41. **(a)** $y_1 = -(2\cos 2t + \sin 2t)e^{-t} + 2$
for $0 \leq t \leq 3\pi/2$
(b) $y_2 = y_h = (c_1\cos 2t + c_2\sin 2t)e^{-t}$
for $t > 3\pi/2$
(c) $c_1 = -2(e^{3\pi/2} + 1)$, $c_2 = -(e^{3\pi/2} + 1)$
43. $y = -\cos t + (1/2)\sin t - (1/2)e^{-3t} + 2e^{-t}$
45. **(a)** $y(t) = \frac{2V\cos(\pi/2V)}{V^2 - 1}\sin t$ for $V \neq 1$;
$y(t) = \frac{\pi}{2}\sin t$ for $V = 1$
(b) $V \approx 0.73$
47. **(a)** $2\sin 3t - \cos 6t$ **(b)** No solution
(c) $c\sin 3t - \cos 6t$, where c is any constant

Exercises 4.6, page 191

1. $c_1\cos 2t + c_2\sin 2t$
$-(1/4)(\cos 2t)\ln|\sec 2t + \tan 2t|$
3. $c_1e^t + c_2te^t + te^t\ln|t|$
5. $c_1\cos 4\theta + c_2\sin 4\theta + (\theta/4)\sin 4\theta$
$+(1/16)(\cos 4\theta)\ln|\cos 4\theta|$
7. $(2\ln t - 3)t^2e^{-2t}/4 + c_1e^{-2t} + c_2te^{-2t}$
9. $y_p = -2t - 4$
11. $y = -(\cos t)\ln|\sec t + \tan t| + (1/10)e^{3t}$
$-1 + c_1\cos t + c_2\sin t$
13. $c_1\cos 2t + c_2\sin 2t + (1/24)\sec^2 2t - 1/8$
$+(1/8)(\sin 2t)\ln|\sec 2t + \tan 2t|$
15. $c_1\cos t + c_2\sin t - t^2 + 3 + 3t\sin t$
$+3(\cos t)\ln|\cos t|$
17. $c_1\cos 2t + c_2\sin 2t - e^t/5$
$-(1/2)(\cos 2t)\ln|\sec 2t + \tan 2t|$
19. $y = e^{1-t} - e^{t-1} + \frac{e^t}{2}\int_1^t \frac{e^{-u}}{u}du - \frac{e^{-t}}{2}\int_1^t \frac{e^u}{u}du$
$[y(2) \approx -1.93]$
21. 0.3785
23. $c_1e^t + c_2(t+1) - t^2$
25. $c_1(5t - 1) + c_2e^{-5t} - t^2e^{-5t}/10$

Exercises 4.7, page 199

1. Unique solution on $(-\pi/2, \pi/2)$
3. Unique solution on $(0, \infty)$
5. Does not apply; $t = 0$ is a point of discontinuity

7. Does not apply; not an initial value problem

9. $c_1 t + c_2 t^{-7}$

11. $c_1 t^{-2} + c_2 t^{-2} \ln t$

13. $c_1 t^{-1/3} + c_2 t^{-1/3} \ln t$

15. $c_1 t \cos\left[2 \ln(-t)\right] + c_2 t \sin\left[2 \ln(-t)\right]$

17. $t^{-4}\left\{c_1 \cos\left[\ln(-t)\right] + c_2 \sin\left[\ln(-t)\right]\right\}$

19. $t - 3t^4$

21. $c_1(t-2) + c_2(t-2)^7$

23. **(c)** $t^{\alpha}\cos(\beta \ln|t|)$, $t^{\alpha}\sin(\beta \ln|t|)$; t^r, $t^r \ln|t|$

25. **(a)** True **(b)** False

27. **(e)** No, because the coefficient of y'' vanishes at $t = 0$ and the equation cannot be written in standard form.

29. Otherwise their Wronskian would be zero at t_0, contradicting linear independence.

31. **(a)** Yes **(b)** No **(c)** Yes **(d)** Yes

33. Cte^{-t}

35. $1 + 2t - t^2$

37. $c_1 \cos(3 \ln t) + c_2 \sin(3 \ln t) + (1/9)\cos(3 \ln t)\ln|\sec(3 \ln t) + \tan(3 \ln t)|$

39. $c_1 t + c_2 t \ln t + (1/2)t(\ln t)^2 + 3t(\ln t)[\ln|\ln t|]$

41. t^4

43. $t + 1$

45. $(t-1)e^{2t}/2$

47. **(a)** $(1 - 2t^2)\displaystyle\int (1 - 2t^2)^{-2} e^{t^2}\,dt$

(b) $(3t - 2t^3)\displaystyle\int (3t - 2t^3)^{-2} e^{t^2}\,dt$

49. $tw'' + 2tw' + (t+1)w = 0$

51. **(a)** $\phi'(t_0) = \lim\limits_{n\to\infty} \dfrac{\phi(t_n) - \phi(t_0)}{t_n - t_0} = \lim\limits_{n\to\infty} \dfrac{0-0}{t_n - t_0} = 0$

Exercises 4.8, page 210

1. Let $Y(t) = y(-t)$. Then $Y'(t) = -y'(-t)$, $Y''(t) = y''(-t)$. But $y''(s) - sy(s) = 0$, so $y''(-t) + ty(-t) = 0$ or $Y''(t) + tY(t) = 0$.

3. The spring stiffness is $(-6y)$, so it opposes negative displacements $(y < 0)$ and reinforces positive displacements $(y > 0)$. Initially $y < 0$ and $y' < 0$, so the (positive) stiffness reverses the negative velocity and restores y to 0. Thereafter, $y > 0$ and the negative stiffness drives y to $+\infty$.

5. **(a)** $y'' = 2y^3 = \dfrac{d}{dy}(y^4/2)$. Thus, by setting $K = 0$ and choosing the $(-)$ sign in equation (11), we get

$$t = -\int \frac{dy}{\sqrt{2y^4/2}} + c = \frac{1}{y} + c, \text{ or } y = 1/(t-c).$$

(b) Linear dependence would imply

$$\frac{y_1(t)}{y_2(t)} = \frac{1/(t-c_1)}{1/(t-c_2)} = \frac{t-c_2}{t-c_1} \equiv \text{ constant}$$

in a neighborhood of 0, which is false if $c_1 \neq c_2$.

(c) If $y(t) = 1/(t-c)$, then $y(0) = -1/c$, $y'(0) = -1/c^2 = -y(0)^2$, which is false for the given data.

7. **(a)** The velocity, which is always perpendicular to the lever arm, is $\ell\, d\theta/dt$. Thus, (lever arm) times (perpendicular momentum) $= \ell m \ell\, d\theta/dt = m\ell^2 d\theta/dt$.

(b) The component of the gravitational force perpendicular to lever arm is $mg \sin\theta$, and is directed toward decreasing θ. Thus, torque $= -\ell mg \sin\theta$.

(c) Torque $= \dfrac{d}{dt}$(angular momentum) or $-\ell mg \sin\theta = (m\ell^2\theta')' = m\ell^2\theta''$.

9. 2 or -2

11. The sign of the damping coefficient $(y')^2 - 1$ indicates that low velocities are boosted by negative damping but that high velocities are slowed. Hence, one expects a limit cycle.

13. **(a)** Airy **(b)** Duffing **(c)** van der Pol

15. **(a)** Yes ($t^2 =$ positive stiffness)

(b) No ($-t^2 =$ negative stiffness)

(c) Yes ($y^4 =$ positive stiffness)

(d) No ($y^5 =$ negative stiffness for $y < 0$)

(e) Yes ($4 + 2\cos t =$ positive stiffness)

(f) Yes (positive stiffness and damping)

(g) No (negative stiffness and damping)

17. $1/4\sqrt{2}$

Exercises 4.9, page 220

1. $y(t) = -(1/4)\cos 5t - (1/5)\sin 5t$

amplitude $= \sqrt{41}/20$; period $= 2\pi/5$;

frequency $= 5/2\pi$; $[\pi - \arctan(5/4)]/5$ sec

3. $b = 0$: $y(t) = \cos 4t$

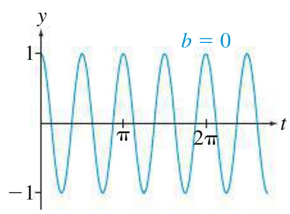


Figure B.11 $b = 0$

$b = 6$: $y(t) = e^{-3t}\cos\sqrt{7}t + (3/\sqrt{7})e^{-3t}\sin\sqrt{7}t$

$= (4/\sqrt{7})e^{-3t}\sin(\sqrt{7}t + \phi)$, where

$\phi = \arctan\sqrt{7}/3 \approx 0.723$

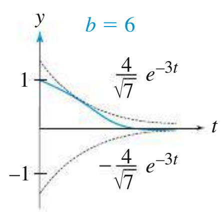


Figure B.12 $b = 6$

$b = 8$: $y(t) = (1 + 4t)e^{-4t}$

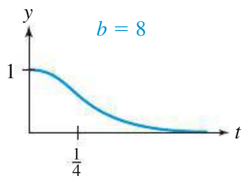


Figure B.13 $b = 8$

$b = 10$: $y(t) = (4/3)e^{-2t} - (1/3)e^{-8t}$

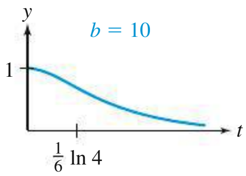


Figure B.14 $b = 10$

5. $k = 20$: $y(t) = [(1 + \sqrt{5})/2]e^{(-5+\sqrt{5})t} + [(1 - \sqrt{5})/2]e^{(-5-\sqrt{5})t}$

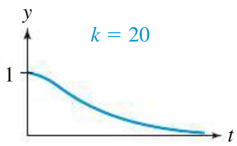


Figure B.15 $k = 20$

$k = 25$: $y(t) = (1 + 5t)e^{-5t}$

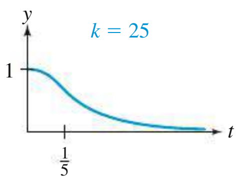


Figure B.16 $k = 25$

$k = 30$: $y(t) = e^{-5t}\cos\sqrt{5}t + \sqrt{5}e^{-5t}\sin\sqrt{5}t$
$= \sqrt{6}e^{-5t}\sin(\sqrt{5}t + \phi)$, where
$\phi = \arctan(1/\sqrt{5}) \approx 0.421$

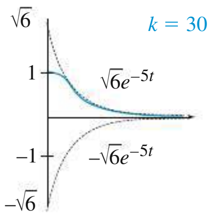


Figure B.17 $k = 30$

7. $y(t) = (-3/4)e^{-8t}\cos 8t - e^{-8t}\sin 8t$
$= (5/4)e^{-8t}\sin(8t + \phi)$, where
$\phi = \pi + \arctan(3/4) \approx 3.785$;
damp. factor $= (5/4)e^{-8t}$;
quasiperiod $= \pi/4$; quasifreq. $= 4/\pi$

9. 0.242 m

11. $(10/\sqrt{9999})\arctan(\sqrt{9999}) \approx 0.156$ sec

13. Relative extrema at
$t = [\pi/3 + n\pi - \arctan(\sqrt{3}/2)]/(2\sqrt{3})$
for $n = 0, 1, 2, \ldots$; but touches curves $\pm\sqrt{7/12}e^{-2t}$
at $t = [\pi/2 + m\pi - \arctan(\sqrt{3}/2)]/(2\sqrt{3})$
for $m = 0, 1, 2, \ldots$.

15. First measure half the quasiperiod P as the time between two successive zero crossings. Then compute the ratio $y(t + P)/y(t) = e^{-(b/2m)P}$.

Exercises 4.10, page 227

1. $M(\gamma) = 1/\sqrt{(1 - 4\gamma^2)^2 + 4\gamma^2}$

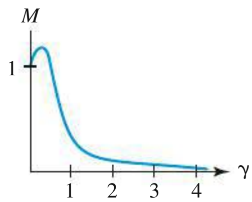


Figure B.18

3. $y(t) = \cos 3t + (1/3)t\sin 3t$

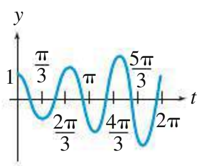


Figure B.19

5. (a) $y(t) = -[F_0/(k - m\gamma^2)]\cos(\sqrt{k/m}\,t) + [F_0/(k - m\gamma^2)]\cos\gamma t$
$= (F_0/[m(\omega^2 - \gamma^2)])(\cos\gamma t - \cos\omega t)$

(c) $y(t) = \sin 8t \sin t$

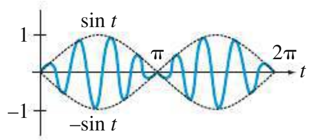


Figure B.20

7. $y(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t} + \dfrac{F_0 \sin(\gamma t + \theta)}{\sqrt{(k - m\gamma^2)^2 + b^2\gamma^2}}$,
where $r_1, r_2 = -(b/2m) \pm (1/2m)\sqrt{b^2 - 4mk}$
and $\tan\theta = (k - m\gamma^2)/(b\gamma)$ as in equation (7)

9. $y_p(t) = (0.08)\cos 2t + (0.06)\sin 2t$
$= (0.1)\sin(2t + \theta)$, where
$\theta = \arctan(4/3) \approx 0.927$

11. $y(t) = -(18/85)e^{-2t}\cos 6t - (22/255)e^{-2t}\sin 6t + (2/\sqrt{85})\sin(2t+\theta)$, where $\theta = \arctan(9/2) \approx 1.352$;
res. freq. $= 2\sqrt{2}/\pi$ cycles/sec

13. $y_p(t) = (3/185)(8\sin 4t - 11\cos 4t)$

15. Amp $= \sqrt{\left(-\frac{11}{986}\right)^2 + \left(-\frac{3}{1972}\right)^2} \approx 0.01\,(\text{m})$,
freq. $= 2/\pi$

Review Problems, page 231

1. $c_1e^{-9t} + c_2e^t$

3. $c_1e^{t/2}\cos(3t/2) + c_2e^{t/2}\sin(3t/2)$

5. $c_1e^{3t/2} + c_2e^{t/3}$

7. $c_1e^{-t/3}\cos(t/6) + c_2e^{-t/3}\sin(t/6)$

9. $c_1e^{7t/4} + c_2te^{7t/4}$

11. $t^{1/2}\{c_1\cos[(\sqrt{19}/2)\ln t] + c_2\sin[(\sqrt{19}/2)\ln t]\}$

13. $c_1\cos 4t + c_2\sin 4t + (1/17)te^t - (2/289)e^t$

15. $c_1e^{-2t} + c_2e^{-t} + c_3e^{-t/3}$

17. $c_1e^t + c_2e^{-t/2}\cos(\sqrt{43}t/2) + c_3e^{-t/2}\sin(\sqrt{43}t/2)$

19. $c_1e^{-3t} + c_2e^{t/2} + c_3te^{t/2}$

21. $c_1e^{3t/2}\cos(\sqrt{19}t/2) + c_2e^{3t/2}\sin(\sqrt{19}t/2) - e^t/5 + t^2 + 6t/7 + 4/49$

23. $c_1\cos 4\theta + c_2\sin 4\theta - (1/16)(\cos 4\theta)\ln|\sec 4\theta + \tan 4\theta|$

25. $c_1e^{3t/2} + c_2te^{3t/2} + e^{3t}/9 + e^{5t}/49$

27. $c_1x + c_2x^{-2} - 2x^{-2}\ln x + x\ln x$

29. $e^{-2t}\cos(\sqrt{3}t)$

31. $2e^t\cos 3t - (7/3)e^t\sin 3t - \sin 3t$

33. $-e^{-t} - 3e^{5t} + e^{8t}$

35. $\cos\theta + 2\sin\theta + \theta\sin\theta + (\cos\theta)\ln|\cos\theta|$

37. **(a)**, **(c)**, **(e)**, and **(f)** have all solutions bounded as $t \to +\infty$

39. $y_p(t) = (1/4)\sin 8t$; $\sqrt{62}/2\pi$

CHAPTER 5

Exercises 5.2, page 249

1. **(a)** $-t^3 + 3t^2 + 8$
(b) $-2t^3 + 3t^2 + 6t + 16$
(c) $2t^3 + 3t^2 - 16$
(d) $-2t^3 + 3t^2 + 6t + 16$
(e) $-2t^3 + 3t^2 + 6t + 16$

3. $x = c_1 + c_2e^{-2t}$; $y = c_2e^{-2t}$

5. $x \equiv -5$; $y \equiv 1$

7. $u = c_1 - (1/2)c_2e^{-t} + (1/2)e^t + (5/3)t$;
$v = c_1 + c_2e^{-t} + (5/3)t$

9. $x = c_1e^t + (1/4)\cos t - (1/4)\sin t$;
$y = -3c_1e^t - (3/4)\cos t - (1/4)\sin t$

11. $u = c_1\cos 2t + c_2\sin 2t + c_3e^{\sqrt{3}t} + c_4e^{-\sqrt{3}t} - (3/10)e^t$;
$v = c_1\cos 2t + c_2\sin 2t - (2/5)c_3e^{\sqrt{3}t} - (2/5)c_4e^{-\sqrt{3}t} + (1/5)e^t$

13. $x = 2c_2e^t\cos 2t - 2c_1e^t\sin 2t$;
$y = c_1e^t\cos 2t + c_2e^t\sin 2t$

15. $w = -(2/3)c_1e^{2t} + c_2e^{7t} + t + 1$;
$z = c_1e^{2t} + c_2e^{7t} - 5t - 2$

17. $x = c_1\cos t + c_2\sin t - 4c_3\cos(\sqrt{6}t) - 4c_4\sin(\sqrt{6}t)$;
$y = c_1\cos t + c_2\sin t + c_3\cos(\sqrt{6}t) + c_4\sin(\sqrt{6}t)$

19. $x = 2e^{3t} - e^{2t}$;
$y = -2e^{3t} + 2e^{2t}$

21. $x = e^t + e^{-t} + \cos t + \sin t$;
$y = e^t + e^{-t} - \cos t - \sin t$

23. Infinitely many solutions satisfying $x + y = e^t + e^{-2t}$

25. $x(t) = -c_1e^t - 2c_2e^{2t} - c_3e^{3t}$,
$y(t) = c_1e^t + c_2e^{2t} + c_3e^{3t}$,
$z(t) = 2c_1e^t + 4c_2e^{2t} + 4c_3e^{3t}$

27. $x(t) = c_1e^{8t} + c_2e^{4t} + c_3$,
$y(t) = \frac{1}{2}(c_1e^{8t} - c_2e^{4t} + c_3)$,
$z(t) = -c_1e^{8t} + c_3$

