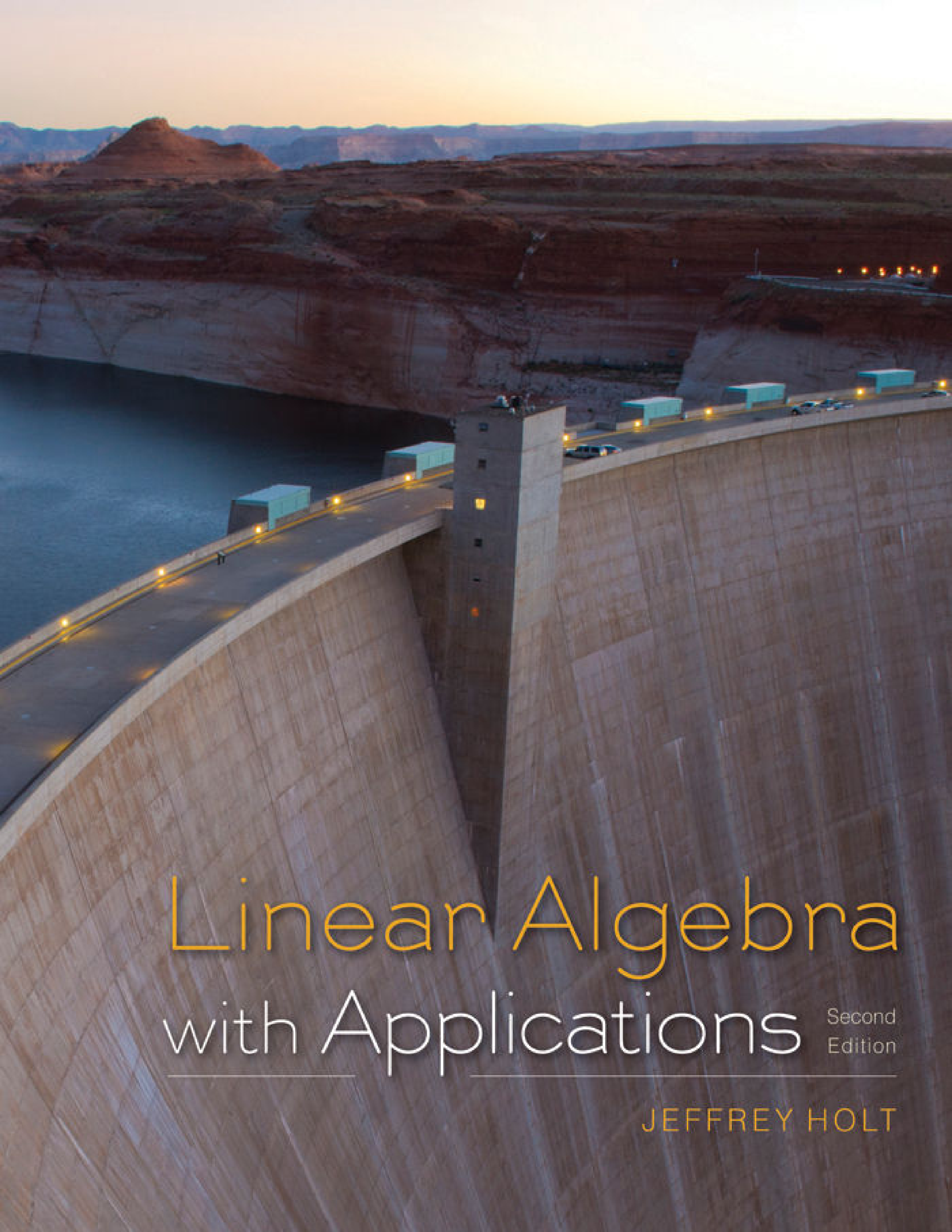

Linear Algebra
with Applications
Second
Edition
JEFFREY HOLT

Linear Algebra

with Applications

Second Edition

JEFFREY HOLT

University of Virginia

w. h. freeman
Macmillan Learning
New York

Vice President, STEM: Ben Roberts
Publisher: Terri Ward
Acquisitions Editor: Nikki Miller
Senior Marketing Manager: Tom DeMarco
Marketing Assistant: Cate McCaffery
Development Editor: Jorge Amaral
Director, Content Management Enhancement: Tracey Kuehn
Managing Editor: Lisa Kinne
Executive Media Editor: Catriona Kaplan
Director of Digital Production: Keri deManigold
Senior Media Producer: Alison Lorber
Associate Editor: Victoria Garvey
Assistant Media Editor: Emily Tenenbaum
Editorial Assistant: Katharine Munz
Photo Editor: Sheena Goldstein
Cover Designer: Blake Logan
Interior Designer: Lee McKevitt
Senior Project Editor: Jodi Isman
Illustrations: MPS Limited
Senior Production Supervisor: Paul Rohloff
Composition: MPS Limited
Printing and Binding: LSC Communications
Cover photo: Cultura RM Exclusive/Pete Saloutos/Getty Images

Library of Congress Control Number: 2016937575

Instructor Complimentary Copy:
ISBN-13: 978-1-319-05771-8
ISBN-10: 1-319-05771-3

Student Edition Cloth:
ISBN-13: 978-1-4641-9334-7
ISBN-10: 1-4641-9334-7

Student Edition Loose-leaf:
ISBN-13: 978-1-4641-9369-9
ISBN-10: 1-4641-9369-X

Printed in the United States of America

First printing

W. H. Freeman and Company
One New York Plaza
Suite 4500
New York, NY 10004-1562
www.macmillanlearning.com

To Kathy, Laura, Mike, and Tom

CONTENTS

Sections marked with an asterisk are optional but may be required for later optional sections. See the start of each optional section for dependency information.

PREFACE

Welcome to *Linear Algebra with Applications*, Second Edition. This book is designed for use in a standard linear algebra course for an applied audience, usually populated by sophomores and juniors. While the majority of students in this type of course are majoring in engineering, some also come from the sciences, economics, and other disciplines. To accommodate a broad audience, applications covering a variety of topics are included.

Although this book is targeted toward an applied audience, full development of the theoretical side of linear algebra is included, so this textbook also can be used as an introductory course for mathematics majors. I have designed this book so that instructors can teach from it at a conceptual level that is appropriate for their individual course.

There is a collection of core topics that appear in virtually all linear algebra texts, and these are included in this text. In particular, the core topics recommended by the Linear Algebra Curriculum Study Group are covered here. The organization of these core topics varies from text to text, with the recent trend being to introduce more of the "abstract" material earlier rather than later. The organization here reflects this trend, with the chapters (approximately) alternating between computational and conceptual topics.

New to this Edition

- Practice Problems have been added to the end of each section, with full solutions available in the back of the book. This allows students to check their understanding of the material as they progress through the chapter and also can be used by instructors as extra examples.
- Supplementary Exercises have been added to the end of each chapter to provide instructors and students with an even greater

variety of exercises.

- The section on Change of Basis has been moved from Chapter 6 to Chapter 4. Reviewers suggested that the Change of Basis material would be useful earlier, so it is now introduced closer to the section on Basis and Dimension.
- Examples have been revised and added throughout the book.
- The subsection on the Shortcut Method has been moved up in section 5.1, as suggested by reviewers.
- In Chapters 1 and 6, the sections on Approximation Methods have been moved to the end of the chapter for a more streamlined presentation.
- In Chapters 4 and 9, the notation for coordinate vectors has been updated.

Text Features

Early Presentation of Key Concepts. Traditional linear algebra texts initially focus on computational topics, then treat more conceptual subjects soon after introducing abstract vector spaces. As a result, at the point abstract vector spaces are introduced, students face two simultaneous challenges:

(a) A change in mode of thinking, from largely mechanical and computational (solving systems of equations, performing matrix arithmetic) to wrestling with conceptual topics (span, linear independence).

(b) A change in context from the familiar and concrete $\mathbf{R}^n$ to abstract vector spaces.

Many students cannot effectively meet both these challenges at the same time. The organization of topics in this book is designed to address this significant problem.

In *Linear Algebra with Applications*, we first address challenge (a). Conceptual topics are explored early and often, blended in with topics that are more computational. This spreads out the impact of conceptual topics, giving students more time to digest them. The first six chapters are presented solely in the context of Euclidean space, which is relatively familiar to students. This defers challenge (b) and also allows for a treatment of eigenvalues and eigenvectors that comes earlier than in other texts.

Challenge (b) is taken up in Chapter 7, where abstract vector spaces are introduced. Here, many of the conceptual topics explored in the context of Euclidean space are revisited in this more general setting. Definitions and theorems presented are similar to those given earlier (with explicit references to reinforce connections), so students have less trouble grasping them and can focus more attention on the new concept of an abstract vector space.

From a mathematical standpoint, there is a certain amount of redundancy in this book. Quite a lot of the material in Chapters 7, 9, and 10, where the majority of the development of abstract vector spaces resides, has close analogs in earlier chapters. This is a deliberate part of the book design, to give students a second pass through key ideas to reinforce understanding and promote success.

Topics Introduced and Motivated Through Applications. To provide understanding of why a topic is of interest, when it makes sense I use applications to introduce and motivate new topics, definitions, and concepts. In particular, many sections open with an application. Applications are also distributed in other places, including the exercises. In a few instances, entire sections are devoted to applications.

Extensive Exercise Sets. I recall reading a review of a text in which the reviewer stated, in essence, “This text has a very nice collection of exercises, which is the only thing I care about. When will textbook authors learn that the most important consideration is the exercises?” Although this sentiment might be extreme, most instructors share it to some degree, and it is certainly true that a text with inadequate problem sets can be frustrating. *Linear Algebra with Applications*, Second Edition, contains over 2600 exercises, covering a wide range of types (computational to conceptual to proofs) and difficulty levels. The Second Edition now includes Supplementary

Exercises at the end of each chapter, providing an even greater variety of exercises from which to choose.

Ample Instructional Examples. For many students, a primary use of a mathematics text is to learn by studying examples. Besides those examples used to introduce new topics, this text contains a large number of additional representative examples. Perhaps the number one complaint from students about mathematics texts is that there are not enough examples. I have tried to address that in this text. The Second Edition also includes Practice Problems at the end of each section, which can be used as extra examples. Full solutions to all of the Practice Problems are included in the back of the book, allowing students to check their understanding of the material as they progress through the chapter.

Support for Theory and Proofs. Many students in a first linear algebra course are not usually math majors, and many have limited experience with proofs. Proofs of most theorems are supplied in this book, but it is possible for a course instructor to vary the level of emphasis given to proofs through choice of lecture topics and homework exercises.

Throughout the book, the goal of proofs is to help students understand why a statement is true. Thus, proofs are presented in different ways. Sometimes a theorem might be proved for a special case, when it is clear that no additional understanding results from presenting the more general case (especially if the general case is more notationally messy). If a proof is difficult and will not help students understand why the theorem is true, then it might be given at the end of the section or omitted entirely. If it provides a source of motivation for the theorem, the proof might come before the statement of the theorem. I also have written an appendix containing an overview of how to read and write proofs to assist those with limited experience. (See the text website at **www.macmillanlearning.com/holt2e**.)

Most linear algebra texts handle theorems and proofs in similar ways, although there is some variety in the level of rigor. However, it seems that often there is not enough concern for whether or not the proof is conveying why the theorem is true, with the goal instead being to keep the proof as short as possible. Sometimes it is worth taking a bit of extra time to give a complete explanation. For example, in Section 1.1, a system of equations is reduced to “0 = 8.” At this point, most texts would state that this shows the system has no solutions, and it is likely that most students would agree. However, it is also possible that many students will not know why the system has no solutions, so a brief explanation is included.

Organization of Material

Roughly speaking, the chapters alternate between computational and conceptual topics. This is deliberate, in order to spread out the challenge of the conceptual topics and to give students more time to digest them. The material in Chapters 1–6 and 8 is exclusively in the context of Euclidean space and includes the core topics recommended by the Linear Algebra Curriculum Study Group. Chapters 7, 9, and 10 cover topics in the context of abstract vector space, and Chapter 11 contains a collection of optional topics that can be included at the end of a course.

Those sections marked with an asterisk (*) are optional, but in some cases they may be assumed in other optional sections that come later. See the start of each optional section for dependency information.

Chapter 1 is fairly computational, providing a comprehensive introduction to systems of linear equations and their solutions. Iterative solutions to systems are also treated. The chapter includes a section containing in-depth descriptions of several applications of linear systems. By the end of this chapter, students should be proficient in using augmented matrices and row operations to find the set of solutions to a linear system.

Chapter 2 shifts from mechanical to conceptual material. This chapter is devoted to introducing vectors and the important concepts of span and linear independence, all in the concrete context of $\mathbf{R}^n$. These topics appear early so that students have more time to absorb these important concepts.

Chapter 3 shifts from conceptual back to (mostly) mechanical material, starting with a treatment of linear transformations from $\mathbf{R}^n$ to $\mathbf{R}^m$. This is used to motivate the definition of matrix multiplication, which is covered in the next section along with other matrix arithmetic. This is followed by a section on computing the inverse of a matrix, motivated by finding the inverse of a linear transformation. Matrix factorizations, arguably related to numerical methods, provide an alternate way of organizing computations. The chapter closes with Markov chains, a topic not typically covered until after discussing eigenvalues and eigenvectors. But this subject easily can be covered earlier, and as there are a number of interesting applications of Markov chains, they are included here.

In Chapter 4, we again shift back to a more conceptual topic, subspaces in $\mathbf{R}^n$. The first section provides the definition of subspace along with examples. The second section develops the notion of basis and dimension for subspaces in $\mathbf{R}^n$, and the third section thoroughly treats row and column spaces. The last section is optional and covers change of basis in Euclidean space. By the end of this chapter, students will have been exposed to many of the central conceptual topics typically covered in a linear algebra course. These are revisited (and eventually generalized) throughout the remainder of the book.

Chapter 5 develops the usual properties of determinants. This topic has moved around in texts in recent years. For some time, the trend was to reduce the emphasis on determinants, but lately they have made something of a comeback. This chapter is relatively short and is introduced at this point in the text to support the introduction of eigenvalues and eigenvectors in the next chapter. Those who want only enough of determinants for eigenvalues can cover only Section 5.1 and lightly treat Section 5.2.

Chapter 6 provides a treatment of eigenvalues and eigenvectors that comes earlier than in most books. Diagonalization is presented here and is revisited for symmetric matrices in Chapter 8. The remaining sections are optional and can be covered as needed.

Abstract vector spaces are first introduced in Chapter 7. This relatively late introduction allows students time to internalize key concepts such as span, linear independence, and subspaces before being presented with the challenge of abstract vector spaces. To further smooth this transition, definitions and theorems in this chapter typically include specific references to analogs in earlier chapters to reinforce connections. Since most proofs are similar to those given in Euclidean space, many are left as homework exercises. Making the parallels between Euclidean space and abstract vector spaces very explicit helps students more easily assimilate this material.

The order of Chapter 7 and Chapter 8 can be reversed, so if time is limited, Chapter 8 can be covered immediately after Chapter 6. However, if both Chapters 7 and 8 are going to be covered, it is recommended that Chapter 7 be covered first so that this abstract material is not appearing at the end of the course.

In Chapter 8, the context shifts back to Euclidean space and treats topics that are more computational than conceptual. Chapter 7 is placed before Chapter 8 to allow for an introduction to abstract vector spaces that does not come at the end of the term, and to a degree preserves the chapter alternation between computational and conceptual. However, the two chapters are interchangeable if necessary.

9. Linear Transformations

The focus of Chapter 9 shifts back to abstract vector spaces, with a general development of linear transformations. As in Chapter 7, there is some deliberate redundancy between the material in Chapter 9 and that presented in earlier chapters. Explicit references to earlier analogous definitions and theorems are provided to reinforce connections and improve understanding.

10. Inner Product Spaces

Chapter 10 is in the context of abstract vector spaces. The content is somewhat parallel to the first two sections of Chapter 8, with explicit analogs noted. The first section defines the inner product and inner product spaces and gives numerous examples of each. The second section generalizes the notion of projection and the Gram–Schmidt process to inner product spaces, and the last section provides applications of inner products. For the most part, Chapter 10 is independent of Chapter 9 (except for a small number of exercise references to linear transformations), so Chapter 10 can be covered without covering Chapter 9.

11. Additional Topics and Applications*

Chapter 11 provides a collection of topics and applications that most instructors consider optional but that are nonetheless important and interesting. These can be inserted at the end of a course as desired.

Course Coverage

Most schools teach linear algebra as a semester-long course that meets 3 hours per week. This usually does not allow enough time to cover everything in this book, so decisions about coverage are required.

The dependencies among chapters are fairly straightforward.

- The first six chapters are designed to be covered in order, although there are some optional sections (flagged in the table of contents) that can be skipped.
- The order of Chapter 7 and Chapter 8 can be interchanged.
- The order of Chapter 9 and Chapter 10 can be interchanged (except for a small number of exercises in Chapter 10 that use linear transformations).
- Chapter 9 assumes Chapter 7, and Chapter 10 assumes Chapter 7 and Chapter 8.

Below are a few options for course coverage. Note that some sections or even subsections can be omitted to fine-tune the course to local needs.

- **Modest Pace:** Chapters 1–8. This course covers all key concepts in the context of Euclidean space and provides an introduction to abstract vector spaces.
- **Intermediate Pace:** Chapters 1–9 or Chapters 1–8 and 10. This includes everything from the Modest Pace course and either linear transformations on abstract vector spaces (Chapter 9) or inner product spaces (Chapter 10).
- **Brisk Pace:** Chapters 1–10. This will include everything from the Modest Pace course, as well as both linear transformations on abstract vector spaces and inner product spaces. This is roughly the syllabus we follow here at the University of Virginia,

although we omit a few optional sections and we give exams in the evening, which makes available more lecture time. (A detailed list of sections that we cover is available on request from the author.)

Chapter Transitions

Each chapter opens with a photo of an example of green engineering, including dams, wind turbines, solar panels, and more. Although linear algebra has applications in many fields, engineering is perhaps one of the most visible. With our society becoming more and more environmentally conscious, these photos demonstrate ways in which engineering is used in attempts to provide environmentally friendly benefits. Although the mathematics behind these examples is not discussed, we hope the chapter opener photos and captions serve as a small reminder to students of the importance and relevance of linear algebra to their everyday lives.

Supplements for Instructors

Instructor's Guide with Full Solutions (includes Instructor's Resource material with all solutions)
Test Bank
Lecture Slides and Image Slides
iClicker Slides
Practice Quizzes
MATLAB Manual
Maple Manual
Mathematica Manual

Supplements for Students

Student Study Guide with Selected Solutions
MATLAB Manual
Maple Manual

Mathematica Manual

Media

WebAssign

WebAssign offers algorithmic questions from *Linear Algebra with Applications,* Second Edition, in a powerful online instructional system. WebAssign lets you easily create assignments, grade homework, and give your students instant feedback. Along with flexible features, class and question-level analytics are available for instructors and students.

WeBWorK

http://webwork.maa.org W. H. Freeman offers algorithmically generated questions (with full solutions) through this free open source online homework system developed at the University of Rochester.

iclicker

The hassle-free solution created for educators by educators. For more information, visit www.iclicker.com.

ACKNOWLEDGMENTS

I'd like to give a special thank you to Roger Lipsett for accuracy reviewing the book and the Answers to Selected Exercises. I'd also like to thank the instructors who authored the supplements that accompany the Second Edition: Matthew Clay at University of Arkansas (Instructor's Resource Manual); Tim Flaherty at Carnegie Mellon University (Solutions); Walden Freedman at Humboldt State University (Test Bank); Mark Gockenbach at Michigan Technological University (MatLab Manual); Mark Hunacek at Iowa State University (Practice Quizzes); Will Murray at California State University, Long Beach (iclicker Questions and Lecture Slides); Steven Pankavich at Colorado School of Mines (Mathematica Manual); Jianzhong Wang at Sam Houston State University (Maple Manual); and Amy Yielding at Eastern Oregon University (Study Guide). Thank you to Melanie Fulton for accuracy reviewing the Test Bank, Mark Hughes for accuracy reviewing the Solutions, and Paul Lorzsak for accuracy reviewing the Practice Quizzes and iclicker Questions.

A large number of experienced mathematics faculty were generous in sharing their thoughts as this Second Edition was developed. Their input contributed in countless ways to the improvement of this book. I gratefully acknowledge the comments and suggestions from the following individuals:

Trefor Bazett, *University of Toronto*
Yuri Berest, *Cornell University*
Papiya Bhattacharjee, *Penn State Erie, The Behrend College*
Darin Brezeale, *University of Texas at Arlington*
Terry Bridgman, *Colorado School of Mines*
Robert G. Brown, *Old Dominion University*
Dietrich Burbulla, *University of Toronto*
Robert Byerly, *Texas Tech University*
Thomas R. Cameron, *Washington State University*
Jen-Mei Chang, *California State University, Long Beach*

Edward Curtis, *University of Washington*
Stefaan De Winter, *Michigan Technological University*
Kosmas Diveris, *St. Olaf College*
Edward Dobson, *Mississippi State University*
Matthew Fahy, *Northern Arizona University*
Walden Freedman, *Humboldt State University*
Jonathan Ginsberg, *New York City College of Technology*
Jason Gregersen, *Michigan Technological University*
Barry Griffiths, *University of Central Florida*
Mark Hughes, *Brigham Young University*
Mark Hunacek, *Iowa State University*
Robert Jackson, *Pellissippi State Community College*
Sophia Jang, *Texas Tech University*
Bin Jiang, *Portland State University*
Mohammad Kazemi, *The University of North Carolina at Charlotte*
Monika Keindl, *Northern Arizona University*
Nataliya Laptyeva, *University of Toronto*
James Lark, *University of Virginia*
Saburo Matsumoto, *College of the Canyons*
Mariana Montiel, *Georgia State University*
Will Murray, *California State University, Long Beach*
Jeff O'Connell, *Ohlone College*
Joyce O'Halloran, *Portland State University*
Antonio Palacios, *San Diego State University*
Stephen Pankavich, *Colorado School of Mines*
Emma Previato, *Boston University*
Marju Purin, *St. Olaf College*
Ernest Ratliff, *Texas State University*
Lih-Ing Wu Roeger, *Texas Tech University*
Hadi Salmasian, *University of Ottawa*
Eric Swenson, *Brigham Young University*
Satoshi Takahashi, *Rochester Institute of Technology*
Nicoleta Tarfulea, *Purdue University Calumet*
Ngo Viet, *California State University, Long Beach*
Haiyan Wang, *Arizona State University*
Tamas Wiandt, *Rochester Institute of Technology*
Xiao-Dong Zhang, *Florida Atlantic University*

I’d also like to thank the instructors who offered development of the first edition:

Lowell Abrams, *The George Washington University*
Maria Theodora Acosta, *Texas State University*
Ulrich Albrecht, *Auburn University*
John Alongi, *Northwestern University*
Paolo Aluffi, *Florida State University*
Dorothy C. Attaway, *Boston University*
Chris Bernhardt, *Fairfield University*
Eddie Boyd Jr., *University of Maryland Eastern Shore*
Natasha Bozovic, *San Jose State University*
Mary E. Bradley, *University of Louisville*
Fernando Burgos, *University of South Florida*
Nancy Childress, *Arizona State University*
Peter Cholak, *University of Notre Dame*
Matthew T. Clay, *University of Arkansas*
Adam Coffman, *Indiana University-Purdue University Fort Wayne*
Ray E. Collings, *Georgia Perimeter College*
Ben W. Crain, *George Mason University*
Curtis Crawford, *Columbia Basin College*
Xianzhe Dai, *University of California, Santa Barbara*
Alain D’Amour, *Southern Connecticut State University*
James W. Daniel, *University of Texas at Austin*
Gregory Daubenmire, *Las Positas College*
Donald Davis, *Lehigh University*
Tristan Denley, *Austin Peay State University*
Yssa DeWoody, *Texas Tech University*
Caren Diefenderfer, *Hollins University*
Javid Dizgam, *University of Illinois Urbana-Champaign*
Neil Malcolm Donaldson, *University of California, Irvine*
Alina Raluca Dumitru, *University of North Florida*
Della Duncan-Schnell, *California State University, Fresno*
Alexander Dynin, *The Ohio State University*
Daniel J. Endres, *University of Central Oklahoma*
Alex Feingold, *SUNY Binghamton*
John Fink, *Kalamazoo College*

Timothy J. Flaherty, *Carnegie Mellon University*
Bill Fleissner, *The University of Kansas*
Chris Francisco, *Oklahoma State University*
Natalie P. Frank, *Vassar College*
Chris Frenzen, *Naval Postgraduate School*
Anda Gadidov, *Kennesaw State University*
Scott Glasgow, *Brigham Young University*
Jay Gopalakrishnan, *Portland State University*
Anton Gorodetski, *University of California-Irvine*
John Goulet, *Worcester Polytechnic Institute*
William Hager, *University of Florida*
Patricia Hale, *California State Polytechnic University, Pomona*
Chungsim Han, *Baldwin Wallace College*
James Hartsman, *Colorado School of Mines*
Willy Hereman, *Colorado School of Mines*
Konrad J. Heuvers, *Michigan Technological University*
Allen Hibbard, *Central College*
Rudy Horne, *Morehouse College*
Kevin James, *Clemson University*
Naihuan Jing, *North Carolina State University*
Raymond L. Johnson, *Rice University*
Thomas W. Judson, *Stephen F. Austin State University*
Steven Kahan, *CUNY Queens College*
Jennifer D. Key, *Clemson University*
In-Jae Kim, *Minnesota State University*
Alan Koch, *Agnes Scott College*
Joseph D. Lakey, *New Mexico State University*
Namyong Lee, *Minnesota State University*
Luen-Chau Li, *Penn State University*
Lucy Lifschitz, *University of Oklahoma*
Roger Lipsett, *Brandeis University*
Xinfeng Liu, *University of South Carolina*
Satyagopal Mandal, *The University of Kansas*
Aldo J. Manfroi, *University of Illinois at Urbana-Champaign*
Douglas Bradley Meade, *University of South Carolina*
Valentin Milanov, *Fayetteville State University*
Mona Mocanasu, *Northwestern University*

Carrie Muir, *University of Colorado at Boulder*
Shashikant Mulay, *The University of Tennessee*
Bruno Nachtergaele, *University of California, Davis*
Ralph Oberste-Vorth, *Marshall University*
Timothy E. Olson, *University of Florida*
Boon Wee Ong, *Penn State Erie, The Behrend College*
Seth F. Oppenheimer, *Mississippi State University*
Bonsu M. Osei, *Eastern Connecticut State University*
Allison Pacelli, *Williams College*
Richard O. Pellerin, *Northern Virginia Community College*
Jack Porter, *The University of Kansas*
Chuanxi Qian, *Mississippi State University*
David Richter, *Western Michigan University*
John Rossi, *Virginia Tech*
Matthew Saltzman, *Clemson University*
Alicia Sevilla, *Moravian College*
Alexander Shibakov, *Tennessee Tech University*
Rick L. Smith, *University of Florida*
Katherine F. Stevenson, *California State University, Northridge*
Allan Struthers, *Michigan Technological University*
Alexey Sukhinin, *University of New Mexico*
Gnana Bhaskar Tenali, *Florida Institute of Technology*
Magdalena Toda, *Texas Tech University*
Mark Tomforde, *University of Houston*
Douglas Torrance, *University of Idaho*
Scott Wilson, *CUNY Queens College*
Amy Yielding, *Eastern Oregon University*
Jeong-Mi Yoon, *University of Houston-Downtown*
John Zerger, *Catawba College*
Jianqiang Zhao, *Eckerd College*

A large round of thanks is due to all of the people at Macmillan Learning for their assistance and guidance during this project. These include Nikki Miller, Jorge Amaral, Jodi Isman, and Victoria Garvey. I particularly want to thank Terri Ward, who has been with this project since the beginning and displayed a remarkable level of patience with my consistently inconsistent progress.

I gratefully acknowledge the support of the University of Virginia, where I class tested portions of this book. I also thank Simon Fraser University and the IRMACS Centre, for their warm hospitality during my sabbatical visit while I completed the final draft of the first edition. Last but certainly not least, I thank my family for their ongoing and unconditional support as I wrote this book.

CHAPTER 1

Systems of Linear Equations

Cultura RM Exclusive/Pete Saloutos/Getty Images

Linear algebra has applications in science, computer science, social science, business, and other fields. Engineering is filled with visible applications of linear algebra, with examples all around us of infrastructure that affects our everyday lives. For example, Glen Canyon Dam, located in northern Arizona, was built to provide flood control, store water to be used during droughts, and produce hydroelectric power for several surrounding states. The construction of Glen Canyon Dam was an impressive feat of engineering, but dams also can be controversial. Even though dams can be used to create environmentally friendly hydroelectric power, they are also criticized for harmful ecological and environmental consequences. Throughout the rest of the chapter openers, we will take a look at other examples of engineering marvels that seek to provide environmentally friendly benefits.

There are endless applications of linear algebra in the sciences, social sciences, and business, and many are included throughout this book. Chapter 1 begins our tour of linear algebra in territory that may be familiar, systems of linear equations. In Section 1.1 and Section 1.2 we develop a systematic method for finding the set of solutions to a linear system. Section 1.3 focuses on a number of applications. In Section 1.4 we consider methods for approximating solutions that can be applied to large systems.

1.1 Lines and Linear Equations

The goal of this section is to provide an introduction to systems of linear equations. The following example is a good place to start. Although not complicated, it contains the essential elements of other applications and also serves as a gateway to our treatment of more general systems of linear equations.

Example 1

A discount movie theater sells 100 tickets for an afternoon matinee. Tickets for children cost \$3 and tickets for adults cost \$5. If the total ticket revenue is \$422, how many of each type of ticket was sold?

Solution Let x be the number of children's tickets and y the number of adult tickets. A total of 100 tickets were sold, so it follows that

$$x+y=100 \tag{1}$$

Since each child's ticket cost \$3, the revenue from children's tickets is equal to $3x$. By the same reasoning, the revenue from adult tickets is $5y$. The total revenue is \$422, so we have

$$3x+5y=422 \tag{2}$$

Both (1) and (2) are equations of lines, shown in Figure 1. The pair (x, y) that satisfies (1) and (2) will lie on *both* lines, so must be at the intersection of the lines. We cannot precisely identify x and y from the graph, so we use algebraic methods. The two equations are

$$x+y=1003x+5y=422 \tag{3}$$

We can “eliminate” x by multiplying the top equation by −3 and adding it to the bottom equation.

$$-3x-3y=-300 \ +(3x+5y=422) \Rightarrow \quad 2y=122$$

Solving for y in $2y = 122$ gives $y = 61$. Substituting $y = 61$ back into the top equation of (3) gives us

$$x+61=100$$

which simplifies to $x = 39$. We can check that $x = 39$ and $y = 61$ works by plugging them into each equation in (3).

$$x+y=39+61=100$$

and

$$3x+5y=3(39)+5(61)=422$$

Note also that (39, 61) looks like the point of intersection in Figure 1. We conclude that tickets were sold for 39 children and 61 adults.

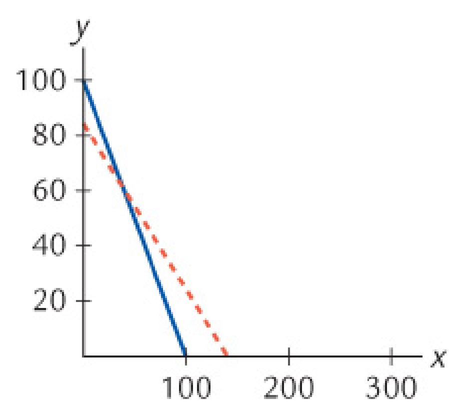


Figure 1 Graphs of $x + y = 100$ (solid) and $3x + 5y = 422$ (dashed) from Example 1.

▶ Substituting into the equations is not part of finding the solution, but it is a way to check that the solution is correct.

Systems of Linear Equations

Linear Equation

The equations in the preceding problem are examples of **linear equations**. In general, a linear equation has the form

a1x1+a2x2+a3x3+⋯+anxn=b (4)

where a_1, a_2, ⋯, a_n and b are constants and x_1, x_2, … , x_n are variables or unknowns. For example, the equations in (3) are both linear, and so is

3x1+4x2+7x3−2x4=−19 (5)

Solution of Linear Equation

A **solution** (s_1, s_2, … , s_n) to (4) is an ordered set of n numbers (sometimes called an *n-tuple*) such that if we set $x_1 = s_1$, $x_2 = s_2$, ⋯ , $x_n = s_n$, then (4) is satisfied. For example, (−2, 5, 1, 13) is a solution to (5) because

3(−2)+4(5)−7(1)−2(13)=−19

Solution Set, Hyperplane

The **solution set** for a linear equation such as (4) consists of the set of all solutions to the equation. When the equation has two variables, the graph of the solution set is a line. In three variables, the graph of a solution set is a plane. (See Figure 2 for an example.) If $n \geq 4$, then the solution set of all n-tuples that satisfy (4) is called a **hyperplane**.

The set of two linear equations in (3) is an example of a *system of linear equations.* Other examples of systems of linear equations are

−3x1+5x2−x3=4 4x1−2x2−8x3+5x4=−1 −x2−9x3=−4and−x1+7x2 +2
x4=136x1+4x2−8x2=11 x3−2x4=5−5x1−9x2 =0 (6)

Our usual practice will be to write all systems of linear equations as shown above, aligning the variables vertically and with x_1, x_2 , … appearing in order from left to right.

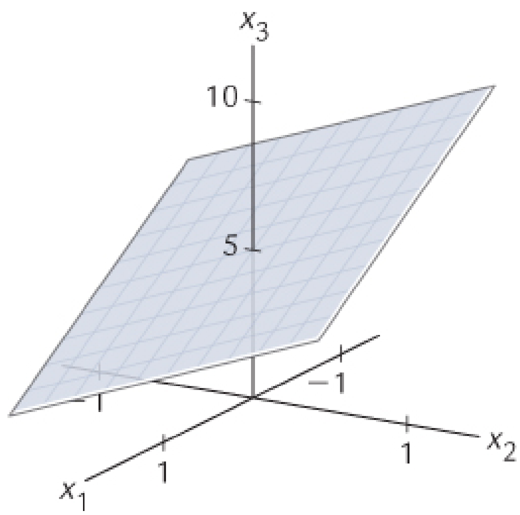


Figure 2 Graph of the solutions to $3x_1 - 2x_2 + x_3 = 5$.

DEFINITION 1.1 ▶

System of Linear Equations

A system of linear equations is a collection of equations of the form

a11x1+a12x2+a13x3+⋯+a1nxn=b1a21x1+a22x2+a23x3+⋯+a2nxn=b2a31x+a32x2+a33x3+⋯+a3nxn=b3 ⋮ ⋮ ⋮ ⋮am1x1+am2x2+am3x3+⋯+amnxn=bm (7)

▶ For brevity, we sometimes use “linear system” or “system” when referring to a system of linear equations.

When reading the coefficients, for a_{32} we say “a-three-two” instead of “a-thirty-two” because the “32” indicates that a_{32} is the coefficient from the third equation that is multiplied by x_2. For example, in the system on the right of (6) we have $a_{14} = 5$, $a_{22} = 7$,

$a_{34} = -2$, and $a_{32} = 0$. Here $a_{32} = 0$ because there is no x_2 term in the third equation.

Solution for Linear System, Solution Set for a Linear System

The system (7) has m equations with n unknowns. It is possible for m to be greater than, equal to, or less than n, and we will encounter all three cases. A **solution** to the linear system (7) is an n-tuple $(s_1, s_2, \cdots, s_n)$ that satisfies every equation in the system. The collection of all solutions to a linear system is called the **solution set** for the system.

In Example 1, there was exactly one solution to the linear system. This is not always the case.

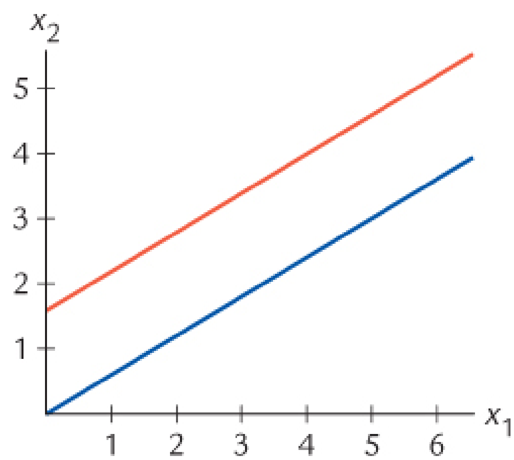


Figure 3 Graphs of $6x_1 - 10x_2 = 0$ (blue) and $-3x_1 + 5x_2 = 8$ (red) from Example 2.

Example 2

Find all solutions to the system of linear equations

$$6x1-10x2=0-3x1+5x2=8 \qquad (8)$$

Solution We will proceed as we did in Example 1, by eliminating a variable. This time we multiply the first equation by 12 and then add it to the second,

$$3x1-5x2=0+(-3x1+5x2=8\ ) \quad \Rightarrow 0=8$$

The equation 0 = 8 tells us that there are *no* solutions to the system. Why? Because if there were values of x_1 and x_2 that satisfied both the equations in (8), then we could plug them in, work through the above algebraic steps with these values in place, and *prove* that 0 = 8, which we know is not true. So, it must be that our original assumption that there are values of x_1 and x_2 that satisfy (8) is false, and therefore we can conclude that the system has no solutions.

This explanation gives an example of a mathematical proof technique called “proof by contradiction.” You can read about this and other methods of proof in the appendix “Reading and writing proofs” posted on the text website. (See the Preface for the Web address.)

The graphs of the two equations in Example 2 are parallel lines (see Figure 3). Since the lines do not have any points in common, there cannot be values that satisfy both equations, matching what we found algebraically.

Consistent Linear System, Inconsistent Linear System

If a linear system has at least one solution, then we say that it is **consistent**. If not (as in Example 2), then it is **inconsistent**.

Example 3

Find all solutions to the system of linear equations

$$4x1+10x2=\ \ 14-6x1-15x2=-21 \tag{9}$$

Solution This time, we multiply the first equation by 32 and then add,

$$6x1-15x2=21\ +(-6x1+15x2=-21) \quad \Rightarrow 0=0$$

Unlike Example 2, where we ended up with an equation that had no solutions, here we find ourselves with the equation 0 = 0 that is satisfied by *any* choices of x_1 and x_2. This tells us that the relationship between x_1 and x_2 is the same in both equations. In this case we select one of the equations (either will work) and solve for x_1 in terms of x_2, which gives us

x1=72−52x2

For *every* choice of x_2 there will be a corresponding choice of x_1 that satisfies the original system (9). Therefore there are infinitely many solutions. To avoid confusing variables with values satisfying the linear system, we describe the solutions to (9) by

x1=72−52s1 x2=s1 (10)

Free Parameter, General Solution

where s_1 is called a **free parameter** and can be any real number. This is known as the **general solution** because it gives all solutions to the system of equations.

We note that (10) is not the only way to describe the solutions. If we solve for x_2 instead of x_1, then we arrive at the formulation of the general solution

x1= s1x2=75−24s1

where s_1 is any real number.

Figure 4 shows the graphs of the two equations in (9). It looks like something is missing, but there is only one line because the two equations have the same graph. Since the graphs coincide, they have infinitely many points in common, which agrees with our algebraic conclusion that there are infinitely many solutions to (9).

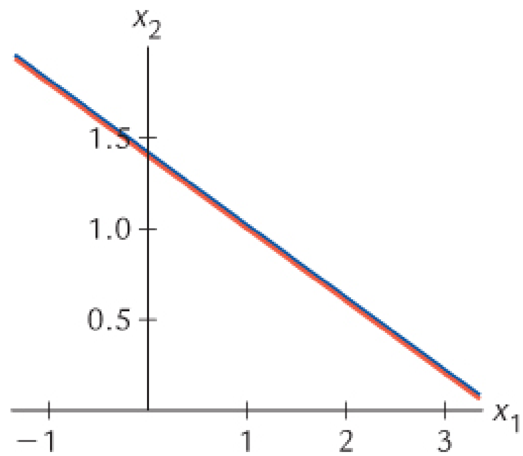


Figure 4 Graphs of $4x_1 + 10x_2 = 14$ (blue) and $-6x_1 - 15x_2 = -21$ (red) from Example 3.

Table 1 Percentage of Glycol Required to Prevent Freezing

Minimum Temp. (F)	Propylene Glycol Volume (%)
20	18
10	29
0	36
−10	42
−20	46
−30	50
−40	54
−50	57

Example 4

Fran is designing a solar hot water system for her home. The system works by circulating a mixture of water and propylene glycol through rooftop solar panels to absorb heat, and then through a heat exchanger to heat household water (Figure 5). The glycol is included in the mixture to prevent freezing during cold weather. Table 1 shows the percentage of glycol required for various minimum temperatures.

The lowest the temperature ever gets at Fran's house is 0° F. She can purchase solutions of water and glycol that contain either 18% glycol or 50% glycol, which she will combine for her 300-liter system. How much of each type of solution is required?

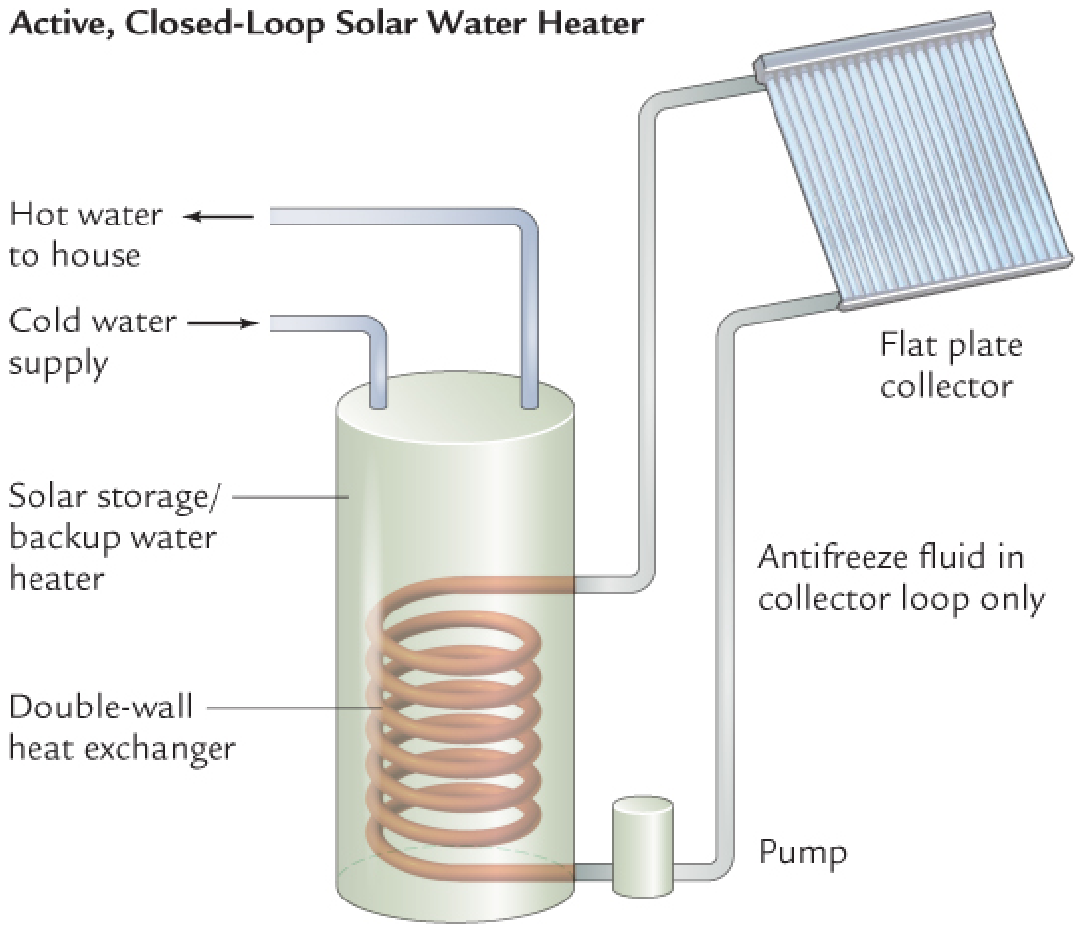


Figure 5 Schematic of a solar hot water system. *(Source: Information from U.S. Dept. of Energy)*

Solution Let x denote the required number of liters of the 18% solution, and y the required number of liters of the 50% solution. Since the system requires a total of 300 liters, it follows that

$$x+y=300$$

From Table 1 we see that we require a mixture that is 36% glycol. Thus the total amount of glycol in the system must be (0.36)(300) = 108 liters. We will get $0.18x$ liters of glycol from the 18% solution and 0.50*y* liters of glycol from the 50% solution. This leads to a second equation

$$0.18x+0.50y=108$$

Thus to find *x* and *y* we need to solve the linear system

$$x+y=3000.18x+0.50y=108 \tag{11}$$

We eliminate *x* by multiplying the first equation by −0.18 and then adding it to the second equation,

$$-0.18x-0.18y=-54+(0.18x+0.50y=108)\Rightarrow \quad 0.32y=54$$

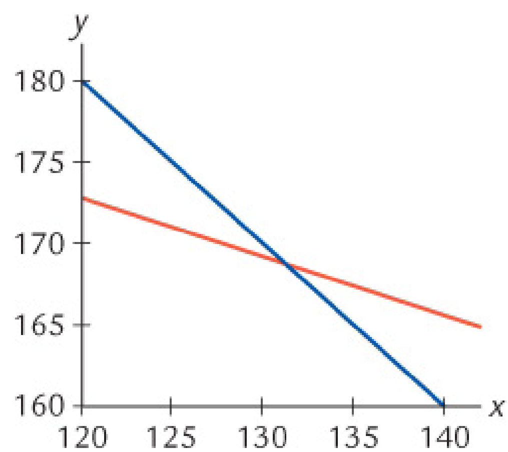


Figure 6 Graphs of $x + y = 300$ (blue) and $0.18x + 0.50y = 108$ (red) from Example 4.

Thus $y = 54/0.32 = 168.75$. Substituting this back into the top equation in (11) gives

$$x+168.75=300$$

which simplifies to $x = 131.25$. Therefore a combination of 131.25 liters of the 18% solution and 168.75 liters of the 50% solution should be used in the solar system.

Finding Solutions: Triangular Systems

Here we begin developing a method for finding the solutions for a linear system. For the remainder of this section we concentrate on special types of systems, and will take on general systems in the next section.

Consider the two systems below. Although not obvious, these systems have exactly the same solution set.

$$-2x1+4x2+11x3-4x4=4\ x1-2x2-5x3+3x4=23x1-6x2-15x3+10x4=11\ x2+3x3-4x4=72x1-4x2-10x3+6x4=4\ x3+2x4=-4-3x1+7x2+18x2-13\ x4=1\ x4=5$$

The one on the right looks easier to solve, so let’s find its solutions.

Example 5

Find all solutions to the system of linear equations

$$x1-2x2-5x3+3x4=2x2+3x3-4x4=7x3+2x4=-4x4=5 \qquad (12)$$

Back Substitution

Solution The method that we use here is called **back substitution**. Looking at the system, we see that the easiest place to start is at the bottom. Since $x_4 = 5$, substituting this back (hence the name for the method) into the next equation up gives us

$$x3+2(5)=-4$$

which simplifies to $x_3 = -14$. Now we know the values of x_3 and x_4. Substituting these back into the next equation up (second from the top) gives

$$x2+3(-14)-4(5)=7$$

so that $x_2 = 69$. Finally, we substitute the values of x_2, x_3, and x_4, back into the top equation to get

$$x_1 - 2(69) - 5(-14) + 3(5) = 2$$

which simplifies to $x_1 = 55$. Thus this system of linear equations has one solution,

$$x_1 = 55, x_2 = 69, x_3 = -14, x_4 = 5$$

Leading Variable

In Example 5, each variable x_1, x_2, x_3, and x_4 appears as the first term of an equation. In a system of linear equations, a variable that appears as the first term in at least one equation is called a **leading variable**. Thus in Example 5 each of x_1, x_2, x_3, and x_4 is a leading variable. In the system

$$\begin{aligned} -4x_1 + 2x_2 - x_3 + 3x_5 &= 7 \\ -3x_4 + 4x_5 &= -7 \\ x_4 - 2x_5 &= 1 \\ 7x_5 &= 2 \end{aligned} \tag{13}$$

x_1, x_4, and x_5 are leading variables, while x_2 and x_3 are not.

A key reason why the system in Example 5 is easy to solve is that every variable is a leading variable in exactly one equation. This feature is useful because as we back substitute from the bottom equation upward, at each step we are working with an equation that has only one unknown variable.

Triangular Form, Triangular System

The system in Example 5 is said to be in *triangular form,* with the name suggested by the triangular shape of the left side of the system. In general, a linear system is in **triangular form** (and is said to be a **triangular system**) if it has the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\ \ddots \qquad \vdots \quad &\ \ \vdots \\ a_{nn}x_n &= b_n \end{aligned}$$

where a_{11}, a_{22}, … , a_{nn} are all nonzero. It is straightforward to verify that triangular systems have the following properties.

PROPERTIES OF TRIANGULAR SYSTEMS

(a) Every variable is the leading variable of exactly one equation.

(b) There are the same number of equations as variables.

(c) There is exactly one solution.

Example 6

A bowling ball dropped off the Golden Gate bridge has height *H* (in meters) above the water at time *t* (in seconds) given by $H(t) = at^2 + bt + c$, where *a*, *b*, and *c* are constants. From calculus it follows that the velocity is $V(t) = 2at + b$ and the acceleration is $A(t) = 2a$. At $t = 2$, it is known that the ball has height 47.4 m, velocity −19.6 m/s, and acceleration $-9.8\ \text{m/s}^2$. (The velocity and acceleration are negative because the ball is moving downward.) What is the height of the bridge and when does the ball hit the water?

▶ Our model ignores forces other than gravity.

Solution We need to find the values of a, b, and *c* in order to answer these questions. At time $t = 2$, we have

$$47.4 = H(2) = 4a + 2b + c, \quad -19.6 = V(2) = 4a + b, \quad -9.8 = A(2) = 2a$$

This gives us the linear system

$$\begin{aligned} 4a + 2b + c &= 47.4 \\ 4a + b \phantom{{}+c} &= -19.6 \\ 2a \phantom{{}+b+c} &= -9.8 \end{aligned} \tag{14}$$

Back substituting as usual, we find that

$$a=-4.9, b=0, c=67$$

so that the height function is $H(t) = -4.9t^2 + 67$. At time $t = 0$ the ball is just starting its descent, so the bridge has height $H(0) = 67$ meters. The ball strikes the water when $H(t) = 0$, which leads to the equation

$$-4.9t^2+67=0$$

The solution is $t64/4.9\approx3.7$ seconds after the ball is released.

Figure 7 Golden Gate Bridge. *John Holt*

Finding Solutions: Echelon Systems

In the next example, we consider a linear system where each variable is a leading variable for *at most* one equation. Although this system is not quite triangular, it is close enough that the solutions can be found using back substitution.

Example 7

Find all solutions to the system of linear equations

$$\begin{aligned} 2x_1-4x_2+2x_3+\ x_4&=11 \\ x_2-\ x_3+2x_4&=\ 5 \\ 3x_4&=\ 9 \end{aligned} \qquad (15)$$

Solution We find the solutions by back substituting, similar to a triangular system. Starting with the bottom equation yields $x_4 = 3$.

The middle equation has x_2 as the leading variable, but we do not yet have a value for x_3. We address this by setting $x_3 = s_1$, where s_1 is a free parameter. We now have values for both x_3 and x_4, which we substitute into the middle equation, giving us

$$x_2 - s_1 + 2(3) = 5$$

Thus $x_2 = -1 + s_1$. Substituting our values for x_2, x_3, and x_4 into the top equation, we have

$$2x_1 - 4(-1 + s_1) + 2s_1 + 3 = 11$$

which simplifies to $x_1 = 2 + s_1$. Therefore the general solution is

$$\begin{aligned} x_1 &= 2 + s_1 \\ x_2 &= -1 + s_1 \\ x_3 &= s_1 \\ x_4 &= 3 \end{aligned}$$

where s_1 can be any real number. Note that each distinct choice for s_1 gives a new solution, so the system has infinitely many solutions.

DEFINITION 1.2

Echelon Form, Echelon System

A linear system is in **echelon form** (and is called an **echelon system**) if the system is organized in a descending "stair step" pattern from left to right, so that the indices of the leading variables are strictly increasing from top to bottom. Equations without variables of the form $0 = c$ are at the bottom, with those where $c \neq 0$ above those where $c = 0$.

All triangular systems are in echelon form.

For example, the system (15) is in echelon form but the system (13) is not, because x_4 is the leading variable of two equations.

PROPERTIES OF ECHELON SYSTEMS

(a) Every variable is the leading variable of at *most* one equation.

(b) There are no solutions, exactly one solution, or infinitely many solutions.

Free Variable

For a system in echelon form, any variable that is not a leading variable is called a **free variable**. For instance, x_3 is a free variable in Example 7.

▶ For a system in echelon form, the total number of variables is equal to the number of leading variables plus the number of free variables.

To find the general solution to a system in echelon form, we use the following two-step procedure.

1. Set each free variable equal to a free parameter.
2. Back substitute to solve for the leading variables.

Example 8

Find all solutions to the system of linear equations

$$x_1+2x_2-x_3+3x_5=7 \quad x_2-4x_3+x_5=-2 \quad x_4-2x_5=1 \qquad (16)$$

Solution In this system x_3 and x_5 are free variables, so we set each equal to a free parameter

$$x_3=s_1 \text{ and } x_5=s_2$$

It remains to determine the values of the leading variables. Substituting x_5 into the bottom equation, we have

$$x_4 - 2s_2 = 1$$

so that $x_4 = 1 + 2s_2$. Substituting our values for x_3 and x_5 into the next equation up gives

$$x_2 - 4s_1 + s_2 = -2$$

so that $x_2 = -2 + 4s_1 - s_2$. Finally, substituting in for x_2, x_3, and x_5 in the top equation, we have

$$x_1 + 2(-2 + 4s_1 - s_2) - s_1 + 3s_2 = 7$$

Hence $x_1 = 11 - 7s_1 - s_2$. Therefore the general solution is

$$\begin{aligned} x_1 &= 11 - 7s_1 - s_2 \\ x_2 &= -2 + 4s_1 - s_2 \\ x_3 &= s_1 \\ x_4 &= 1 + 2s_2 \\ x_5 &= s_2 \end{aligned}$$

where s_1 and s_2 can be any real numbers.

Example 9

Find all solutions to the system of linear equations

$$\begin{aligned} x_1 - 4x_2 + x_3 + 5x_4 - x_5 &= 0 \\ -x_3 + 4x_4 + 3x_5 &= 8 \end{aligned} \tag{17}$$

Solution We see that x_2, x_4, and x_5 are free variables, so we set $x_2 = s_1$, $x_4 = s_2$, and $x_5 = s_3$, where s_1, s_2, and s_3 are free parameters.

Starting with the bottom equation, we substitute in our values for x_4 and x_5, yielding the equation

$$-x_3 + 4s_2 + 3s_3 = 8$$

so that $x_3 = -8 + 4s_2 + 3s_3$. Back substituting into the top equation gives us

$$x_1 - 4s_1 + (-8 + 4s_2 + 3s_3) + 5s_2 - s_3 = 0$$

which simplifies to $x_1 = 8 + 4s_1 - 9s_2 - 2s_3$. Therefore the general solution is

$$\begin{aligned} x_1 &= 8 + 4s_1 - 9s_2 - 2s_3 \\ x_2 &= s_1 \\ x_3 &= -8 + 4s_2 + 3s_3 \\ x_4 &= s_2 \\ x_5 &= s_3 \end{aligned}$$

where s_1, s_2, and s_3 can be any real numbers.

Echelon Systems: Equations with No Variables

The next two examples consider echelon systems that include an equation of the form $0 = c$. We'll see in the next section how this kind of echelon system can arise.

Example 10

Find all solutions to the system of linear equations

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

Solution The equation $0 = 3$ is not satisfied for any choices of x_1, x_2, and x_3, so $0 = 3$ has no solutions. Since the last equation has no solutions, the entire system has no solutions.

Example 11 looks similar to Example 10, but the set of solutions is very different.