29. $-3 \le \lambda \le -1$

31. $x(t) = -\left(10 + \frac{20}{\sqrt{7}}\right)e^{r_1t} - \left(10 - \frac{20}{\sqrt{7}}\right)e^{r_2t} + 20$ kg,
$y(t) = \frac{30}{\sqrt{7}}e^{r_1t} - \frac{30}{\sqrt{7}}e^{r_2t} + 20$ kg, where
$r_1 = \frac{-5-\sqrt{7}}{100} \approx -0.0765$,
$r_2 = \frac{-5+\sqrt{7}}{100} \approx -0.0235$

33. $x = \left[\frac{20 - 10\sqrt{5}}{\sqrt{5}}\right]e^{(-3+\sqrt{5})t/100} - \left[\frac{20 + 10\sqrt{5}}{\sqrt{5}}\right]e^{(-3-\sqrt{5})t/100} + 20$;
$y = -\left(\frac{10}{\sqrt{5}}\right)e^{(-3+\sqrt{5})t/100} + \left(\frac{10}{\sqrt{5}}\right)e^{(-3-\sqrt{5})t/100} + 20$

35. 90.4°F **37.** $460/11 \approx 41.8$°F

Exercises 5.3, page 259

1. $x_1' = x_2$, $x_2' = 3x_1 - tx_2 + t^2$;
$x_1(0) = 3$, $x_2(0) = -6$

3. $x_1' = x_2$, $x_2' = x_3$, $x_3' = x_4$,
$x_4' = x_4 - 7x_1 + \cos t$; $x_1(0) = x_2(0) = 1$,
$x_3(0) = 0$, $x_4(0) = 2$

5. $x_1' = x_2$, $x_2' = x_2 - x_3 + 2t$,
$x_3' = x_4$, $x_4' = x_1 - x_3 - 1$;
$x_1(3) = 5$, $x_2(3) = 2$, $x_3(3) = 1$, $x_4(3) = -1$

7. $x_1' = x_2$, $x_2' = x_3$, $x_3' = x_4 + t$, $x_4' = x_5$,
$x_5' = (2x_4 - 2x_3 + 1)/5$; $x_1(0) = x_2(0) = x_3(0) = 4$,
$x_4(0) = x_5(0) = 1$

9. $t_{n+1} = t_n + h, \quad n = 0, 1, 2, \ldots;$

$$x_{i,\,n+1} = x_{i,n} + \frac{h}{2}\left[f_i(t_n, x_{1,n}, \ldots, x_{m,n}) + f_i(t_n + h, x_{1,n} + hf_1(t_n, x_{1,n}, \ldots, x_{m,n}), \ldots, x_{m,n} + hf_m(t_n, x_{1,n}, \ldots, x_{m,n}))\right]$$

$i = 1, 2, \ldots, m$

11.

i	t_i	$y(t_i)$
1	0.250	0.750000
2	0.500	0.625000
3	0.750	0.573529
4	1.000	0.563603

13.

i	t_i	$y(t_i)$
1	0.250	0.25000
2	0.500	0.50000
3	0.750	0.75000
4	1.000	1.00000

15. $y(1) \approx x_1(1, 2^{-2}) = 1.69, y'(1) \approx 1.82$

17. $u(1; 2^{-2}) = v(1; 2^{-2}) = 0.36789$

19.

		Part (a)		**Part (b)**		**Part (c)**	
i	t_i	$x(t_i)$	$y(t_i)$	$x(t_i)$	$y(t_i)$	$x(t_i)$	$y(t_i)$
1	0.5	1.95247	2.25065	1.48118	2.42311	0.91390	2.79704
2	1.0	3.34588	1.83601	2.66294	1.45358	1.63657	1.13415
3	1.5	4.53662	3.36527	5.19629	2.40348	4.49334	1.07811
4	2.0	2.47788	4.32906	3.10706	4.64923	5.96115	5.47788
5	2.5	1.96093	2.71900	1.92574	3.32426	1.51830	5.93110
6	3.0	2.86412	1.96166	2.34143	2.05910	0.95601	2.18079
7	3.5	4.28449	2.77457	3.90106	2.18977	2.06006	0.98131
8	4.0	3.00965	4.11886	3.83241	3.89043	5.62642	1.38072
9	4.5	2.18643	3.14344	2.32171	3.79362	5.10594	5.10462
10	5.0	2.63187	2.25824	2.21926	2.49307	1.74187	5.02491

21.

i	t_i	$x_1(t_i) \approx H(t_i)$
1	0.5	0.09573
2	1.0	0.37389
3	1.5	0.81045
4	2.0	1.37361
5	2.5	2.03111
6	3.0	2.75497
7	3.5	3.52322
8	4.0	4.31970
9	4.5	5.13307
10	5.0	5.95554

23. Yes, yes

25. $y(1) \approx x_1(1; 2^{-3}) = 1.25958$

27. $y(0.1) \approx 0.00647, \ldots, y(2.0) \approx 1.60009$

29. **(a)** $P_1(10) \approx 0.567, P_2(10) \approx 0.463, P_3(10) \approx 0.463$
(b) $P_1(10) \approx 0.463, P_2(10) \approx 0.567, P_3(10) \approx 0.463$
(c) $P_1(10) \approx 0.463, P_2(10) \approx 0.463, P_3(10) \approx 0.567$
All populations approach 0.5.

Exercises 5.4, page 271

1. $x = y^3, \quad y > 0$

Figure B.21

3. $(1/\sqrt{2}, 1/\sqrt{2}), \quad (-1/\sqrt{2}, -1/\sqrt{2})$ **5.** $(0, 0)$

7. $e^x + ye^{-y} = c$ **9.** $e^x + xy - y^2 = c$

11. $y^2 - x^2 = c$ **13.** $(x - 1)^2 + (y - 1)^2 = c$

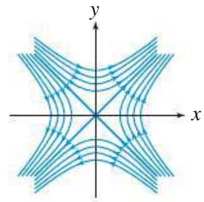


Figure B.22

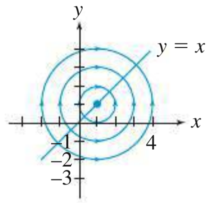


Figure B.23

15. $(-2, 1)$ is a saddle point (unstable).

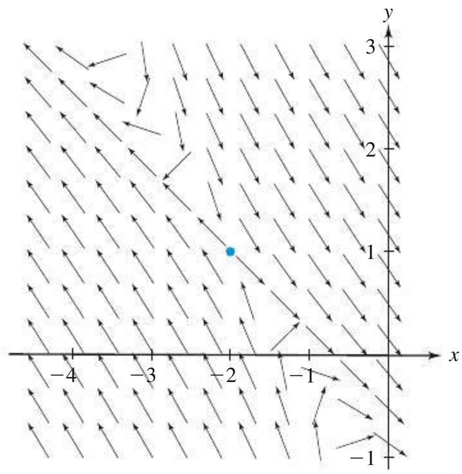


Figure B.24

17. $(0, 0)$ is a center (stable).

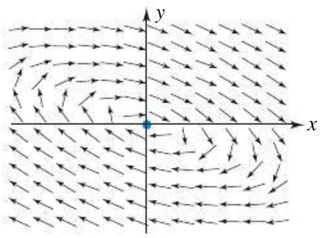


Figure B.25

19. $(0, 0)$ is a saddle point (unstable).

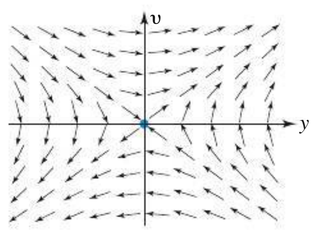


Figure B.26

21. $(0, 0)$ is a center (stable).

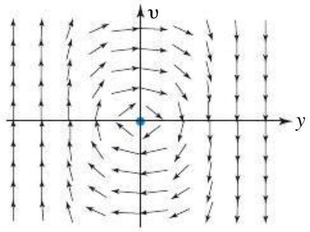


Figure B.27

23. $(0, 0)$ is a center (stable); $(1, 0)$ is a saddle point (unstable).

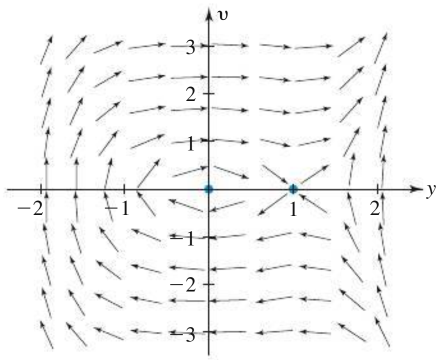


Figure B.28

25. **(a)** Periodic **(b)** Not periodic
(c) Critical point (periodic)

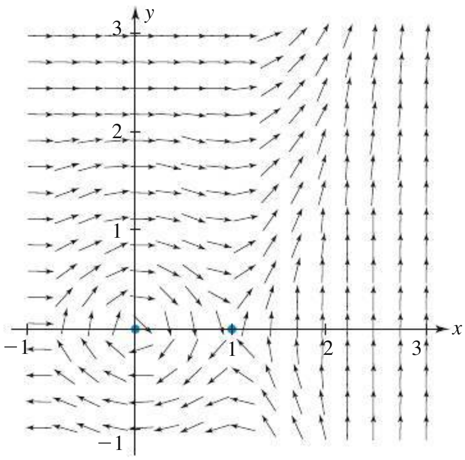


Figure B.29

27. $\big(x(t), y(t)\big)$ approaches $(0, 0)$.

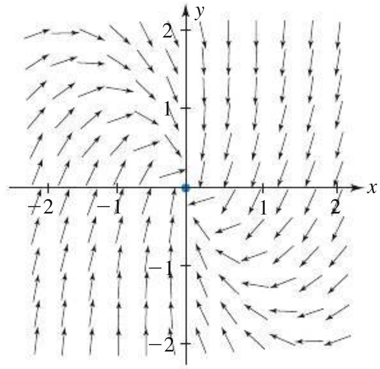


Figure B.30

29. Critical points are $(1, 0)$ and $(x, 1)$ with x arbitrary. Phase plane solutions are $x = 1$, and $y = C(x - 1)$. Solution starting at $(3, 2)$ goes to infinity. The others converge to $(1, 0)$.

31. **(a)** $y' = v, \quad v' = f(y)$

(b) $\dfrac{dv}{dy} = \dfrac{f(y)}{v}$ implies $\displaystyle\int v\,dv = \int f(y)\,dy$ + constant, or $v^2/2 = F(y) + K$

33. **(a)** $x' = v, v' = -x + 1/(\lambda - x)$

(b) $\dfrac{dv}{dx} = \dfrac{-x + 1/(\lambda - x)}{v}$ implies

$$\int v\,dv = \int\left(-x + \frac{1}{\lambda - x}\right)dx + \text{constant, or}$$

$v^2/2 = -x^2/2 - \ln(\lambda - x) +$ constant. The solution equation follows with the choice $C/2$ for the constant.

(c) At a critical point $v = 0$ and $-x + 1/(\lambda - x) = 0$.

Solutions for the latter are $x = \dfrac{\lambda \pm \sqrt{\lambda^2 - 4}}{2}$, and are real only for $\lambda \geq 2$.

(d)

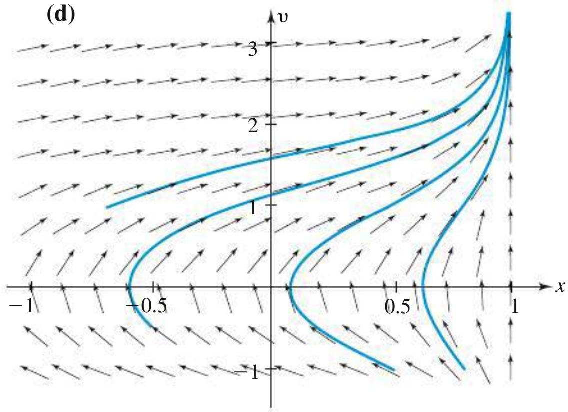


Figure B.31 $\lambda = 1$

(e) See Figures B.31 and B.32. When $\lambda = 3$, one critical point is a center and the other is a saddle. For $\lambda = 1$, the bar is attracted to the magnet. For $\lambda = 3$, the bar may oscillate periodically, or (rarely) come to rest at the saddle critical point.

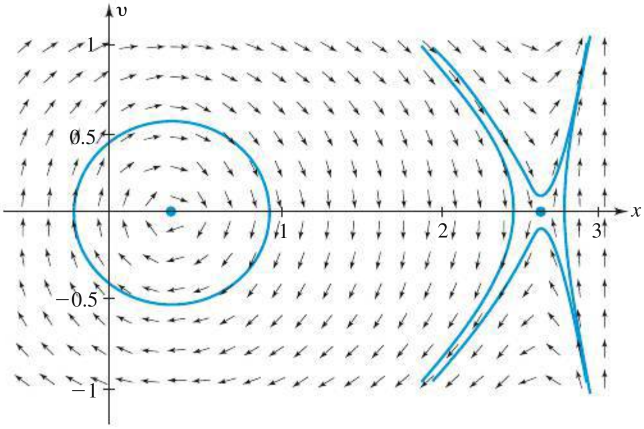


Figure B.32 $\lambda = 3$

35. (c) For upper half-plane, center is at $v = 0$, $y = -1$; for lower half-plane, center is at $v = 0$, $y = 1$.

(d) All points on the segment $v = 0$, $-1 \le y \le 1$

(e) $y = -0.5$; see Figure B.33.

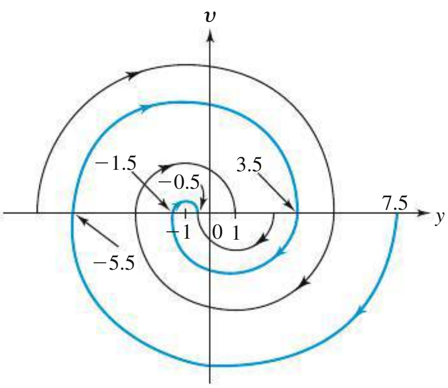


Figure B.33

Exercises 5.5, page 281

1. Runge–Kutta approximations

t_n	$p_n(r = 1.5)$	$p_n(r = 2)$	$p_n(r = 3)$
0.25	1.59600	1.54249	1.43911
0.5	2.37945	2.07410	1.64557
0.75	3.30824	2.47727	1.70801
1.0	4.30243	2.72769	1.72545
1.25	5.27054	2.86458	1.73024
1.5	6.13869	2.93427	1.73156
1.75	6.86600	2.96848	1.73192
2.0	7.44350	2.98498	1.73201
2.25	7.88372	2.99286	1.73204
2.5	8.20933	2.99661	1.73205
2.75	8.44497	2.99839	1.73205
3.0	8.61286	2.99924	1.73205
3.25	8.73117	2.99964	1.73205
3.5	8.81392	2.99983	1.73205
3.75	8.87147	2.99992	1.73205
4.0	8.91136	2.99996	1.73205
4.25	8.93893	2.99998	1.73205
4.5	8.95796	2.99999	1.73205
4.75	8.97107	3.00000	1.73205
5.0	8.98010	3.00000	1.73205

Limiting population is $3^{1/(r-1)}$.

3. $$x(t) = \frac{1}{2} - \frac{2\cos(\pi t/12)}{4 + (\pi/12)^2} - \frac{(\pi/12)\sin(\pi t/12)}{4 + (\pi/12)^2} + \left(\frac{19}{2} + \frac{2}{4 + (\pi/12)^2}\right)e^{-2t}$$

5. (a)

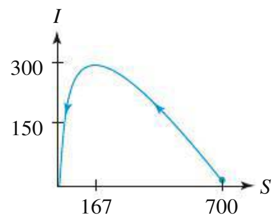


Figure B.34

(b) 295 persons

7. $$P(t) = \exp\left\{\frac{c[1 - \exp(-bt)]}{b} - bt\right\},$$
$$Q(t) = \exp\left\{\frac{c[1 - \exp(-bt)]}{b} - bt\right\}[\exp(bt) - 1]$$

9. $$N(t) = \frac{N_0\,(c + s)}{N_0 s + (c + s - N_0 s)e^{-(c+s)t}},\ N_0 := N(0)$$

11. Roughly, 2

Exercises 5.6, page 287

1. $m_1x'' = k_1(y-x)$,
$m_2y'' = -k_1(y-x) - k_2y$;
$x(0) = -1$, $x'(0) = 0$, $y(0) = 0$, $y'(0) = 0$
$x(t) = -(8/17)\cos t - (9/17)\cos\left(\sqrt{20/3}\,t\right)$,
$y(t) = -(6/17)\cos t + (6/17)\cos\left(\sqrt{20/3}\,t\right)$

3. $mx'' = -kx + k(y-x)$,
$my'' = -k(y-x) + k(z-y)$,
$mz'' = -k(z-y) - kz$;
The normal frequency $(1/2\pi)\sqrt{\left(2+\sqrt{2}\right)(k/m)}$ has the mode $x(t) = z(t) = -\left(1/\sqrt{2}\right)y(t)$; the normal frequency $(1/2\pi)\sqrt{\left(2-\sqrt{2}\right)(k/m)}$ has the mode $x(t) = z(t) = \left(1/\sqrt{2}\right)y(t)$; and the normal frequency $(1/2\pi)\sqrt{2k/m}$ has the mode $x(t) = -z(t), y(t) \equiv 0$.

5. $x(t) = -e^{-t} - te^{-t} - \cos t$; $y(t) = e^{-t} + te^{-t} - \cos t$

7. $x(t) = (2/5)\cos t + (4/5)\sin t - (2/5)\cos\sqrt{6}t$
$\quad + \left(\sqrt{6}/5\right)\sin\sqrt{6}t - \sin 2t$;
$y(t) = (4/5)\cos t + (8/5)\sin t + (1/5)\cos\sqrt{6}t$
$\quad - \left(\sqrt{6}/10\right)\sin\sqrt{6}t - (1/2)\sin 2t$

9. $(1/2\pi)\sqrt{g/l}$; $(1/2\pi)\sqrt{(g/l) + (2k/m)}$

Exercises 5.7, page 294

1. $I(t) = \left(19/\sqrt{21}\right)\left[e^{\left(-25-5\sqrt{21}\right)t/2} - e^{\left(-25+5\sqrt{21}\right)t/2}\right]$

3. $I_p(t) = (4/51)\cos 20t - (1/51)\sin 20t$; resonance frequency is $5/\pi$.