Example 11

Find all solutions to the system of linear equations

$$x1+2x2-5x3=-4-2x2+4x3=\ \ 80=\ \ 0$$

Solution The equation 0 = 0 is satisfied for every combination of x_1, x_2, and x_3, so it imposes no constraints on the solutions for the system. Thus we can ignore the bottom equation, and find the solutions to the system

$$x1+2x2-5x3=-4-2x2+4x3=8$$

The only free variable is $x_3 = s_1$. Solving the last equation for x_2 and back substituting, we have

$$x2=-4+2x3=-4+2s1$$

Moving up to the top equation and solving for x_1 gives

$$x1=-4-2x2+5x3=-4-2(-4+2s1)+5s1=4+s1$$

Thus the general solution is

$$x1=4+s1x2=-4+2s1x3=s1$$

where s_1 can be any real number.

To sum up, if an echelon system includes an equation $0 = c$, where c is a nonzero constant (as in Example 10), then the system has no solutions. Otherwise, there are two possibilities:

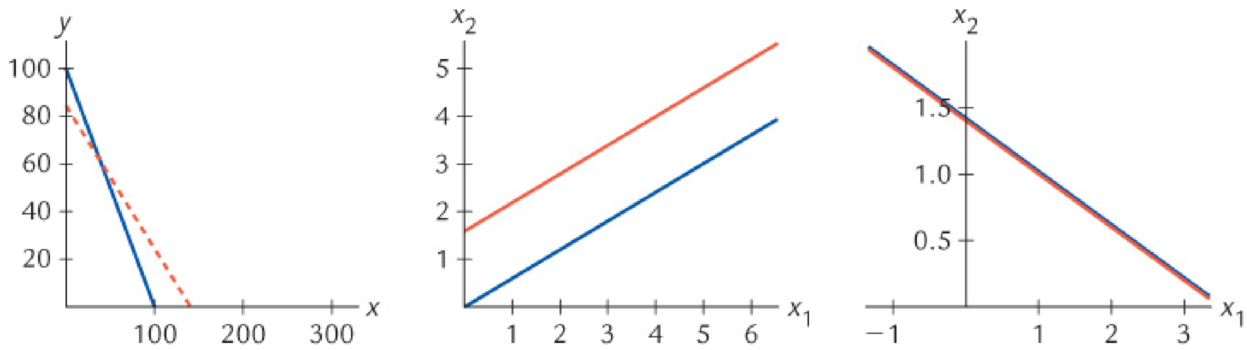


Figure 8 Graphs of equations in Examples 1–3.

1. If the system has no free variables, then there is exactly one solution.
2. If the system has at least one free variable, then the general solution has free parameters and there are infinitely many solutions.

Number of Solutions: Geometry

In Examples 1–3 we saw that a linear system can have a single solution, no solutions, or infinitely many solutions.

Figure 8 shows that our examples illustrate all possibilities for two lines: They can intersect in exactly one point, they can be parallel and have no points in common, or they can coincide and have infinitely many points of intersection. Thus it follows that a system of two linear equations with two variables can have zero, one, or infinitely many solutions.

Now consider systems of linear equations with three variables. Recall that the graph of the solutions for each equation is a plane. To explore the solutions such a system can have, you can experiment by using a few pieces of cardboard to represent planes.

▶ A **theorem** is a mathematical statement that has been rigorously proved to be true. As we progress through this book, theorems will serve to organize our expanding body of linear algebra knowledge. The appendix "Reading and writing proofs" provides an introduction for how to prove theorems.

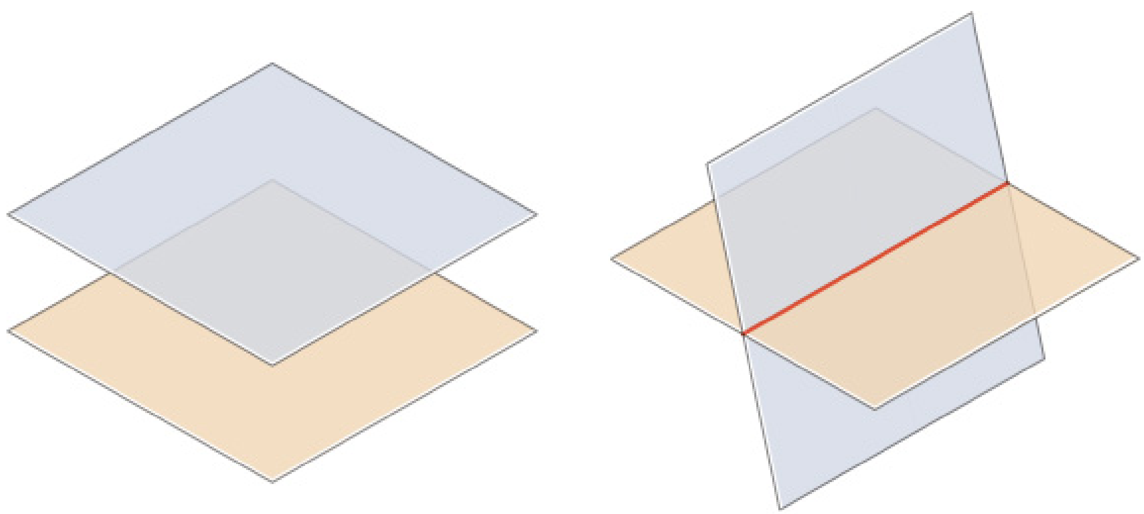
Starting with two pieces, you will quickly discover that the only two possibilities for the number of points of intersection is either none or infinitely many. (See Figure 9.) This geometric observation is equivalent to the algebraic statement that a system of two linear equations in three variables has either no solutions or infinitely many solutions.

Now try three pieces of cardboard. More configurations are possible, with some shown in Figure 10. This time, we see that the number of points of intersection can be zero, one, or infinitely many. (Note that this also held for a pair of lines.) In fact, this turns out to be true in general, not only for planes but also for solution sets in higher dimensions. The equivalent statement for systems of linear equations is contained in Theorem 1.3.

THEOREM 1.3 ▶

A system of linear equations has no solutions, exactly one solution, or infinitely many solutions.

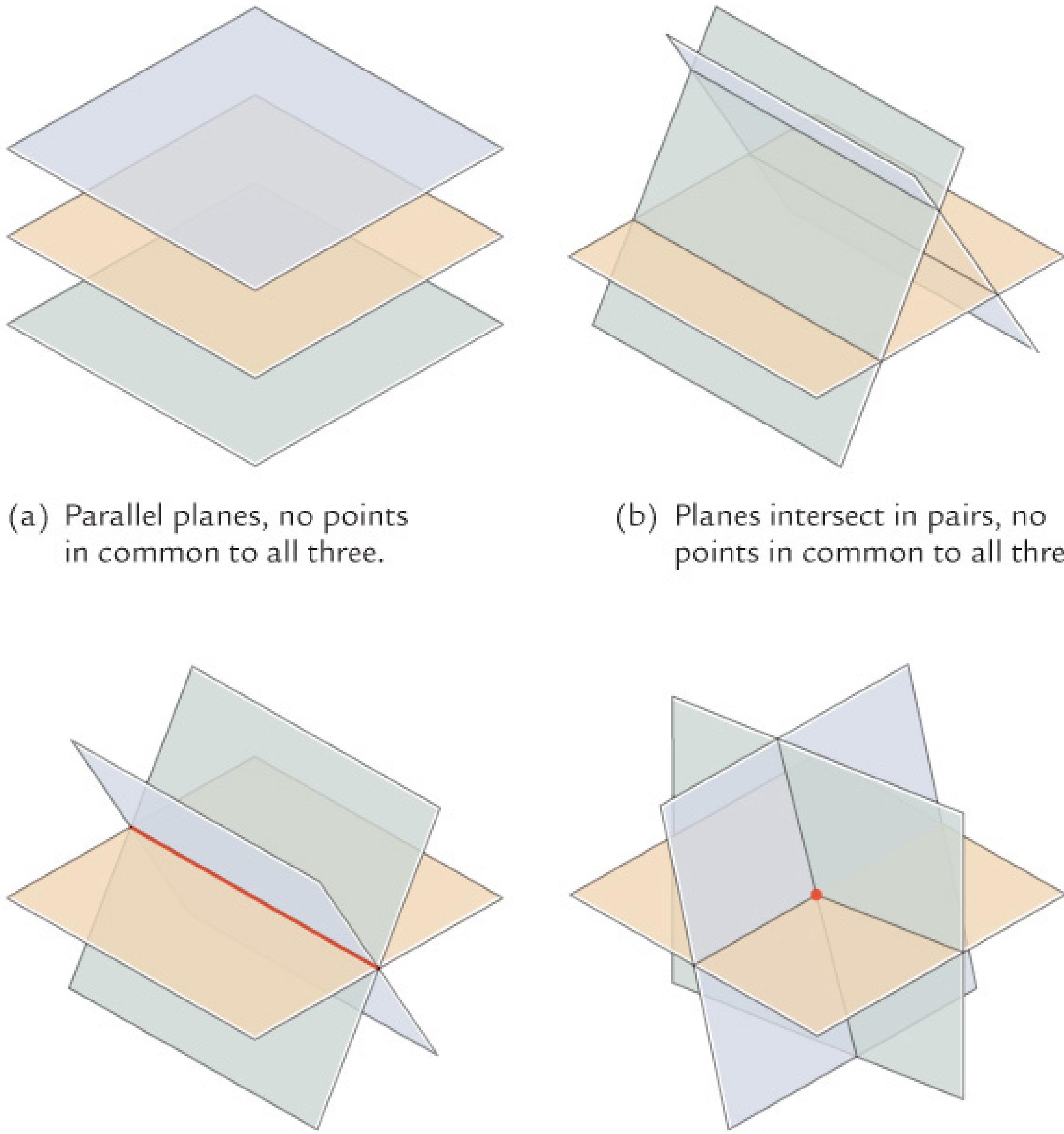
We will prove this theorem at the end of the next section.



(a) Parallel planes, no points in common.

(b) Planes intersect, infinitely many points in common.

Figure 9 Graphs of systems of two equations with three variables.



(a) Parallel planes, no points in common to all three.

(b) Planes intersect in pairs, no points in common to all three.

(c) Planes intersect in a line, infinitely many points in common.

(d) Planes intersect at a point, unique point in common.

Figure 10 Graphs of systems of three equations with three variables.

▶ Practice problems can also be used as additional examples.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the solutions to each linear system.
   (a) $-2x_1 + 8x_2 = 5 \quad 3x_1 - 12x_2 = 4$
   (b) $x_1 - 2x_2 = 3 - 3x_1 + 6x_2 = -9$
2. Find the solutions to each linear system.
   (a) $2x_1 + x_3 = 5 - x_2 + x_4 = 2 \quad x_3 - 2x_4 = 1 \quad 0 = -2$
   (b) $x_1 - x_2 - 2x_3 + x_4 - 2x_5 = 1 \quad x_3 - 2x_4 + x_5 = 2 \quad 3x_5 = 12$
3. Determine if each statement is true or false, and justify your answer.
   (a) A triangular system can have infinitely many solutions.
   (b) An echelon system with five equations (all with variables) and eight variables must have three free variables.
   (c) A linear system with two variables and three equations cannot have one solution.
   (d) Every variable in an echelon system is either a free variable or a free parameter.
4. Suppose that an echelon system has four equations (all with variables) and nine variables.
   (a) How many leading variables are there?
   (b) How many free variables are there?
   (c) How many free parameters are in the solution?
   (d) How many solutions are there?
5. A play is sold out in a theater with a capacity of 280. Tickets cost \$22 for the floor and \$14 for the balcony. The total ticket sales is \$5320. How many of each type of ticket was sold? (You may want to use a calculator for the computations in this problem.)
6. Consider a pile of nickels, dimes, and quarters. The quarters are worth \$2.75, and the dimes and quarters together are worth \$3.65. If there are a total of 31 coins, how many of each type of coin are there?

EXERCISES

In each exercise set, problems marked with C are designed to be solved using a programmable calculator or computer algebra system.

1. Determine which of the points (1, −2), (− 3, −3), and (− 2, − 3) lie on the line $2x_1 - 5x_2 = 9$.

2. Determine which of the points (1, − 2, 0), (4, 2, 1), and (2, −5, 1) lie in the plane $x_1 - 3x_2 + 4x_3 = 7$.

3. Determine which of the points (− 1, 2), (−2, 5), and (1, −5) lie on both the lines $3x_1 + x_2 = -1$ and $-5x_1 + 2x_2 = 20$.

4. Determine which of the points (3, 1), (2, −4), and (−4, 5) lie on both the lines $2x_1 - 5x_2 = 1$ and $-4x_1 + 10x_2 = -2$.

5. Determine which of the points (1, 2, 3), (1, − 1, 1), and (−1, −2, −6) satisfy the linear system

−2x1+9x2− x3=−10 x1−5x2+2x3= 4

6. Determine which of the points (1, −2, −1, 3), (− 1, 0, 2, 1), and (−2, −1, 4, −3) satisfy the linear system

3x1− x2+2x3 = 12x1+3x2 −x4=−3

Exercises 7–8: Determine which of (a)–(d) form a solution to the given system for any choice of the free parameter(s). (HINT: All parameters of a solution must cancel completely when substituted into each equation.)

7. $-2x_1 + 3x_2 + 2x_3 = 6$ (This system has one equation.)

(a) $(-3 + s_1 + s_2, s_1, s_2)$

(b) $(-3 + 3s_1 + s_2, 2s_1, s_2)$

(c) $(3s_1 + s_2, 2s_1 + 2, s_2)$

(d) $(s_1, s_2, 3 - 3s_2/2 + s_1)$

8. 3x1+8x2−14x3=6x1+3x2− 4x3=1

(a) $(5 - 2s_1, 7 + 3s_1, s_1)$

(b) $(-5 - 5s_1, s_1, -(3 + s_1)/2)$

(c) $(10 + 10s_1, -3 - 2s_1, s_1)$

(d) $((6 - 4s_1)/3, s_1, -(5 - s_1)/4)$

Exercises 9–14: Find all solutions to the given system by eliminating one of the variables.

9. 3x1+5x2=42x1−7x2=13

10. −3x1+ 2x2=15x1+ x2=−4

11. −10x1+4x2=215x1−6x2=−3

12. −3x1+ 4x2=09x1− 12x2=−2

13. 7x1−3x2=−1−5x1+8x2=0

14. 6x1−3x2=5−8x1+4x2=1

Exercises 15–22: Determine if the linear system is in echelon form. If so, identify the leading variables and the free variables. If not, explain why not.

15. x1− x2=7 7x2=0

16. 6x1−5x2=12−2x1+7x2= 0

17. −7x1−x2+2x3=116x3=−1

18. 3x1+2x2+7x3=0−3x3=−3− x2+4x3=13

19. 4x1+ 3x2−9x3+2x4=3 6x2+ x3 =−2 −5x2−8x3+ x4=−4

20. 2x1 +2x3 =12 12x2 − 5x4=−19 3x3+11x4=14 −x4=3

21. −2x1−3x2+ x3−13x4= 2 2x3 =−7

22. −7x1+3x2 +8x4−2x5+13x6=−6 −5x3− x4+6x5+ 3x6=0 2x 4+5x5 =0

Exercises 23–30: Find the set of solutions for the linear system.

23. −5x1−3x2= 42x2=10

24. x1+4x2−7x3=−3− x2+4x3=13x3=−9

25. $-3x_1 + 4x_2 = 2$ (This system has one equation.)

26. 3x1−2x2+ x3=4 −6x3=−12

27. x1+5x2−2x3 =0 −2x2+ x3−x4=−1 x4=5

28. $2x_1 - x_2 + 6x_3 = -3$ (This system has one equation.)

29. −2x1+x2+2x3 =1 −3x3+x4=−4

30. −7x1+3x2 +8x4−2x5+13x6=−6 −5x3− x4+6x5+ 3x6=0 2x 4+5x5 =1

Exercises 31–32: Each linear system is not in echelon form, but can be put in echelon form by reordering the equations. Write the system in echelon form, and then find the set of solutions.

31.

(a) $-5x_2=43x_1+2x_2=1$

(b) $-3x_3=-3-x_2-4x_3=133x_1+2x_2+7x_3=0$

32.

(a) $2x_2+\ x_3-5x_4=0x_1+3x_2-2x_3+2x_4=-1$

(b) $x_2-4x_3+3x_4=\ 2x_1-5x_2-6x_3+3x_4=\ 3-3x_4=155x_3-4x_4=10$

Exercises 33–36: Find the set of solutions to the linear system.

33. $x_1+2x_2-\ x_3+\ x_4=1 \quad x_2+2x_3-2x_4=2 \qquad x_4=0$

34. $x_1-x_2+2x_3=-2x_2+3x_3=\ 10=\ 1$

35. $x_1+x_2-x_3=4x_2+x_3=10=0$

36. $x_1+3x_2+x_3-x_4+\ \ x_5=2x_3+x_4-2x_5=-30=-5$

37. Find value(s) of *k* so that the linear system is consistent.

(a) $6x_1-5x_3=49x_1+kx_2=-1$

(b) $6x_1-\ 8x_2=k-9x_1+12x_2=-1$

38. Find values of *h* and *k* so that the linear system has no solutions.

(a) $2x_1+5x_2=-1hx_1+5x_2=k$

(b) $2x_1+5x_2=-1hx_1+kx_2=3$

Exercises 39–42: Answer the given question, including a justification. For each, assume that all equations have variables.

39. A linear system is in echelon form. If there are four free variables and five leading variables, how many variables are there?

40. Suppose that a linear system with five equations and eight unknowns is in echelon form. How many free variables are there?

41. Suppose that a linear system with seven equations and thirteen unknowns is in echelon form. How many leading variables are there?

42. A linear system is in echelon form. There are a total of nine variables, of which four are free variables. How many equations does the system have?

FIND AN EXAMPLE Exercises 43–50: Find an example that meets the given specifications.

43. A linear system with three equations and three variables that has exactly one solution.

44. A linear system with three equations and three variables that has infinitely many solutions.

45. A linear system with four equations and three variables that has infinitely many solutions.

46. A linear system with three equations and four variables that has no solutions.

47. Come up with an application that has a solution found by solving an echelon linear system. Then solve the system to find the solution.

48. A linear system with two equations and two variables that has $x_1 = -1$ and $x_2 = 3$ as the only solution.

49. A linear system with two equations and three variables that has solutions $x_1 = 1$, $x_2 = 4$, $x_3 = -1$ and $x_1 = 2$, $x_2 = 5$, $x_3 = 2$.

50. A linear system with two equations and two variables that has the line $x_1 = 2x_2$ for solutions.

TRUE OR FALSE Exercises 51–56: Determine if the statement is true or false, and justify your answer.

51.

(a) A linear system with three equations and two variables must be inconsistent.

(b) A linear system with three equations and five variables must be consistent.

52.

(a) There is only one way to express the general solution for a linear system.

(b) A triangular system always has exactly one solution.

53.

(a) All triangular systems are in echelon form.

(b) All systems in echelon form are also triangular systems.

54.

(a) A system in echelon form can be inconsistent.

(b) A system in echelon form can have more equations than variables.

55.

(a) If a triangular system has integer coefficients (including the constant terms), then the solution consists of rational numbers.

(b) A system in echelon form can have more variables than equations.

56.

(a) If a general solution has free parameters, then there must be infinitely many solutions.

(b) The pairs (1, 2), (4, 8), and (−1, 5) can all be solutions to a linear system with two equations and two variables. (Assume both equations have variables.)

57. A total of 385 people attend the premiere of a new movie. Ticket prices are $11 for adults and $8 for children. If the total revenue is $3974, how many adults and children attended?

58. A plane holds 150 passengers. Tickets for a flight are $160 for coach and $220 for business class. Suppose that total ticket sales for a full flight is $24,960. How many of each type of ticket were sold?

59. **Calculus required** Suppose that $f(x) = a_1e^{2x} + a_2e^{-3x}$ is a solution to a differential equation. If we know that $f(0) = 5$ and $f'(0) = -1$ (these are the *initial conditions),* what are the values of a_1 and a_2? (Hint: $f'(x) = 2a_1e^{2x} - 3a_2e^{-3x}$.)

60. **Calculus required** Suppose that $f(x) = a_1e^{-5x} + a_2e^{2x}$ is a solution to a differential equation. If we know that $f(0) = 3$ and $f'(0) = -1$ (these are the *initial conditions),* what are the values of a_1 and a_2? (Hint: $f'(x) = -5a_1e^{-5x} + 2a_2e^{2x}$.)

61. Referring to Example 4, suppose that the minimum outside temperature is 10°F. In this case, how much of each type of solution is required?

62. Referring to Example 4, suppose that the minimum outside temperature is − 20°F. In this case, how much of each type of solution is required?

63. An investor has $100,000 and can invest in any combination of two types of bonds, one that is safe and pays 3% annually, and one that carries risk and pays 9% annually. The investor wants to keep risk as low as possible while realizing a 7% annual return. How much should be invested in each type of bond?

64. An investor has $200,000 and can invest in any combination of two types of bonds, one that is safe and pays 4% annually, and one that carries risk and pays 11% annually. The investor wants to keep risk as low as possible while realizing an 8% annual return. How much should be invested in each type of bond?

65. A 60-gallon bathtub is to be filled with water that is exactly 100° F. The hot water supply is 125° F and the cold water supply is 60° F. When mixed, the temperature will be a weighted average based on the amount of each water source in the mix. How much of each should be used to fill the tub as specified?

66. A 50-gallon bathtub is to be filled with water that is exactly 105° F. The hot water supply is 115° F and the cold water supply is 70° F. When mixed, the temperature will be a weighted average based on the amount of each water source in the mix. How much of each should be used to fill the tub as specified?

67. Degrees Fahrenheit (F) and Celsius (C) are related by a linear equation $C = aF + b$. Pure water freezes at 32° F and 0° C, and boils at 212° F and 100° C. Use this information to find a and b.

68. For tax and accounting purposes, corporations depreciate the value of equipment each year. One method used is called “linear depreciation,’’ where the value decreases over time in a linear manner. Suppose that two years after purchase, an industrial milling machine is worth $800,000, and five years after purchase, the machine is worth $440,000. Find a formula for the machine value at time $t \geq 0$ after purchase.

69. This problem requires 8 nickels, 8 quarters, and a sheet of 8.5-by-11-inch paper. The goal is to estimate the diameter of each type of coin as follows: Using trial and error, find a combination of nickels and quarters that, when placed side by side, extend the height (long side) of the paper. Then do the same along the width (short side) of the paper. Use the information obtained to write two linear equations involving the unknown diameters of each type of coin, then solve the resulting system to find the diameter for each type of coin.

Bixby Creek Bridge. *Dennis Frates/Alamy*

70. The Bixby Creek Bridge is located along California’s Big Sur coast and has been featured in numerous television commercials. Suppose that a bag of concrete is projected downward from the bridge deck at an initial rate of 5 meters per second. After 3 seconds, the bag is 25.9 meters from the Bixby Creek, has a velocity of −34.4 m/s, and has an acceleration of $-9.8\ \text{m/s}^2$. Use the model in Example 6 to find a formula for $H(t)$, the height at time t.

Exercises 71–78: Use computational assistance to find the set of solutions to the linear system.

71. $-4x_1+7x_2=-13 \quad 3x_1-5x_2=11$

72. $3x_1+5x_2=0 \quad -7x_1-2x_2=-2$

73. $2x_1-5x_2+3x_3=10 \quad 4x_2-9x_3=-7$

74. $-x_1+4x_2+7x_3=6 \quad -3x_2=1$

75. $-2x_1-x_2+5x_3+x_4=20 \quad 3x_2+6x_4=13 \quad -4x_3+7x_4=-6$

76. $3x_1+5x_2-x_3-x_4=17 \quad -x_2-6x_3+11x_4=5 \quad 2x_3+x_4=11$

77. x1+13x2−8x3+9x4−15x5=−124x2− x3+8x4+ 2x5=02x3+ x4+1 4x5=19−11x5=−2

78. 3x1−x2+2x3 − 5x5−3x6=−1 4x2 +x4+13x5+x6=3 6x4−8x5 +4x6=2 2x5−5x6= 11 x5−x6= 3

1.2 Linear Systems and Matrices

Systems of linear equations arise naturally in many applications, but the systems rarely come in echelon form. In this section, we develop a method for converting any linear system into a system in echelon form, so that we can apply back substitution.

To get us started, consider the following projectile motion problem. Suppose that a cannon sits on a hill and fires a ball across a flat field below. The path of the ball is known to be approximately parabolic and so can be modeled by a quadratic function $E(x) = ax^2 + bx + c$, where E is the elevation (in feet) over position x, and a, b, and c are constants.

Figure 1 shows the elevation of the ball at three separate places. Since every point on its path is given by $(x, E(x))$, the data can be converted into three linear equations

$$\begin{aligned} 100a+10b+c&=117 \\ 900a+30b+c&=171 \\ 2500a+50b+c&=145 \end{aligned} \tag{1}$$

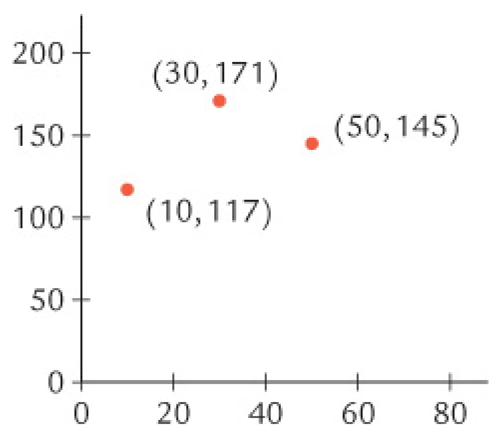


Figure 1 Positions and elevations $(x, E(x))$ of an airborne cannonball.

This system is not in echelon form, so back substitution is not easy to use here. We will return to this system shortly.

The primary goal of this section is to develop a systematic procedure for transforming *any* linear system into a system that is in echelon form. The key feature of our transformation procedure is that it produces a new linear system that is in echelon form (hence solvable using back substitution) and has exactly the same set of solutions as the original system.

▶ Two linear systems are said to be **equivalent** if they have the same set of solutions.

Elementary Operations

Elementary Operations

We can transform a linear system using a sequence of **elementary operations**. Each operation produces a new system that is equivalent to the old one, so the solution set is unchanged. There are three types of elementary operations.

1. Interchange the position of two equations.

This amounts to nothing more than rewriting the system of equations. For example, we exchange the places of the first and second equations in the following system.

3x1−5x2−8x3=−4 x1+2x2−4x3=5 x1+2x2−4x3=5∼ 3x1−5x2−8x 3=−4−2x1+6x2+ x3=3 −2x1+6x2+ x3=3

▶ The symbol ~ indicates the transformation from one linear system to an equivalent linear system.

2. Multiply an equation by a nonzero constant.

For example, here we multiply the third equation by −2.

x1+ 2x2−4x3=5 x1+ 2x2−4x3=53x1−5x2−8x3=−4∼3x1− 5x2−8x3=−4−2x1+6x2+ x3=3 4x1−12x2−2x3=−6

▶ Proving that each elementary operation produces an equivalent linear system is left as Exercise 56.

3. Add a multiple of one equation to another.

For this operation, we multiply one of the equations by a constant and then add it to another equation, replacing the latter with the result. For example, below we multiply the top equation by −4 and add it to the bottom equation, replacing the bottom equation with the result.

x1+ 2x2−4x3=5 x1+ 2x2− 4x3= 53x1− 5x2−8x3=−4∼3x1− 5x2− 8x3= −44x1−12x2−2x3=−6 −20x2+14x3=−26

The third operation may look familiar. It is similar to the method used in the first three examples of Section 1.1 to eliminate a variable. Note that this is exactly what happened here, with the lower left coefficient becoming zero, transforming the system closer to echelon form. This illustrates a single step of our basic strategy for transforming any linear system into a system that is in echelon form.

Example 1

Find the set of solutions to the system of linear equations

x1− 3x2+2x3=−12x1− 5x2− x3=2−4x1+13x2−12x3= 11

Solution We begin by focusing on the variable x_1, in each equation. Our goal is to transform the system to echelon form, so we want to eliminate the x_1 terms in the second and third equations. This will leave x_1, as the leading variable in only the top equation.

▶ Going forward we identify coefficients using the notation for a generic system of equations introduced in Section 1.1,

a11x1+a12x2+⋯+a1nxn=b1a21x1+a22x2+⋯+a2nxn=b2a31x1+a32x2+⋯+a3nxn=b3⋮ ⋮ ⋮am1x1+am2x2+⋯+amnxn=bm

- **Add a multiple of one equation to another** (focus on x_1).

We need to transform a_{21} and a_{31} to 0. We do this in two parts. Since a_{21} = 2, if we take −2 times the first equation and add it to the second, then the resulting coefficient on x_1, will be $(-2) \cdot 1 + 2 = 0$, which is what we want.

x1−3x2+2x3=−1 x1−3x2+2x3=−12x1−5x2−x3=2∼ x2−5x3=4−4x1+13x2−12x3=11 −4x1+13x2−12x3=11

The second part is similar. This time, since $(4) \cdot 1 - 4 = 0$, we multiply 4 times the first equation and add it to the third.

x1−3x2+ 2x3=−1 x1−3x2+2x3=−1 x2− 5x3=4∼ x2−5x3=4−4x1+13x2−12x3= 11 x2−4x3=7

With these steps complete, the x_1 terms in the second and third equations are gone, exactly as we want.

Next, we focus on the x_2 coefficients. Since our goal is to reach echelon form, we do not care about the coefficient on x_2 in the top equation, so we concentrate on the second and third equations.

- **Add a multiple of one equation to another** (focus on x_2).

Here we need to transform a_{32} to 0. Since $(-1) \cdot 1 + 1 = 0$, we multiply − 1 times the second equation and add the result to the third equation.

x1−3x2+2x3=−1 x1−3x2+2x3=−1 x2−5x3=4∼ x2−5x3=4 x2−4x3=7 x3=3

▶ Using only the second and third equations avoids reintroducing x_1 into the third equation.

The system is now in echelon (indeed, triangular) form, and using back substitution we can easily show that the solution (we know there is only one) is $x_1 = 50$, $x_2 = 19$, and $x_3 = 3$. To check our solution, we plug these values into the original system.

1(50)−3(19)+2(3)=−12(50)−5(19)−1(3)=2−4(50)+13(19)−12(3)=11

Example 2

Find the set of solutions to the linear system (1) from the start of the section,

$$100a+10b+c=117900a+30b+c=1712500a+50b+c=145$$

Solution We follow the same procedure as in the previous example.

- **Add a multiple of one equation to another** (focus on *a*).

  We need to transform a_{21} and a_{31} to 0. Since $a_{21} = 900$, we multiply the first equation by − 9 and add it to the second, so that

$$100a+10b+c=117\ 100a\ +10b+\ c=117900a+30b+c=171\sim-60a-8c=-8822500a+50b+c=145\ 2500a\ +50b+\ c=145$$

The second part is similar. We multiply the first equation by −25 and add it to the third.

$$100a+10b+c=117\ 100a+\ 10b+c=117\ -60b-8c=-882\sim-60b-8c=-8822500a+50b+c=145\ \ -200b-24c=-2780$$

- **Multiply an equation by a nonzero constant** (focus on *b*).

  Here we multiply the third equation by −0.3, so that the coefficients on *b* match up (other than sign).

$$100a+\ 10b+c=117\ 100a+10b+\ c=117-60b-8c=-882\sim-60b-8c=-882\ -200b-24c=-2780\ \ 60b+7.2c=\ 834$$

- **Add a multiple of one equation to another** (focus on *b*).

  Thanks to the previous step, we need only add the second equation to the third to transform a_{32} to 0.

$$100a+10b+\ c=117\ 100a+10b+\ c=117\ -60b-8c=-882\sim-60b-8c=-882\ \ 60b+7.2c=834\ \ -0.8c=-48$$

The system is now in triangular form. Using back substitution, we can show that the solution is $a = -0.1$, $b = 6.7$, and $c = 60$, which gives us $E(x) = -0.1x^2 + 6.7x + 60$. Figure 2 shows a graph of the model together with the known points.

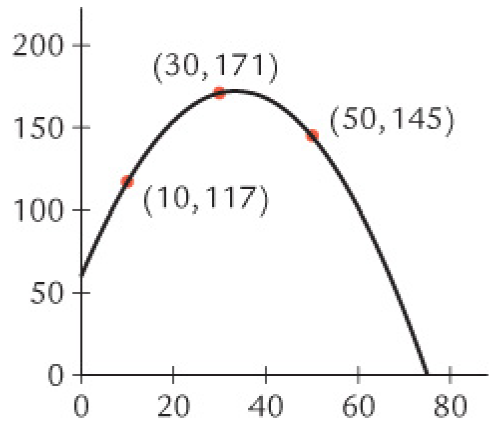


Figure 2 Cannonball data and the graph of the model.

The Augmented Matrix

Matrix

Augmented Matrix

When manipulating systems of equations, the coefficients change but the variables do not. We can simplify notation by transferring the coefficients to a **matrix**, which for the moment we can think of as a rectangular table of numbers. When a matrix contains all the coefficients of a linear system, including the constant terms on the right side of each equation, it is called an **augmented matrix**. For instance, the system in Example 1 is transferred to an augmented matrix by

Linear System

x1− 3x2+ 2x3=−12x1− 5x2−   x3=2−4x1+13x2−12x3=11

Augmented Matrix

[12−4−3−513   2 −1−12|−1211]

▶ Augmented matrices include a vertical line separating the left and right sides of the equations.

Elementary Row Operations

The three elementary operations that we performed on equations can be translated into equivalent **elementary row operations** for matrices.[1]

ELEMENTARY ROW OPERATIONS

1. Interchange two rows.
2. Multiply a row by a nonzero constant.
3. Replace a row with the sum of that row and the scalar multiple of another row.

Equivalent Matrices

Borrowing from the terminology for systems of equations, we say that two matrices are **equivalent** if one can be obtained from the other through a sequence of elementary row operations. Hence equivalent augmented matrices correspond to equivalent linear systems.

Zero Row, Zero Column

When discussing matrices, the rows are numbered from top to bottom, and the columns are numbered from left to right. A **zero row** is a row consisting entirely of zeros, and a **nonzero row** contains at least one nonzero entry. The terms **zero column** and **nonzero column** are similarly defined.

In the examples that follow, we transfer the system of equations to an augmented matrix, but our goal is the same as before, to find an equivalent system in echelon form.

Example 3

Find all solutions to the system of linear equations

2x1−3x2 + 10x3=−2x2− 2x2+ 3x3=−2−x1+ 3x3+ x3=4

Solution We begin by converting the system to an augmented matrix.

[21−1−3−2310 3 1|−2−24]

• **Interchange rows** (focus on column 1).

We focus on the first column of the matrix, which contains the coefficients of x_1. Although this step is not required, exchanging Row 1 and Row 2 will move a 1 into the upper left position and avoid the early introduction of fractions.

[21−1−3−231031|−2−24] ~ [12−1−2−333101|−2−24]

▶ As with linear systems, we use the symbol ~ to indicate that two matrices are equivalent.

▶ The compact notation for this operation is R1↔R2.

• **Add a multiple of one row to another** (focus on column 1).

To transform the system to echelon form, we need to introduce zeros in the first column below Row 1. This requires two operations. Focusing first on Row 2, since $(-2)(1) + 2 = 0$, we add −2 times Row 1 to Row 2 and replace Row 2 with the result.

[12−1−2−333101|−2−24] ~ [10−1−213341|−224]

▶ The compact notation for this operation is $-2R_1 + R_2 \to R_2$.

Focusing now on Row 3, since $(1)(1) + (-1) = 0$ we add 1 times Row 1 to Row 3 and replace Row 3 with the result.

$$[10-1-213341|-224] \sim [100-211344|-222]$$

▶ The compact notation for this operation is $R_2 + R_3 \to R_3$.

• **Add a multiple of one row to another** (focus on column 2).
With the first column complete, we move down to the second row and to the right to the second column. Since (−1)(1) + (1) = 0, we add −1 times Row 2 to Row 3 and replace Row 3 with the result.

$$[100-211344|-222] \sim [100-210340|-220]$$

▶ The compact notation for this operation is $-R_2 + R_3 \to R_3$.

We now extract the transformed system of equations from the matrix. The row of zeros indicates that one of the equations in the transformed system is 0 = 0. Since any choice of values for the variables will satisfy 0 = 0, this equation contributes no information about the solution set so can be ignored. The new equivalent system is therefore

$$x1-2x2+3x3=-2x2+4x3=2$$

Back substitution can be used to show that the general solution is

$$x1=2-11s1x2=2-4s1x3=s1$$

where s_1 can be any real number. We can substitute into the original system to verify our solution.

$$2(2-11s1)-3(2-4s1)+10s1=4-22s1-6+12s1+10s1=-2,(2-11s1)-2(2-4s1)+3s1=2-11s1-4+8s1+3s1=-2,-(2-11s1)+3(2-4s1)+s1=-2+11s1+6-12s1+s1=4$$

Gaussian Elimination

Gaussian Elimination, Echelon Form, Leading Term

The procedure that we used in Example 3 is known as **Gaussian elimination**. The resulting matrix is said to be in **echelon form** (or **row echelon form**) and will have the properties given in Definition 1.4 below. In the definition, the **leading term** of a row is the leftmost nonzero term in that row, and a row of all zeros has no leading term.

DEFINITION 1.4 ▶

A matrix is in **echelon form** if

(a) Every leading term is in a column to the left of the leading term of the row below it.

(b) Any zero rows are at the bottom of the matrix.

▶ Echelon form is the counterpart to echelon systems from Section 1.1.

Note that the first condition in the definition implies that a matrix in echelon form will have zeros filling out the column below each of the leading terms. Examples of matrices in echelon form are

[51−4 09 202−3−673100 0−24 9]and[−123−217 9 7 009−626 3−6 000 0 0−3 0 000 0 0 0 4 000 0 0 0 0] (2)

Pivot Positions, Pivot Columns, Pivot

For a matrix in echelon form, the **pivot positions** are those that contain a leading term. The entries in the pivot positions for the matrices in (2) are shown in boldface. The **pivot columns** are the columns that contain pivot positions, and a **pivot** is a nonzero number in a pivot position.

In what follows, it will be handy to have a general augmented matrix when referring to entries in specific positions. We adopt a notation similar to that for a general system of equations given in (7) of Section 1.1,

[a11a21a31⋮am1a12a22a32⋮am2a13a23a33⋮am3⋯⋯⋯ ⋯
a1na2na3n⋮anm|b1b2b3⋮bm]

Example 4

Use Gaussian elimination to find all solutions to the linear system

2x1−2x2−6x3+x4=3−x1+ x2+3x3−x4=−3x1−2x2 − x3+x4=2

▶ Gaussian elimination was originally discovered by Chinese mathematicians over 2000 years ago. It is named in honor of German mathematician Carl Friedrich Gauss, who independently discovered the method and introduced it to the West in the nineteenth century.

Solution The augmented matrix for this system is

[2−11−21−2−63−11−11|3−32]

- **Identify pivot position for Row 1.**

We see that $a_{11} \neq 0$ so this can serve as a pivot. However, the arithmetic will be easier if there is a one in that position, so let’s start by interchanging Row 1 and Row 3.

[2−11−21−2−63−11−11|3−32] ∼ [1−12−21−2−1361−11|2−33]

▶ The operation is R1↔R3.

- **Elimination.**

Next, we need zeroes down the first column below the pivot position. We arrange for $a_{21} = 0$ and $a_{31} = 0$ using the operations shown in the margin.

[1−12−21−2−13−61−11|2−33] ∼ [100−2−12−12410−1|2−1−1]

▶ The *elimination* steps are used to “eliminate” coefficients by transforming them to zero.

▶ The operations are
R1+R2→R2−R1+R3→R3

- **Identify pivot position for Row 2.**

Moving from a_{11} down one row and to the right one column, we find $a_{22} = -1$. This makes a good pivot, so we move to the elimination step.

- **Elimination.**

Down the remainder of the second column, we need only introduce a zero at a_{32} by using the operation shown in the margin.

[100−2−12−12−4101|2−11] ~ [100−2−10−12010−1|2−1−3]

▶ The operation is $2R_2 + R_3 \to R_3$

The matrix is now in echelon form. The corresponding echelon system is

x1−2x2−x3+x4=2 −x2+2x3 =−1   −x4=−3

Applying our usual back substitution procedure gives us the general solution

x1=1+5s1x2=1+2s1x3=s1x4=3

where s_1 can be any real number.

Example 5

Find all solutions to the system of linear equations

6x3 +19x5+11x6=−273x1+12x2+9x3−6x4+26x5+31x6=−63x1+4
x2+3x3−2x4+10x5+9x6=−17−x1−4x2−4x3+2x4−13x5−11x6=22

Solution The augmented matrix for this system is

[031−10124−4693−40−6−22192610−13 11 31   9−11|
−27−63−1722]

• **Identify pivot position for Row 1.**

Starting with the first column, we see that $a_{11} = 0$, which will not work for a pivot. However, there are nonzero terms down the first column, so we interchange Row 1 and Row 3 to place a 1 in the pivot position.

[031−11124−4693−40−6−22192610−1311319−11| −27−63−1722] ~ [130−14120−4396−4−2−602102619−13 93111−11|−17−63−2722]

▶ The operation is R1↔R3.

• **Elimination.**

Next, we need zeros down the first column below the pivot position. We already have $a_{31} = 0$, and we arrange for $a_{21} = 0$ and $a_{41} = 0$ by using the operations shown in the margin.

[130−14120−4396−4−2−602102619−13 93111−11| −17−63−2722] ~ [10004000306−1−200010−419−139411−2| −17−12−27 5]

▶ The operations are
−3R1+R1→R2 R1+R4→R4

▶ Do not be tempted to perform the operation R1↔R2. This will undo the zeros in the first column.

• **Identify pivot position for Row 2.**

Moving down one row and to the right one column from a_{11}, we find $a_{22} = 0$. Since all the entries below a_{22} are also zero, interchanging with lower rows will not put a nonzero term in the a_{22} position. Thus a_{22} cannot be a pivot position, so we move to the right to the third column to determine if a_{23} is a suitable pivot position. Although a_{23} is also zero, there are nonzero terms below, so we interchange Row 2 and Row 4, putting a − 1 in the pivot position.

[10004000306−1−200010−419−39411−2|
−17−12−27   5] ~ [100040003−160−2000 10 −319−49−2114|
−17   5−27−12]

▶ The operation is R2↔R4.

- **Elimination.**

Down the remainder of the third column, we already have a_{43} = 0, so we need only introduce a zero at a_{33} by using the operation shown in the margin.

[100040003−160−200010−319−49−2114|
−17   5−27−12] ~ [100040003−100−200010−31−49−2−14|
−17   5   3−12]

▶ The operation is $6R_2 + R_3 \rightarrow R_3$

- **Identify pivot position for Row 3.**

From the Row 2 pivot position, we move down one row and to the right one column to a_{34}. This entry is 0, as is the entry below, so interchanging rows will not yield an acceptable pivot. As we did before, we move one column to the right. Since a_{35} = 1 is nonzero, this becomes the pivot for Row 3.

▶ The operation is $4R_3 + R_4 \rightarrow R_4$

- **Elimination.**

We introduce a zero in the a_{45} position by using the operation shown in the margin.

[100040003−100−200010−31−49−2−14|
−17   5   3−12] ~ [100040003−100−200010−3109−2−10|
−17   5   3   0]

- **Identify pivot position for Row 4.**

Since Row 4 is the only remaining row and consists entirely of zeros, it has no pivot position. The matrix is now in echelon form, so no additional row operations are required. Converting the augmented matrix back to a linear system gives us

x1+4x2+3x3−2x4+10x5+9x6=−17 −x3−3x5−2x6=5 x5−x6=3

Using back substitution, we arrive at the general solution

x1=−5−4s1+2s2−4s3x2=s1x3=−14−5s3x4=s2x5=3+s3x6=s3

where s_1, s_2, and s_3 can be any real numbers.

Gaussian elimination can be applied to any matrix to find an equivalent matrix that is in echelon form. If matrix *A* is equivalent to matrix *B* that is in echelon form, we say that *B* is an echelon form of A. Different sequences of row operations can produce different echelon forms of the same starting matrix, but all echelon forms of a given matrix will have the same pivot positions.

Example 6

Use Gaussian elimination to find all solutions to the system of linear equations

x1+ 4x2−3x3=23x1− 2x2− x3=−1−x1+10x2−5x3=3

Solution The augmented matrix for this system is

[13−14−210−3−1−5|2−13]

• **Identify pivot position for Row 1, then elimination.**
We have a_{11} = 1, so this is the pivot position for Row 1. We introduce zeros down the first column with the row operations shown in the margin.

[13−14−210−3−1−5|2−13] ~ [100 4−1414−38−8|2−75]

▶ The operations are
$-3R1+R2\to R2 \quad R1+R3\to R3$

• **Identify pivot position for Row 2, then elimination.**

$$[000 \quad 4-1414-38-8|2-75] \sim [100 \quad 4-14 \quad 0-380|2-7-2]$$

▶ The operation is $R_2 + R_3 \to R_3$

Let's consider what we have. We find ourselves with a matrix in echelon form, but when we translate the last row back into an equation, we get 0 = − 2, which clearly has no solutions. Thus this system has no solutions, and so is inconsistent.

The preceding example illustrates a general principle. When applying row operations to an augmented matrix, if at any point in the process the matrix has a row of the form

$$[000\ldots0|c] \tag{3}$$

where c is nonzero, then stop. The system is inconsistent.

Gauss–Jordan Elimination

Let's return to the echelon form of the augmented matrix from Example 5,

$$[100040003-100-200010-3109-2-10|-17 \quad 5 \quad 3 \quad 0]$$

▶ Gauss–Jordan elimination is named for the previously encountered C. F. Gauss, and Wilhelm Jordan (1842–1899), a German engineer who popularized this method for finding solutions to linear systems in his book on geodesy (the science of measuring earth shapes).

After extracting the linear system from this matrix, we back substituted and simplified to find the general solution. We can make it easier to find the general solution by performing additional row operations on the matrix. Specifically, we do the following:

1. Multiply each nonzero row by the reciprocal of the pivot so that we end up with a 1 as the leading term in each nonzero row.
2. Use row operations to introduce zeros in the entries *above* each pivot position.

Picking up with our matrix, we see that the first and third rows already have a 1 in the pivot position. Multiplying the second row by −1 takes care of the remaining nonzero row.

[100040003−100−200010−3109−2−10|
−17 5 3 0] ~ [100040003100−20001031092−10|−17 −5 3 0]

▶ The operation is $-R_2 \to R_2$.

When implementing Gaussian elimination, we worked from left to right. To put zeros above pivot positions, we work from right to left, starting with the rightmost pivot, which in this case appears in the fifth column. Two row operations are required to introduce zeros above this pivot.

[100040003100−20001031092−10|
−17 −5 3 0] ~ [100040003100−20000010195−10|
−47−14 3 0]

▶ The operations are
−3R3+R2→R2−10R3+R1→R1

Next we move up to the pivot in the second row, located in the third column. One row operation is required to introduce a zero in the a_{13} position.

[100040003100−20000010195−10|
−47−14 3 0] ~ [100040000100−2000001045−10|−5−14 3 0]
(4)

▶ The operations is $-\,3R_2 + R_1 \to R_1$

Naturally there are no rows above the pivot position in the first row, so we are done. Now when we extract the linear system, it has the form

$$x_1+4x_2 -2x_4 +4x_6=-5 \quad x_3 +5x_6=-14 \quad x_5- x_6=3$$

▶ When using Gaussian and Gauss–Jordan elimination, do not yield to the temptation to alter the order of the row operations. Changing the order can result in a circular sequence of operations that lead to endless misery.

Note that when the system is expressed in this form, the leading variables appear *only* in the equation that they lead. Thus during back substitution we need only plug in free parameters and then subtract to solve for the leading variables, simplifying the process considerably.

The matrix on the right in (4) is said to be in *reduced echelon form.*

DEFINITION 1.5 ▶

Reduced Echelon Form

A matrix is in **reduced echelon form** (or **reduced row echelon form**) if

(a) It is in echelon form.

(b) All pivot positions contain a 1.

(c) The only nonzero term in a pivot column is in the pivot position.

Examples of matrices in reduced echelon form include

[01−200017001−603−600001−250000000] and [1−300−721001021300015−19]

Forward Phase, Backward Phase, Gauss–Jordan Elimination

Transforming a matrix to reduced echelon form can be viewed as having two parts: The **forward phase** is Gaussian elimination, transforming the matrix to echelon form, and the **backward phase**, which completes the transformation to reduced echelon form. The combination of the forward and backward phases is referred to as **Gauss–Jordan elimination**. Although a given matrix can be equivalent to many different echelon form matrices, the same is not true of reduced echelon form matrices.

THEOREM 1.6 ▶

A given matrix is equivalent to a unique matrix that is in reduced echelon form.

The proof of Theorem 1.6 is available on the text website.

Example 7

Use Gauss–Jordan elimination to find all solutions to the system of linear equations

x1−2x2−3x3=−1x1− x2−2x3=1−x1+3x2+5x3=2

Solution The augmented matrix and row operations are shown below.

[1−2−31−1−2−135|
−112]−R1+R2→R2 R1+R3→R3~[1−2−3011012|
−112] −R2+R3→R3~[1−2−3011001|−12−1]

That completes the forward phase, yielding a matrix in echelon form. Next, we implement the backward phase to transform the matrix to reduced echelon form.

[1−2−301100−1|−12−1]−R3+R2→R23R3+R1→R1~[1−20010002|
−43−1] −2R2+R1→R1~[100010001|23−1]

▶ From here on row operations are shown between the matrices.

The reduced echelon form is equivalent to the linear system

x1 =2 x2 =3 x3=−1

We see immediately that the system has the unique solution $x_1 = 2$, $x_2 = 3$, and $x_3 = -1$.

Example 8

Use Gauss–Jordan elimination to find all solutions to the system of linear equations

2x1− 2x2−6x3+x4=3−x1+   x2+3x3−x4=−3x1− 2x2− x3+x4=2

Solution This is the system in Example 4, where we transformed the augmented matrix to echelon form

[100−2−10−12010−1|2−1−3]

With the forward phase already done, it remains to complete the backward phase. We have

[1−2−110−120000−1|2−1−3]−R2→R2−R3→R3∼[1−2−1101−20000
1|213] −R3+R1→R1∼[1−2−1001−200001|
−113] 2R2+R1→R1∼[10−5001−200001|113]

Each pivot column has a one in the pivot position and zeroes elsewhere, so the matrix has been transformed to reduced echelon form. The corresponding linear system is

x1 −5x3 =1 x2−2x3 =1   x4=3

Back substitution gives us the general solution

x1=1+5s1x2=1+2s1x3=s1x4=3

where s_1 can be any real number, which is the same solution as found in Example 4.

Homogeneous Linear Systems

Homogeneous Equation, Homogeneous System

A linear equation is **homogeneous** if it has the form

$$a_1x_1+a_2x_2+\cdots+a_nx_n=0$$

Homogeneous linear systems are an important class of systems that are made up of homogeneous linear equations.

$$\begin{array}{c}a_{11}x_1+a_{12}x_2+a_{13}x_3+\cdots+a_{1n}x_n=0\\ a_{21}x_1+a_{22}x_2+a_{23}x_3+\cdots+a_{2n}x_n=0\\ \vdots\quad\vdots\quad\vdots\quad\vdots\\ a_{m1}x_1+a_{m2}x_2+a_{m3}x_3+\cdots+a_{mn}x_n=0\end{array}$$

All homogeneous systems are consistent, because there is always one easy solution, namely,

$$x_1=0,x_2=0,\ldots,x_n=0$$

Trivial Solution, Nontrivial Solution

This is called the **trivial solution**. If there are additional solutions, they are called **nontrivial solutions**. We determine if there are nontrivial solutions in the usual way, using elimination methods.

Example 9

Use Gauss–Jordan elimination to find all solutions to the homogeneous system of linear equations

$$\begin{array}{c}2x_1-6x_2-x_3+8x_4=0\\ x_1-3x_2-x_3+6x_4=0\\ -x_1+3x_2-x_3+2x_4=0\end{array}$$

Solution The system is homogeneous, so we know that it has the trivial solution. To find the other solutions, we load the system into an augmented matrix and transform to reduced echelon form.

[2−6−181−3−16−13−12|000]R1↔R2∼[1−3−162−6−18−13−12|000] −2R1+R2→R2 R1+R3→R3∼[1−3−16001−400−28|000] 2R2+R3 →R3∼[1−3−16001−40000|000] R2+R1→R1∼[1−302001−40000|000]

The last matrix is in reduced echelon form. The corresponding linear system is

x1−3x2 +2x4=0 x3−4x4=0

Back substituting yields the general solution

x1=3s1−2s2x2=s1x3=4s2x4=s2

where s_1 and s_2 can be any real numbers.

Proof of Theorem 1.3

We are now in a position to revisit and prove Theorem 1.3 from Section 1.1. Recall the statement of the theorem.

THEOREM 1.3 ▶

A system of linear equations has no solutions, exactly one solution, or infinitely many solutions.

Proof We can take any linear system, form the augmented matrix, use Gaussian elimination to reduce to echelon form, and extract the transformed system. This process can lead to one of three possible outcomes:

(a) The system has an equation of the form $0 = c$ for $c \neq 0$. In this case, the system has no solutions.

If (a) does not occur, then one of (b) or (c) must:

(b) The transformed system has no free variables and hence exactly one solution.

(c) The transformed system has one or more free variables and hence infinitely many solutions.

Homogeneous linear systems are even simpler. Since all such systems have the trivial solution, (a) cannot happen. Therefore a homogeneous linear system has either a unique solution or infinitely many solutions. ■ ■

COMPUTATIONAL COMMENTS

- We can find the solutions to any system by using either Gaussian elimination or Gauss–Jordan elimination. Which is better? For a system of n equations with n unknowns, Gaussian elimination requires approximately 23n3 flops (i.e., arithmetic operations) and Gauss–Jordan requires about n^3 flops. Back substitution is slightly more complicated for Gaussian elimination than for Gauss–Jordan, but overall Gaussian elimination is more efficient and is the method that is usually implemented in computer software.
- When elimination methods are implemented on computers, to control round-off error they typically include an extra step called "partial pivoting" which involves selecting the entry having the largest absolute value to serve as the pivot. When performing row operations by hand, partial pivoting tends to introduce fractions and leads to messy calculations, so we avoided the topic. However, it is discussed Section 1.4.

- There are various similar definitions for what constitutes a “flop.” Here we take a “flop” to be one arithmetic operation, either addition or multiplication. Counting flops gives a measure of algorithm efficiency.

- Practice problems can also be used as additional examples.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Transform the following matrices to echelon form.

(a) [1−2122−355−132−5]

(b) [13−112126−1536−1−336−4413−2−141]

2. Transform the following matrices to reduced row echelon form.

(a) [1−232−35−14−5]

(b) [1421−1−4−102864]

3. Convert the system to an augmented matrix, transform to echelon form, then back substitute to find the solutions to the system.

(a) −x1+2x2− 3x3=−1−x1+3x2− x3=−32x1−2x2+10x2=−2

(b) x1−x2−2x3+ x4−2x5=0x1−x2− x3− x4− x5=0−x1+x2+ x3−3x4+6x5=0

4. Convert the system to an augmented matrix, transform to reduced row echelon form, then back substitute to find the solutions to the system.

(a) x1+2x2+ x3=4x1+ x2+ 2x3=1

(b) x1− x2+3x3=−12x1− x2+4x3=−1−x1+3x3−6x3=4

5. Determine if each statement is true or false, and justify your answer.

(a) A matrix that has more columns than rows cannot be transformed to reduced row echelon form.

(b) A matrix that has more rows than columns and is in echelon form must have a row of zeros.

(c) Every elementary row operation is reversible.

(d) A linear system with free variables cannot have one solution.

6. Suppose that a matrix with four rows and nine columns is in echelon form.
   (a) If the matrix has no row of all zeros, then how many pivot positions are there?
   (b) What is the minimum number of zeros in the bottom row?
   (c) What is the minimum number of zeros in the matrix?
   (d) If this is the augmented matrix for a linear system, then what is the minimum number of free variables?

EXERCISES

In each exercise set, problems marked with are designed to be solved using a programmable calculator or computer algebra system.

Exercises 1–4: Convert the augmented matrix to the equivalent linear system.

1. [4−120−15|27]
2. [−2 13−11 1−37|06−5]
3. [0−12 61712 5 8 0−3−320−9111013|170−8−1]
4. [−153|2−70]

Exercises 5–10: Determine those matrices that are in echelon form, and those that are also in reduced echelon form.

5. [13−2026000]
6. [1306−20015300000]
7. [3−311000−24000000]
8. [1−31−701−24001−20002]
9. [1005−1020−20001−32]
10. [1−10900018000001]

Exercises 11–14: The matrix on the right results after performing a single row operation on the matrix on the left. Identify the row operation.