5. $M(\gamma) = 1/\sqrt{(100-4\gamma^2)^2 + 100\gamma^2}$

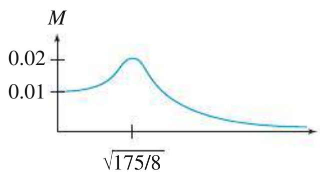


Figure B.35

7. $L = 35$ H, $R = 10\ \Omega$, $C = 1/15$ F, and $E(t) = 50\cos 10t$ V

11. $I_1 = (3/5)e^{-3t/2} - (8/5)e^{-2t/3} + 1$,
$I_2 = (1/5)e^{-3t/2} - (6/5)e^{-2t/3} + 1$,
$I_3 = (2/5)e^{-3t/2} - (2/5)e^{-2t/3}$

13. $(1/2)I_1' + 2q_3 = \cos 3t$ (where $I_3 = q_3'$),
$(1/2)I_1' + I_2 = 0, I_1 = I_2 + I_3$:
$I_1(0) = I_2(0) = I_3(0) = 0$;
$I_1 = -(36/61)e^{-t}\cos\sqrt{3}t$
$\quad - \left(42\sqrt{3}/61\right)e^{-t}\sin\sqrt{3}t + (36/61)\cos 3t$
$\quad + (30/61)\sin 3t$,
$I_2 = (45/61)e^{-t}\cos\sqrt{3}t - \left(39\sqrt{3}/61\right)e^{-t}\sin\sqrt{3}t$
$\quad - (45/61)\cos 3t + (54/61)\sin 3t$,
$I_3 = -(81/61)e^{-t}\cos\sqrt{3}t$
$\quad - \left(3\sqrt{3}/61\right)e^{-t}\sin\sqrt{3}t + (81/61)\cos 3t$
$\quad - (24/61)\sin 3t$

Exercises 5.8, page 303

1. For $\omega = 3/2$: The Poincaré map alternates between the points $(0.8, 1.5)$ and $(0.8, -1.5)$. There is a subharmonic solution of period 4π. For $\omega = 3/5$: The Poincaré map cycles through the points $(-1.5625, 0.6)$, $(-2.1503, -0.4854)$, $(-0.6114, 0.1854)$, $(-2.5136, 0.1854)$, and $(-0.9747, -0.4854)$. There is a subharmonic solution of period 10π.

3. The points become unbounded.

5. The attractor is the point $(-1.0601, 0.2624)$.

9. **(a)** $\{1/7, 2/7, 4/7, 1/7, \ldots\}$,
$\{3/7, 6/7, 5/7, 3/7, \ldots\}$
(b) $\{1/15, 2/15, 4/15, 8/15, 1/15, \ldots\}$,
$\{1/5, 2/5, 4/5, 3/5, 1/5, \ldots\}$,
$\{1/3, 2/3, 1/3, \ldots\}$,
$\{7/15, 14/15, 13/15, 11/15, 7/15, \ldots\}$
(c) $x_n = 0$ for $n \geq j$

13.

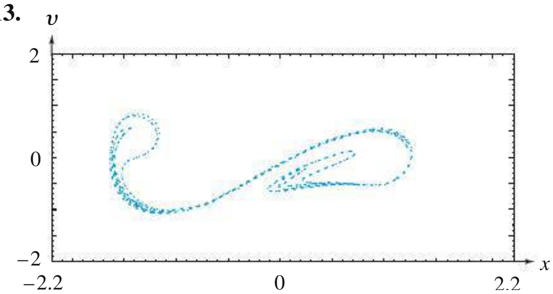


Figure B.36

Review Problems, page 306

1. $x(t) = -(c_1/3)t^3 - (c_2/2)t^2 - (c_3 + 2c_1)t + c_4$,
$y(t) = c_1t^2 + c_2t + c_3$

3. $x(t) = c_1\cos 3t + c_2\sin 3t + e^t/10$,
$y(t) = (3/2)(c_1 + c_2)\cos 3t - (3/2)(c_1 - c_2)\sin 3t$
$\quad - (11/20)e^t - (1/4)e^{-t}$

5. $x(t) = 2\sin t$, $y(t) = e^t - \cos t + \sin t$,
$z(t) = e^t + \cos t + \sin t$

7. $x(t) = -(13.9/4)e^{-t/6} - (4.9/4)e^{-t/2} + 4.8$,
$y(t) = -(13.9/2)e^{-t/6} + (4.9/2)e^{-t/2} + 4.8$

9. $x_1' = x_2$, $x_2' = x_3$, $x_3' = (1/3)(5 + e^tx_1 - 2x_2)$

11. $x_1' = x_2$, $x_2' = x_3$, $x_3' = t - x_5 - x_6$,
$x_4' = x_5$, $x_5' = x_6$, $x_6' = x_2 - x_3$

13. $x^2 - (y-1)^2 = c$; critical point $(0, 1)$ is saddle (unstable).

15. Asymptotically stable spiral point

17. $I_1 + I_2 + I_3 = 0$, $q/C = R_2I_2$,
$R_2I_2 = R_1I_3 + L\,dI_3/dt$, where q is the charge on the capacitor $(I_1 = dq/dt)$;
$I_3 = e^{-t}(A\cos t + B\sin t)$,
$I_2 = e^{-t}(B\cos t - A\sin t)$,
$I_1 = e^{-t}[(A-B)\sin t - (A+B)\cos t]$.

CHAPTER 6

Exercises 6.1, page 326

1. $(-\infty, 0)$ **3.** $(3\pi/2, 5\pi/2)$ **5.** $(0, \infty)$
7. Lin. indep. **9.** Lin. dep.
11. Lin. indep. **13.** Lin. indep.
15. $c_1e^{3x} + c_2e^{-x} + c_3e^{-4x}$
17. $c_1x + c_2x^2 + c_3x^3$
19. **(a)** $c_1e^x + c_2e^{-x}\cos 2x + c_3e^{-x}\sin 2x + x^2$
(b) $-e^x + e^{-x}\sin 2x + x^2$
21. **(a)** $c_1x + c_2x\ln x + c_3x(\ln x)^2 + \ln x$
(b) $3x - x\ln x + x(\ln x)^2 + \ln x$
23. **(a)** $2\sin x - x$
(b) $4x - 6\sin x$
29. **(b)** Let $f_1(x) = |x-1|$ and $f_2(x) = x-1$
33. e^{2x}, $(\sin x - 2\cos x)/5$, $-(2\sin x + \cos x)/5$
35. $xy''' - y'' + xy' - y = 0$

Exercises 6.2, page 332

1. $c_1 + c_2e^{2x} + c_3e^{-4x}$
3. $c_1e^{-x} + c_2e^{-2x/3} + c_3e^{x/2}$
5. $c_1e^{-x} + c_2e^{-x}\cos 5x + c_3e^{-x}\sin 5x$
7. $c_1e^{-x} + c_2e^{(3+\sqrt{65})x/4} + c_3e^{(3-\sqrt{65})x/4}$
9. $c_1e^{3x} + c_2xe^{3x} + c_3x^2e^{3x}$
11. $c_1e^{-x} + c_2xe^{-x} + c_3x^2e^{-x} + c_4x^3e^{-x}$
13. $c_1\cos\sqrt{2}x + c_2x\cos\sqrt{2}x + c_3\sin\sqrt{2}x + c_4x\sin\sqrt{2}x$
15. $c_1e^x + c_2xe^x + c_3e^{-3x} + (c_4 + c_5x)e^{-x}\cos 2x$
$+ (c_6 + c_7x)e^{-x}\sin 2x$
17. $c_1e^{-4x} + c_2e^{3x} + (c_3 + c_4x + c_5x^2)e^{-2x}$
$+ (c_6 + c_7x)e^{-2x}\cos x + (c_8 + c_9x)e^{-2x}\sin x + c_{10}$
$+ c_{11}x + c_{12}x^2 + c_{13}x^3 + c_{14}x^4$
19. $e^x - 2e^{-2x} - 3e^{2x}$
21. $e^{2x} - \sqrt{2}e^x\sin\sqrt{2}x$
23. $x(t) = c_1 + c_2t + c_3e^t$, $y(t) = c_1 - c_2 + c_2t$
27. $c_1e^{1.120x} + c_2e^{0.296x} + c_3e^{-0.520x} + c_4e^{-2.896x}$
29. $c_1e^{-0.5x}\cos(0.866x) + c_2e^{-0.5x}\sin(0.866x)$
$+ c_3e^{-0.5x}\cos(1.323x) + c_4e^{-0.5x}\sin(1.323x)$
31. **(a)** $\{x, x^{-1}, x^2\}$
(b) $\{x, x^2, x^{-1}, x^{-2}\}$
(c) $\{x, x^2\cos(3\ln x), x^2\sin(3\ln x)\}$
33. **(b)** $x(t) = c_1\cos t + c_2\sin t + c_3\cos\sqrt{6}t$
$+ c_4\sin\sqrt{6}t$
(c) $y(t) = 2c_1\cos t + 2c_2\sin t - (c_3/2)\cos\sqrt{6}t$
$- (c_4/2)\sin\sqrt{6}t$
(d) $x(t) = (3/5)\cos t + (2/5)\cos\sqrt{6}t$,
$y(t) = (6/5)\cos t - (1/5)\cos\sqrt{6}t$
35. $c_1\cosh rx + c_2\sinh rx + c_3\cos rx + c_4\sin rx$, where $r^4 = k/(EI)$

Exercises 6.3, page 337

1. $c_1xe^x + c_2 + c_3x + c_4x^2$
3. $c_1x^2e^{-2x}$
5. $c_1e^x + c_2e^{3x} + c_3e^{-2x} - (1/6)xe^x + (1/6)x^2$
$+ (5/18)x + 37/108$
7. $c_1e^x + c_2e^{-2x} + c_3xe^{-2x} - (1/6)x^2e^{-2x}$
9. $c_1e^x + c_2xe^x + c_3x^2e^x + (1/6)x^3e^x$
11. D^5
13. $D + 7$
15. $(D-2)(D-1)$
17. $[(D+1)^2 + 4]^3$
19. $(D+2)^2[(D+5)^2 + 9]^2$
21. $c_3\cos 2x + c_4\sin 2x + c_5$
23. $c_3xe^{3x} + c_4x^2 + c_5x + c_6$
25. $c_3 + c_4x + c_5\cos 2x + c_6\sin 2x$
27. $c_3xe^{-x}\cos x + c_4xe^{-x}\sin x + c_5x^2 + c_6x + c_7$
29. $c_2x + c_3x^2 + c_6x^2e^x$
31. $-2e^{3x} + e^{-2x} + x^2 - 1$
33. $x^2e^{-2x} - x^2 + 3$
39. $x(t) = -(1/63)e^{3t} + c_1 + c_2t - c_3e^{\sqrt{2}t} - c_4e^{-\sqrt{2}t}$,
$y(t) = (8/63)e^{3t} + c_1 + c_2t + c_3e^{\sqrt{2}t} + c_4e^{-\sqrt{2}t}$

Exercises 6.4, page 341

1. $(1/6)x^2e^{2x}$
3. $e^{2x}/16$
5. $\ln(\sec x) - (\sin x)\ln(\sec x + \tan x)$
7. $c_1x + c_2x^2 + c_3x^3 - (1/24)x^{-1}$
9. $-(1/2)e^x\int e^{-x}g(x)\,dx + (1/6)e^{-x}\int e^xg(x)\,dx$
$+ (1/3)e^{2x}\int e^{-2x}g(x)\,dx$
11. $c_1x + c_2x^{-1} + c_3x^3 - x\sin x - 3\cos x + 3x^{-1}\sin x$

Review Problems, page 343

1. **(a)** $(0, \infty)$
(b) $(-4, -1)$, $(-1, 1)$, $(1, \infty)$
5. **(a)** $e^{-5x}(c_1 + c_2x) + e^{2x}(c_3 + c_4x + c_5x^2)$
$+ (\cos x)(c_6 + c_7x) + (\sin x)(c_8 + c_9x)$
(b) $c_1 + c_2x + c_3x^2 + c_4x^3 + e^x(c_5 + c_6x)$
$+ (e^{-x}\cos\sqrt{3}x)(c_7 + c_8x)$
$+ (e^{-x}\sin\sqrt{3}x)(c_9 + c_{10}x)$
7. **(a)** D^3 **(b)** $D^2(D-3)$ **(c)** $[D^2+4]^2$
(d) $[(D+2)^2 + 9]^3$
(e) $D^3(D+1)^2(D^2+4)(D^2+9)$
9. $c_1x + c_2x^5 + c_3x^{-1} - (1/21)x^{-2}$

CHAPTER 7

Exercises 7.2, page 360

1. $\dfrac{1}{s^2}$, $s > 0$ **3.** $\dfrac{1}{s-6}$, $s > 6$
5. $\dfrac{s}{s^2+4}$, $s > 0$
7. $\dfrac{s-2}{(s-2)^2+9}$, $s > 2$
9. $e^{-2s}\left(\dfrac{2s+1}{s^2}\right)$, $s > 0$ **11.** $\dfrac{e^{-\pi s}+1}{s^2+1}$, all s

13. $\dfrac{6}{s+3}-\dfrac{2}{s^3}+\dfrac{2}{s^2}-\dfrac{8}{s}, \quad s>0$

15. $\dfrac{6}{s^4}-\dfrac{1}{(s-1)^2}+\dfrac{s-4}{(s-4)^2+1}, \quad s>4$

17. $\dfrac{6}{(s-3)^2+36}-\dfrac{6}{s^4}+\dfrac{1}{s-1}, \quad s>3$

19. $\dfrac{24}{(s-5)^5}-\dfrac{s-1}{(s-1)^2+7}, \quad s>5$

21. Continuous (hence piecewise continuous)

23. Piecewise continuous

25. Continuous (hence piecewise continuous)

27. Neither

29. All but functions (c), (e), and (h)

Exercises 7.3, page 365

1. $\dfrac{2}{s^3}+\dfrac{2}{(s-1)^2+4}$

3. $\dfrac{s+1}{(s+1)^2+9}+\dfrac{1}{s-6}-\dfrac{1}{s}$

5. $\dfrac{4}{(s+1)^3}-\dfrac{1}{s^2}+\dfrac{s}{s^2+16}$

7. $\dfrac{24}{s^5}-\dfrac{24}{s^4}+\dfrac{12}{s^3}-\dfrac{4}{s^2}+\dfrac{1}{s}$

9. $\dfrac{4(s+1)}{[(s+1)^2+4]^2}$ **11.** $\dfrac{s}{s^2-b^2}$

13. $\dfrac{1}{2s}-\dfrac{s}{2(s^2+4)}$ **15.** $\dfrac{3s}{4(s^2+1)}+\dfrac{s}{4(s^2+9)}$

17. $\dfrac{s}{2(s^2+9)}-\dfrac{s}{2(s^2+49)}$

19. $\dfrac{n+m}{2[s^2+(n+m)^2]}+\dfrac{m-n}{2[s^2+(m-n)^2]}$

21. $\dfrac{s-a}{(s-a)^2+b^2}$

25. **(a)** $\dfrac{s^2-b^2}{(s^2+b^2)^2}$ **(b)** $\dfrac{2s^3-6sb^2}{(s^2+b^2)^3}$

29. $\dfrac{1}{s^2+6s+10}$ **33.** e^{-s}/s^2

35. $e^{-(\pi/2)s}/(s^2+1)$

Exercises 7.4, page 374

1. $e^t t^3$ **3.** $e^{-t}\cos 3t$ **5.** $(1/2)e^{-2t}\sin 2t$

7. $2e^{-2t}\cos 3t+4e^{-2t}\sin 3t$

9. $(3/2)e^t\cos 2t-3e^t\sin 2t$

11. $\dfrac{6}{s+5}-\dfrac{1}{s+2}-\dfrac{4}{s-1}$ **13.** $\dfrac{1}{(s+1)^2}-\dfrac{2}{s}$

15. $-\dfrac{3}{s+1}+\dfrac{(s-1)+2}{(s-1)^2+4}$

17. $-\dfrac{5}{6s}+\dfrac{11}{10(s-2)}-\dfrac{4}{15(s+3)}$

19. $\dfrac{1}{17}\left[\dfrac{1}{s-3}-\dfrac{s+1}{(s+1)^2+1}-\dfrac{4}{(s+1)^2+1}\right]$

21. $(1/3)+e^t+(14/3)e^{6t}$

23. $-e^{-3t}+2te^{-3t}+6e^{-t}$

25. $8e^{2t}-e^{-t}\cos 2t+3e^{-t}\sin 2t$

27. $-(5/3)e^{-t}+(5/12)e^{2t}+(5/4)e^{-2t}$

29. $3e^{-2t}+7e^t\cos t+11e^t\sin t$

31. $F_1(s)=F_2(s)=F_3(s)=1/s^2$,
$\mathscr{L}^{-1}\{1/s^2\}(t)=f_3(t)=t$

33. $e^{5t}/t-e^{-2t}/t$

35. $2(\cos t-\cos 3t)/t$

39. $\dfrac{A}{s}+\dfrac{B}{s-1}+\dfrac{C}{s+2}$, where

$$A=\left.\frac{2s+1}{(s-1)(s+2)}\right|_{s=0}=\frac{-1}{2},$$
$$B=\left.\frac{2s+1}{s(s+2)}\right|_{s=1}=1,$$
$$C=\left.\frac{2s+1}{s(s-1)}\right|_{s=-2}=\frac{-1}{2}$$

41. $2e^{-t}-4e^{3t}+5e^{2t}$

43. $\dfrac{4}{s+2}+\dfrac{2(s-1)+2(3)}{(s-1)^2+2^2}$

Exercises 7.5, page 382

1. $2e^t\cos 2t+e^t\sin 2t$ **3.** $-e^{-3t}+3te^{-3t}$

5. $t^2+\cos t-\sin t$ **7.** $\cos t-4e^{5t}+8e^{2t}$

9. $3(t-1)e^{t-1}-e^{-6(t-1)}$

11. $2-t+e^{2-t}+2e^{t-2}$

13. $(7/5)\sin t+(11/5)\cos t+(3/5)e^{2t-\pi}-e^{-(t-\pi/2)}$

15. $\dfrac{-s^2+s-1}{(s^2+1)(s-1)(s-2)}$

17. $\dfrac{s^5+s^4+6}{s^4(s^2+s-1)}$ **19.** $\dfrac{s^3+5s^2-6s+1}{s(s-1)(s^2+5s-1)}$

21. $\dfrac{s^3+s^2+2s}{(s^2+1)(s-1)^2}$ **23.** $\dfrac{-s^3+1+3se^{-2s}-e^{-2s}}{s^2(s^2+4)}$