11. [−210   13−36−117−5] ~ [4−2013−36−117−5]
12. [42−12−1057] ~[121423−1057]
13. [2−13049−23675−1] ~ [2−13049−23−2−119−7]
14. [2−13049−23675−1] ~ [675−149−232−130]

Exercises 15–18: A single row operation was performed on the matrix on the left to produce the matrix on the right. Unfortunately, an error was made when performing the row operation. Identify the operation and fix the error.

15. [37−2−14350−2] ~ [−14−337−250−3]
16. [−2−2164−10−5] ~[−2−2160527]
17. [03−12−1−9415072] ~ [26−24−1−9415072]
18. [172004−8−33001] ~ [1   7200   4−8−31−140   1]

Exercises 19–26: Convert the system to an augmented matrix and then find all solutions by reducing the system to echelon form and back substituting.

19. 2x1+x2=1−4x1−x2=3
20. 3x1−7x2=0x1+4x2=0
21. −2x1+ 5x2−10x3=   4   x1− 2x2+ 3x3=−17x1−17x2+34x3=−16
22. 2x1+8x2−4x3=−10−x1−3x2+5x3=   4
23. 2x1+2x2−x3=8−x1− x2   =−33x1+ 3x2+x3=7
24. −5x1+9x2=133x1−5x2=−9x1−2x2=−2
25. 2x1+   6x2−9x3−4x4=0−3x1−11x2+9x3− x4=0   x1+   4x2−2x3+ x4=0
26.   x1− x2− 3x3− x4=−1−2x1+2x2+ 6x3+2x4=−1−3x1−3x2+10x3   =5

Exercises 27–32: Convert the system to an augmented matrix and then find all solutions by transforming the system to reduced echelon form and back substituting.

27. −2x1−5x2=0x1+3x2=1
28. x1+ x2=13x1+4x2=6−x1+ x2=5

29. 2x1+x2 =2−x1−x2−x3=1

30. −4x1+2x2−2x3=10   x1 +x3=−33x1−x2+x3=−8

31. −3x1+2x2−x3+6x4=−77x1−3x2+2x3−11x4=14   x1   −   x4=1

32. x1+x2+x3−2x4+4x5=−5−x1 −3x3+4x4−5x5=52x1+4x2−2x3+x 4+5x5=−9

Exercise 33–36: Suppose that the row operation is used to transform a matrix. Which row operation will transform the matrix back to its original form?

33.

(a) $5R_1 \to R_1$

(b) $-2R_3 \to R_3$

34.

(a) R1↔R3

(b) R4↔R1

35.

(a) $-5R_2 + R_6 \to R_6$

(b) $-3R_1 + R_3 \to R_3$

36.

(a) $4R_5 + R_1 \to R_1$

(b) $-R_4 + R_2 \to R_2$

FIND AN EXAMPLE Exercises 37–42: Find an example that meets the given specifications.

37. A matrix with three rows and five columns that is in echelon form, but not in reduced echelon form.

38. A matrix with six rows and four columns that is in echelon form, but not in reduced echelon form.

39. An augmented matrix for an inconsistent linear system that has four equations and three variables.

40. An augmented matrix for an inconsistent linear system that has three equations and four variables.

41. A homogeneous linear system with three equations, four variables, and infinitely many solutions.

42. Two matrices that are distinct yet equivalent.

TRUE OR FALSE Exercises 43–46: Determine if the statement is true or false, and justify your answer.

43.

(a) If two matrices are equivalent, then one can be transformed into the other with a sequence of elementary row operations.

(b) Different sequences of row operations can lead to different echelon forms for the same matrix.

44.

(a) Different sequences of row operations can lead to different reduced echelon forms for the same matrix.

(b) If a linear system has four equations and seven variables, then it must have infinitely many solutions.

45.

(a) If a linear system has seven equations and four variables, then it must be inconsistent.

(b) Every linear system with free variables has infinitely many solutions.

46.

(a) Any linear system with more variables than equations cannot have a unique solution.

(b) If a linear system has the same number of equations and variables, then it must have a unique solution.

Exercises 47–50: The row operation shown is not an elementary row operation. If the given operation is the combination of elementary row operations, then provide them. If not, explain why not.

47.

(a) $R_3 - R_2 - R_2$

(b) $3R_1 + R_4 \to R_1$

48.

(a) $3R_2 + 2R_4 \to R_4$

(b) $-R_5 + R_4 \to R_5$

49.

(a) $R_3 - 4R_6 \to R_6$

(b) $2R_1 + 5R_3 \to R_2$

50.

(a) $5R_1 + R_2 \to R_3$

(b) $2R_4 - R_5 \to R_5$

51. Suppose that the echelon form of an augmented matrix has a pivot position in every column except the rightmost one. How many solutions does the associated linear system have? Justify your answer.

52. Suppose that the echelon form of an augmented matrix has a pivot position in every column. How many solutions does the associated linear system have? Justify your answer.

53. Show that if a linear system has two different solutions, then it must have infinitely many solutions.

54. Show that if a matrix has more rows than columns and is in echelon form, then it must have at least one row of zeros at the bottom.

55. Show that a homogeneous linear system with more variables than equations must have an infinite number of solutions.

56. Show that each of the elementary operations on linear systems (see page 16) produces an equivalent linear system. (Recall two linear systems are equivalent if they have the same solution set.)

(a) Interchange the position of two equations.

(b) Multiply an equation by a nonzero constant.

(c) Add a multiple of one equation to another.

C Exercises 57–58: Find the *interpolating polynomial f(x),* which is used to fit a function to a set of data.

57. Figure 3 shows the plot of the points (1, 4), (2, 7), and (3, 14). Find a polynomial of degree 2 of the form $f(x) = ax^2 + bx + c$ whose graph passes through these points.

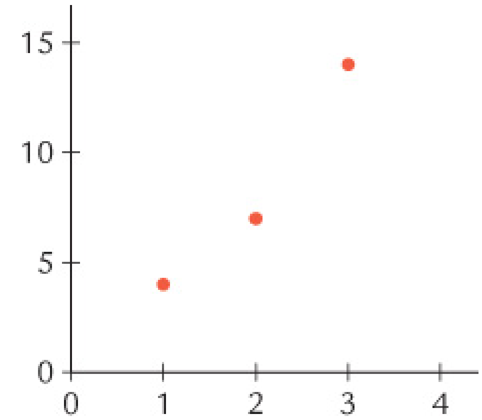


Figure 3 Exercise 57 data.

58. Figure 4 shows the plot of the points (1, 8), (2, 3), (3, 9), (5, 1), and (7, 7). Find a polynomial of degree 4 of the form $f(x) = ax^4 + bx^3 + cx^2 + dx + e$ whose graph passes through these points.

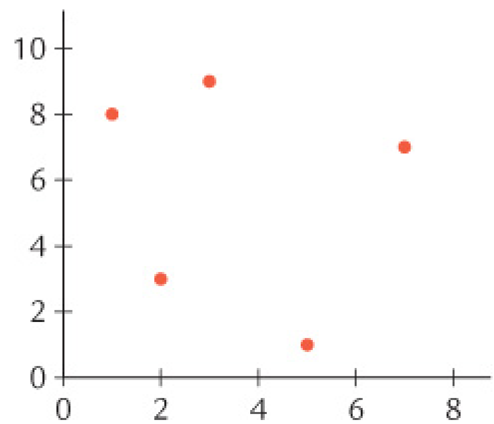


Figure 4 Exercise 58 data.

Exercises 59–60: Refer to the cannonball scenario described at the start of the section. For each problem, the three ordered pairs are $(x, E(x))$, where x is the distance on the ground from the position of the cannon and $E(x)$ is the elevation of the ball. Find a model for the elevation of the ball, and use the model to determine where it hits the ground.

59. (20, 288), (40, 364), (60, 360)

60. (40, 814), (80, 1218), (110, 1311)

Exercises 61–68: Use technology to perform the row operations needed to transform the augmented matrix to reduced echelon form, and then find all solutions to the corresponding system.

61. [2−3−2706−35−5|014]

62. [1127−5−33083|000]

63. [572−21006−3325|9−24]

64. [908−27120−1−6−4−15|63−8]

65. [865−2−82600−1−3−1|−1010−4]

66. [54043−3377−2178|50−212]

67. [63−7135−2121−130080−2−31211−7|0000]

68. [2037151100−1−9508153−16−3−5402|701311]

[1] The plural of matrix is matrices.

1.3 Applications of Linear Systems

In this section we consider applications of linear systems. These are but a few of the many different possible applications that exist.

▶ This section is optional. Some applications presented here are referred to later, but can be reviewed as needed.

Traffic Flow

Example 1

Located on the northern coast of California is Arcata, a charming college town with a popular central plaza. Figure 1 shows a map of the streets surrounding and adjacent to the plaza. As indicated by the arrows, all streets in the vicinity of the plaza are one-way. Traffic flows north and south on G and H streets, respectively, and east and west on 8th and 9th streets, respectively. The number of cars flowing on and off the plaza during a typical 15-minute period on a Saturday morning is also shown. Find x_1, x_2, x_3, and x_4, the volume of traffic along each side of the plaza.

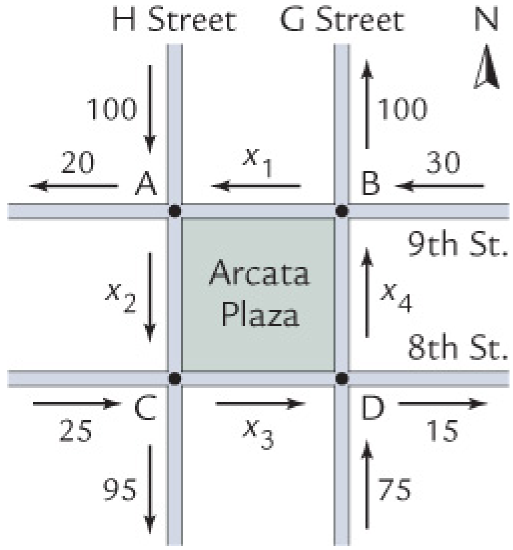


Figure 1 Traffic volumes around the Arcata plaza.

Solution The four intersections are labeled A, B, C, and D. At each intersection, the number of cars entering the intersection must equal the number leaving. For example, the number of cars entering A is $100 + x_1$ and the number exiting is $20 + x_2$. Since these must be equal, this gives us the equation

$$A: 100+x_1=20+x_2$$

Applying the same reasoning to intersections B, C, and D, we arrive at three more equations

$$B: x_4+30=x_1+100 \quad C: x_2+25=x_3+95 \quad D: x_3+75=x_4+15$$

Rewriting the equations in the usual form of a linear system, we obtain

$$x_1-x_2 \quad =-80 \quad x_1 \quad -x_4=-70 \quad x_2-x_3 \quad =70 \quad x_3-x_4=-60$$

To solve the system, we populate an augmented matrix and transform to echelon form.

[1−100100−101−10001−1|
−80−7070−60]−R1+R2→R2~[1−100010−101−10001−1|

−801070−60] −R2+R3→R3∼[1−100010−100−11001−1|
−801060−60] R3+R4→R4∼[1−100010−100−110000|−8010600]

Back substitution yields the general solution

x1=−70+s1,x2=10+s1,x3=−60+s1,s4=s1

where s_1 is a free parameter.

A moment's thought reveals why it makes sense that this system has infinitely many solutions. There can be an arbitrary number of cars simply circling the plaza, perhaps looking for a parking space. Note also that since each of x_1, x_2, x_3, and x_4, must be nonnegative, it follows that the parameter $s_1 \geq 70$.

The analysis performed above can be carried over to much more complex traffic questions, or to other similar settings, such as computer networks.

Equilibrium Temperatures

Example 2

Figure 2 gives a diagram of a piece of heavy wire mesh. Each of the eight wire ends has temperature held fixed as shown. When the temperature of the mesh reaches equilibrium, the temperature at each connecting point will be the average of the temperatures of the adjacent points and fixed ends. Determine the equilibrium temperature at the connecting points x_1, x_2, x_3, and x_4.

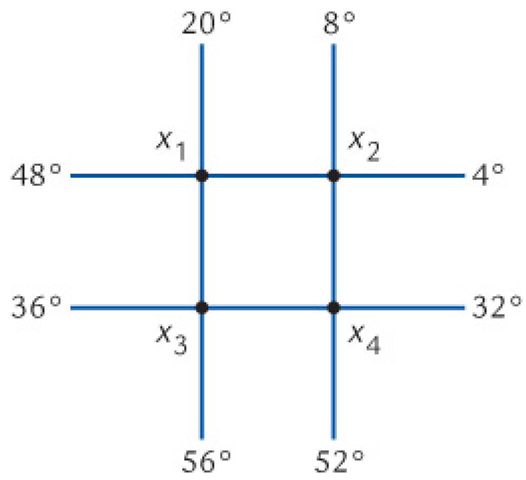


Figure 2 Grid Temperatures for Example 2.

Solution

The temperature of each connecting point depends in part on the temperature of other connecting points. For instance, since x_1 is adjacent to x_2, x_3, and the ends held fixed at 48° and 20°, its temperature at equilibrium will be the average

x1x2+x3+48+204=0.25x2+0.25x3+17 (1)

Similarly, for the other connecting points we have the equations (after simplifying)

x2=0.25x1+0.25x4+3x3=0.25x1+0.25x4+23 x4=0.25x2+0.25x3+21 (2)

Moving the variables in (1) and (2) to the left side, multiplying each by four to clear fractions, and collecting them together gives the system

4x1−x2−x3 =68−x1+4x2 −x4=12−x1 +4x3−x4=92 −x2−x3+4x4=84

Shown below are the augmented matrix and sequence of row operations to transform to echelon form.

[4−1−10−140−1−104−10−1−14|68129284]R1↔R3~[−104−1−140−14−1−100−1−14|92126884] −R1+R2→R24R1+R3→R3~[−104−104−400−115−40−1−14|92−8043684] 14R2↔R2~[−104−101−100−1

15−40−1−14|92−2043684] R2+R3→R3R2+R4→R4~[−104−101−1 00014−400−24|92−2041664] R3↔R412R3→R3~[−104−101−1000 −120014−4|92−2032416] 14R3+R4→R4~[−104−101−1000−12000 24|92−2032846]

▶ A hand-held calculator Is useful for the arithmetic involving larger numbers.

Extracting the triangular system and back substituting gives the equilibrium temperatures $x_1 = 32$, $x_2 = 20$, $x_3 = 40$, and $x_4 = 36$.

Economic Inputs and Outputs

Example 3

Imagine a simple economy that consists of consumers and just three industries, which we refer to as A, B and C. These industries have annual consumer sales of 60, 75, and 40 (in billions of dollars), respectively. In addition, for every dollar of goods A sells, A requires 10 cents of goods from B and 15 cents of goods from C to support production. (For instance, maybe B sells electricity and C sells shipping services.) Similarly, each dollar of goods B sells requires 20 cents of goods from A and 5 cents of goods from C, and each dollar of goods C sells requires 25 cents of goods from A and 15 cents of goods from B. What output from each industry will satisfy both consumer and between-industry demand?

▶ Example 3 is based on the work of Nobel prize winning economist Wassily Leontief (1906–1999). He divided the economy into 500 sectors in developing his input-output model.

Solution Let a, b, and c denote the total output from each of A, B, and C, respectively. The entire output for A is 60 for consumers, $0.20b$ for B, and $0.25c$ for C. Totaling this up yields the equation

$$a=60+0.20b+0.25c$$

Similar reasoning applied to industries B and C yields the equations

$$\begin{aligned} b &= 75 + 0.10a + 0.15c \\ c &= 40 + 0.15a + 0.05b \end{aligned}$$

These equations lead to the linear system

$$\begin{aligned} a - 0.20b - 0.25c &= 60 \\ -0.10a + b - 0.15c &= 75 \\ -0.15a - 0.05b + c &= 40 \end{aligned}$$

Converting the system to an augmented matrix and transforming as usual, we have

$$\left[\begin{array}{ccc|c} 1.00 & -0.20 & -0.25 & 60 \\ -0.10 & 1.00 & -0.15 & 75 \\ -0.15 & -0.05 & 1.00 & 40 \end{array}\right] \begin{array}{l} 0.1R_1 + R_2 \to R_2 \\ 0.15R_1 + R_3 \to R_3 \end{array} \sim \left[\begin{array}{ccc|c} 1 & -0.20 & -0.25 & 60 \\ 0 & 0.98 & -0.175 & 81 \\ 0 & -0.08 & 0.963 & 49 \end{array}\right] \frac{0.08}{0.98}R_2 + R_3 \to R_3 \sim \left[\begin{array}{ccc|c} 1 & -0.20 & -0.25 & 60.000 \\ 0 & 0.98 & -0.175 & 81.000 \\ 0 & 0 & 0.948 & 55.612 \end{array}\right]$$

▶ A hand-held calculator was used for this problem. Some values are rounded.

Extracting the triangular system and back substituting gives the solution *a* = 93.29, *b* = 93.13, and *c* = 58.65.

Planetary Orbital Periods

Example 4

Planets that are closer to the sun take less time than those farther out to make one orbit around the sun. Table 1 gives the average distance from the sun and the number of Earth days required to make one orbit for each planet. Develop an equation that describes the relationship between the distance from the sun and the length of the orbital period.

Table 1 **Planetary Orbital Distances and Periods**

Planet	Distance from Sun ($\times 10^6$ km)	Orbital Period (days)
Mercury	57.9	88
Venus	108.2	224.7
Earth	149.6	365.2
Mars	227.9	687
Jupiter	778.6	4331
Saturn	1433.5	10747
Uranus	2872.5	30589
Neptune	4495.1	59800

Solution As a starting point, consider the scatter plot of the data given in Figure 3. There seems to be a pattern to the data. The points do not lie on a line, but the curved shape suggests that for constants *a* and b, the data may come close to satisfying the equation

$$p=adb \tag{3}$$

where *p* is the orbital period and *d* is the distance from the sun. Here we proceed by substituting data to create a system of equations to solve. However, before doing that we note (3) is not linear in *a* and *b,* but is if we apply the logarithm function to both sides. Doing this gives us

$$\ln(p)=\ln(adb)=\ln(a)+b\ln(d)$$

If we let a_1 = ln(a) and substitute the data from Mercury and Venus, we get the system of two equations and two unknowns

$$a1+ \ b\ln(57.9)=\ln(88)a1+b\ln(108.2)=\ln(224.7)$$

The solution to this system is $a_1 \approx -1.60771$ and $b \approx 1.49925$. Since $a_1 = \ln(a)$, we have $a \approx e^{-1.60771} = 0.200346$, yielding the formula

$$p=(0.200346)d1.49925$$

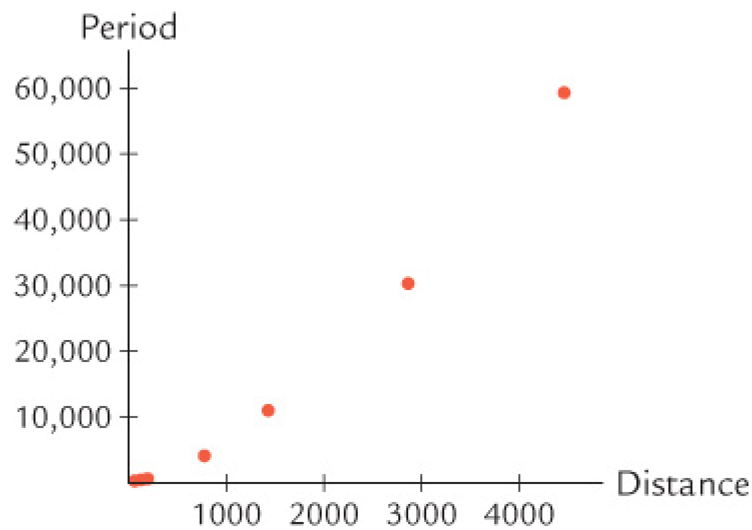


Figure 3 Orbital Distance vs. Orbit Period.

Table 2 Planetary Orbital Distances and Periods

Planet	Distance	Actual Period	Predicted Period
Mercury	57.9	88	87.9988
Venus	108.2	224.7	224.696
Earth	149.6	365.2	365.214
Mars	227.9	687	686.482
Jupiter	778.6	4331	4330.96
Saturn	1433.5	10,747	10,814.6
Uranus	2872.5	30,589	30,644.3
Neptune	4495.1	59,800	59,999.8

Table 2 gives the actual and predicted (using the above formula) orbital period for each planet.

The predictions are fairly good, suggesting that our formula is on the right track. However, the predictions become less accurate for those planets farther from the sun. Because we used the data for Mercury and Venus to develop our formula, perhaps this is not surprising. If instead we use Uranus and Neptune, we arrive at the formula

$$p=(0.20349)d^{1.497}$$

Table 3 shows that this formula produces better predictions.

Table 3 **Predicted Orbital Periods**

Planet	Predicted Period
Mercury	88.6
Venus	225.8
Earth	366.8
Mars	688.8
Jupiter	4334
Saturn	10,806
Uranus	30,589
Neptune	59,799

A natural idea is to incorporate more data into our formula, by using more planets to generate a larger system of equations. Unfortunately, if we use more than two planets, we end up with a system that has no solutions. (Try it for yourself.) Thus there are limitations to what we can do with the tools we currently have available. In Chapter 8 we develop a more sophisticated method that allows us to use all of our data simultaneously to come up with a formula that provides a good estimate for a range of distances from the sun.

Balancing Chemical Equations

Example 5

A popular chemical among college students is caffeine, which has chemical composition $C_8H_{10}N_4O_2$. When heated and combined with oxygen (O_2), the ensuing reaction produces carbon dioxide (CO_2), water (H_2O), and nitrogen dioxide (NO_2). This chemical reaction is indicated using the notation

$$x_1C_8H_{10}N_4O_2+x_2O_2\rightarrow x_3CO_2+x_4H_2O+x_5NO_2 \tag{4}$$

where the subscripts on the elements indicate the number of atoms. (No subscript indicates one atom.) Balancing the equation involves finding values for x_1, x_2, x_3, x_4, and x_5 so that the number of atoms of each element is the same before and after the reaction. Most chemistry texts describe a method of solution that might be described as trial and error. However, there is no need for a haphazard approach. Use linear algebra to balance the equation.

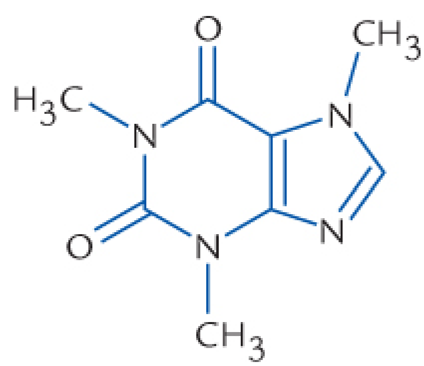


Figure 4 The caffeine molecule.

Solution

Let's start with carbon in the reaction in (4). On the left side there are $8x_1$ carbon atoms, while on the right side there are x_3 carbon atoms. This yields the equation

$$8x_1=x_3$$

For oxygen, we see that there are $2x_1 + 2x_2$ atoms on the left side and $2x_3 + x_4 + 2x_5$ on the right, producing another equation,

$$2x_1+2x_2=2x_3+x_4+2x_5$$

Similar analysis on nitrogen and hydrogen results in two additional equations,

$$4x_1 = x_5 \quad \text{and} \quad 10x_1 = 2x_4$$

To balance the chemical equation, we must find a solution that satisfies all four equations. That is, we need to find the solution set to the linear system

$$\begin{aligned} 2x_1 + 2x_2 - 2x_3 - x_4 - 2x_5 &= 0 \\ 4x_1 \qquad\qquad\qquad\quad - x_5 &= 0 \\ 8x_1 \qquad - x_3 \qquad\qquad &= 0 \\ 10x_1 \qquad\qquad - 2x_4 \qquad &= 0 \end{aligned}$$

The augmented matrix and row operations are

$$\left[\begin{array}{ccccc|c} 2 & 2 & -2 & -1 & -2 & 0 \\ 4 & 0 & 0 & 0 & -1 & 0 \\ 8 & 0 & -1 & 0 & 0 & 0 \\ 10 & 0 & 0 & -2 & 0 & 0 \end{array}\right] \begin{array}{l} -2R_1 + R_2 \to R_2 \\ -4R_1 + R_3 \to R_3 \\ -5R_1 + R_4 \to R_4 \end{array} \sim \left[\begin{array}{ccccc|c} 2 & 2 & -2 & -1 & -2 & 0 \\ 0 & -4 & 4 & 2 & 3 & 0 \\ 0 & -8 & 7 & 4 & 8 & 0 \\ 0 & -10 & 10 & 3 & 10 & 0 \end{array}\right] \begin{array}{l} -2R_2 + R_3 \to R_3 \\ \frac{5}{2}R_2 + R_4 \to R_4 \end{array} \sim \left[\begin{array}{ccccc|c} 2 & 2 & -2 & -1 & -2 & 0 \\ 0 & -4 & 4 & 2 & 3 & 0 \\ 0 & 0 & -1 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 & \frac{5}{2} & 0 \end{array}\right]$$

Back substituting and scaling the free parameter gives the general solution

$$x_1 = 2s_1 \quad x_2 = 27s_1 \quad x_3 = 16s_1 \quad x_4 = 10s_1 \quad x_5 = 8s_1$$

where s_1 can be any real number. Any choice of s_1 yields constants that balance our chemical equation, but it is customary to select the solution that makes each of the coefficients x_1, x_2, x_3, x_4, and x_5 integers that have no common factors. Setting $s_1 = 1$ accomplishes this, yielding the balanced equation

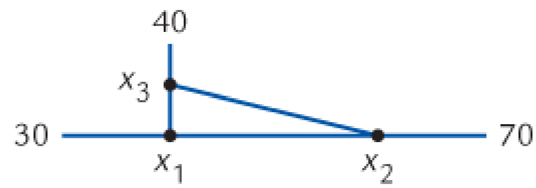
$$2C_8H_{10}N_4O_2 + 27O_2 \to 16CO_2 + 10H_2O + 8NO_2$$

▶ Practice problems can also be used as additional examples.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the equilibrium temperatures at x_1, x_2, and x_3 for the heavy wires with endpoints held at the given temperatures.



2. Economy has two industries A and B. These industries have annual consumer sales of 50 and 80 (in billions of dollars), respectively. For every dollar of goods A sells, A requires 35 cents of goods from B. For each dollar of goods B sells, B requires 20 cents of goods from A. Determine the total output required from each industry in order to meet both consumer and between-industry demand.

3. The equation below describes how carbon dioxide and water combine to produce glucose and oxygen during the process of photosynthesis. Balance the equation.

CO2+H2O→C6H12O6+O2

4. Find a model for planetary orbital period using the data from Earth and Neptune.

5. Find the values of the unknown constants in the given decomposition.

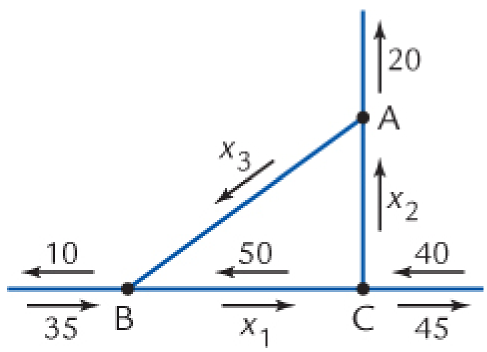
x+5(2x+1)(x−1)=A2x+1+Bx−1

6. Determine if each statement is true or false, and justify your answer.
   (a) Traffic flow problems must have a unique solution.
   (b) There are infinitely many ways to balance a chemical equation.
   (c) There is a unique parabola passing through any three distinct points in the plane.
   (d) There is exactly one function $f(x) = ae^{x} + be^{-2x}$ such that $f(0) = 5$.

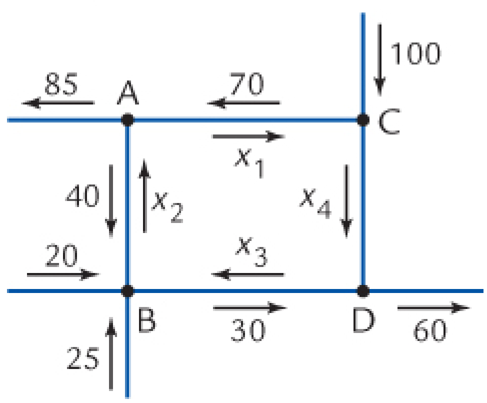
EXERCISES

In each exercise set, problems marked with C are designed to be solved using a programmable calculator or computer algebra system.

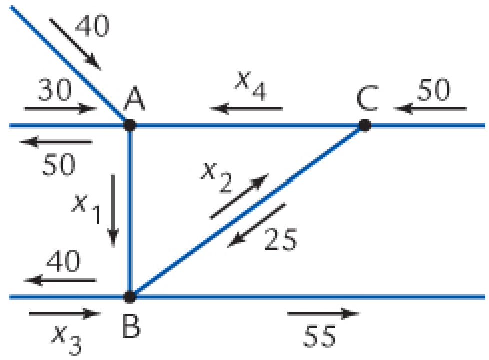
1. The volume of traffic for a collection of intersections is shown in the figure below. Find all possible values for x_1, x_2, and x_3. What is the minimum volume of traffic from C to A?



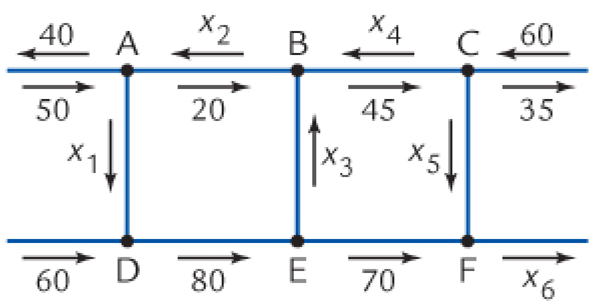
2. C The volume of traffic for a collection of intersections is shown in the figure below. Find all possible values for x_1, x_2, x_3, and x_4. What is the minimum volume of traffic from C to D?



3. C The volume of traffic for a collection of intersections is shown in the figure below. Find all possible values for x_1, x_2, x_3, and x_4. What is the minimum volume of traffic from C to A?

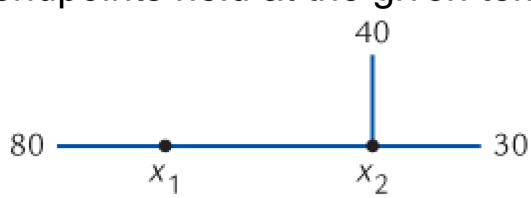


4. The volume of traffic for a collection of intersections is shown in the figure below. Find all possible values for x_1, x_2, x_3, x_4, x_5, and x_6.

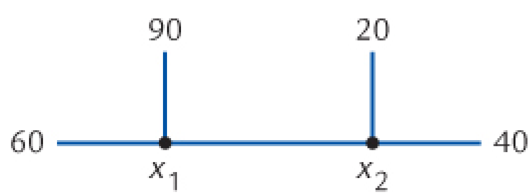


Exercises 5–8: Find the equilibrium temperatures for the heavy wires with endpoints held at the given temperatures.

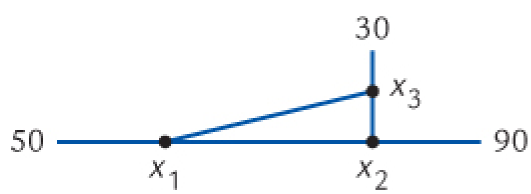
5.



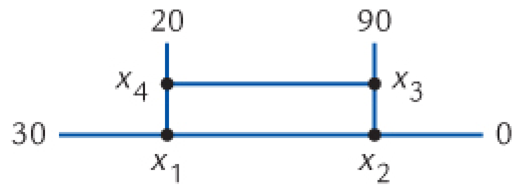
6.



7.



8.



Exercises 9–12: Determine the total output required from each industry in order to meet both consumer and between-industry demand. (See Example 3 for background information.)

9. Economy has two industries A and B. These industries have annual consumer sales of 60 and 40 (in billions of dollars), respectively. For every dollar of goods A sells, A requires 20 cents of goods from B. For each dollar of goods B sells, B requires 30 cents of goods from A.

10. Economy has two industries A and B. These industries have annual consumer sales of 80 and 50 (in billions of dollars), respectively. For every dollar of goods A sells, A requires 25 cents of goods from B. For each dollar of goods B sells, B requires 15 cents of goods from A.

11. Economy has three industries A, B, and C. These industries have annual consumer sales of 30, 50, and 60 (in billions of dollars), respectively. For every dollar of goods A sells, A requires 10 cents of goods from B and 15 cents of goods from C. For each dollar of goods B sells, B requires 15 cents of goods from A and 20 cents of goods from C. For each dollar of goods C sells, C requires 20 cents of goods from A and 10 cents of goods from B.

12. Economy has three industries A, B, and C. These industries have annual consumer sales of 40, 30, and 70 (in billions of dollars), respectively. For every dollar of goods A sells, A requires 20 cents of goods from B and 10 cents of goods from C. For each dollar of goods B sells, B requires 25 cents of goods from A and 10 cents of goods from C. For each dollar of goods C sells, C requires 10 cents of goods from A and 15 cents of goods from B.

Exercises 13–20: Balance the chemical equation.

13. Hydrogen burned in oxygen forms steam:

$$H_2 + O_2 \rightarrow H_2O$$

14. Hydrogen and nitrogen combine to form ammonia:

$$H_2 + N_2 \rightarrow NH_3$$

15. Iron and oxygen combine to form iron oxide:

$$Fe + O_2 \rightarrow Fe_2O_3$$

16. Sodium and water react to form sodium hydroxide (lye) and hydrogen:

$$Na + H_2O \rightarrow NaOH + H_2$$

17. When propane burns in oxygen, it produces carbon dioxide and water:

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

18. When acetylene burns in oxygen, it produces carbon dioxide and water:

$$C_2H_2 + O_2 \rightarrow CO_2 + H_2O$$

19. Potassium superoxide and carbon dioxide react to form potassium carbonate and oxygen:

$$KO_2 + CO_2 \rightarrow K_2CO_3 + O_2$$

20. Manganese dioxide and hydrochloric acid combine to form manganese chloride, water, and chlorine gas:

$$MnO_2 + HCl \rightarrow MnCl_2 + H_2O + Cl_2$$

Exercises 21–24: Find a model for planetary orbital period using the data for the given planets.

21. Earth and Mars.

22. Mercury and Uranus.

23. Venus and Neptune.

24. Jupiter and Saturn.

Exercises 25–26: The data given provides the distance required for a particular type of car to stop when traveling at various speeds. A reasonable model for braking distance is, $d = as^k$, where d is distance, s is speed, and a and k are constants. Use the data in the table to find values for a and k, and test your model. (HINT: Methods similar to those used to find a model for planetary orbital periods can be applied here.)

25.

Speed (MPH)	10	20	30	40
Distance (Feet)	4.5	18	40.5	72

26.

Speed (MPH)	10	20	30	40
Distance (Feet)	20	80	180	320

Exercises 27–30: When using partial fractions to find antiderivatives in calculus, we decompose complicated rational expressions into the sum of simpler expressions that can be integrated individually. Find the values of the missing constants in the provided decomposition.

27. 1x(x+1)=Ax+Bx+1

28. 3x−1(x−1)(x+1)=Ax−1+Bx+1

29. 1x2(x−1)=Ax+Bx2+Cx−1

30. 1x(x2+1)=Ax+Bx+Cx2+1

31. The points (1, 3) and (−2, 6) lie on a line. Where does the line cross the *x*-axis?

32. The points (5, −1) and (− 8, 3) lie on a line. Where does the line cross the *y*-axis?

33. The points (2, −1, −2), (1, 3, 12), and (4, 2, 3) lie on a unique plane. Where does this plane cross the *z*-axis?

34. The points (2, 2, −1), (− 1, −1, 0), and (2, 1, 1) lie on a unique plane. Where does this plane cross the *z*-axis?

Exercises 35–36: A parabola has the form $y = ax^2 + bx + c$, where a, b, and c are constants and $a \neq 0$. Find an equation for the parabola that passes through the given points.

35. $(-1, -2)$, $(1, 4)$, and $(2, 4)$.

36. $(-1, -10)$, $(1, -4)$, and $(2, -7)$.

37. Find a polynomial of the form

$$f(x)=ax3+bx2+cx+d$$

such that $f(0) = -3$, $f(1) = 2$, $f(3) = 5$, and $f(4) = 0$.

38. Find a polynomial of the form

$$g(x)=ax4+bx3+cx2+dx+e$$

such that $g(-2) = -17$, $g(-1) = 6$, $g(0) = 5$, $g(1) = 4$, and $g(2) = 3$.

Calculus required Exercises 39–40: Find the values of the coefficients a, b, and c so the given conditions for the function f and its derivatives are met. (This type of problem arises in the study of differential equations.)

39. $f(x) = ae^{x} + be^{2x} + ce^{-3x}$; $f(0) = 2$, $f'(0) = 1$, and $f''(0) = 19$.

40. $f(x) = ae^{-2x} + be^{x} + cxe^{x}$; $f(0) = -1$, $f'(0) = -2$, and $f''(0) = 3$.

1.4 Numerical Solutions

In theory the elimination methods developed in Section 1.2 can be used to find the solutions to *any* system of linear equations. And in practice elimination methods work fine as long as the system is not too large. However, when implemented on a computer, elimination methods can lead to the wrong answer due to round-off error. Furthermore, for very large systems elimination methods may not be efficient enough to be practical. In this section we consider some shortcomings of elimination methods, and develop alternative solution methods.

▶ This section is optional and can be omitted without loss of continuity.

▶ *In theory there is no difference between theory and practice. In practice there is.*–Yogi Berra (Also attributed to computer scientist Jan L. A. van de Snepscheut and physicist Albert Einstein.)

Round-off Error

Sensible people do not spend their time solving complicated systems of linear equations by hand—they use computers. Although computers are fast, they have drawbacks, one being the round-off errors that can arise when using floating-point representations for numbers.

For example, suppose that we have a simple computer that has only four digits of accuracy. Using this computer and Gauss–Jordan elimination to solve the system

$$7x_1 - 3x_2 + 2x_3 + 6x_4 = 13 \quad -3x_1 + 9x_2 + 5x_3 - 2x_4 = 9 \quad x_1 - 13x_2 - 3x_3 + 8x_4 = -13 \quad 2x_1 \quad - x_3 + 3x_4 = -6$$

yields the solution $x_1 = 2$, $x_2 = -0.999$, $x_3 = 3.998$, and $x_4 = -1.999$. This differs from the exact solution $x_1 = 2$, $x_2 = -1$, $x_3 = 4$, and $x_4 = -$

2 because of round-off error occurring while performing row operations. The degree of error here is not too large, but this is a small system. Elimination methods applied to larger systems will require many more arithmetic operations, which can result in accumulation of round-off errors, even on a high-precision computer.

If the right combination of conditions exists, even small systems can generate significant round-off errors.

Example 1

Suppose that we are using a computer with four digits of accuracy. Apply Gaussian elimination to find the solution to the system

3x1+1000x2= 7006 42x1−36x2=−168 (1)

▶ The choice of four digits of accuracy does not restrict us to numbers less than 10,000. For instance, a number such as 973,400 can be represented as 9.734×10^5.

Solution We need only one row operation to put the system in triangular form. The exact computations are

[3100042−36|7006−168]−14R1+R2→R2~[310000−14036|7006−98252]

Since our computer only carries four digits of accuracy, the number − 14,036 is rounded to −14,040 and −98,252 is rounded to −98,250. Thus the triangular system we end up with is

3x1+1000x2=7006 −14,040x2=−98,250

Solving for x_2, we get

x2=−98,250−14,040:≈6.998

Back substituting to solve for x_1 gives us x1=7006−1000(6.998)3:≈2.667

▶ The notation :≈ means that rounding has occurred, and that the value on the right is being assigned to the indicated variable.

The exact solution to the system is x_1 = 2 and x_2 = 7. Although the approximation for x_2 is fairly good, the approximation for x_1 is off by quite a bit. The source of the problem is that the coefficients in the equation

3x1+1000x2=7006

differ dramatically in size. During back substitution into this equation, the error in x_2 is magnified by the coefficient 1000 and only can be compensated for by the $3x_1$ term. But since the coefficient on this term is so much smaller, the error in x_1, is forced to be large.

Partial Pivoting

One way to combat round-off error is to use **partial pivoting**, which adds a step to the usual elimination algorithms. With partial pivoting, when starting on a new column we first switch the row with the largest leading entry (compared using absolute values) to the pivot position before beginning the elimination process.

For instance, with the system (1) we interchange the position of the two rows because because |42| > |3|, which gives us

42x1−36x2=−1683x1+1000x2=7006

This time the single elimination step is (shown with four digits of accuracy)

[42−3631000|−1687006]−114R1+R2→R2~[42−3601003|−1687018]

From this we have x_2 = 7018/1003 = 6.997, which is slightly less accurate than before. However, when we back substitute this value of x_2 into the equation

$$42x1-36x2=-168$$

we get x_1 = 1.997, a much better approximation for the exact value of x_1.

Full Pivoting

When Gaussian and Gauss–Jordan elimination are implemented in computer software, partial pivoting is often used to help control round-off errors. It is also possible to implement **full pivoting**, where both rows and columns are interchanged to arrange for the largest possible leading coefficient. However, full pivoting is slower and so is employed less frequently than partial pivoting.

Jacobi Iteration

It is not at all unusual for an application to yield a system of linear equations with thousands of equations and variables. In such a case, even if round-off error is controlled, elimination methods may not be efficient enough to be practical.

Diverge, Converge

Here we turn our attention to a pair of related *iterative methods* that attempt to find the solution to a system of equations through a sequence of approximations. These methods do not suffer from the round-off problems described earlier, and in many cases they are faster than elimination methods. However, they only work on systems where the number of equations equals the number of variables, and sometimes they **diverge**—that is, they fail to reach the solution. In the cases where a solution is found, we say that the method **converges**.

Our first approximation method is called **Jacobi iteration.** We illustrate this method by using it to find the solution to the system

$$10x1+4x2-x3=3\ 2x1+10x2+x3=-19\ x1-x2+5x3=-2 \qquad (2)$$

▶ Jacobi iteration is named for German mathematician Karl Gustav Jacobi (1804–1851).

Step 1: Solve the first equation of the system for x_1, the second equation for x_2, and so on,

$$x1= 0.3-0.4x2+0.1x3 \quad x2=-1.9-0.2x1-0.1x3 \quad x3=-0.4-0.2x1+0.2x2 \tag{3}$$

Step 2: Make a guess at the values of x_1, x_2, and x_3 that satisfy the system. If we have no idea about the solution, then set each equal to 0,

$$x1=0,x2=0,x3=0$$

Step 3: Substitute the values for x_1, x_2, and x_3 into (3). This is Iteration 1, and it gives the updated values:

$$\text{Iteration 1:}x1= 0.3-0.4(0)+0.1(0)= 0.3 \quad x2=-1.9-0.2(0)-0.1(0)=-1.9 \quad x3=-0.4-0.2(0)+0.2(0)=-0.4$$

Now repeat the process, substituting the new values for x_1, x_2, and x_3 into the equations in Step 1.

$$\text{Iteration 2:}x1=0.3-0.4(-1.9)+0.1(-0.4)= 1.02 \quad x2=-1.9-0.2(0.3)-0.1(-0.4)=-1.92 \quad x3=-0.4-0.2(0.3)+0.2(-1.9)=-0.84$$

We keep repeating this procedure until we have two consecutive iterations where each value differs from its predecessor by no more than the accuracy desired. Table 1 shows the outcome from the first nine iterations, each rounded to four decimal places. We see that the values have converged to $x_1 = 1$, $x_2 = -2$, and $x_3 = -1$, which is the exact solution to the system.

Table 1 **Jacobi Iterations (*n* is the iteration number)**

n	x_1	x_2	x_3
0	0	0	0

1	0.3000	−1.9000	−0.4000
2	1.0200	−1.9200	−0.8400
3	0.9840	−2.0200	−0.9880
4	1.0092	−1.9980	−1.0008
5	0.9991	−2.0018	−1.0014
6	1.0006	−1.9997	−1.0002
7	0.9999	−2.0001	−1.0001
8	1.0000	−2.0000	−1.0000
9	1.0000	−2.0000	−1.0000

▶ Table values are rounded to four decimal places, and the rounded values are carried to the next iteration.

In the next example we revisit an application introduced in Section 1.3.

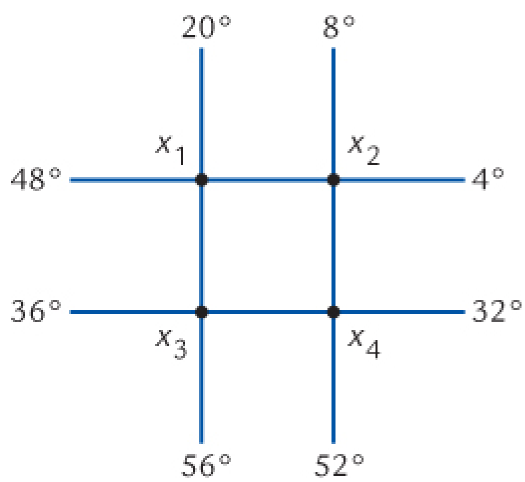


Figure 1 Grid Temperatures for Example 2.

Example 2

Figure 1 gives a diagram of a piece of heavy wire mesh. Each of the eight wire ends has temperature held fixed as shown. When the temperature of the mesh reaches equilibrium, the temperature at each connecting point will be the average of the temperatures of the adjacent points and fixed ends. Determine the equilibrium temperature at the connecting points x_1, x_2, x_3, and x_4.

Solution The temperature of each connecting point depends in part on the temperature of other connecting points. For instance, since x_1 is adjacent to x_2, x_3, and the ends held fixed at 48° and 20°, its temperature at equilibrium will be the average

$$x1x2+x3+48+204=0.25x2+0.25x3+17$$

Similarly, for the other connecting points we have the equations (after simplifying)

$$x2=0.25x1+0.25x4+3x3=0.25x1+0.25x4+23x4=0.25x2+0.25x3+21$$

In Section 1.3 these four equations were reorganized into the usual form of a linear equation and solved using elimination methods. But since each equation has one variable written in terms of the other variables, the problem sets up perfectly for Jacobi iteration. Starting with initial choices $x_1 = x_2 = x_3 = x_4 = 0$, the first two iterations are

$$\text{Iteration 1:}x1=0.25(0)+0.25(0)+17=17\ x2=0.25(0)+0.25(0)+3=3\ x3=0.25(0)+0.25(0)+23=23\ x4=0.25(0)+0.25(0)+21=21$$

$$\text{Iteration 2:}x1=0.25(3)+0.25(23)+17=23.5\ x2=0.25(17)+0.25(21)+3=12.5\ x3=0.25(17)+0.25(21)+23=32.5\ x4=0.25(3)+0.25(23)+21=27.5$$

Table 2 shows additional Jacobi iterations for this example.

Table 2 **Jacobi Iterations for Example 2**

n	x_1	x_2	x_3	x_4
4	29.8750	18.1250	38.1250	33.8750
8	31.8672	19.8828	39.8828	35.8672
12	31.9917	19.9927	39.9927	35.9917
16	31.9995	19.9995	39.9995	35.9995

20	32.0000	20.0000	40.0000	36.0000
24	32.0000	20.0000	40.0000	36.0000

This suggests equilibrium temperatures of $x_1 = 32$, $x_2 = 20$, $x_3 = 40$, and $x_4 = 36$, which matches the solution found in Section 1.3.

▶ To save space, only every fourth iteration is given in Table 2.

Gauss–Seidel Iteration

At each step of Jacobi iteration we take the values from the previous step and plug them into the set of equations, updating the values of all variables at the same time. We modify this approach with a variant of Jacobi iteration called Gauss–Seidel iteration. With this method, we always use the current value of each variable.

▶ We encountered C. F. Gauss earlier. Ludwig Philipp von Seidel (1821–1896) was a German mathematician. Interestingly, Gauss discovered the method long before Seidel but discarded it as worthless. Nonetheless, the Gauss name was attached to the algorithm along with that of Seidel, who independently discovered and published it after Gauss died.

To illustrate how Gauss–Seidel works, let's return to the system (2) considered before.

Step 1: As with Jacobi, we start by solving for x_1, x_2, and x_3,

x1= 0.3−0.4x2+0.1x3x2=−1.9−0.2x1−0.1x3x3=−0.4−0.2x1+0.2x2

Step 2: We set initial values for x_1, x_2, and x_3. In the absence of any approximation for the solution, we use

x1=0,x2=0,x3=0

Step 3: For the first part of Iteration 1, we have (again as with Jacobi)

$$x_1=0.3-0.4(0)+0.1(0)=0.3$$

It is at this point that the Jacobi and Gauss–Seidel methods begin to differ. To calculate the updated value of x_2, we use the most current variable values, which are $x_1 = 0.3$ and $x_3 = 0$.

$$x_2=-1.9-0.2(0.3)-0.1(0)=-1.96$$

We finish this iteration by updating the value of x_3, using the current values $x_1 = 0.3$ and $x_2 = -1.96$, giving us

$$x_3=-0.4-0.2(0.3)+0.2(-1.96)=-0.852$$

Subsequent iterations proceed in the same way, always incorporating the most current variable values. The second iteration is

$$\text{Iteration 2:} \quad x_1=0.3-0.4(-1.96)+0.1(-0.852)=0.9988 \quad x_2=-1.9-0.2(0.9988)-0.1(-0.852) \approx -2.0146 \quad x_3=-0.4-0.2(0.9988)+0.2(-2.01456) \approx -1.0027$$

As with Jacobi, we continue until reaching the point where two consecutive iterations yield values sufficiently close together. Table 3 gives the first six iterations of Gauss–Seidel applied to our system.

Table 3 Gauss–Seidel Iterations

n	x_1	x_2	x_3
0	0	0	0
1	0.3000	−1.9600	−0.8520
2	0.9988	−2.0146	−1.0027
3	1.0056	−2.0008	−1.0013
4	1.0002	−1.9999	−1.0000
5	1.0000	−2.0000	−1.0000
6	1.0000	−2.0000	−1.0000

Note that Gauss–Seidel converged to the solution faster than Jacobi. Since Gauss–Seidel immediately incorporates new values into the computations, it seems reasonable to expect that it would converge faster than Jacobi. Most of the time this is true, but surprisingly not always—there are systems where Jacobi iteration converges more rapidly.

Example 3

Imagine a simple economy that consists of consumers and just three industries, which we refer to as A, B, and C. These industries have annual consumer sales of 60, 75, and 40 (in billions of dollars), respectively. In addition, for every dollar of goods A sells, A requires 10 cents of goods from B and 15 cents of goods from C to support production. (For instance, maybe B sells electricity and C sells shipping services.) Similarly, each dollar of goods B sells requires 20 cents of goods from A and 5 cents of goods from C, and each dollar of goods C sells requires 25 cents of goods from A and 15 cents of goods from B. What output from each industry will satisfy both consumer and between-industry demand?

▶ This application is also discussed in Section 1.3.

Solution Let *a*, *b*, and *c* denote the total output from each of A, B, and C, respectively. The entire output for A is 60 for consumers, 0.20b for B, and 0.25c for C. Totaling this up yields the equation

$$a=60+0.20b+0.25c$$

Similar reasoning applied to industries B and C yields the equations

$$b=75+0.10a+0.15c \qquad c=40+0.15a+0.05b$$

Here we apply Gauss–Seidel iteration to find a solution. Since we are given the consumer demand for each industry, we take that as our starting point, initially setting $a = 60$, $b = 75$, and $c = 40$. For the first two iterations, we have

$$\text{Iteration 1:} \quad a = 60 + 0.20(75) + 0.25(40) = 85 \quad b = 75 + 0.10(85) + 0.15(40) = 89.5 \quad c = 40 + 0.15(85) + 0.05(89.5) = 57.225$$

$$\text{Iteration 2:} \quad a = 60 + 0.20(89.5) + 0.25(57.225) \approx 92.2063 \quad b = 75 + 0.10(92.2063) + 0.15(57.225) \approx 92.8044 \quad c = 40 + 0.15(92.2063) + 0.05(92.8044) \approx 58.4712$$

Additional iterations are shown in Table 4, and suggest convergence to $a = 93.2876$, $b = 93.1262$, and $c = 58.6494$. These match the solution found in Section 1.3.

Table 4 Gauss–Seidel Iterations for Example 3

n	*a*	*b*	*c*
0	60	75	40
1	85	89.5	57.225
2	92.2063	92.8044	58.4712
3	93.1787	93.0885	58.6312
4	93.2755	93.1222	58.6474
5	93.2863	93.1257	58.6492
6	93.2875	93.1261	58.6494
7	93.2876	93.1262	58.6494
8	93.2876	93.1262	58.6494

Table 5 Gauss–Seidel Iterations

n	x_1	x_2
0	0	0
1	6	11
2	−27	−55
3	171	341
4	−1017	−2035

5	6111	12221

Convergence

Gaussian and Gauss–Jordan elimination are called *direct methods,* because they will always yield the solution in a finite number of steps (ignoring the potential problems brought about by round-off). On the other hand, as noted earlier, Jacobi and Gauss–Seidel are iterative methods and do not converge to a solution in all cases. For instance, applying Gauss–Seidel iteration starting at $x = x_2 = 0$ to the system

$$x1+3x2=62x1-x2=1 \tag{4}$$

yields the sequence shown in Table 5. The values grow quickly in absolute value, and do not converge.

Diagonally Dominant

One case where we are guaranteed convergence is if the coefficients of the system are **diagonally dominant**. This means that for each equation of the system, the coefficient a_{ii} (in equation i) along the diagonal has absolute value larger than the sum of the absolute values of the other coefficients in the equation. For example, the system

$$7x1-3x2+2x3=6\ x1+5x2-2x3=1\ -3x1+x2-6x3=-4 \tag{5}$$

is diagonally dominant because

$$|\ 7|>|-3|+|\ 2||\ 5|>|\ 1|+|-2||-6|>|-3|+|\ 1|$$

- Diagonal dominance is not required in order for the iterative methods to converge. In some instances convergence occurs without diagonal dominance.

On the other hand, the system

$$-2x1+x2-9x3=0\ 6x1-x2+4x3=-12\ -x1+4x2-x3=3 \tag{6}$$

is not diagonally dominant as expressed, but reordering the equations to

$$6x1-x2+4x3=-12\ -x1+4x2-x3=3\ -2x1+x2-9x3=0 \qquad (7)$$

makes it diagonally dominant.

Example 4

Reverse the order of the equations in (4) to make the system diagonally dominant, and then find the solution using Gauss–Seidel iteration.

Solution Reversing the order of the equations gives us

$$2x1-x2=1x1+3x2=6$$

which is diagonally dominant. Next, we solve for x_1 and x_2 (rounded to four decimal places),

$$x1=0.5+0.5x2x2=2-0.3333x1$$

Table 6 **Gauss–Seidel Iterations for Example 4**

n	x_1	x_2
0	0	0
1	0.5	1.8333
2	1.4167	1.5278
3	1.2639	1.5787
4	1.2894	1.5702
5	1.2851	1.5716
6	1.2858	1.5714
7	1.2857	1.5714
8	1.2857	1.5714

Starting with $x_1 = 0$ and $x_2 = 0$, we have

Iteration 1:x1=0.5+0.5(0)=0.5 x2=2−0.3333(0.5):≈1.8333

Iteration 2:x1=0.5+0.5(1.8333):≈1.4167 x2=2−0.3333(1.4167):≈1.5278

The first eight iterations are shown in Table 6, which shows convergence to $x_1 = 1.2857$ and $x_2 = 1.5714$. These match the exact solutions, which are $x_1 = 9/7$ and $x_2 = 11/7$.

COMPUTATIONAL COMMENTS

- Our iterative methods do not suffer from the round-off errors that can afflict elimination methods. The values from one iteration can be thought of as an initial guess for the next, so no accumulation of errors occurs. For the same reason, if a computation error is made, the result still can be used in the next iteration. By contrast, if a computation error is made when using elimination methods, the end result is almost always wrong.
- For a system of n equations with n unknowns, Jacobi and Gauss–Seidel both require about $2n^2$ flops per iteration. As mentioned earlier, Gauss–Seidel usually converges in fewer iterations than Jacobi, so Gauss–Seidel is typically the preferred method.
- If we ignore potential round-off issues and go solely by the number of flops, then Gaussian elimination requires about $2n^3/3$ flops, versus $2n^2$ flops per iteration for Gauss–Seidel. Thus as long as Gauss–Seidel converges in fewer than $n/3$ iterations, this will be the more efficient method.
- The rate of convergence of our iterative methods is influenced by the degree of diagonal dominance of the

system. If the diagonal terms are much larger than the others, then iterative methods generally will converge relatively quickly. If the diagonal terms are only slightly dominant, then although iterative methods eventually will converge, they can be too slow to be practical. There are other iterative methods besides those presented here that are designed to have better convergence properties.