25. $2e^t-\cos t-\sin t$ **27.** $(t^2-4)e^{-t}$

29. $(3a-b)e^t/2+(b-a)e^{3t}/2$

31. $5/2+(a-5/2)e^{-t}\cos t+(a+b-5/2)e^{-t}\sin t$

35. $t^2/2$

37. $\cos t+t\sin t+c(\sin t-t\cos t)$, ($c$ arbitrary)

39. $e(t)=-a\cos\left(\sqrt{k/I}\,t\right)$

41. $e(t)=\left(-2aI/\sqrt{4Ik-\mu^2}\right)e^{-\mu t/(2I)}\sin\left(\sqrt{4Ik-\mu^2}\,t/(2I)\right)$

Exercises 7.6, page 390

1. $2e^{-s}/s^3$ **3.** $e^{-2s}(4s^2+4s+2)/s^3$

5. $(2e^{-s}-e^{-2s}+2e^{-3s})/s$

7. $[(e^{-s}-e^{-2s})(s+1)]/s^2$

9. $(e^{-s}-2e^{-2s}+e^{-3s})/s^2$

11. $e^{t-2}u(t-2)$

13. $e^{-2(t-2)}u(t-2)-3e^{-2(t-4)}u(t-4)$

15. $e^{-2(t-3)}[\cos(t-3)-2\sin(t-3)]u(t-3)$

17. $(7e^{6-2t}-6e^{3-t})u(t-3)$

19. $10-10u(t-3\pi)[1+e^{-(t-3\pi)}(\cos t+\sin t)]$
$+10u(t-4\pi)[1-e^{-(t-4\pi)}(\cos t+\sin t)]$

21. $\sin t + [1 - \cos(t-3)]u(t-3)$

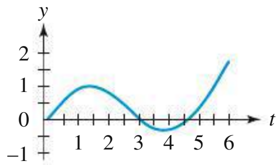


Figure B.37

23. $t + [4 - t + \sin(t-2) - 2\cos(t-2)]u(t-2)$

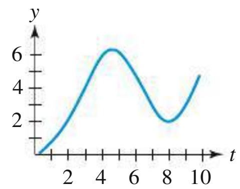


Figure B.38

25. $e^{-t}\cos t + 2e^{-t}\sin t$
$+ (1/2)[1 - e^{2\pi - t}(\cos t + \sin t)]u(t - 2\pi)$
$- (1/2)[1 - e^{4\pi - t}(\cos t + \sin t)]u(t - 4\pi)$

27. $e^{-t} + e^{-2t} + (1/2)[e^{-3t} - 2e^{-2(t+1)} + e^{-(t+4)}]u(t-2)$

29. $\cos 2t + (1/3)[1 - u(t - 2\pi)]\sin t$
$+ (1/6)[8 + u(t - 2\pi)]\sin 2t$

31. $2e^{-2t} - 2e^{-3t}$
$+ [1/36 + (1/6)(t-1) - (1/4)e^{-2(t-1)}$
$+ (2/9)e^{-3(t-1)}]u(t-1)$
$- [19/36 + (1/6)(t-5)$
$- (7/4)e^{-2(t-5)} + (11/9)e^{-3(t-5)}]u(t-5)$

33. $0.04 - 0.02\, e^{-3t/125} + 0.02\, u(t-10)[1 - e^{-3(t-10)/125}]$

35. The resulting differential equation has polynomial coefficients, so the Laplace transform method will result in a differential equation for the transform.

Exercises 7.7, page 396

1. $\dfrac{1 - 2se^{-2s} - e^{-2s}}{s^2(1 - e^{-2s})}$

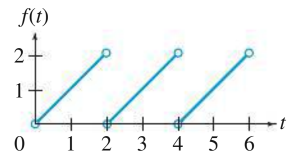


Figure B.39

3. $\dfrac{1}{1 - e^{-2s}}\left[\dfrac{1 - e^{-s-1}}{s+1} + \dfrac{e^{-s} - e^{-2s}}{s}\right]$

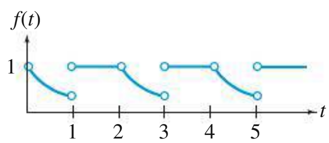


Figure B.40

5. $\dfrac{1}{s(1 + e^{-as})}$ **7.** $\dfrac{1 - e^{-as}}{as^2(1 + e^{-as})}$

13. $\dfrac{e^{t-n}}{6} - \dfrac{e^{-t}}{2}(1 + e - e^{n+1})$
$+ \dfrac{e^{-2t}}{3}\left[\dfrac{1 + e + e^2 - e^{2n+2}}{e+1}\right],$
for $n < t < n+1$

15. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{s^n} = \frac{1}{s-1}$

17. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{2ns^{2n}} = \frac{1}{2}\ln\left(1 + \frac{1}{s^2}\right)$

19. **(a)** $\sqrt{\pi/s}$
(b) $105\sqrt{\pi}/(16s^{9/2})$

27. (a) and (b), but not (c)

Exercises 7.8, page 404

1. $2te^t - e^t + \displaystyle\int_0^t e^{t-v}(t-v)g(v)\,dv$

3. $\displaystyle\int_0^t g(v)e^{2v-2t}\sin(t-v)\,dv + e^{-2t}\cos t + 3e^{-2t}\sin t$

5. $1 - \cos t$ **7.** $2e^{5t} - 2e^{-2t}$ **9.** $(t/2)\sin t$

11. $(2/3)e^{-2t} + (1/3)e^t$ **13.** $s^{-2}(s-3)^{-1}$

15. $t/4 + (3/8)\sin 2t$ **17.** $\cos t$ **19.** 3

21. $e^{-t/2}\cos(\sqrt{3}t/2) - (1/\sqrt{3})e^{-t/2}\sin(\sqrt{3}t/2)$

23. $H(s) = (s^2 + 9)^{-1}$; $h(t) = (1/3)\sin 3t$;
$y_k(t) = 2\cos 3t - \sin 3t$;
$y(t) = (1/3)\displaystyle\int_0^t [\sin 3(t-v)]g(v)\,dv + 2\cos 3t - \sin 3t$

25. $H(s) = (s^2 - s - 6)^{-1}$; $h(t) = (e^{3t} - e^{-2t})/5$;
$y_k(t) = 2e^{3t} - e^{-2t}$;
$y(t) = (1/5)\displaystyle\int_0^t [e^{3(t-v)} - e^{-2(t-v)}]g(v)\,dv + 2e^{3t} - e^{-2t}$

27. $H(s) = (s^2 - 2s + 5)^{-1}$; $h(t) = (1/2)e^t\sin 2t$;
$y_k(t) = e^t\sin 2t$;
$y(t) = (1/2)\displaystyle\int_0^t e^{(t-v)}[\sin 2(t-v)]g(v)\,dv + e^t\sin 2t$

29. $(1/30)\displaystyle\int_0^t e^{-2(t-v)}[\sin 6(t-v)]e(v)\,dv - e^{-2t}\cos 6t$
$+ e^{-2t}\sin 6t$

31. $t^2/2$

Exercises 7.9, page 410

1. -1 **3.** -1 **5.** e^{-2} **7.** $e^{-s} - e^{-3s}$

9. e^{-s} **11.** 0 **13.** $-(\sin t)u(t - \pi)$

15. $e^t + e^{-3t} + (1/4)(e^{t-1} - e^{3-3t})u(t-1)$
$- (1/4)(e^{t-2} - e^{6-3t})u(t-2)$

17. $2(e^{t-2} - e^{-(t-2)})u(t-2) + 2e^t - t^2 - 2$

19. $e^{-t} - e^{-5t} + (e/4)(e^{1-t} - e^{5-5t})u(t-1)$

21. $\sin t + (\sin t)u(t-2\pi)$; see Fig. B.41.

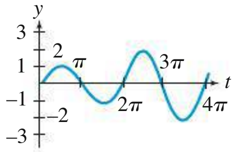


Figure B.41

23. $\sin t + (\sin t)u(t-\pi) + (\sin t)u(t-2\pi)$

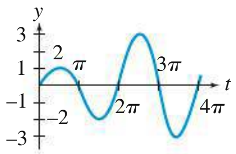


Figure B.42

25. $(1/2)e^{-2t}\sin 2t$ **27.** $(1/2)e^{t}\sin 2t$

29. The mass remains stopped at $x(t) \equiv 0$, $t > \pi/2$

35. $\dfrac{L}{6EI}[3\lambda x^2 - x^3 + (x-\lambda)^3 u(x-\lambda)]$

Exercises 7.10, page 413

1. $x = e^t$; $y = e^t$ **3.** $z = e^t$; $w \equiv 0$

5. $x = (7/4)e^t + (7/4)e^{-t} - (3/2)\cos t$;
$y = (7/4)e^t - (7/4)e^{-t} + (1/2)\sin t$

7. $x = -(150/17)e^{5t/2}\cos(\sqrt{15}t/2) - (334\sqrt{15}/85)e^{5t/2}\sin(\sqrt{15}t/2) - (3/17)e^{-3t}$;
$y = (46/17)e^{5t/2}\cos(\sqrt{15}t/2) - (146\sqrt{15}/85)e^{5t/2}\sin(\sqrt{15}t/2) + (22/17)e^{-3t}$

9. $x = 4e^{-2t} - e^{-t} - \cos t$; $y = 5e^{-2t} - e^{-t}$

11. $x = (e^t - e^{-t})/2 - (1/2)[e^{t-2} - e^{-(t-2)}]u(t-2)$;
$y = 1 - (e^t + e^{-t})/2 - [1 - (e^{t-2} + e^{-(t-2)})/2]u(t-2)$

13. $x = e^{-t} - (1/2)[e^{-(t-\pi)} + \cos t - \sin t]u(t-\pi)$;
$y = e^{-t} + [1 - (1/2)e^{-(t-\pi)} + (1/2)\cos t + (1/2)\sin t]u(t-\pi)$

15. $x = t^2$; $y = t - 1$

17. $x = (t-2)e^{t-2}$; $y = e^{t-2}$

19. $x = -7e^{-t} + e^t$; $y = 2e^{-t}$; $z = -13e^{-t} + e^t$

21. $x = 0.01(-e^{-t/2} + e^{-t/6}) + [0.48 - 0.36e^{-(t-5)/6} - 0.12e^{-(t-5)/2}]\,u(t-5)$;
$y = 0.02(e^{-t/2} + e^{-t/6}) + [0.48 - 0.72e^{-(t-5)/6} + 0.24e^{-(t-5)/2}]\,u(t-5)$

23. $2I_1 + (0.1)I_3' + (0.2)I_1' = 6$, $(0.1)I_3' - I_2 = 0$,
$I_1 = I_2 + I_3$; $I_1(0) = I_2(0) = I_3(0) = 0$;
$I_1 = -e^{-20t} - 2e^{-5t} + 3$,
$I_2 = -2e^{-20t} + 2e^{-5t}$,
$I_3 = e^{-20t} - 4e^{-5t} + 3$

Review Problems, page 415

1. $\dfrac{3}{s} + e^{-2s}\left[\dfrac{1}{s} - \dfrac{1}{s^2}\right]$ **3.** $\dfrac{2}{(s+9)^3}$

5. $\dfrac{1}{s-2} - \dfrac{6}{s^4} + \dfrac{2}{s^3} - \dfrac{5}{s^2+25}$ **7.** $\dfrac{s^2-36}{(s^2+36)^2}$

9. $2e^{-4s}\left[\dfrac{1}{s^3} + \dfrac{4}{s^2} + \dfrac{8}{s}\right]$ **11.** $(7/2)t^2e^{-3t}$

13. $2e^t + 2e^{-2t}\cos 3t + e^{-2t}\sin 3t$

15. $e^{-2t} + e^{-t} - 2te^{-t}$

17. $[2e^{t-2} + 2e^{4-2t}]u(t-2)$

19. $e^{2t} - e^{5t}$

21. $-(3/2) + t + t^2/2 + (3/2)e^{-t}\cos t - (1/2)e^{-t}\sin t$

23. $(2/\sqrt{7})e^{-3t/2}\sin(\sqrt{7}t/2) + \{(1/4) - [3/(4\sqrt{7})]e^{-3(t-1)/2}\sin[\sqrt{7}(t-1)/2] - (1/4)e^{-3(t-1)/2}\cos[\sqrt{7}(t-1)/2]\}u(t-1)$

25. $c[t + te^{-2t} + e^{-2t} - 1]$

27. $(9/10)e^{-3t} + (1/10)\cos t - (3/10)\sin t$

29. $(s^2 - 5s + 6)^{-1}$; $e^{3t} - e^{2t}$

31. $x = 1 - (e^t + e^{-t})/2 - [1 - (e^{t-2} + e^{-(t-2)})/2]u(t-2)$,
$y = (e^t - e^{-t})/2 - (1/2)[e^{t-2} - e^{-(t-2)}]u(t-2)$

CHAPTER 8

Exercises 8.1, page 425

1. $1 + x + x^2 + \cdots$

3. $x + x^2 + (1/2)x^3 + \cdots$

5. $1 - (1/6)t^3 + (1/180)t^6 + \cdots$

7. $(1/6)\theta^3 - (1/120)\theta^5 + (1/5040)\theta^7 + \cdots$

9. **(a)** $p_3(x) = (x-1) - \dfrac{1}{2}(x-1)^2 + \dfrac{1}{3}(x-1)^3$
(b) $|\varepsilon_3(1.5)| = |f^{(4)}(\xi)(1.5-1)^4/4!| \le 6(0.5)^4/4!$
(c) $|\ln(1.5) - p_3(1.5)| = 0.011202\ldots$
(d)

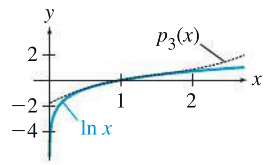


Figure B.43

13. $t + (1/2)t^2 - (1/6)t^3 + \cdots$ **15.** $1 - x^2/6$

Exercises 8.2, page 433

1. $[-1, 3)$ **3.** $(-4, 0)$ **5.** $[1, 3]$

9. $\displaystyle\sum_{n=0}^{\infty}\left[\frac{1}{n+1} + 2^{-n-1}\right]x^n$

11. $x + x^2 + (1/3)x^3 + \cdots$

13. $1 - 2x + (5/2)x^2 + \cdots$

15. (c) $1 - (1/2)x + (1/4)x^2 - (1/24)x^3 + \cdots$

17. $\sum_{n=1}^{\infty} (-1)^n n x^{n-1}$ **19.** $\sum_{k=1}^{\infty} a_k 2k x^{2k-1}$

21. $\ln(x+1) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1}$

23. $\sum_{k=0}^{\infty} (k+1) a_{k+1} x^k$

25. $\sum_{k=1}^{\infty} a_{k-1} x^k$ **29.** $\sum_{n=0}^{\infty} \frac{(-1)^{n+1}(x-\pi)^{2n}}{(2n)!}$

31. $1 + \sum_{n=1}^{\infty} 2x^n$

33. $6(x-1) + 3(x-1)^2 + (x-1)^3$

35. (a) $\sum_{n=0}^{\infty} (-1)^n (x-1)^n$ **(b)** $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} (x-1)^n$

Exercises 8.3, page 443

1. -1 **3.** $\pm\sqrt{2}$ **5.** $-1, 2$

7. $x = n\pi$, n an integer

9. $\theta \le 0$ and $\theta = n\pi$, $n = 1, 2, 3, \ldots$

11. $y = a_0(1 - 2x + (3/2)x^2 - x^3/3 + \cdots)$

13. $a_0(1 + x^4/12 + \cdots) + a_1(x + x^5/20 + \cdots)$

15. $a_0(1 - x^2/2 - x^3/6 + \cdots)$
$+ a_1(x + x^2/2 - x^3/6 + \cdots)$

17. $a_0(1 - x^2/2 + \cdots) + a_1(x - x^3/6 + \cdots)$

19. $a_0 \sum_{n=0}^{\infty} \frac{1}{n!} x^{2n} = a_0 e^{x^2}$

21. $a_0(1 - 2x^2 + x^4/3)$
$+ a_1\left[x + \sum_{k=1}^{\infty} \frac{(-3)(-1)\cdots(2k-5)}{(2k+1)!} x^{2k+1}\right]$

23. $a_{3k+2} = 0$, $k = 0, 1, \ldots$

$$a_0\left(1 + \sum_{k=1}^{\infty} \frac{[1\cdot 4\cdot 7\cdots(3k-2)]^2}{(3k)!} x^{3k}\right) + a_1\left(x + \sum_{k=1}^{\infty} \frac{[2\cdot 5\cdot 8\cdots(3k-1)]^2}{(3k+1)!} x^{3k+1}\right)$$

25. $2 + x^2 - (5/12)x^4 + (11/72)x^6 + \cdots$

27. $x + (1/6)x^3 - (1/12)x^4 + (7/120)x^5 + \cdots$

29. $1 - 2x + x^2 - (1/6)x^3$

31. $1 + 2x - (3/4)x^2 - (5/6)x^3$

35. $10 - 25t^2 + (250/3)t^3 - (775/4)t^4 + \cdots$

Exercises 8.4, page 449

1. 2 **3.** $\sqrt{3}$ **5.** $\pi/2$

7. $a_0[1 - (x-1)^2 + (1/2)(x-1)^4$
$- (1/6)(x-1)^6 + \cdots]$

9. $a_0[1 + (x-1)^2 + \cdots]$
$+ a_1[(x-1) + (1/3)(x-1)^3 + \cdots]$

11. $a_0[1 - (1/8)(x-2)^2 + (1/32)(x-2)^3 + \cdots]$
$+ a_1[(x-2) + (1/8)(x-2)^2$
$-(7/96)(x-2)^3 + \cdots]$

13. $1 - (1/2)t^2 + (1/6)t^4 - (31/720)t^6 + \cdots$

15. $1 + x + (1/24)x^4 + (1/60)x^5 + \cdots$

17. $1 - (1/6)(x-\pi)^3 + (1/120)(x-\pi)^5$
$+ (1/180)(x-\pi)^6 + \cdots$

19. $-1 + x + x^2 + (1/2)x^3 + \cdots$

21. $a_0[1 + (1/2)x^2 + (1/8)x^4 + (1/48)x^6 + \cdots]$
$+ [(1/2)x^2 + (1/12)x^4 + (11/720)x^6 + \cdots]$

23. $a_0[1 - (1/2)x^2 + \cdots] + a_1[x - (1/3)x^3 + \cdots]$
$+ [(1/2)x^2 + (1/3)x^3 + \cdots]$

25. $a_0[1 - (1/2)x^2 + \cdots] + a_1[x + \cdots]$
$+ [(1/2)x^2 - (1/6)x^3 + \cdots]$

27. $a_0[1 - (1/2)x^2 - (1/6)x^3 + \cdots]$
$+ a_1[x + (1/2)x^2 + \cdots] + [(1/6)x^3 + \cdots]$

---

29. (a)
$$a_0\left[1 + \sum_{k=1}^{\infty} (-1)^k \frac{n(n-2)(n-4)\cdots(n-2k+2)(n+1)(n+3)\cdots(n+2k-1)}{(2k)!} x^{2k}\right] + a_1\left[x + \sum_{k=1}^{\infty} (-1)^k \frac{(n-1)(n-3)\cdots(n-2k+1)(n+2)(n+4)\cdots(n+2k)}{(2k+1)!} x^{2k+1}\right]$$

(c) $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = (1/2)(3x^2 - 1)$

31. (a) $1 - (1/2)t^2 + (\eta/6)t^3 + [(1 - \eta^2)/24]t^4 + \cdots$ **(b)** Yes

Exercises 8.5, page 453

1. $c_1x^{-2} + c_2x^{-3}$
3. $c_1x\cos(4\ln x) + c_2x\sin(4\ln x)$
5. $c_1x^3\cos(2\ln x) + c_2x^3\sin(2\ln x)$
7. $c_1x + c_2x^{-1}\cos(3\ln x) + c_3x^{-1}\sin(3\ln x)$
9. $c_1x + c_2x^{-1/2}\cos[(\sqrt{19}/2)\ln x] + c_3x^{-1/2}\sin[(\sqrt{19}/2)\ln x]$
11. $c_1(x-3)^{-2} + c_2(x-3)^{1/2}$
13. $c_1x + c_2x^2 + (4/15)x^{-1/2}$
15. $2t^4 + t^{-3}$
17. $(31/17)x + (3/17)x^{-2}\cos(5\ln x) - (76/85)x^{-2}\sin(5\ln x)$

Exercises 8.6, page 464

1. ± 1 are regular.
3. $\pm i$ are regular.
5. 1 is regular, -1 is irregular.
7. 2 is regular, -1 is irregular.
9. -4 is regular and 2 is irregular.
11. $r^2 - 3r - 10 = 0$; $r_1 = 5$, $r_2 = -2$
13. $r^2 - 5r/9 - 4/3 = 0$; $r_1 = (5 + \sqrt{457})/18$, $r_2 = (5 - \sqrt{457})/18$
15. $r^2 - 1 = 0$; $r_1 = 1$, $r_2 = -1$
17. $r^2 + r - 12 = 0$; $r_1 = 3$, $r_2 = -4$
19. $a_0[x^{2/3} - (1/2)x^{5/3} + (5/28)x^{8/3} - (1/21)x^{11/3} + \cdots]$
21. $a_0[1 - (1/4)x^2 + (1/64)x^4 - (1/2304)x^6 + \cdots]$
23. $a_0[x - (1/3)x^2 + (1/12)x^3 - (1/60)x^4 + \cdots]$
25. $a_0 \sum_{n=0}^{\infty} \frac{(-1)^n x^{n+(3/2)}}{2^{n-1}(n+2)!}$
27. $a_0 \sum_{n=0}^{\infty} \frac{x^{2n+2}}{2^{2n}(n+1)!n!}$
29. a_0x^2
31. $a_0 \sum_{n=0}^{\infty} \frac{x^n}{n!} = a_0e^x$; yes, $a_0 < 0$
33. $a_0\left[x^{1/3} + \sum_{n=1}^{\infty} \frac{2^{n-2}(3n+4)x^{n+(1/3)}}{3^n n!}\right]$; yes, $a_0 < 0$
35. $a_0[x^{5/6} - (1/11)x^{11/6} + (1/374)x^{17/6} - (1/25{,}806)x^{23/6} + \cdots]$
37. $a_0[x^{4/3} + (1/17)x^{7/3} + (1/782)x^{10/3} + (1/68{,}034)x^{13/3} + \cdots]$
39. The expansion $\sum_{n=0}^{\infty} n!x^n$ diverges for $x \neq 0$.
41. The transformed equation is $z\,d^2y/dz^2 + 3\,dy/dz - y = 0$. Also $zp(z) = 3$ and $z^2q(z) = -z$ are analytic at $z = 0$; hence, $z = 0$ is a regular singular point.
$y_1(x) = a_0[1 + (1/3)x^{-1} + (1/24)x^{-2} + (1/360)x^{-3} + \cdots]$
45. $a_0[x^{-3} - (1/2)x^{-1} - (1/8)x - (1/144)x^3 + \cdots] + a_3[1 + (1/10)x^2 + (1/280)x^4 + (1/15{,}120)x^6 + \cdots]$

Exercises 8.7, page 473

1. $c_1y_1(x) + c_2y_2(x)$, where
$y_1(x) = x^{2/3} - (1/2)x^{5/3} + (5/28)x^{8/3} + \cdots$ and
$y_2(x) = x^{1/3} - (1/2)x^{4/3} + (1/5)x^{7/3} + \cdots$
3. $c_1y_1(x) + c_2y_2(x)$, where
$y_1(x) = 1 - (1/4)x^2 + (1/64)x^4 + \cdots$ and
$y_2(x) = y_1(x)\ln x + (1/4)x^2 - (3/128)x^4 + (11/13{,}824)x^6 + \cdots$
5. $c_1[x - (1/3)x^2 + (1/12)x^3 + \cdots] + c_2[x^{-1} - 1]$
7. $c_1y_1(x) + c_2y_2(x)$, where
$y_1(x) = x^{3/2} - (1/6)x^{5/2} + (1/48)x^{7/2} + \cdots$ and
$y_2(x) = x^{-1/2} - (1/2)x^{1/2}$
9. $c_1w_1 + c_2w_2$, where
$w_1(x) = x^2 + (1/8)x^4 + (1/192)x^6 + \cdots$ and
$w_2(x) = w_1(x)\ln x + 2 - (3/32)x^4 - (7/1152)x^6 + \cdots$
11. $c_1y_1(x) + c_2y_2(x)$, where $y_1(x) = x^2$ and
$y_2(x) = x^2\ln x - 1 + 2x - (1/3)x^3 + (1/24)x^4 + \cdots$
13. $c_1y_1(x) + c_2y_2(x)$, where
$y_1(x) = 1 + x + (1/2)x^2 + \cdots$ and
$y_2(x) = y_1(x)\ln x - [x + (3/4)x^2 + (11/36)x^3 + \cdots]$
15. $c_1y_1(x) + c_2y_2(x)$, where
$y_1(x) = x^{1/3} + (7/6)x^{4/3} + (5/9)x^{7/3} + \cdots$ and
$y_2(x) = 1 + 2x + (6/5)x^2 + \cdots$; all solutions are bounded near the origin.
17. $c_1y_1(x) + c_2y_2(x) + c_3y_3(x)$, where
$y_1(x) = x^{5/6} - (1/11)x^{11/6} + (1/374)x^{17/6} + \cdots$,
$y_2(x) = 1 - x + (1/14)x^2 + \cdots$, and
$y_3(x) = y_2(x)\ln x + 7x - (117/196)x^2 + (4997/298116)x^3 + \cdots$
19. $c_1y_1(x) + c_2y_2(x) + c_3y_3(x)$, where
$y_1(x) = x^{4/3} + (1/17)x^{7/3} + (1/782)x^{10/3} + \cdots$,
$y_2(x) = 1 - (1/3)x - (1/30)x^2 + \cdots$, and
$y_3(x) = x^{-1/2} - (1/5)x^{1/2} - (1/10)x^{3/2} + \cdots$
21. **(b)** $c_1y_1(t) + c_2y_2(t)$, where
$y_1(t) = 1 - (1/6)(\alpha t)^2 + (1/120)(\alpha t)^4 + \cdots$, and
$y_2(t) = t^{-1}[1 - (1/2)(\alpha t)^2 + (1/24)(\alpha t)^4 + \cdots]$
(c) $c_1y_1(x) + c_2y_2(x)$, where
$y_1(x) = 1 - (1/6)(\alpha/x)^2 + (1/120)(\alpha/x)^4 + \cdots$, and
$y_2(x) = x[1 - (1/2)(\alpha/x)^2 + (1/24)(\alpha/x)^4 + \cdots]$
23. $c_1y_1(x) + c_2y_2(x)$, where
$y_1(x) = 1 + 2x + 2x^2$ and
$y_2(x) = -(1/6)x^{-1} - (1/24)x^{-2} - (1/120)x^{-3} + \cdots$
25. For $n = 0$,
$c_1 + c_2[\ln x + x + (1/4)x^2 + (1/18)x^3 + \cdots]$;
and for $n = 1$,
$c_1(1 - x) + c_2[(1 - x)\ln x + 3x - (1/4)x^2 - (1/36)x^3 + \cdots]$

Exercises 8.8, page 485

1. $c_1F\left(1, 2; \frac{1}{2}; x\right) + c_2x^{1/2}F\left(\frac{3}{2}, \frac{5}{2}; \frac{3}{2}; x\right)$
3. $c_1F\left(1, 1; \frac{1}{2}; x\right) + c_2x^{1/2}F\left(\frac{3}{2}, \frac{3}{2}; \frac{3}{2}; x\right)$

5. $F(1,1;2;x)=\sum_{n=0}^{\infty}\dfrac{1}{(n+1)}x^n=-x^{-1}\ln(1-x)$

7. $F\left(\frac{1}{2},1;\frac{3}{2};x^2\right)=\sum_{n=0}^{\infty}(1/2)_n x^{2n}/(3/2)_n$

$=\sum_{n=0}^{\infty}x^{2n}/(2n+1)$

$=\frac{1}{2}x^{-1}\ln\left(\frac{1+x}{1-x}\right)$

9. $(1-x)^{-1},\quad (1-x)^{-1}\ln x$

13. $c_1J_{1/2}(x)+c_2J_{-1/2}(x)$ **15.** $c_1J_1(x)+c_2Y_1(x)$

17. $c_1J_{2/3}(x)+c_2J_{-2/3}(x)$

19. $J_1(x)\ln x-x^{-1}+(3/64)x^3-(7/2304)x^5+\cdots$

21. $x^{3/2}J_{3/2}(x)$

27. $J_0(x)\ln x+\sum_{n=1}^{\infty}\dfrac{(-1)^{n+1}}{2^{2n}(n!)^2}\left[1+\frac{1}{2}+\cdots+\frac{1}{n}\right]x^{2n}$

29. $1,\quad x,\quad (3x^2-1)/2,$
$(5x^3-3x)/2,\quad (35x^4-30x^2+3)/8$

37. $1,\quad 2x,\quad 4x^2-2,\quad 8x^3-12x$

39. $1,\quad 1-x,\quad (2-4x+x^2)/2,$
$(6-18x+9x^2-x^3)/6$

41. (b) $z=c_1\cos x+c_2\sin x$, so for x large, $y(x)\approx c_1x^{-1/2}\cos x+c_2x^{-1/2}\sin x$

Review Problems, page 489

1. (a) $1-x+(3/2)x^2-(5/3)x^3+\cdots$
(b) $-1+x-(1/6)x^3+(1/6)x^4+\cdots$

3. (a) $a_0[1+x^2+\cdots]+a_1[x+(1/3)x^3+\cdots]$
(b) $a_0[1+(1/2)x^2-(1/6)x^3+\cdots]$
$+a_1[x-(1/2)x^2+(1/2)x^3+\cdots]$

5. $a_0[1+(1/2)(x-2)^2-(1/24)(x-2)^4+\cdots]$
$+a_1(x-2)$

7. (a) $a_0[x^3+x^4+(1/4)x^5+(1/36)x^6+\cdots]$
(b) $a_0[x-(1/4)x^2+(1/20)x^3-(1/120)x^4+\cdots]$

9. (a) $c_1y_1(x)+c_2y_2(x)$, where
$y_1(x)=x+x^2+(1/2)x^3+\cdots=xe^x$ and
$y_2(x)=y_1(x)\ln x-x^2-(3/4)x^3-(11/36)x^4+\cdots$
(b) $c_1y_1(x)+c_2y_2(x)$, where
$y_1(x)=1+2x+x^2+\cdots$ and
$y_2(x)=y_1(x)\ln x-[4x+3x^2+(22/27)x^3+\cdots]$
(c) $c_1y_1(x)+c_2y_2(x)$, where
$y_1(x)=1-(1/6)x+(1/96)x^2+\cdots$ and
$y_2(x)=y_1(x)\ln x-8x^{-2}-4x^{-1}+(29/36)+\cdots$
(d) $c_1y_1(x)+c_2y_2(x)$, where
$y_1(x)=x^2-(1/4)x^3+(1/40)x^4+\cdots$ and
$y_2(x)=x^{-1}+(1/2)+(1/4)x+\cdots$

CHAPTER 9

Exercises 9.1, page 500

1. $\begin{bmatrix}x\\y\end{bmatrix}'=\begin{bmatrix}7&2\\3&-2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$

3. $\begin{bmatrix}x\\y\\z\end{bmatrix}'=\begin{bmatrix}1&1&1\\-1&0&2\\0&4&0\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$

5. $\begin{bmatrix}x\\y\end{bmatrix}'=\begin{bmatrix}\sin t&e^t\\\cos t&a+bt^3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$

7. $\begin{bmatrix}y\\v\end{bmatrix}'=\begin{bmatrix}0&1\\-k/m&-b/m\end{bmatrix}\begin{bmatrix}y\\v\end{bmatrix}$

9. $\begin{bmatrix}x_1\\x_2\end{bmatrix}'=\begin{bmatrix}0&1\\t&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix};\quad x_1=y,\quad x_2=y'$

11. $\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}'=\begin{bmatrix}0&1&0&0\\-3&0&-2&0\\0&0&0&1\\2&0&0&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix};$
$x_1=x,\quad x_2=x',\quad x_3=y,\quad x_4=y'$

13. $\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}'=\begin{bmatrix}0&1&0&0&0\\\cos t&3&-t^2&0&0\\0&0&0&1&0\\0&0&0&0&1\\-e^t&t&0&-1&-1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix};$
$x_1=x,\quad x_2=x',\quad x_3=y,\quad x_4=y',\quad x_5=y''$

Exercises 9.2, page 504

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

3. $x_1=-s+t,\quad x_2=s,\quad x_3=t$

5. $x_1=0,\quad x_2=0$

7. $x_1=3s,\quad x_2=s\quad(-\infty<s<\infty)$

9. $x_1=2s,\quad x_2=(-1+i)s$, with s any complex number

11. $x_1=-(s+1)/2,\quad x_2=-(11s+1)/2,\quad x_3=s$
$(-\infty<s<\infty)$

13. For $r=2$, unique solution is $x_1=x_2=0$.
For $r=1$, solutions are $x_1=3s$, $x_2=s$, $-\infty<s<\infty$.

Exercises 9.3, page 513

1. (a) $\begin{bmatrix}1&1\\5&2\end{bmatrix}$ **(b)** $\begin{bmatrix}7&3\\7&18\end{bmatrix}$

3. (a) $\begin{bmatrix}18&14\\4&5\end{bmatrix}$ **(b)** $\begin{bmatrix}8&12\\3&5\end{bmatrix}$ **(c)** $\begin{bmatrix}16&3\\5&19\end{bmatrix}$

5. (a) $\begin{bmatrix}-1&-2\\-1&-3\end{bmatrix}$ **(b)** $\begin{bmatrix}-5&-1\\-8&-1\end{bmatrix}$ **(c)** $\begin{bmatrix}-6&-3\\-9&-4\end{bmatrix}$

7. (c) Yes **(d)** Yes **9.** $\begin{bmatrix}4/9&-1/9\\1/9&2/9\end{bmatrix}$ **11.** Doesn't exist

13. $\begin{bmatrix}1&0&1\\-2&1&-2\\1&1&0\end{bmatrix}$ **17.** $\begin{bmatrix}(4/3)e^{-t}&-(1/3)e^{-t}\\-(1/3)e^{-4t}&(1/3)e^{-4t}\end{bmatrix}$

19. $\begin{bmatrix}e^{-t}&(1/2)e^{-t}&-(1/2)e^{-t}\\(1/3)e^t&-(1/2)e^t&(1/6)e^t\\-(1/3)e^{-2t}&0&(1/3)e^{-2t}\end{bmatrix}$

21. 11 **23.** −12 **25.** 25 **27.** 2, 3

29. 0, 1, 1 **31.** $\begin{bmatrix}3e^{3t}\\6e^{3t}\\-3e^{3t}\end{bmatrix}$ **33.** $\begin{bmatrix}5e^{5t}&6e^{2t}\\-10e^{5t}&-2e^{2t}\end{bmatrix}$

39. (a) $\begin{bmatrix} (1/2)t^2 + c_1 & e^t + c_2 \\ t + c_3 & e^t + c_4 \end{bmatrix}$

(b) $\begin{bmatrix} \sin 1 & -1 + \cos 1 \\ 1 - \cos 1 & \sin 1 \end{bmatrix}$

(c) $\begin{bmatrix} (1+e^t)\cos t + (e^t - t)\sin t & (e^t - t)\cos t - (e^t + 1)\sin t \\ (e^t - 1)\sin t + e^t \cos t & (e^t - 1)\cos t - e^t \sin t \end{bmatrix}$

Exercises 9.4, page 521

1. $\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} + \begin{bmatrix} t^2 \\ e^t \end{bmatrix}$

3. $\begin{bmatrix} dx/dt \\ dy/dt \\ dz/dt \end{bmatrix} = \begin{bmatrix} t^2 & -1 & -1 \\ 0 & 0 & e^t \\ t & -1 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t \\ 5 \\ -e^t \end{bmatrix}$

5. $\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 10 & 3 \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \sin t \end{bmatrix}$;
$x_1 = y, \quad x_2 = y'$

7. $\begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t) \\ x_4'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ t^2 \end{bmatrix}$;
$x_1 = w, \quad x_2 = w', \quad x_3 = w'', \quad x_4 = w'''$

9. $x_1'(t) = 5x_1(t) + 2e^{-2t}$;
$x_2'(t) = -2x_1(t) + 4x_2(t) - 3e^{-2t}$

11. $x_1'(t) = x_1(t) + x_3(t) + e^t$;
$x_2'(t) = -x_1(t) + 2x_2(t) + 5x_3(t) + t$;
$x_3'(t) = 5x_2(t) + x_3(t)$

13. LI **15.** LD **17.** LI **19.** LI

21. Not a fundamental solution set

23. Yes; $\begin{bmatrix} e^{-t} & e^t & e^{3t} \\ 2e^{-t} & 0 & -e^{3t} \\ e^{-t} & e^t & 2e^{3t} \end{bmatrix}$;
$c_1\begin{bmatrix} e^{-t} \\ 2e^{-t} \\ e^{-t} \end{bmatrix} + c_2\begin{bmatrix} e^t \\ 0 \\ e^t \end{bmatrix} + c_3\begin{bmatrix} e^{3t} \\ -e^{3t} \\ 2e^{3t} \end{bmatrix}$

25. $\frac{3}{2}\begin{bmatrix} te^t \\ te^t \end{bmatrix} - \frac{1}{4}\begin{bmatrix} e^t \\ 3e^t \end{bmatrix} + \begin{bmatrix} t \\ 2t \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} + c_1\begin{bmatrix} e^t \\ e^t \end{bmatrix} + c_2\begin{bmatrix} e^{-t} \\ 3e^{-t} \end{bmatrix}$

29. $\mathbf{X}^{-1}(t) = \begin{bmatrix} 0 & (1/4)e^t & -(1/4)e^t \\ -(1/5)e^{2t} & (4/5)e^{2t} & -(2/5)e^{2t} \\ (1/5)e^{-3t} & (9/20)e^{-3t} & (3/20)e^{-3t} \end{bmatrix}$;
$\mathbf{x}(t) = \begin{bmatrix} -(3/2)e^{-t} + (3/5)e^{-2t} - (1/10)e^{3t} \\ (1/4)e^{-t} - (1/5)e^{-2t} - (1/20)e^{3t} \\ (5/4)e^{-t} - (1/5)e^{-2t} - (1/20)e^{3t} \end{bmatrix}$

37. (b) $2e^{6t}$ **(e)** $2e^{6t}$; same

Exercises 9.5, page 531

1. Eigenvalues are $r_1 = 0$ and $r_2 = -5$ with associated eigenvectors $\mathbf{u}_1 = \begin{bmatrix} s \\ 2s \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2s \\ -s \end{bmatrix}$.