Sparse System, Sparse Matrix

- Iterative methods are particularly useful for solving **sparse systems**, which are linear systems where most of the coefficients are zero. The augmented matrix of such a system has mostly zero entries and is said to be a **sparse matrix**. Elimination methods applied to sparse systems have a tendency to change the zeros to nonzero terms, removing the sparseness.

▶ See *Matrix Computations* by G. Golub and C. Van Loan for a more extensive discussion of iterative methods and an explanation of why those described here work.

▶ Practice problems can also be used as additional examples.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Solve the system as given with Gaussian elimination with three significant digits of accuracy. Then solve the system again, incorporating partial pivoting.

(a) x1+563x2=5249x1−78x2=−11

(b) 2x1−8x2+598x3=15−3x1+7x2+913x3=567x1−39x2+84x3=11

2. Compute the first three Jacobi iterations for the given system, using 0 as the initial value for each variable. Then find the exact solution and compare.

(a) −4x1+x2=92x1+13x2=5

(b) 23x1+2x2+4x3=21−x1−9x2+4x3=272x1−3x2−8x3=−11

3. Compute the first three Gauss–Seidel iterations for the given system, using 0 as the initial value for each variable. Then compare to the results found using Jacobi iteration above.

(a) −4x1+x2=92x1+13x2=5

(b) 23x1+2x2+4x3=21−x1−9x2+4x3=272x1−3x2−8x3=−11

EXERCISES

In each exercise set, problems marked with are designed to be solved using a programmable calculator or computer algebra system.

Exercises 1–4: Use partial pivoting with Gaussian elimination to find the solutions to the system.

1. −2x1+3x2=45x1−2x2=1

2. x1−2x2=−1−3x1+7x2=5

3. x1+x2−2x3=−33x1−2x2+2x3=96x1−7x2−x3=4

4. x1−3x2+2x3=4−2x1+7x2−2x3=−74x1−13x2+7x3=12

Exercises 5–8: Solve the system as given using Gaussian elimination with three significant digits of accuracy. Then solve the system again, incorporating partial pivoting.

5. 2x1+975x2=4153x1−82x2=−13

6. 3x1−813x2=3271x1−93x2=−5

7. 3x1−7x2+639x3=12−2x1+5x2+803x3=756x1−41x2+79x3=10

8. 2x1−5x2+802x3=−1−x1+3x2−789x3=−840x1+34x2+51x3=19

Exercises 9–12: Compute the first three Jacobi iterations for the system, using 0 as the initial value for each variable. Then find the exact solution and compare.

9. $-5x1+2x2=63x1+10x2=2$
10. $2x1-x2=-4-4x1+5x2=11$
11. $20x1+3x2+5x3=-26-2x1-10x2+3x3=-23x1-2x2-5x3=-13$
12. $-2x1\ \ +x3=5-x1+5x2-x3=82x1-6x2+10x3=16$

Exercises 13–16: Compute the first three Gauss–Seidel iterations for the system, using 0 as the initial value for each variable. Then find the exact solution and compare.

13. The system given in Exercise 9.
14. The system given in Exercise 10.
15. The system given in Exercise 11.
16. The system given in Exercise 12.

Exercises 17–20: Determine if the system is diagonally dominant. If not, then (if possible) rewrite the system so that it is diagonally dominant.

17. $2x1-5x2=73x1+7x2=4$
18. $4x1+2x2-x3=13-2x1+7x2+2x3=-9x1+3x2-5x3=6$
19. $3x1+6x2-x3=0-x1-2x2+4x3=-17x1+5x2-3x3=3$
20. $-2x1+6x2=125x1-x2=-4$

Exercises 21–24: Compute the first four Jacobi iterations for the system as written, with the initial value of each variable set equal to 0. Then rewrite the system so that it is diagonally dominant, set the value of each variable to 0, and compute 4 Jacobi iterations.

21. $x1-2x2=-12x1-x2=1$
22. $x1-3x2=-23x1-x2=2$
23. $x1-2x2+5x3=-15x1+x2-2x3=82x1-10x2+3x3=-1$
24. $2x1+4x2-10x3=-33x1-x2+x3=7-x1+6x2-2x3=-6$

Exercises 25–28: Compute the first four Gauss–Seidel iterations for the system as written, with the initial value of each variable set equal to 0. Then rewrite the system so that it is diagonally dominant, set the value of each variable to 0, and compute four Gauss–Seidel iterations.

25. The system given in Exercise 21.

26. The system given in Exercise 22.

27. The system given in Exercise 23.

28. The system given in Exercise 24.

Exercises 29–30: The values from the first few Jacobi iterations are given for an unknown system. Find the values for the next iteration.

29.

n	x_1	x_2
0	0	0
1	1	−2
2	5	2
3	?	?

30.

n	x_1	x_2	x_3
0	0	0	0
1	−2	−1	1
2	−4	−4	5
3	−11	−4	5
4	?	?	?

Exercises 31–32: The values from the first few Gauss–Seidel iterations are given for an unknown system. Find the values for the next iteration.

31.

n	x_1	x_2
0	0	0

n	x_1	X_2
1	3	4
2	− 5	−12
3	?	?

32.

n	x_1	x_2	x_3
0	0	0	0
1	3	4	12
2	7	−24	− 76
3	− 25	176	556
4	?	?	?

SUPPLEMENTARY EXERCISES

Exercises 1–6: Find all solutions to the system by eliminating one of the variables.

1. 2x1−5x2=13x1+7x2=3
2. 3x1−x2=2−5x1+2x2=−1
3. x1−4x2=1−2x1+8x2=−2
4. 4x1−2x2=66x1+3x2=9
5. x1−3x2=53x1−9x2=7
6. −6x1+2x2=315x1−5x2=4

Exercises 7–14: Find the set of solutions for the linear system.

7. x1+2x2−4x3=0 −x2+3x3=−2 2x3=6
8. x1 −4x3=3 2x2−6x3=4
9. $-x_1 - 5x_2 + x_3 = -2$
10. $x_1 + 2x_2 + 4x_3 - x_4 = -2$
11. −x1−2x2+7x3−3x4=7 −x2 +x4=0 3x4=−15
12. 2x1−x2 −x4=−2 −x3+3x4=1
13. x1+x2+3x3−x4+x5=7 x2−4x3 +2x5=−3 −2x4−x5=0 −2x 5=−8
14. 2x1 +4x3 +3x5=−1 x2−x3+x4 =2

Exercises 15–18: Convert the augmented matrix to the equivalent linear system.

15. [2−43−3511|10]
16. [322−5030−2|76]
17. [4257−20312|11−4]
18. [13−220−5044322|110−21]

Exercises 19–22: Transform the matrix to echelon form.

19. [1−2132−655−16−73]

20. [−11−212−56−1−1−866]

21. [−32−202−96−3256−4106−7]

22. [2−84211−3022−122−45−311−422]

Exercises 23–26: Transform the matrix to reduced row echelon form.

23. [1−34−26−72−66]

24. [12−3−1−2−343−3−455]

25. [12−321213−511222−2418]

26. [1−27133−51912−26−18−15−17103−13−10]

Exercises 27–30: Find the general solution for the linear system.

27. 2x1−5x2+x3=−1−x1+3x2−x3=1

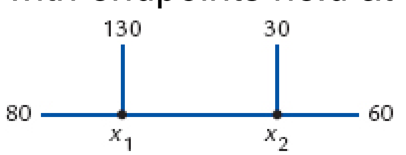
28. x1−3x2+4x3=1−2x1+5x2−7x3=1x1−5x2+8x3=5

29. x1−3x2+x3+2x4=2−x1+4x2−4x3−x4=−42x1−3x2−7x3+8x4=−5

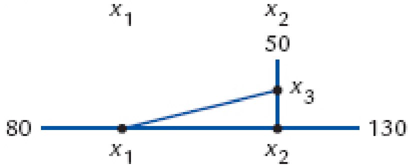
30. 2x1+4x2+9x3−5x4+2x5=−5x1+2x2+4x3−x4+2x5=−1−3x1−6x2−14x3+9x4−3x5=14

Exercises 31–32: Find the equilibrium temperatures for the heavy wires with endpoints held at the given temperatures.

31.



32.



Exercises 33–34: Determine the total output required from each industry in order to meet both consumer and between-industry demand.

33. Economy has two industries A and B. These industries have annual consumer sales of 50 and 20 (in billions of dollars), respectively. For every dollar of goods A sells, A requires 30 cents of goods from B. For each dollar of goods B sells, B requires 50 cents of goods from A.

34. Economy has three industries A, B, and C. These industries have annual consumer sales of 40, 70, and 90 (in billions of dollars), respectively. For every dollar of goods A sells, A requires 20 cents of goods from B and 25 cents of goods from C. For each dollar of goods B sells, B requires 25 cents of goods from A and 30 cents of goods from C. For each dollar of goods C sells, C requires 10 cents of goods from A and 30 cents of goods from B.

Exercises 35–36: Balance the given chemical equation.

35. The reaction for one form of anaerobic respiration:

$$C6H12O6 \rightarrow C2H5OH+CO2$$

36. The reaction between baking powder and vinegar when mixed:

$$HC2H3O2+NaHCO3 \rightarrow NaC2H3O2+H2O+CO2$$

Exercises 37–38: Solve the system as given with Gaussian elimination with three significant digits of accuracy. Then solve the system again, incorporating partial pivoting.

37. $3x1+891x2=3748x1-91x2=-12$

38. $x1-6x2+745x3=17-3x1+4x2+902x3=849x1-39x2+81x3=10$

Exercises 39–40: Compute the first three Jacobi iterations for the given system, using 0 as the initial value for each variable. Then find the exact solution and compare.

39. $-10x1+3x2=82x1+5x2=3$

40. $10x1+2x2+3x3=29-7x1-20x2+2x3=212x1-x2-5x3=-11$

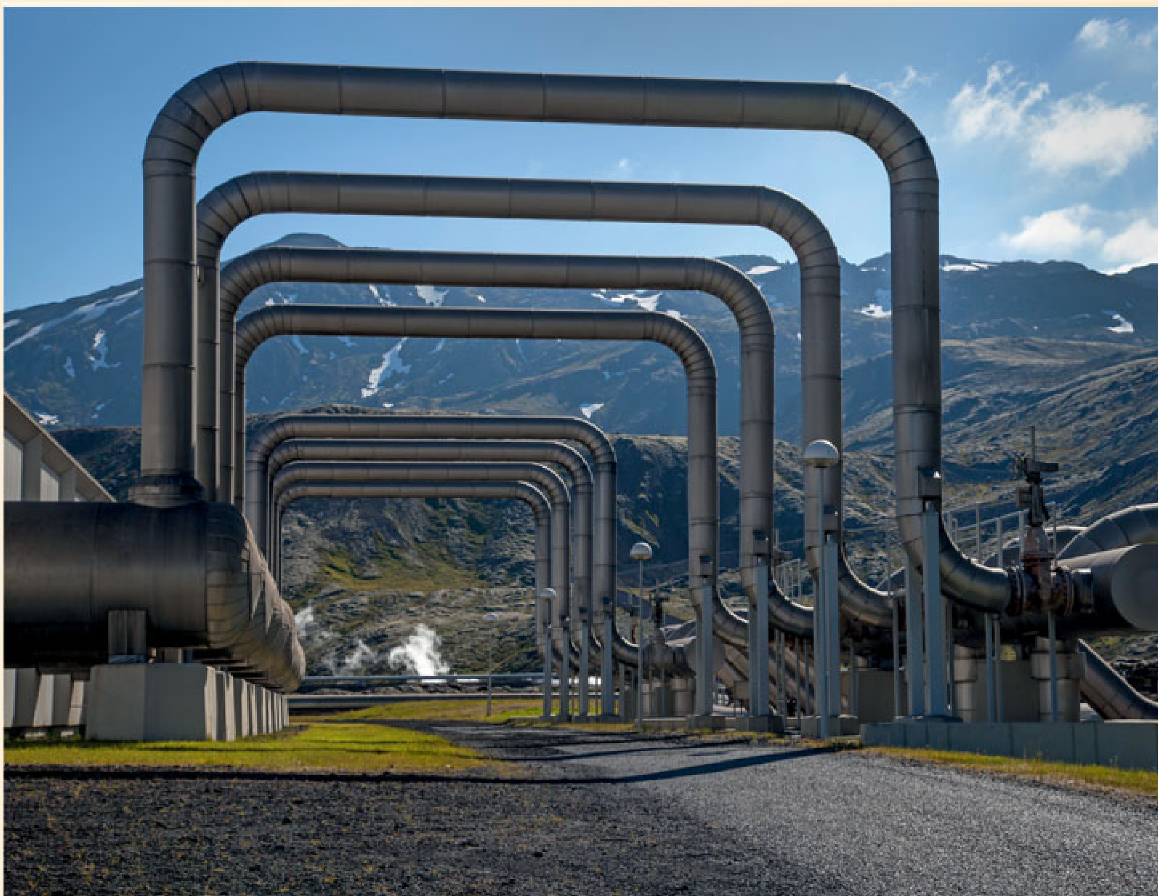
Exercises 41–42: Compute the first three Gauss–Seidel iterations for the given system, using 0 as the initial value for each variable. Then find the exact solution and compare.

41. The system given in Exercise 39.

42. The system given in Exercise 40.

CHAPTER 2

Euclidean Space



Gunnar Örn Árnason/Moment/Getty Images

Shown here is a geothermal power station located in Iceland. Geothermal power uses steam generated and stored in the earth to produce electricity. The three basic types of geothermal power plants are dry steam, flash steam, and binary cycle. Geothermal power is considered a sustainable, renewable source of energy that is cleaner than burning fossil fuels. However, gases from inside the earth can contribute to global warming and acid rain if they are released into the air, but with the binary cycle system there are no emissions.

We can think of algebra as the study of the properties of arithmetic performed on numbers. In linear algebra, we study the properties of arithmetic performed on objects called *vectors*. As we shall see, we can use vectors to express a system of linear equations and to compactly describe the set of solutions of a linear system. But vectors have many other applications as well. Section 2.1 gives an introduction to vectors, arithmetic with vectors, and the geometry of vectors. Section 2.2 and Section 2.3 describe important properties of sets of vectors.

2.1 Vectors

Fertilizer is sold in bags labelled with three numbers that indicate the amount of nitrogen (N), phosphoric acid (P_2O_5), and potash (K_2O) present. The mixture of these nutrients varies from one type of fertilizer to the next. For example, a bag of Vigoro Ultra Turf has the numbers "29–3–4" which means that 100 pounds of this fertilizer contains 29 pounds of nitrogen, 3 pounds of phosphoric acid, and 4 pounds of potash. Organizing these quantities vertically in a matrix, we have

$$[2934]$$

This representation is an example of a *vector*. Using a vector provides a convenient way to record the amounts of each nutrient and also lends itself to compact forms of algebraic operations that arise naturally. For instance, if we want to know the amount of nitrogen, phosphoric acid, and potash contained in a ton (2000 pounds) of Ultra Turf, we just multiply each vector entry by 20. Thinking of this as multiplying the vector by 20, we represent this operation by

$$20[2934]$$

so that we have

$$20[2934]=[20\cdot2920\cdot320\cdot4]=[5806080]$$

The "=" between vectors means that the entries in corresponding positions are equal.

Another type of fertilizer, Parker's Premium Starter, has 18 pounds of nitrogen, 25 pounds of phosphoric acid, and 6 pounds of potash per 100 pounds, which is represented in vector form by

[18256]

If we mix together 100 pounds of each type of fertilizer, then we can find the total amount of each nutrient in the mixture by adding entries in each of the vectors. Thinking of this as addition of vectors, we have

[2934]+[18256]=[29+183+254+6]=[472810]

Vectors and $\mathbf{R}^n$

We formalize our notion of vector with the following definition.

DEFINITION 2.1 ▶

Vector

A **vector** is an ordered list of real numbers $u_1, u_2, \ldots, u_n$ expressed as

u=[u1u2:un]

$\mathbf{R}^n$

or as $\mathbf{u} = (u_1, u_2, \cdots, u_n)$. The set of all vectors with n entries is denoted by $\mathbf{R}^n$.

Component, Column Vector, Row Vector

Our convention will be to denote vectors using boldface, such as $\mathbf{u}$. Each of the entries $u_1, u_2, \ldots, u_n$ is called a **component** of the vector. A vector expressed in the vertical form is also called a **column vector**, and a vector expressed in horizontal form is also called a **row vector**. It is customary to express vectors in column form, but we will occasionally use row form to save space.

The fertilizer discussion provides a good model for how vector arithmetic works. Here we formalize the definitions.

DEFINITION 2.2 ▶

Let **u** and **v** be vectors in $\mathbf{R}^n$ given by

u=[u1u2⋮un]andv=[v1v2⋮vn]

Vector Arithmetic, Scalar, Euclidean Space

Suppose that *c* is a real number, called a **scalar** in this context. Then we have the following definitions:

Equality:u=v if and only if u1=v1, u2=v2, … , un=vn.Addition:u+v=[u1u2⋮un]+[v1v2⋮vn]=[u1+v1u2+v2 ⋮ un+vn]Scalar Multiplication:cu=c[u1u2⋮un]=[c·u1c·u2 ⋮ c·un]

The set of all vectors in $\mathbf{R}^n$, taken together with these definitions of addition and scalar multiplication, is called **Euclidean space**.

▶ Euclidean space is named for the Greek mathematician Euclid, the father of geometry. Euclidean space is an example of a *vector space*, discussed in Chapter 7.

Although vectors with negative components and negative scalars do not make sense in the fertilizer discussion, they do in other contexts and are included in Definition 2.2.

Example 1

Suppose that we have the vectors in $\mathbf{R}^4$

u=[2-30-1]andv=[-46-27]

Find **u** + **v**, −4**v**, and 2**u** − 3**v**.

▶ Two vectors can be equal only if they have the same number of components. Similarly, there is no way to add two vectors that have a different number of components.

Solution The solutions to the first two parts are

u+v=[2-30-1]+[-46-27]=[2-4-3+60-2-1+7]=[-23-26]

-4v=-4[-46-27]=[-4(-4)-4(6)-4(-2)-4(7)]=[16-248-28]

The third computation has a slight twist because we have not yet defined the difference of two vectors. But subtraction works exactly as we would expect and follows from the natural interpretation that 2**u** − 3**v** = 2**u** + (−3)**v**.

2u-3v=2[2-30-1]-3[-46-27]=
[2(2)-2(-4)2(-3)-3(6)2(0)-3(-2)2(-1)-3(7)]=[16-246-23]

Many of the properties of arithmetic of real numbers, such as the commutative, distributive, and associative laws, carry over as properties of vector arithmetic. These are summarized in the next theorem.

THEOREM 2.3 ▶

(ALGEBRAIC PROPERTIES OF VECTORS)

Let a and b be scalars, and **u**, **v**, and **w** be vectors in $\mathbf{R}^n$. Then

(a) **u** + **v** = **v** + **u**

(b) a(**u** + **v**) = a**u** + a**v**

(c) (a + b)**u** = a**u** + b**u**

(d) (**u** + **v**) + **w** = **u** + (**v** + **w**)

(e) $a(b\mathbf{u}) = (ab)\mathbf{u}$

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

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

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

▶ The **zero vector** is given by
0=[00⋮0]
and $-\mathbf{u} = (-1)\mathbf{u}$.

Proof Let

u=[u1u2⋮un]andv=[v1v2⋮vn]

The components of each vector are real numbers, so we have

u1+v1=v1+u1,…un+vn=vn+un

Thus

u+v=[u1+v1u2+v2 ⋮ un+vn]=[v1+u1v2+u2 ⋮ vn+un]=v+u

which proves (a). For (b), suppose that *a* is a scalar. Since

a(u1+v1)=au1+av1,…a(un+vn)=aun+avn

it follows that

a(u + v)=a[u1+v1u2+v2⋮un+vn]=[a(u1+v1)a(u2+v2)⋮a(un+vn)]=
[au1+av1au2+av2⋮aun+avn]=av+au

Therefore (b) is true. Proofs of the remaining properties have a similar flavor and are left as exercises. ■■

Linear Combinations and Systems of Equations

Let’s return to the fertilizer example from the beginning of this section. We have two different kinds, Vigoro and Parker’s, with nutrient vectors given by

$$\text{Vigoro: } \mathbf{v} = \begin{bmatrix} 29 \\ 3 \\ 4 \end{bmatrix} \qquad \text{Parker's: } \mathbf{p} = \begin{bmatrix} 18 \\ 25 \\ 6 \end{bmatrix}$$

By using vector arithmetic, we can find the nutrient vector for combinations of the two fertilizers. For example, if 500 pounds of Vigoro and 300 pounds of Parker’s are mixed, then the total amount of each nutrient is given by

$$5\mathbf{v} + 3\mathbf{p} = 5\begin{bmatrix} 29 \\ 3 \\ 4 \end{bmatrix} + 3\begin{bmatrix} 18 \\ 25 \\ 6 \end{bmatrix} = \begin{bmatrix} 145 + 54 \\ 15 + 75 \\ 20 + 18 \end{bmatrix} = \begin{bmatrix} 199 \\ 90 \\ 38 \end{bmatrix}$$

The sum $5\mathbf{v} + 3\mathbf{p}$ is an example of a *linear combination* of vectors.

DEFINITION 2.4 ▶

If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are vectors and $c_1, c_2, \ldots, c_m$ are scalars, then

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m$$

Linear Combination

is a **linear combination** of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$. Note that it is possible for scalars to be negative or equal to zero.

Linear combinations provide an alternate way to express a system of linear equations. For example, consider

$$x_1\begin{bmatrix} 5 \\ -1 \\ 6 \end{bmatrix} + x_2\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} 7 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 12 \\ -4 \\ 11 \end{bmatrix} \quad \text{and} \quad \begin{array}{rcrcrcr} 5x_1 & - & 3x_2 & + & 7x_3 & = & 12 \\ -x_1 & + & 2x_2 & + & 5x_3 & = & -4 \\ 6x_1 & + & x_2 & & & = & 11 \end{array}$$

These are equivalent, meaning that both are satisfied by the same choices of x_1, x_2, and x_3. The equation on the left is an example of a **vector equation**.

Example 2

If possible, find the amount of Vigoro and Parker's required to create the following mixtures containing:

(a) 148 pounds of nitrogen, 131 pounds of phosphoric acid, 38 pounds of potash

(b) 294 pounds of nitrogen, 168 pounds of phosphoric acid, 60 pounds of potash

Solution We can formulate each problem in terms of a linear combination. For part (a) we need to find scalars x_1 and x_2 such that

x1[2934]+x2[18256]=[14813138] (1)

This vector equation is equivalent to the system

29x1+18x2=1483x1+25x2=1314x1+6x2=38

▶ The row operations used in (2) and (4) are (in order performed):
R1↔R3-(3/4)R1+R2→R2-
(29/4)R1+R3→R3(51/41)R2+R3→R3(2/41)R2→R2

The augmented matrix and echelon form are

[291832546|14813135]~[460100|3850] (2)

Back substitution gives the solution x_1 = 2 and x_2 = 5, so combining 200 pounds of Vigoro and 500 pounds of Parker's will provide the desired mixture.

For part (b), we need to find x_1 and x_2 such that x1[2934]+x2[18256]=[29416860] (3)

The corresponding augmented matrix and echelon form are

[291832546|29416860]~[460100|60612] (4)

The only difference between (2) and (4) is the rightmost column. The row operations are exactly the same. This time the last row corresponds to the equation 0 = 12, which has no solutions. Thus (3) has no solutions, so the combination in (b) is not possible.

Solutions as Linear Combinations

The solution to Example 2 can be expressed in the form of a vector,

x=[x1x2]=[25]

Vector Form

In fact, the general solution to any system of linear equations can be expressed as a linear combination of vectors, called the **vector form** of the general solution.

Example 3

Express the general solution to the linear system

2x1-3x2+10x3=-2x1-2x2+3x3=-2-x1+3x2+x3=4

in vector form.

Solution In Example 3 of Section 1.2, we found the general solution to this system. Separating each part of the general solution into the constant term and the term multiplied by the parameter s_1, we have x1=2 -11s1=2-11s1x2=2 -4s1=2-4s1x3=s1 =0+1s1

Thus the vector form of the general solution is

x=[x1x2x3]=[220]+s1[-11-41]

where s_1 can be any real number.

A more complicated general solution arises in Example 5 of Section 1.2. There we found the general solution

x1=-5-4s1+2s2-4s3=-5-4s1+2s2-4s3x2= s1 =0+1s1+0s2+0s3x3=-14 -5s3=-14+0s1+0s2-5s3x4= s2 =0+0s1+1s2+0s3x5=3 +s3=3+0s1+0s2+1s3x6= s3=0+0s1+0s2+1s3

where s_1, s_2, and s_3 can be any real numbers. In vector form, the general solution is given by

x=[x1x2x3x4x5x6]=[-50-14030]+s1[-410000]+s2[200100]+s3[-40-5011]

Geometry of Vectors

Tip, Tail of Vector

Vectors have a geometric interpretation that is most easily understood in $\mathbf{R}^2$. We plot the vector [x1x2] by drawing an arrow from the origin to the point (x_1, x_2) in the plane. For example, the vectors (2, 3) and (3, −1) are illustrated in Figure 1. Using an arrow to denote a vector suggests a direction, which is a common interpretation in physics and other sciences, and will frequently be useful for us as well. We call the end of the vector with the arrow the **tip**, and the end at the origin the **tail**.

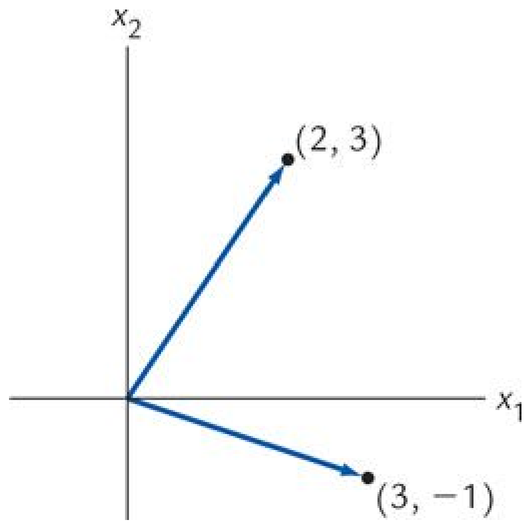


Figure 1 Vectors in $\mathbf{R}^2$.

Note that the ordered pair for the point (x_1, x_2) looks the same as the row vector (x_1, x_2). The difference between the two is that vectors have an algebraic and geometric structure that is not associated with points. Most of the time we focus on vectors, so use that interpretation unless the alternative is clearly appropriate.

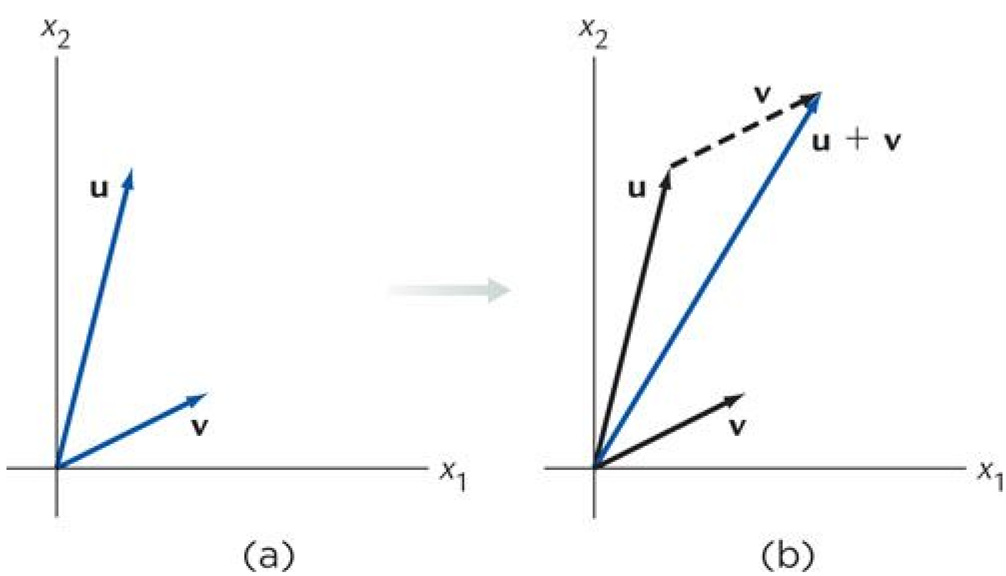


Figure 2 (a) The vectors **u** and **v**. (b) The vector **u** + **v**.