3. Eigenvalues are $r_1 = 2$ and $r_2 = 3$ with associated eigenvectors $\mathbf{u}_1 = s\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = s\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

5. Eigenvalues are $r_1 = 1$, $r_2 = 2$, and $r_3 = -2$ with associated eigenvectors $\mathbf{u}_1 = s\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = s\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{u}_3 = s\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$.

7. Eigenvalues are $r_1 = 1$, $r_2 = 2$, and $r_3 = 5$ with associated eigenvectors $\mathbf{u}_1 = s\begin{bmatrix} 2 \\ -3 \\ 2 \end{bmatrix}$, $\mathbf{u}_2 = s\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, and $\mathbf{u}_3 = s\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$.

9. Eigenvalues are $r_1 = i$ and $r_2 = -i$, with associated eigenvectors $\mathbf{u}_1 = s\begin{bmatrix} i \\ 1 \end{bmatrix}$, and $\mathbf{u}_2 = s\begin{bmatrix} -i \\ 1 \end{bmatrix}$,

11. $c_1 e^{3t/2}\begin{bmatrix} 3 \\ 10 \end{bmatrix} + c_2 e^{t/2}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

13. $c_1 e^{-t}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} + c_2 e^{-3t}\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} + c_3 e^{-5t}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

15. $c_1 e^{-t}\begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} + c_2 e^{t}\begin{bmatrix} 1 \\ -6 \\ 4 \end{bmatrix} + c_3 e^{4t}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

17. (b)

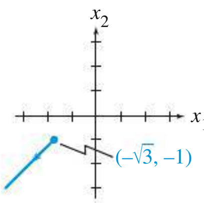


Figure B.43

(c)

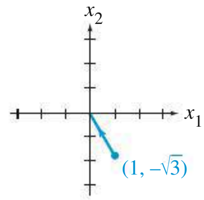


Figure B.44

(d)

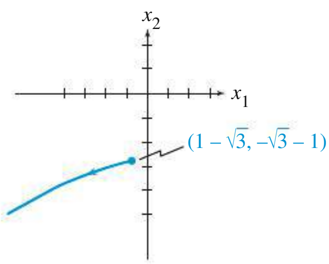


Figure B.45

19. $\begin{bmatrix} e^{3t} & e^{-3t} \\ 4e^{3t} & -2e^{-3t} \end{bmatrix}$

21. $\begin{bmatrix} e^t & e^{2t} & e^{4t} \\ e^t & 2e^{2t} & 4e^{4t} \\ e^t & 4e^{2t} & 16e^{4t} \end{bmatrix}$

23. $\begin{bmatrix} e^{2t} & e^{-t} & e^{3t} & -e^{7t} \\ 0 & -3e^{-t} & 0 & e^{7t} \\ 0 & 0 & e^{3t} & 2e^{7t} \\ 0 & 0 & 0 & 8e^{7t} \end{bmatrix}$

25. $x = -c_1e^t - 2c_2e^{2t} - c_3e^{3t}$,
$y = c_1e^t + c_2e^{2t} + c_3e^{3t}$,
$z = 2c_1e^t + 4c_2e^{2t} + 4c_3e^{3t}$

27. $\begin{bmatrix} e^{-0.3473t} & e^{0.5237t} & 0.0286e^{-7.0764t} \\ -0.3157e^{-0.3473t} & 0.4761e^{0.5237t} & -0.1837e^{-7.0764t} \\ 0.0844e^{-0.3473t} & 0.1918e^{0.5237t} & e^{-7.0764t} \end{bmatrix}$

29. $\begin{bmatrix} 0.0251e^{3.4142t} & -0.2361e^{-1.6180t} & e^{0.6180t} & e^{0.5858t} \\ 0.0858e^{3.4142t} & 0.3820e^{-1.6180t} & 0.6180e^{0.6180t} & 0.5858e^{0.5858t} \\ 0.2929e^{3.4142t} & -0.6180e^{-1.6180t} & 0.3820e^{0.6180t} & 0.3431e^{0.5858t} \\ e^{3.4142t} & e^{-1.6180t} & 0.2361e^{0.6180t} & 0.2010e^{0.5858t} \end{bmatrix}$

31. $\begin{bmatrix} 2e^{4t} + e^{-2t} \\ 2e^{4t} - e^{-2t} \end{bmatrix}$ **33.** $\begin{bmatrix} -3e^{-t} + e^{5t} \\ -2e^{-t} - e^{5t} \\ e^{-t} + e^{5t} \end{bmatrix}$

35. (b) $\mathbf{x}_1(t) = e^{-t}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(c) $\mathbf{x}_2(t) = te^{-t}\begin{bmatrix} 1 \\ 2 \end{bmatrix} + e^{-t}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

37. (b) $\mathbf{x}_1(t) = e^{2t}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

(c) $\mathbf{x}_2(t) = te^{2t}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + e^{2t}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

(d) $\mathbf{x}_3(t) = \dfrac{t^2}{2}e^{2t}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + te^{2t}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + e^{2t}\begin{bmatrix} 0 \\ -6/5 \\ 1/5 \end{bmatrix}$

39. (b) $\mathbf{x}_1(t) = e^t\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$; $\mathbf{x}_2(t) = e^t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

(c) $\mathbf{x}_3(t) = te^t\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + e^t\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ **(d) 0**

43. $c_1\begin{bmatrix} 3t^2 \\ t^2 \end{bmatrix} + c_2\begin{bmatrix} t^4 \\ t^4 \end{bmatrix}$

45. $x_1(t) = (2.5/2)(e^{-3t/25} + e^{-t/25})$,
$x_2(t) = (2.5)(e^{-t/25} - e^{-3t/25})$

47. (a) $r_1 = 2.39091$, $r_2 = -2.94338$, $r_3 = 3.55247$

(b) $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -2.64178 \\ -9.31625 \end{bmatrix}$; $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1.16825 \\ 0.43862 \end{bmatrix}$;

$\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0.81004 \\ -0.12236 \end{bmatrix}$

(c) $c_1e^{r_1t}\mathbf{u}_1 + c_2e^{r_2t}\mathbf{u}_2 + c_3e^{r_3t}\mathbf{u}_3$, where the r_i's and the $\mathbf{u}_i$'s are given in parts (a) and (b).

Exercises 9.6, page 537

1. $c_1\begin{bmatrix} 2\cos 2t \\ \cos 2t + \sin 2t \end{bmatrix} + c_2\begin{bmatrix} 2\sin 2t \\ \sin 2t - \cos 2t \end{bmatrix}$

3. $c_1e^t\cos t\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} - c_1e^t\sin t\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} + c_2e^t\sin t\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + c_2e^t\cos t\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} + c_3e^t\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

Note: Equivalent answer is obtained if col $(-1, 1, 0)$ and col $(-2, 0, 1)$ are replaced by col $(-5, 1, 2)$ and col $(0, -2, 1)$, respectively.

5. $\begin{bmatrix} e^{-t}\cos 4t & e^{-t}\sin 4t \\ 2e^{-t}\sin 4t & -2e^{-t}\cos 4t \end{bmatrix}$ **7.** $\begin{bmatrix} 1 & \cos t & \sin t \\ 0 & -\cos t & -\sin t \\ 0 & -\sin t & \cos t \end{bmatrix}$

9. $\begin{bmatrix} \cos t & \sin t & 0 & 0 \\ -\sin t & \cos t & 0 & 0 \\ 0 & 0 & \cos t & \sin t \\ 0 & 0 & -\sin t & \cos t \end{bmatrix}$ **11.** $\begin{bmatrix} e^t\sin t - e^t\cos t & -e^t\sin t - e^t\cos t & -\cos t & -\sin t \\ 2e^t\sin t & -2e^t\cos t & \sin t & -\cos t \\ 2e^t\sin t + 2e^t\cos t & 2e^t\sin t - 2e^t\cos t & \cos t & \sin t \\ 4e^t\cos t & 4e^t\sin t & -\sin t & \cos t \end{bmatrix}$

13. (a) $\begin{bmatrix} e^{-2t}(\sin t - \cos t) \\ -2e^{-2t}\sin t \end{bmatrix}$ **(b)** $\begin{bmatrix} -e^{-2(t-\pi)}\cos t \\ e^{-2(t-\pi)}(\cos t - \sin t) \end{bmatrix}$

(c) $\begin{bmatrix} e^{-2(t+2\pi)}(2\cos t - 3\sin t) \\ e^{-2(t+2\pi)}(\cos t + 5\sin t) \end{bmatrix}$ **(d)** $\begin{bmatrix} e^{\pi-2t}\cos t \\ e^{\pi-2t}(\sin t - \cos t) \end{bmatrix}$

17. $c_1\begin{bmatrix} t^{-1} \\ 0 \\ -2t^{-1} \end{bmatrix} + c_2\begin{bmatrix} t^{-1}\cos(\ln t) \\ t^{-1}\sin(\ln t) \\ -t^{-1}\cos(\ln t) \end{bmatrix} + c_3\begin{bmatrix} t^{-1}\sin(\ln t) \\ -t^{-1}\cos(\ln t) \\ -t^{-1}\sin(\ln t) \end{bmatrix}$

19. $\dfrac{\sqrt{9-\sqrt{17}}}{2\sqrt{2}\pi} \approx 0.249$; $\dfrac{\sqrt{9+\sqrt{17}}}{2\sqrt{2}\pi} \approx 0.408$

21. $I_1 = (5/6)e^{-2t}\sin 3t$,
$I_2 = -(10/13)e^{-2t}\cos 3t + (25/78)e^{-2t}\sin 3t$,
$I_3 = (10/13)e^{-2t}\cos 3t + (20/39)e^{-2t}\sin 3t$

Exercises 9.7, page 542

1. $c_1 e^{7t}\begin{bmatrix}1\\1\end{bmatrix} + c_2 e^{2t}\begin{bmatrix}1\\-4\end{bmatrix} + \begin{bmatrix}2\\-1\end{bmatrix}$

3. $c_1 e^{-3t}\begin{bmatrix}1\\1\\-1\end{bmatrix} + c_2 e^{3t}\begin{bmatrix}1\\0\\1\end{bmatrix} + c_3 e^{3t}\begin{bmatrix}-1\\1\\0\end{bmatrix} + e^{t}\begin{bmatrix}1\\0\\-1\end{bmatrix}$

5. $t\begin{bmatrix}0\\0\\-1\end{bmatrix} + \begin{bmatrix}0\\0\\-1\end{bmatrix} + \dfrac{e^{2t}}{3}\begin{bmatrix}2\\-1\\0\end{bmatrix} + \dfrac{\sin t}{2}\begin{bmatrix}1\\0\\0\end{bmatrix} + \dfrac{\cos t}{2}\begin{bmatrix}0\\-1\\0\end{bmatrix}$

$+ c_1 e^{-t}\begin{bmatrix}1\\1\\0\end{bmatrix} + c_2 e^{t}\begin{bmatrix}-1\\1\\0\end{bmatrix} + c_3 e^{t}\begin{bmatrix}0\\0\\1\end{bmatrix}$

7. $\mathbf{x}_p = t\mathbf{a} + \mathbf{b} + \sin 3t\,\mathbf{c} + \cos 3t\,\mathbf{d}$

9. $\mathbf{x}(t) = t\mathbf{a} + \mathbf{b} + e^{2t}\mathbf{c} + (\sin t)\,\mathbf{d} + (\cos t)\,\mathbf{e}$

11. $c_1\begin{bmatrix}\sin t\\ \cos t\end{bmatrix} + c_2\begin{bmatrix}\cos t\\ -\sin t\end{bmatrix} + \begin{bmatrix}0\\-1\end{bmatrix}$

13. $c_1\begin{bmatrix}1\\2\end{bmatrix} + c_2 e^{6t}\begin{bmatrix}2\\1\end{bmatrix} - \begin{bmatrix}t^{-1}/2\\ t^{-1}\end{bmatrix}$

15. $c_1\begin{bmatrix}1\\2\end{bmatrix} + c_2 e^{-5t}\begin{bmatrix}-2\\1\end{bmatrix} + \begin{bmatrix}\ln|t| + (8/5)t - 8/25\\ 2\ln|t| + (16/5)t + 4/25\end{bmatrix}$

17. $c_1 e^{2t}\begin{bmatrix}1\\1\\1\end{bmatrix} + c_2 e^{-t}\begin{bmatrix}-1\\0\\1\end{bmatrix} + c_3 e^{-t}\begin{bmatrix}-1\\1\\0\end{bmatrix} + e^{t}\begin{bmatrix}1\\-1\\-1\end{bmatrix}$

19. $\begin{bmatrix}c_1\cos t + c_2\sin t + 1\\ -c_1\sin t + c_2\cos t - t\\ c_3e^{t} + c_4e^{-t} - (1/4)e^{-t} - t + (1/2)te^{-t}\\ c_3e^{t} - c_4e^{-t} - (1/4)e^{-t} - 1 - (1/2)te^{-t}\end{bmatrix}$

21. (a) $\begin{bmatrix}4te^{t} + 5e^{t}\\ 2te^{t} + 4e^{t}\end{bmatrix}$

(b) $\begin{bmatrix}-2e^{t-1} + 2e^{2(t-1)} - 3e^{2t-1} + (4t-1)e^{t}\\ -e^{t-1} + 2e^{2(t-1)} - 3e^{2t-1} + (2t+1)e^{t}\end{bmatrix}$

(c) $\begin{bmatrix}(-20 + 2e^{-5})e^{t} + (-3e^{-5} - e^{-10})e^{2t} + (4t+3)e^{t}\\ (-10 + e^{-5})e^{t} + (-3e^{-5} - e^{-10})e^{2t} + (2t+3)e^{t}\end{bmatrix}$

(d) $\begin{bmatrix}-18e^{t+1} + 14e^{2t+2} - 3e^{2t+1} + (4t+7)e^{t}\\ -9e^{t+1} + 14e^{2t+2} - 3e^{2t+1} + (2t+5)e^{t}\end{bmatrix}$

23. $x = (c_1 + c_2 t)e^{t} + c_3 e^{-t} - t^2 - 4t - 6\,;$
$y = -c_2 e^{t} - 2c_3 e^{-t} - t^2 - 2t - 3$

25. (a) $\left\{\begin{bmatrix}e^{t}\\ e^{t}\end{bmatrix}, \begin{bmatrix}e^{2t}\\ 2e^{2t}\end{bmatrix}\right\}$ **(c)** $te^{t}\begin{bmatrix}2\\2\end{bmatrix} + e^{t}\begin{bmatrix}0\\1\end{bmatrix}$

(d) $c_1 e^{t}\begin{bmatrix}1\\1\end{bmatrix} + c_2 e^{2t}\begin{bmatrix}1\\2\end{bmatrix} + te^{t}\begin{bmatrix}2\\2\end{bmatrix} + e^{t}\begin{bmatrix}0\\1\end{bmatrix}$

27. $c_1 e^{2t}\begin{bmatrix}1\\1\\1\end{bmatrix} + c_2 e^{-t}\begin{bmatrix}-1\\0\\1\end{bmatrix} + c_3 e^{-t}\begin{bmatrix}-1\\1\\0\end{bmatrix}$

$+ te^{-t}\begin{bmatrix}1\\0\\-1\end{bmatrix} + e^{-t}\begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\1\end{bmatrix}$

29. $c_1 t\begin{bmatrix}1\\1\end{bmatrix} + c_2 t^{-1}\begin{bmatrix}1\\3\end{bmatrix}$

$+ \begin{bmatrix}-(3/4)t^{-1} - (1/2)t^{-1}\ln t + 1\\ -(3/4)t^{-1} - (3/2)t^{-1}\ln t + 2\end{bmatrix}$

33. $I_1 = I_2 + 3I_5$, $I_3 = 3I_5$, $I_4 = 2I_5$, where
$I_2 = \left(1/\left(20\sqrt{817}\right)\right)\left[\left(13 - \sqrt{817}\right)e^{-(31+\sqrt{817})5t/2}\right.$
$\left. - \left(13 + \sqrt{817}\right)e^{(-31+\sqrt{817})5t/2}\right] + 1/10\,,$
$I_5 = \left(3/\left(40\sqrt{817}\right)\right)\left[\left(31 - \sqrt{817}\right)e^{-(31+\sqrt{817})5t/2}\right.$
$\left. - \left(31 + \sqrt{817}\right)e^{(-31+\sqrt{817})5t/2}\right] + 3/20$

Exercises 9.8, page 551

1. (a) $r = 3\,;\quad k = 2$ **(b)** $e^{3t}\begin{bmatrix}1 & -2t\\ 0 & 1\end{bmatrix}$

3. (a) $r = -1\,;\quad k = 3$ **(b)** $e^{-t}\begin{bmatrix}1 + 3t - (3/2)t^2 & t & -t + (1/2)t^2\\ -3t & 1 & t\\ 9t - (9/2)t^2 & 3t & 1 - 3t + (3/2)t^2\end{bmatrix}$

5. (a) $r = -2\,;\quad k = 2$ **(b)** $e^{-2t}\begin{bmatrix}1 & 0 & 0\\ 4t & 1 & 0\\ t & 0 & 1\end{bmatrix}$ **7.** $\begin{bmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{bmatrix}$

9. $\dfrac{1}{2}\begin{bmatrix}e^{t} + \cos t - \sin t & 2\sin t & e^{t} - \cos t - \sin t\\ e^{t} - \sin t - \cos t & 2\cos t & e^{t} + \sin t - \cos t\\ e^{t} - \cos t + \sin t & -2\sin t & e^{t} + \cos t + \sin t\end{bmatrix}$

11. $\dfrac{1}{25}\begin{bmatrix}-4 + 29e^{5t} - 20te^{5t} & 20 - 20e^{5t} & -8 + 8e^{5t} - 40te^{5t}\\ -5 + 5e^{5t} & 25 & -10 + 10e^{5t}\\ 2 - 2e^{5t} + 10te^{5t} & -10 + 10e^{5t} & 4 + 21e^{5t} + 20te^{5t}\end{bmatrix}$

13. $\begin{bmatrix}(1 - t + t^2/2)e^{t} & (t - t^2)e^{t} & (t^2/2)e^{t} & 0 & 0\\ (t^2/2)e^{t} & (1 - t + t^2)e^{t} & (t + t^2/2)e^{t} & 0 & 0\\ (t + t^2/2)e^{t} & (-3t - t^2)e^{t} & (1 + 2t + t^2/2)e^{t} & 0 & 0\\ 0 & 0 & 0 & \cos t & \sin t\\ 0 & 0 & 0 & -\sin t & \cos t\end{bmatrix}$

15. $\begin{bmatrix} (1+t+t^2/2)e^{-t} & (t+t^2)e^{-t} & (t^2/2)e^{-t} & 0 & 0 \\ (-t^2/2)e^{-t} & (1+t-t^2)e^{-t} & (t-t^2/2)e^{-t} & 0 & 0 \\ (-t+t^2/2)e^{-t} & (-3t+t^2)e^{-t} & (1-2t+t^2/2)e^{-t} & 0 & 0 \\ 0 & 0 & 0 & (1+2t)e^{-2t} & te^{-2t} \\ 0 & 0 & 0 & -4te^{-2t} & (1-2t)e^{-2t} \end{bmatrix}$

17. $c_1e^{-t}\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + c_2e^{-t}\begin{bmatrix} -t \\ -1+t \\ 2-t \end{bmatrix} + c_3e^{-2t}\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}$ **19.** $c_1e^{2t}\begin{bmatrix} 3 \\ 6 \\ 1 \\ 1 \end{bmatrix} + c_2e^{t}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_3e^{t}\begin{bmatrix} 1 \\ t \\ 0 \\ 0 \end{bmatrix} + c_4e^{t}\begin{bmatrix} 2t \\ t+t^2 \\ 0 \\ 1 \end{bmatrix}$

21. $\begin{bmatrix} e^{-2t} \\ 4te^{-2t}+e^{-2t} \\ te^{-2t}-e^{-2t} \end{bmatrix}$. **23.** $e^{-t}\begin{bmatrix} 3t \\ 3 \\ 9t \end{bmatrix} + e^{\mathbf{A}t}\begin{bmatrix} 2-e^t(t^2-2t+2) \\ 1+e^t(t-1) \\ 6-3e^t(t^2-2t+2) \end{bmatrix}$, where $e^{\mathbf{A}t}$ is the matrix in the answer to Problem 3.

29. $\mathbf{x}(t) = \begin{bmatrix} (4/13)e^{-3t}+6te^{3t}+(2/13)(3\sin 2t-2\cos 2t) \\ e^{3t} \\ (8/13)e^{-3t}+(3t-1)e^{3t}+(4/13)(3\sin 2t-2\cos 2t) \end{bmatrix}$.

Review Problems, page 555

1. $c_1e^{3t}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2e^{4t}\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ **3.** $c_1e^{-t}\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + c_2e^{-t}\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} + c_3e^{3t}\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_4e^{3t}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ **5.** $\begin{bmatrix} e^{2t} & e^{3t} \\ -e^{2t} & -2e^{3t} \end{bmatrix}$

7. $c_1e^{-t}\begin{bmatrix} 1 \\ -2 \end{bmatrix} + c_2e^{3t}\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -1/3 \\ -14/3 \end{bmatrix}$ **9.** $c_1e^{-t}\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + c_2e^{-t}\begin{bmatrix} t \\ 1 \\ 3t \end{bmatrix} + c_3e^{-t}\begin{bmatrix} -t+(1/2)t^2 \\ t \\ 1-3t+(3/2)t^2 \end{bmatrix} + \begin{bmatrix} 2+t \\ 7-3t \\ 10 \end{bmatrix}$

11. $\begin{bmatrix} 3e^t-2e^{2t} \\ 3e^t-4e^{2t} \end{bmatrix}$ **13.** $c_1t^{-1}\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} + c_2t^{-1}\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + c_3t^{4}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

15. $\begin{bmatrix} (1/2)e^t+(2/3)e^{5t}-(1/6)e^{-t} & -e^t+(2/3)e^{5t}+(1/3)e^{-t} & -(1/2)e^t+(2/3)e^{5t}-(1/6)e^{-t} \\ (1/3)e^{5t}-(1/3)e^{-t} & (1/3)e^{5t}+(2/3)e^{-t} & (1/3)e^{5t}-(1/3)e^{-t} \\ -(1/2)e^t+(1/2)e^{-t} & e^t-e^{-t} & (1/2)e^t+(1/2)e^{-t} \end{bmatrix}$

17. **(a)** false **(b)** false **(c)** false **(d)** true **(e)** true **(f)** false **(g)** true

---

CHAPTER 10

Exercises 10.2, page 570

1. $y = [4/(e-e^{-1})](e^{-x}-e^x)$ **3.** $y \equiv 0$

5. $y = e^x + 2x - 1$ **7.** $y = \cos x + c\sin x$; c arbitrary

9. $\lambda_n = (2n-1)^2/4$ and $y_n = c_n \sin[(2n-1)x/2]$, where $n = 1, 2, 3, \ldots$ and the c_n's are arbitrary.

11. $\lambda_n = n^2, n = 0, 1, 2, \ldots; y_0 = a_0$ and $y_n = a_n \cos nx + b_n \sin nx$, $n = 1, 2, 3, \ldots$, where a_0, a_n, and b_n are arbitrary.

13. The eigenvalues are the roots of $\tan(\sqrt{\lambda_n}\,\pi) + \sqrt{\lambda_n} = 0$, where $\lambda_n > 0$. For n large, $\lambda_n \approx (2n-1)^2/4$, n is a positive integer. The eigenfunctions are $y_n = c_n[\sin(\sqrt{\lambda_n}x) + \sqrt{\lambda_n}\cos(\sqrt{\lambda_n}x)]$, where the c_n's are arbitrary.

15. $u(x,t) = e^{-3t}\sin x - 6e^{-48t}\sin 4x$

17. $u(x,t) = e^{-3t}\sin x - 7e^{-27t}\sin 3x + e^{-75t}\sin 5x$

19. $u(x,t) = 3\cos 6t\sin 2x + 12\cos 39t\sin 13x$

21. $u(x,t) = 6\cos 6t\sin 2x + 2\cos 18t\sin 6x + (11/27)\sin 27t\sin 9x - (14/45)\sin 45t\sin 15x$

23. $u(x,t) = \sum_{n=1}^{\infty} n^{-2}e^{-2\pi^2 n^2 t}\sin n\pi x$

25. If $K > 0$, then $T(t)$ becomes unbounded as $t \to \infty$, and so the temperature $u(x,t) = X(x)T(t)$ becomes unbounded at each position x. Since the temperature must remain bounded for all time, $K \not> 0$.

33. **(a)** $u(x) \equiv 50$ **(b)** $u(x) = 30x/L + 10$

Exercises 10.3, page 584

1. Odd **3.** Even **5.** Neither

9. $f(x) \sim \sum_{n=1}^{\infty} \frac{2(-1)^{n+1}}{n} \sin nx$

11. $f(x) \sim 1 + \sum_{n=1}^{\infty} \left[\frac{2}{\pi^2 n^2}\left(-1 + (-1)^n\right) \cos \frac{n\pi x}{2} + \frac{1}{\pi n}\left((-1)^{n+1} - 1\right) \sin \frac{n\pi x}{2} \right]$

13. $f(x) \sim \frac{1}{3} + \sum_{n=1}^{\infty} \frac{4(-1)^n}{n^2 \pi^2} \cos n\pi x$

15. $f(x) \sim [(\sinh \pi)/\pi]\left(1 + 2\sum_{n=1}^{\infty}\left[\frac{(-1)^n}{1+n^2}\cos nx + \frac{(-1)^{n+1} n}{1+n^2} \sin nx\right]\right)$

17. The 2π-periodic function $g(x)$, where

$$g(x) = \begin{cases} x & (-\pi < x < \pi), \\ 0 & (x = \pm\pi) \end{cases}$$

19. The 4-periodic function $g(x)$, where

$$g(x) = \begin{cases} 1 & (-2 < x < 0), \\ x & (0 < x < 2), \\ 1/2 & (x = 0), \\ 3/2 & (x = \pm 2) \end{cases}$$

21. The 2-periodic function $g(x)$, where
$g(x) = x^2 \quad (-1 \le x \le 1)$

23. The 2π-periodic function $g(x)$, where

$$g(x) = \begin{cases} e^x & (-\pi < x < \pi), \\ (e^{\pi} + e^{-\pi})/2 & (x = \pm\pi) \end{cases}$$

25. **(a)** $F(x) = (x^2 - \pi^2)/2$ **(b)** $F(x) = |x| - \pi$

27. $f(x) \sim \sum_{n=1}^{\infty} \frac{2}{(2n-1)\pi}\left[(-1)^{n+1}\cos\left(\frac{(2n-1)\pi x}{2}\right) + \sin\left(\frac{(2n-1)\pi x}{2}\right)\right]$

29. $c_0 = 0; \quad c_1 = 3/2; \quad c_2 = 0$

Exercises 10.4, page 591

1. **(a)** The π-periodic function $\tilde{f}(x)$, where
$\tilde{f}(x) = x^2 \quad (0 < x < \pi)$

(b) The 2π-periodic function $f_o(x)$, where

$$f_o(x) = \begin{cases} x^2 & (0 < x < \pi), \\ -x^2 & (-\pi < x < 0) \end{cases}$$

(c) The 2π-periodic function $f_e(x)$, where

$$f_e(x) = \begin{cases} x^2 & (0 < x < \pi), \\ x^2 & (-\pi < x < 0) \end{cases}$$

3. **(a)** The π-periodic function $\tilde{f}(x)$, where

$$\tilde{f}(x) = \begin{cases} 0 & (0 < x < \pi/2), \\ 1 & (\pi/2 < x < \pi) \end{cases}$$

(b) The 2π-periodic function $f_o(x)$, where

$$f_o(x) = \begin{cases} -1 & (-\pi < x < -\pi/2), \\ 0 & (-\pi/2 < x < 0), \\ 0 & (0 < x < \pi/2), \\ 1 & (\pi/2 < x < \pi) \end{cases}$$

(c) The 2π-periodic function $f_e(x)$, where

$$f_e(x) = \begin{cases} 1 & (-\pi < x < -\pi/2), \\ 0 & (-\pi/2 < x < 0), \\ 0 & (0 < x < \pi/2), \\ 1 & (\pi/2 < x < \pi) \end{cases}$$

5. $f(x) \sim -\frac{4}{\pi}\sum_{k=1}^{\infty} \frac{1}{2k-1} \sin(2k-1)\pi x$

7. $f(x) \sim \sum_{n=1}^{\infty}\left[\frac{2\pi(-1)^{n+1}}{n} + \frac{4}{\pi n^3}\left((-1)^n - 1\right)\right] \sin nx$

9. $f(x) \sim \sum_{k=0}^{\infty} \frac{8}{(2k+1)^3 \pi^3} \sin(2k+1)\pi x$

11. $f(x) \sim \frac{\pi}{2} + \frac{4}{\pi}\sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} \cos(2k-1)x$

13. $f(x) \sim e - 1 + 2\sum_{n=1}^{\infty} \frac{(-1)^n e - 1}{1 + \pi^2 n^2} \cos n\pi x$

15. $f(x) \sim \frac{2}{\pi} + \frac{2}{\pi}\sum_{k=1}^{\infty}\left(\frac{1}{2k+1} - \frac{1}{2k-1}\right)\cos 2kx$

17. $u(x,t) =$
$\frac{2}{\pi}\sum_{k=1}^{\infty}\left[\frac{2}{2k-1} - \frac{1}{2k+1} - \frac{1}{2k-3}\right] e^{-5(2k-1)^2 t} \sin(2k-1)x$

19. $u(x,t) = \frac{4}{\pi}\sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{(2k+1)^2} e^{-5(2k+1)^2 t} \sin(2k+1)x$

Exercises 10.5, page 603

1. $u(x,t) = \sum_{n=1}^{\infty} \frac{8(-1)^{n+1} - 4}{\pi^3 n^3} e^{-5\pi^2 n^2 t} \sin n\pi x$

3. $u(x,t) = \frac{\pi}{2} - \sum_{k=0}^{\infty} \frac{4}{\pi(2k+1)^2} e^{-3(2k+1)^2 t} \cos(2k+1)x$

5. $u(x,t) = \frac{2(e^{\pi} - 1)}{\pi} + \sum_{n=1}^{\infty} \frac{2e^{\pi}(-1)^n - 2}{\pi(1+n^2)} e^{-n^2 t} \cos nx$

7. $u(x,t) = 5 + \frac{5}{\pi}x - \frac{30}{\pi}e^{-2t}\sin x + \frac{5}{\pi}e^{-8t}\sin 2x + \left(1 - \frac{10}{\pi}\right)e^{-18t}\sin 3x + \frac{5}{2\pi}e^{-32t}\sin 4x - \left(1 + \frac{6}{\pi}\right)e^{-50t}\sin 5x + \sum_{n=6}^{\infty} \frac{10}{\pi n}[2(-1)^n - 1]e^{-2n^2 t}\sin nx$

9. $u(x,t) = \dfrac{e^{-\pi}-1}{\pi}x - e^{-x} + 1 + \sum_{n=1}^{\infty} c_n e^{-n^2 t}\sin nx\,,$

where

$$c_n = \begin{cases} \dfrac{2e^{-\pi}-2}{\pi n}(-1)^n + \dfrac{2n}{\pi(1+n^2)}[(-1)^{n+1}e^{-\pi}+1] \\ \quad + \dfrac{2}{\pi n}[(-1)^n - 1] \quad (n \neq 2)\,, \\ \dfrac{e^{-\pi}-1}{\pi} + \dfrac{4}{5\pi}(1-e^{-\pi}) + 1 \qquad (n = 2) \end{cases}$$

11. $u(x,t) = \sum_{n=0}^{\infty} a_n e^{-4(n+1/2)^2 t}\cos(n+1/2)x$, where

$$f(x) = \sum_{n=0}^{\infty} a_n \cos(n+1/2)x$$

13. $u(x,t) = \dfrac{\pi^2}{3}x - \dfrac{1}{3}x^3 - 3e^{-2t}\sin x$

$$+ \sum_{n=2}^{\infty} \frac{4(-1)^n}{n^3} e^{-2n^2 t}\sin nx$$

15. $u(x,y,t) = e^{-52t}\cos 6x \sin 4y - 3e^{-122t}\cos x \sin 11y$

17. $u(x,y,t) = 2\sum_{n=1}^{\infty} \dfrac{(-1)^{n+1}}{n} e^{-n^2 t}\sin ny$

19. $C(x,t) = \sum_{n=1}^{\infty} c_n e^{-[L+kn^2\pi^2/a^2]t}\sin\left(\dfrac{n\pi x}{a}\right)$, where

$$c_n = \frac{2}{a}\int_0^a f(x)\sin\left(\frac{n\pi x}{a}\right)dx$$

Concentration goes to zero as $t \to +\infty$.

Exercises 10.6, page 614

1. $u(x,t) = \dfrac{1}{7\pi}\sin 7\pi t \sin 7\pi x$

$$+ \sum_{k=0}^{\infty} \frac{8}{((2k+1)\pi)^3}\cos(2k+1)\pi t \sin(2k+1)\pi x$$

3. $u(x,t) = \sum_{n=1}^{\infty} \dfrac{4}{n^3}[2(-1)^{n+1} - 1]\cos 2nt \sin nx$

5. $u(x,t) = \dfrac{2h_0 L^2}{\pi^2 a(L-a)} \sum_{n=1}^{\infty} \dfrac{1}{n^2}\sin\dfrac{n\pi a}{L}\sin\dfrac{n\pi x}{L}\cos\dfrac{n\pi\alpha t}{L}$

7. $u(x,t) = \cos t \sin x + \dfrac{5}{2}\sin 2t \sin 2x - \dfrac{3}{5}\sin 5t \sin 5x$

$$+ 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^3}\left[t - \frac{\sin nt}{n}\right]\sin nx$$

9. $u(x,t) = \sum_{n=0}^{\infty}\left[a_n \cos\dfrac{(2n+1)\pi\alpha t}{2L}\right.$

$$\left. + b_n \sin\frac{(2n+1)\pi\alpha t}{2L}\right]\sin\frac{(2n+1)\pi x}{2L}, \text{ where}$$

$$f(x) = \sum_{n=0}^{\infty} a_n \sin\frac{(2n+1)\pi x}{2L} \quad \text{and}$$

$$g(x) = \sum_{n=0}^{\infty} b_n \frac{(2n+1)\pi\alpha}{2L}\sin\frac{(2n+1)\pi x}{2L}$$

11. $u(x,t) = \sum_{n=1}^{\infty} a_n T_n(t)\sin\dfrac{n\pi x}{L}$, where

$$a_n = \frac{2}{L}\int_0^L f(x)\sin\frac{n\pi x}{L}\,dx\,, \text{ and}$$

$$T_n(t) = e^{-t/2}\left(\cos\beta_n t + \frac{1}{2\beta_n}\sin\beta_n t\right), \text{ where}$$

$$\beta_n = \frac{1}{2L}\sqrt{3L^2 + 4\alpha^2\pi^2 n^2}$$

13. $u(x,t) = \dfrac{1}{2\alpha}[\sin(x+\alpha t) - \sin(x-\alpha t)]$

$$= \frac{1}{\alpha}\sin\alpha t \cos x$$

15. $u(x,t) = x + tx$

17. $u(x,t) = \dfrac{1}{2}\left[e^{-(x+\alpha t)^2} + e^{-(x-\alpha t)^2}\right.$

$$\left. + \frac{\cos(x-\alpha t) - \cos(x+\alpha t)}{\alpha}\right]$$

21. $u(r,t) = \sum_{n=1}^{\infty}[a_n\cos(k_n\alpha t) + b_n\sin(k_n\alpha t)]J_0(k_n r)\,,$

where

$$a_n = \frac{1}{c_n}\int_0^1 f(r)J_0(k_n r)r\,dr\,, \text{ and}$$

$$b_n = \frac{1}{\alpha k_n c_n}\int_0^1 g(r)J_0(k_n r)r\,dr\,, \text{ with}$$

$$c_n = \int_0^1 J_0^2(k_n r)r\,dr$$

Exercises 10.7, page 626

1. $u(x,y) = \dfrac{4\cos 6x \sinh[6(y-1)]}{\sinh(-6)}$

$$+ \frac{\cos 7x \sinh[7(y-1)]}{\sinh(-7)}$$

3. $u(x,y) = \sum_{n=1}^{\infty} A_n \sin nx \sinh(ny - n\pi)$, where

$$A_n = \frac{2}{\pi\sinh(-n\pi)}\int_0^{\pi} f(x)\sin nx\,dx$$

5. $u(x,y) = \dfrac{\cos x \sinh(y-1)}{\sinh(-1)} - \dfrac{\cos 3x \sinh(3y-3)}{\sinh(-3)}$

$$+ \frac{\cos 2x \sinh 2y}{\sinh(2)}$$

7. $u(r,\theta) = \dfrac{\pi}{2} - \sum_{k=0}^{\infty}\dfrac{r^{2k+1}}{(2k+1)^2\pi 2^{2k-1}}\cos(2k+1)\theta$

9. $u(r,\theta) = \dfrac{a_0}{2} + \sum_{n=1}^{\infty}\left(\dfrac{r}{a}\right)^n(a_n\cos n\theta + b_n\sin n\theta)$, where a_0 is arbitrary, and for $n = 1, 2, 3, \ldots$

$$a_n = \frac{a}{\pi n}\int_{-\pi}^{\pi} f(\theta)\cos n\theta\,d\theta\,,$$

$$b_n = \frac{a}{n\pi}\int_{-\pi}^{\pi} f(\theta)\sin n\theta\,d\theta$$

11. $u(r, \theta) = \left(\frac{1}{3}r - \frac{4}{3}r^{-1}\right)\cos\theta + \left(\frac{2}{3}r - \frac{2}{3}r^{-1}\right)\sin\theta$
$+ \left(-\frac{1}{255}r^4 + \frac{256}{255}r^{-4}\right)\sin 4\theta$

13. $u(r, \theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} r^{-n}(a_n \cos n\theta + b_n \sin n\theta)$,

where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)\cos n\theta\, d\theta \qquad (n = 0, 1, 2, \ldots),$$

and

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)\sin n\theta\, d\theta \qquad (n = 1, 2, 3, \ldots)$$

15. $u(r, \theta) = r^3 \sin 3\theta$

17. $u(r, \theta) = a_0 + b_0 \ln r + \sum_{n=1}^{\infty}\left[(a_n r^n + b_n r^{-n})\cos n\theta + (c_n r^n + d_n r^{-n})\sin n\theta\right]$,

where

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\, d\theta, \qquad b_0 = \frac{3}{2\pi}\int_{-\pi}^{\pi} g(\theta)\, d\theta,$$

and for $n = 1, 2, 3, \ldots$

$$a_n + b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)\cos n\theta\, d\theta,$$

$$n3^{n-1}a_n - n3^{-n-1}b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(\theta)\cos n\theta\, d\theta,$$

and

$$c_n + d_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)\sin n\theta\, d\theta,$$

$$n3^{n-1}c_n - n3^{-n-1}d_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(\theta)\sin n\theta\, d\theta$$

21. $u(r, \theta, z) = [I_0(r)/I_0(\pi)]\sin z$

23. **(c)** $\phi(x, y) = 2xy$

APPENDIX A

Exercises A, page A-1

1. $(x^2+1)^{5/2}(5x^2-2)/35 + C$ **3.** $\frac{2}{15}e^{6r^5} + C$

5. $\frac{1}{2}\theta + \frac{1}{4\pi}\sin(2\pi\theta) + C$

7. $\frac{1}{2}\ln(t^2+4) + \frac{1}{2}\arctan(t/2) + C$

9. $(1/\sqrt{2})\ \arcsin\sqrt{2}x + C$

11. $2\ln|x-1| - 4\ln|x-2| + 2\ln|x-3| + C$

13. $\sqrt{9x^2-1} - \arccos\left(\frac{1}{3x}\right) + C$

15. $\frac{1}{2}\ln(x^2+4) - \frac{2}{x} + C$ **17.** $\ln|\ln x| + C$

19. $y\cosh y - \sinh y + C$ **21.** $\frac{x}{4\sqrt{x^2+4}} + C$

23. $\frac{1}{2}(t^2-9)\ln(t+3) - \frac{t^2}{4} + \frac{3t}{2} + C$

25. $-\arctan(\cos t) + C$

27. $\frac{5}{7}\sin^{7/5}x - \frac{5}{17}\sin^{17/5}x + C$

29. $\frac{\sin(10x)}{20} + \frac{\sin(4x)}{8} + C$

31. $\frac{1}{9}\tan^9\theta + \frac{2}{7}\tan^7\theta + \frac{1}{5}\tan^5\theta + C$

33. $\frac{1}{8}x - \frac{1}{96}\sin(12x) + C$

Index

A BRIEF TABLE OF INTEGRALS*

$\int (f(u) + g(u))\,du = \int f(u)\,du + \int g(u)\,du.$ $\int cf(u)\,du = c\int f(u)\,du.$

$\int u\,dv = uv - \int v\,du.$ $\int u^n du = \frac{u^{n+1}}{n+1}, \quad n \neq -1.$ $\int \frac{du}{u} = \ln|u|.$

$\int e^u du = e^u.$ $\int ue^u du = (u-1)e^u.$ $\int u^n e^u du = u^n e^u - n\int u^{n-1}e^u du.$

$\int a^u du = \frac{a^u}{\ln a}, \quad a > 0, a \neq 1.$ $\int \ln u\,du = u\ln u - u.$ $\int \frac{du}{u\ln u} = \ln|\ln u|.$

$\int u^n \ln u\,du = u^{n+1}\left(\frac{\ln u}{n+1} - \frac{1}{(n+1)^2}\right), \quad n \neq -1.$ $\int \frac{du}{a^2+u^2} = \frac{1}{a}\arctan\frac{u}{a}.$

$\int \frac{du}{u^2-a^2} = \frac{1}{2a}\ln\left|\frac{u-a}{u+a}\right|.$ $\int \frac{du}{a^2-u^2} = \frac{1}{2a}\ln\left|\frac{a+u}{a-u}\right|.$

$\int \frac{du}{(a+bu)(\alpha+\beta u)} = \frac{1}{a\beta-\alpha b}\ln\left|\frac{\alpha+\beta u}{a+bu}\right|$

$\int \frac{u\,du}{(a+bu)(\alpha+\beta u)} = \frac{1}{a\beta-\alpha b}\left[\frac{a}{b}\ln|a+bu| - \frac{\alpha}{\beta}\ln|\alpha+\beta u|\right].$

$\int \frac{du}{\sqrt{u^2+a^2}} = \ln\left|u+\sqrt{u^2+a^2}\right|.$ $\int \frac{du}{\sqrt{u^2-a^2}} = \ln\left|u+\sqrt{u^2-a^2}\right|, \quad u^2 \geq a^2.$

$\int \frac{du}{\sqrt{a^2-u^2}} = \arcsin\frac{u}{a}, \quad a^2 \geq u^2.$ $\int \frac{du}{u\sqrt{u^2-a^2}} = \frac{1}{a}\arccos\frac{a}{u}, \quad u > a > 0.$

$\int \sin u\,du = -\cos u.$ $\int \cos u\,du = \sin u.$ $\int \tan u\,du = -\ln|\cos u|.$

$\int \cot u\,du = \ln|\sin u|.$ $\int \sec u\,du = \ln|\sec u + \tan u|.$

$\int \csc u\,du = -\ln|\csc u + \cot u| = \ln|\csc u - \cot u|.$

$\int \sec^2 u\,du = \tan u.$ $\int \csc^2 u\,du = -\cot u.$ $\int \sec u\tan u\,du = \sec u.$

$\int \sin^2 u\,du = \frac{1}{2}u - \frac{1}{4}\sin 2u.$ $\int \cos^2 u\,du = \frac{1}{2}u + \frac{1}{4}\sin 2u.$ $\int \tan^2 u\,du = \tan u - u.$

$\int \sin^n u\,du = -\frac{\sin^{n-1}u\cos u}{n} + \frac{n-1}{n}\int \sin^{n-2}u\,du.$

$\int \cos^n u\,du = \frac{\cos^{n-1}u\sin u}{n} + \frac{n-1}{n}\int \cos^{n-2}u\,du.$

**Note:* An arbitrary constant is to be added to each indefinite integral.

A BRIEF TABLE OF INTEGRALS* (continued)

$$\int u \sin u \, du = \sin u - u \cos u. \qquad \int u^n \sin u \, du = -u^n \cos u + n \int u^{n-1} \cos u \, du.$$

$$\int u \cos u \, du = \cos u + u \sin u. \qquad \int u^n \cos u \, du = u^n \sin u - n \int u^{n-1} \sin u \, du.$$

$$\int e^{au} \sin nu \, du = \frac{e^{au}(a \sin nu - n \cos nu)}{a^2 + n^2}. \qquad \int e^{au} \cos nu \, du = \frac{e^{au}(a \cos nu + n \sin nu)}{a^2 + n^2}.$$

$$\int \sin au \sin bu \, du = -\frac{\sin(a+b)u}{2(a+b)} + \frac{\sin(a-b)u}{2(a-b)}, \qquad a^2 \neq b^2.$$

$$\int \cos au \cos bu \, du = \frac{\sin(a+b)u}{2(a+b)} + \frac{\sin(a-b)u}{2(a-b)}, \qquad a^2 \neq b^2.$$

$$\int \sin au \cos bu \, du = -\frac{\cos(a+b)u}{2(a+b)} - \frac{\cos(a-b)u}{2(a-b)}, \qquad a^2 \neq b^2.$$

$$\int \sinh u \, du = \cosh u. \qquad \int \cosh u \, du = \sinh u.$$

$$\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du, \quad t > 0; \qquad \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}; \quad \text{and} \quad \Gamma(n+1) = n!, \text{ if } n \text{ is a positive integer.}$$

SOME POWER SERIES EXPANSIONS

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots \quad \text{(Taylor series)}$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \qquad \sin x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} \qquad \cos x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

$$(1-x)^{-1} = \sum_{n=0}^{\infty} x^n \qquad (1-x)^{-2} = \sum_{n=0}^{\infty} (n+1)x^n \qquad \ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}$$

$$\tan x = x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 + \tfrac{17}{315}x^7 + \tfrac{62}{2835}x^9 + \cdots$$

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$$

$$\arcsin x = x + \frac{x^3}{2 \cdot 3} + \frac{1 \cdot 3}{2 \cdot 4 \cdot 5} x^5 + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6 \cdot 7} x^7 + \cdots$$

$$J_0(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(k!)^2 2^{2k}} \qquad J_1(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{k!(k+1)! 2^{2k+1}} \qquad J_n(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+n}}{k! \Gamma(n+k+1) 2^{2k+n}}$$

**Note:* An arbitrary constant is to be added to each indefinite integral.

LINEAR FIRST-ORDER EQUATIONS

The solution to $y' + P(x)y = Q(x)$ taking the value $y(x_0)$ at $x = x_0$ is

$$y(x) = e^{-\int_{x_0}^{x} P(\xi)d\xi}\left[\int_{x_0}^{x} e^{\int_{x_0}^{\xi} P(\zeta)d\zeta}\, Q(\xi)d\xi + y(x_0)\right].$$

METHOD OF UNDETERMINED COEFFICIENTS

To find a particular solution to the constant-coefficient differential equation

$$ay'' + by' + cy = P_m(t)e^{rt},$$

where $P_m(t)$ is a polynomial of degree m, use the form

$$y_p(t) = t^s(A_m t^m + \cdots + A_1 t + A_0)e^{rt};$$

if r is not a root of the associated auxiliary equation, take $s = 0$; if r is a simple root of the associated auxiliary equation, take $s = 1$; and if r is a double root of the associated auxiliary equation, take $s = 2$.

To find a particular solution to the differential equation

$$ay'' + by' + cy = P_m(t)e^{\alpha t}\cos\beta t + Q_n(t)e^{\alpha t}\sin\beta t\,,\ \ \beta \neq 0\,,$$

where $P_m(t)$ is a polynomial of degree m and $Q_n(t)$ is a polynomial of degree n, use the form

$$\begin{aligned} y_p(t) = {} & t^s(A_k t^k + \cdots + A_1 t + A_0)e^{\alpha t}\cos\beta t \\ & + t^s(B_k t^k + \cdots + B_1 t + B_0)e^{\alpha t}\sin\beta t\,, \end{aligned}$$

where k is the larger of m and n. If $\alpha + i\beta$ is not a root of the associated auxiliary equation, take $s = 0$; if $\alpha + i\beta$ is a root of the associated auxiliary equation, take $s = 1$.

VARIATION OF PARAMETERS FORMULA

If y_1 and y_2 are two linearly independent solutions to $ay'' + by' + cy = 0$, then a particular solution to $ay'' + by' + cy = f$ is $y = v_1 y_1 + v_2 y_2$, where

$$v_1(t) = \int \frac{-f(t)y_2(t)}{aW[y_1, y_2](t)}\,dt, \qquad v_2(t) = \int \frac{f(t)y_1(t)}{aW[y_1, y_2](t)}\,dt,$$

and $W[y_1, y_2](t) = y_1(t)y_2'(t) - y_1'(t)y_2(t)$.

A TABLE OF LAPLACE TRANSFORMS

$f(t)$	$F(s) = \mathscr{L}\{f\}(s)$	$f(t)$	$F(s) = \mathscr{L}\{f\}(s)$
1. $f(at)$	$\frac{1}{a}F\left(\frac{s}{a}\right)$	**20.** $\frac{1}{\sqrt{t}}$	$\frac{\sqrt{\pi}}{\sqrt{s}}$
2. $e^{at}f(t)$	$F(s-a)$	**21.** $\sqrt{t}$	$\frac{\sqrt{\pi}}{2s^{3/2}}$
3. $f'(t)$	$sF(s)-f(0)$	**22.** $t^{n-(1/2)}, \quad n=1,2,\ldots$	$\frac{1\cdot 3\cdot 5\cdots(2n-1)\sqrt{\pi}}{2^n s^{n+(1/2)}}$
4. $f^{(n)}(t)$	$s^nF(s)-s^{n-1}f(0)-s^{n-2}f'(0)$	**23.** $t^r, \quad r>-1$	$\frac{\Gamma(r+1)}{s^{r+1}}$
	$-\cdots-sf^{(n-2)}(0)-f^{(n-1)}(0)$	**24.** $\sin bt$	$\frac{b}{s^2+b^2}$
5. $t^nf(t)$	$(-1)^nF^{(n)}(s)$	**25.** $\cos bt$	$\frac{s}{s^2+b^2}$
6. $\frac{1}{t}f(t)$	$\int_s^\infty F(u)\,du$	**26.** $e^{at}\sin bt$	$\frac{b}{(s-a)^2+b^2}$
7. $\int_0^t f(v)\,dv$	$\frac{F(s)}{s}$	**27.** $e^{at}\cos bt$	$\frac{s-a}{(s-a)^2+b^2}$
8. $(f*g)(t)$	$F(s)G(s)$	**28.** $\sinh bt$	$\frac{b}{s^2-b^2}$
9. $f(t+T)=f(t)$	$\frac{\int_0^T e^{-st}f(t)\,dt}{1-e^{-sT}}$	**29.** $\cosh bt$	$\frac{s}{s^2-b^2}$
10. $f(t-a)u(t-a), a\ge 0$	$e^{-as}F(s)$	**30.** $\sin bt - bt\cos bt$	$\frac{2b^3}{(s^2+b^2)^2}$
11. $g(t)u(t-a), \quad a\ge 0$	$e^{-as}\mathscr{L}\{g(t+a)\}(s)$	**31.** $t\sin bt$	$\frac{2bs}{(s^2+b^2)^2}$
12. $u(t-a), \quad a\ge 0$	$\frac{e^{-as}}{s}$	**32.** $\sin bt + bt\cos bt$	$\frac{2bs^2}{(s^2+b^2)^2}$
13. $\prod_{a,b}(t), \quad 0<a<b$	$\frac{e^{-sa}-e^{-sb}}{s}$	**33.** $t\cos bt$	$\frac{s^2-b^2}{(s^2+b^2)^2}$
14. $\delta(t-a), \quad a\ge 0$	e^{-as}	**34.** $\sin bt\cosh bt - \cos bt\sinh bt$	$\frac{4b^3}{s^4+4b^4}$
15. e^{at}	$\frac{1}{s-a}$	**35.** $\sin bt\sinh bt$	$\frac{2b^2s}{s^4+4b^4}$
16. $t^n, \quad n=1,2,\ldots$	$\frac{n!}{s^{n+1}}$	**36.** $\sinh bt - \sin bt$	$\frac{2b^3}{s^4-b^4}$
17. $e^{at}t^n, \quad n=1,2,\ldots$	$\frac{n!}{(s-a)^{n+1}}$	**37.** $\cosh bt - \cos bt$	$\frac{2b^2s}{s^4-b^4}$
18. $e^{at}-e^{bt}$	$\frac{(a-b)}{(s-a)(s-b)}$	**38.** $J_\nu(bt), \nu>-1$	$\frac{\left(\sqrt{s^2+b^2}-s\right)^\nu}{b^\nu\sqrt{s^2+b^2}}$
19. $ae^{at}-be^{bt}$	$\frac{(a-b)s}{(s-a)(s-b)}$		