There are two related geometric procedures for adding vectors.

1. The Tip-to-Tail Rule: Let **u** and **v** be two vectors. Translate the graph of **v**, preserving direction, so that its tail is at the tip of **u**. Then the tip of the translated **v** is at the tip of **u** + **v**.

Figure 2(a) shows vectors **u** and **v**, and Figure 2(b) shows **u**, the translated **v** (dashed), and **u** + **v**.

The Tip-to-Tail Rule makes sense from an algebraic standpoint. When we add **v** to **u**, we add each component of **v** to the corresponding component of **u**, which is exactly what we are doing geometrically. We also see in Figure 2(b) that we get to the same place if we translate **u** instead of **v**.

The second rule follows easily from the first.

2. The Parallelogram Rule: Let vectors **u** and **v** form two adjacent sides of a parallelogram with vertices at the origin, the tip of **u**, and the tip of **v**. Then the tip of **u** + **v** is at the fourth vertex.

Figure 3 illustrates the Parallelogram Rule. We see that the third and fourth sides of the parallelogram are translated copies of **u** and **v**, which shows the connection to the Tip-to-Tail Rule.

Scalar multiplication and subtraction also have nice geometric interpretations.

Scalar Multiplication: If a vector **u** is multiplied by a scalar c, then the new vector c**u** points in the same direction as **u** when $c > 0$ and in the opposite direction when $c < 0$. The length of c**u** is equal to the length of **u** multiplied by $|c|$.

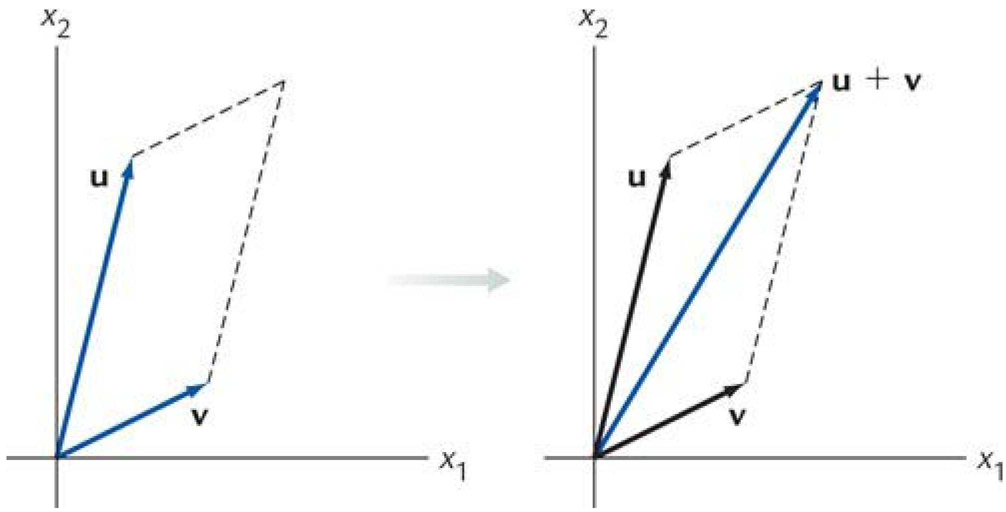


Figure 3 The Parallelogram Rule for vector addition.

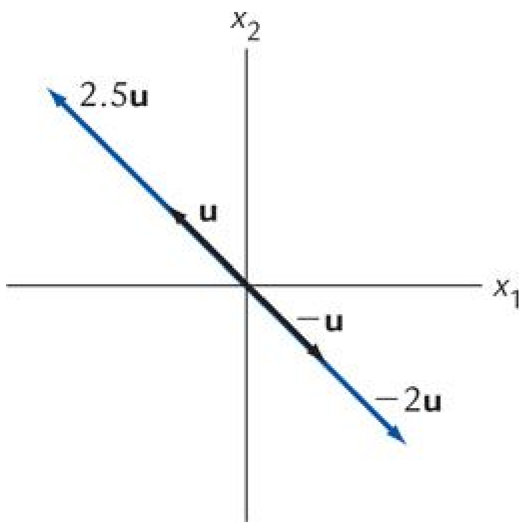


Figure 4 Scalar multiples of the vector **u**.

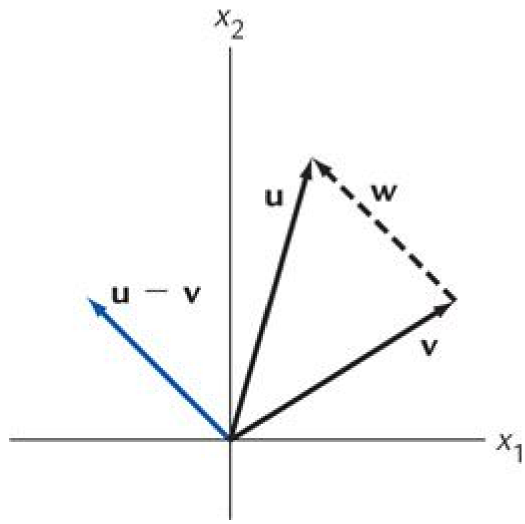


Figure 5 Subtracting vectors.

For example, −2**u** points in the opposite direction of **u** and is twice as long. (We will consider how to find the length of a vector later in the book.) A few examples of scalar multiples, starting with **u** = (−2, 3), are shown in Figure 4.

> **Subtraction:** Draw a vector **w** from the tip of **v** to the tip of **u**. Then translate **w**, preserving direction and placing the tail at the origin. The resulting vector is **u** − **v**.
>
> The subtraction procedure is illustrated in Figure 5 and is considered in more detail in Exercise 80.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Compute **u** − **w**, **v** + 3**u**, and −2**w** + **u** + 3**v** for

u=[-434]v=[-162]andw=[50-2]

2. Express the vector equation as a linear system.
   (a) x1[-172]+x2[46-6]=[3105]

(b) x1[3402]+x2[0-2-56]+x3[-1295]=[4711-6]

3. Express the linear system as a vector equation.

(a) x1+x2-2x3=3-5x1+7x2+6x3=124x1 -8x3=0

(b) 4x1-3x2-x3+5x4=0 2x2+5x3-2x4=63x1+12x2+6x3 =10

4. Express the general solution as a linear combination of vectors.

(a) x1=5+3s1x2=7-2s1x3=s1

(b) x1=1+2s1+13s2x2=s2x3=17+s1x4=s1

5. Determine if **b** is a linear combination of the other vectors. If so, express **b** as a linear combination.

(a) a1=[1-5],a2=[36],b=[59]

(b) a1=[1-38],a2=[-23-3],b=[75-4]

6. Determine if each statement is true or false, and justify your answer.

(a) Addition of vectors is commutative but not associative.

(b) The scalars in a linear combination can be negative or zero.

(c) The set of solutions to a vector equation always can be expressed as a linear combination of vectors with the same number of components.

(d) The Parallelogram Rule gives a geometric interpretation of scalar multiplication.

EXERCISES

Exercises 1–6: Let

u=[3-20]v=[-415]andw=[2-7-1]

1. Compute **u** − **v** and 6**w**.

2. Compute **w** − **u** and −5**v**.

3. Compute **w** + 3**v** and 2**w** − 7**v**.

4. Compute 4**w** − **u** and −2**v** + 5**w**.

5. Compute −**u** + **v** + **w** and 2**u** − **v** + 3**w**.

6. Compute 3**u** − 2**v** + 5**w** and −4**u** + 3**v** − 2**w**.

Exercises 7–10: Express the vector equation as a system of linear equations.

7. x1[32]+x2[-15]=[813]

8. x1[-16-4]+x2[9-50]=[-7-113]

9. x1[-65]+x2[5-3]+x3[02]=[416]

10. x1[2783]+x2[0242]+x3[5161]+x4[4570]=[0435]

Exercises 11–14: Express the system of linear equations as a vector equation.

11. 2x1+8x2-4x3=-10-x1-3x2+5x3=4

12. -2x1+5x2-10x3=4x1-2x2+3x3=-17x1-17x2+34x3=-16

13. x1-x2-3x3-x4=-1-2x1+2x2+6x3+2x4=-1-3x1-3x2+10x3 =5

14. -5x1+9x2=133x1-5x2=-9x1-2x2=-2

Exercises 15–20: The general solution to a linear system is given. Express this solution as a linear combination of vectors.

15. x1=-4+3s1x2=s1

16. x1=-2s1x2=s1

17. x1=7-2s1x2=-3x3=s1

18. x1=1+3s1-4s2x2=-2+5s2x3=s2x4=s1

19. x1=4+6s1-5s2x2=s2x3=-9+3s1x4=s1

20. x1=1-7s1+14s2-s3x2=s3x3=s2x4=-12+s1x5=s1

Exercises 21–24: Find three different vectors that are a linear combination of the given vectors.

21. u=[3-2],v=[-1-4]

22. u=[71-13],v=[5-32]

23. u=[-40-3],v=[-2-15],w=[9611]

24. u=[1822],v=[4-25-5],w=[9901]

Exercises 25–30: Find the unknown entries in the vector equation.

25. -3[a3]+4[-1b]=[-1019]

26. 4[4a]+3[-35]-2[b8]=[-17]

27. -[-1a2]+2[3-2b]=[c-78]

28. -[a-30]-[1b5]=[42c]

29. -[12a1]+2[b1-23]-[2c50]=[-3-43d]

30. -[a4-2-1]+2[51b3]-[2c-3-6]=[11-43d]

Exercises 31–36: Determine if **b** is a linear combination of the other vectors. If so, express **b** as a linear combination.

31. a1=[-25],a2=[7-3],b=[89]

32. a1=[4-6],a2=[-69],b=[1-5]

33. a1=[2-31],a2=[03-3],b=[1-5-2]

34. a1=[2-31],a2=[03-3],b=[63-9]

35. a1=[121],a2=[-35-3],a3=[224],b=[1-23]

36. a1=[2-31],a2=[03-3],a3=[-2-13],b=[2-45]

Exercises 37–38: Refer to Vigoro and Parker's fertilizers described at the beginning of the section. Determine the total amount of nitrogen, phosphoric acid, and potash in the given mixture.

37. 200 pounds of Vigoro, 100 pounds of Parker's.

38. 400 pounds of Vigoro, 700 pounds of Parker's.

Exercises 39–42: Refer to Vigoro and Parker's fertilizers described at the beginning of the section. Determine the amount of each type required to produce a mixture containing the given amounts of nitrogen, phosphoric acid, and potash.

39. 112 pounds of nitrogen, 81 pounds of phosphoric acid, and 26 pounds of potash.

40. 285 pounds of nitrogen, 284 pounds of phosphoric acid, and 78 pounds of potash.

41. 123 pounds of nitrogen, 59 pounds of phosphoric acid, and 24 pounds of potash.

42. 159 pounds of nitrogen, 109 pounds of phosphoric acid, and 36 pounds of potash.

Exercises 43–46: Refer to Vigoro and Parker's fertilizers described at the beginning of the section. Show that it is not possible to combine Vigoro and Parker's to obtain the specified mixture of nitrogen, phosphoric acid, and potash.

43. 148 pounds of nitrogen, 131 pounds of phosphoric acid, and 40 pounds of potash.

44. 100 pounds of nitrogen, 120 pounds of phosphoric acid, and 40 pounds of potash.

45. 25 pounds of nitrogen, 72 pounds of phosphoric acid, and 14 pounds of potash.

46. 301 pounds of nitrogen, 8 pounds of phosphoric acid, and 38 pounds of potash.

Exercises 47–50: One 8.3-ounce can of Red Bull contains energy in two forms: 27 grams of sugar and 80 milligrams of caffeine. One 23.5-ounce can of Jolt Cola contains 94 grams of sugar and 280 milligrams of caffeine. Determine the number of cans of each drink that when combined will contain the specified heart-pounding combination of sugar and caffeine.

47. 148 grams sugar, 440 milligrams caffeine.

48. 309 grams sugar, 920 milligrams caffeine.

49. 242 grams sugar, 720 milligrams caffeine.

50. 457 grams sugar, 1360 milligrams caffeine.

Exercises 51–54: One serving of Lucky Charms contains 10% of the percent daily values (PDV) for calcium, 25% of the PDV for iron, and 25% of the PDV for zinc, and one serving of Raisin Bran contains 2% of the PDV for calcium, 25% of the PDV for iron, and 10% of the PDV for zinc. Determine the number of servings of each cereal required to create the given mix of nutrients.

51. 40% of the PDV for calcium, 200% of the PDV for iron, and 125% of the PDV for zinc.

52. 34% of the PDV for calcium, 125% of the PDV for iron, and 95% of the PDV for zinc.

53. 26% of the PDV for calcium, 125% of the PDV for iron, and 80% of the PDV for zinc.

54. 38% of the PDV for calcium, 175% of the PDV for iron, and 115% of the PDV for zinc.

55. An electronics company has two production facilities, denoted A and B. During an average week, facility A produces 2000 computer monitors and 8000 flat panel televisions, and facility B produces 3000 computer monitors and 10,000 flat panel televisions.

(a) Find vectors **a** and **b** that give the weekly production amounts at A and B, respectively.

(b) Compute 8**b**, and then describe what the entries tell us.

(c) Determine the combined output from A and B over a 6-week period.

(d) Determine the number of weeks of production from A and B required to produce 24,000 monitors and 92,000 televisions.

56. An industrial chemical company has three facilities, denoted A, B, and C. Each facility produces polyethylene (PE), polyvinyl chloride (PVC), and polystyrene (PS). The table below gives the daily production output (in metric tons) for each facility:

	Facility		
Product	**A**	**B**	**C**
PE	10	20	40
PVC	20	30	70
PS	10	40	50

(a) Find vectors **a**, **b**, and **c** that give the daily production amounts at each facility.

(b) Compute 20**c**, and describe what the entries tell us.

(c) Determine the combined output from all three facilities over a 2-week period. (Note: The facility does not operate on weekends.)

(d) Determine the number of days of production from each facility required to produce 240 metric tons of polyethylene, 420 metric tons of polyvinyl chloride, and 320 metric tons of polystyrene.

Exercises 57–60: Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be vectors, and suppose that a *point mass* of $m_1, \ldots, m_k$ is located at the tip of each vector. The *center of* *mass* for this set of point masses is equal to

$$v¯=m1v1+\cdots+mkvkm$$

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

57. Let $\mathbf{u}_1 = (3, 2)$ have mass 5kg, $\mathbf{u}_2 = (-1, 4)$ have mass 3kg, and $\mathbf{u}_3 = (2, 5)$ have mass 2kg. Graph the vectors and determine the center of mass.

58. Determine the center of mass for the vectors $\mathbf{u}_1 = (-1, 0, 2)$ (mass 4kg), $\mathbf{u}_2 = (2, 1, -3)$ (mass 1kg), $\mathbf{u}_3 = (0, 4, 3)$ (mass 2kg), and $\mathbf{u}_4 = (5, 2, 0)$ (mass 5kg).

59. Determine how to divide a total mass of 11kg among the vectors $\mathbf{u}_1 = (-1, 3)$, $\mathbf{u}_2 = (3, -2)$, and $\mathbf{u}_3 = (5, 2)$ so that the center of mass is (1311, 1611).

60. Determine how to divide a total mass of 11kg among the vectors $\mathbf{u}_1 = (1, 1, 2)$, $\mathbf{u}_2 = (2, -1, 0)$, $\mathbf{u}_3 = (0, 3, 2)$, and $\mathbf{u}_4 = (-1, 0, 1)$ so that the center of mass is (411, 511, 1211).

FIND AN EXAMPLE Exercises 61–70: Find an example that meets the given specifications.

61. Two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbf{R}^3$ such that $\mathbf{u} + \mathbf{v} = (3, 2, -1)$.

62. Two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbf{R}^4$ such that $\mathbf{u} - \mathbf{v} = (4, -2, 0, -1)$.

63. Three nonzero vectors in $\mathbf{R}^3$ whose sum is the zero vector.

64. Three nonzero vectors in $\mathbf{R}^4$ whose sum is the zero vector.

65. Two vectors in $\mathbf{R}^2$ that point in the same direction.

66. Two vectors in $\mathbf{R}^2$ that point in opposite directions.

67. Three vectors in $\mathbf{R}^3$ that point in the same direction.

68. Three vectors in $\mathbf{R}^4$ that point in the same direction.

69. A linear system with two equations and two variables that has x=[3-2] as the only solution.

70. A linear system with two equations and three variables that has x=[101]+s[21-1] as the general solution.

TRUE OR FALSE Exercises 71–78: Determine if the statement is true or false, and justify your answer.

71.

(a) if u=[-35], then -2u=[6-10].

(b) if u=[13] and v=[-42], then u-v=[-31].

72.

(a) A vector can have positive or negative components, but a scalar must be positive.

(b) Vector components and scalars can be positive or negative.

73.

(a) If $\mathbf{u}$ and $\mathbf{v}$ are vectors and c is a scalar, then $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.

(b) If c_1 and c_2 are scalars and $\mathbf{u}$ is a vector, then $c_1c_2 + c_2\mathbf{u} = (c_1 + \mathbf{u})c_2$.

74.

(a) The initial point for every vector is the origin.

(b) The vectors [1-24] and [-248] point in opposite directions.

75.

(a) [-21] and (−2, 1) are the same when both are considered as vectors.

(b) The zero vector is a multiple of every other vector.

76.

(a) The vector $2\mathbf{u}$ is the same as $-2(-\mathbf{u})$.

(b) All vector equations have a solution.

77.

(a) The Parallelogram Rule for adding vectors only works in the first quadrant.

(b) The order of the vectors matters for the Tip-to-Tail Rule.

78.

(a) The difference $\mathbf{u} - \mathbf{v}$ is found by adding $-\mathbf{u}$ to $\mathbf{v}$.

(b) Any vector can be subtracted from any other vector.

79. Prove the following parts of Theorem 2.3:

(a) Part (c)

(b) Part (d)

(c) Part (e)
(d) Part (f)
(e) Part (g)
(f) Part (h)

80. In this exercise we verify the geometric subtraction rule shown in Figure 5 by combining the identity **u** − **v** = **u** + (−**v**) and the Tip-to-Tail Rule for addition. Draw a set of coordinate axes, and then sketch and label each of the following:

(a) Vectors **u** and **v** of your choosing.
(b) The vector −**v**.
(c) The translation of −**v** so that its tail is at the tip of **u**.
(d) Using the Tip-to-Tail Rule, the vector **u** + (−**v**).

Explain why the vector you get is the same as the one obtained using the subtraction rule shown in Figure 5.

Exercises 81–82: Sketch the graph of the vectors **u** and **v** and then use the Tip-to-Tail Rule to sketch the graph of **u** + **v**.

81. u=[-23],v=[14]

82. u=[-1-2],v=[31]

Exercises 83–84: Sketch the graph of the vectors **u** and **v** and then use the Parallelogram Rule to sketch the graph of **u** + **v**.

83. u=[0-3],v=[22]

84. u=[42],v=[20]

Exercises 85–86: Sketch the graph of the vectors **u** and **v**, and then use the subtraction procedure shown in Figure 5 to sketch the graph of **u** − **v**.

85. u=[32],v=[1-1]

86. u=[13],v=[2-3]

Exercises 87–88: Find the solutions to the vector equation.

87. x1[273]+x2[242]+x3[516]=[035]

88. x1[1-320]+x2[4321]+x3[-42-31]+x4[52-40]=[172-6]

2.2 Span

We open this section with a fictitious situation. Imagine that you live in the two-dimensional plane $\mathbf{R}^2$ and have just purchased a new car, the VecMobile II. The VecMobile II is delivered at the origin (0, 0) and is a simple vehicle. At any given time, it can be pointed in either direction

$$\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{or} \quad \mathbf{u}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

shown in Figure 1. The VecMobile II can go in forward or reverse.

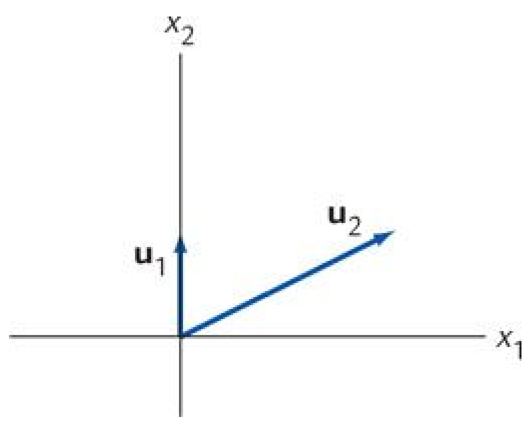


Figure 1 The VecMobile II vectors.

Despite its simplicity there are many places in $\mathbf{R}^2$ that we can go in the VecMobile II. For instance, the point (2, 3) can be reached by first traversing $2\mathbf{u}_1$, changing direction, and then traversing $\mathbf{u}_2$, as shown in Figure 2. The trip also could be made in the reverse order, first taking $\mathbf{u}_2$ and then $2\mathbf{u}_1$. Since we are traversing vectors in a “tip-to-tail” manner, the entire trip can be summarized by the sum

$$2\mathbf{u}_1 + \mathbf{u}_2 = 2\begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

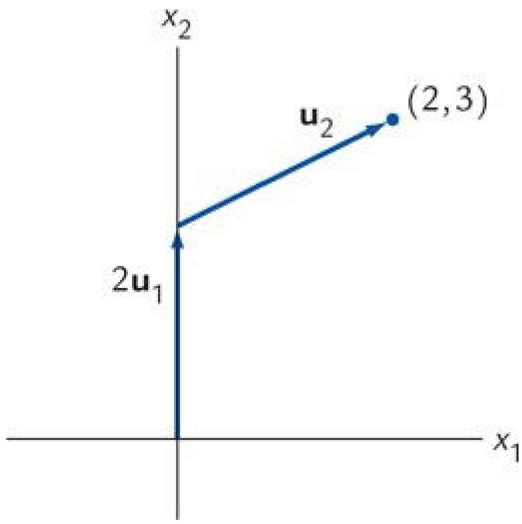


Figure 2 2u1+u2=[23].

Figure 3 depicts a more complicated path that arises from traversing $3\mathbf{u}_1$, then $-\mathbf{u}_2$, then $-2\mathbf{u}_1$, and finally $2\mathbf{u}_2$. This simplifies algebraically to

3u1-u2-2u1+2u2=u1+u2

Any path taken in the VecMobile II can be similarly simplified, so that the set of all possible destinations can be expressed as

x1u1+x2u2

where x_1 and x_2 can be any real numbers. This set of linear combinations is called the *span* of the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$. Although the VecMobile II is simple, we can go anywhere within $\mathbf{R}^2$.

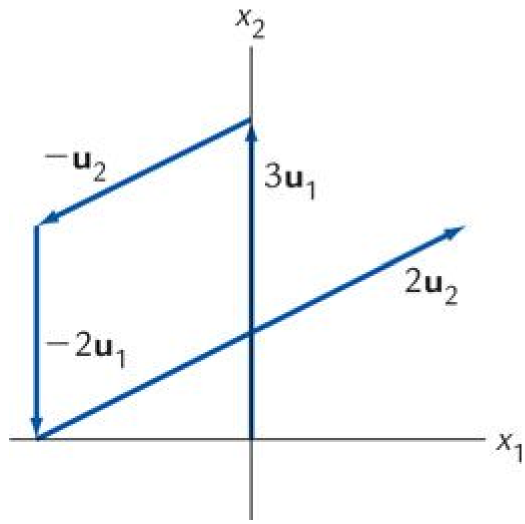


Figure 3 Depiction of 3u1-u2-2u1+2u2=[22].

Example 1

Show algebraically that the VecMobile II can reach any position in $\mathbf{R}^2$.

Solution Suppose that we want to reach an arbitrary point (*a*, *b*). To do so, we need to find scalars x_1 and x_2 such that

x1[01]+x2[21]=[ab]

This vector equation translates into the system of equations

2x2=ax1+ x2=b

which has the unique solution x1=b-a2 and x2=a2. We now know exactly how to find the scalars x_1 and x_2 required to reach any point (*a*, *b*), and so we can conclude that the VecMobile II can get anywhere in $\mathbf{R}^2$.

Example 1 shows that the span of $\mathbf{u}_1$ and $\mathbf{u}_2$ equals all of $\mathbf{R}^2$. The notion of span generalizes to sets of vectors in $\mathbf{R}^n$.

DEFINITION 2.5 ▸

Span

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbf{R}^n$. The **span** of this set is denoted span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ and is defined as the set of all linear combinations

$$x_1 u_1 + x_2 u_2 + \cdots + x_m u_m$$

where $x_1, x_2, \ldots, x_m$ can be any real numbers.

If span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\} = \mathbf{R}^n$, then we say that the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ spans $\mathbf{R}^n$.

Span in $\mathbf{R}^3$

Now let's turn our attention to $\mathbf{R}^3$. Suppose

$$u_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \text{and} \, u_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

as shown in Figure 4. Then span$\{\mathbf{u}_1, \mathbf{u}_2\}$ is the set of linear combinations of the form

$$x_1 u_1 + x_2 u_2$$

Some vectors in R^3 are in span$\{\mathbf{u}_1, \mathbf{u}_2\}$ and some are not.

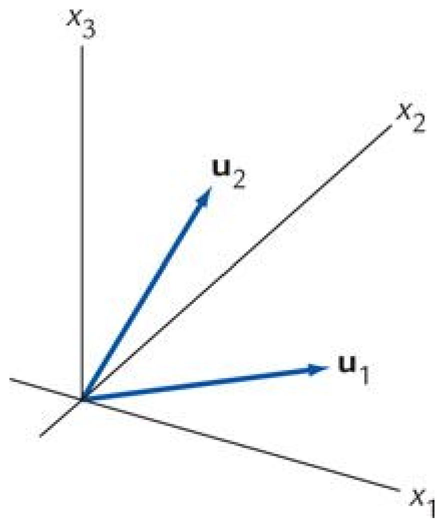


Figure 4 The vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in $\mathbf{R}^3$.

Example 2

Show that v1=[-147] is in S = span{$\mathbf{u}_1$, $\mathbf{u}_2$}, and that v2=[821] is not.

Solution To show that $\mathbf{v}_1$ is in S, we need to find scalars x_1 and x_2 that satisfy the equation

x1=[211]+x2[123]=[-147]

This is equivalent to the linear system

2x1+x2=-1x1+2x2=4x1+3x2=7

Transferring to an augmented matrix and performing row operations gives us

[211213|-147]~[120100|430] (1)

Extracting the echelon system and back substituting produces $x_1 = -2$ and $x_2 = 3$. Thus $\mathbf{v}_1 = -2\mathbf{u}_1 + 3\mathbf{u}_2$, so $\mathbf{v}_1$ is in S = span{$\mathbf{u}_1$, $\mathbf{u}_2$}.

▶ The row operations used in both (1) and (2) are (in order performed):
R1↔R2-2R1+R2→R2-R1+R3→R3R2↔R33R2+R3→R3

The approach for $\mathbf{v}_2$ is the same. This time we need to determine if there exist scalars x_1 and x_2 such that

$$x_1\begin{bmatrix}2\\1\\1\end{bmatrix}+x_2\begin{bmatrix}1\\2\\3\end{bmatrix}=\begin{bmatrix}8\\2\\1\end{bmatrix}$$

This is equivalent to the linear system

$$\begin{aligned}2x_1+x_2&=8\\x_1+2x_2&=2\\x_1+3x_2&=1\end{aligned}$$

The augmented matrix and corresponding echelon form are

$$\left[\begin{array}{cc|c}2&1&8\\1&2&2\\1&3&1\end{array}\right]\sim\left[\begin{array}{cc|c}1&2&2\\0&1&-1\\0&0&1\end{array}\right] \qquad (2)$$

The third row of the echelon matrix corresponds to the equation $0 = 1$. Thus the system has no solutions and $\mathbf{v}_2$ is not in S.

Here is another way to visualize the span of two vectors in $\mathbf{R}^3$. Get two pieces of string, about 3 feet long each, and tie them both to some solid object (like a refrigerator). Get a friend to pull the strings tight and in different directions. These are your vectors.

Next get a light-weight flat surface (a pizza box works well) and gently rest it on the strings (see Figure 5). Think of the surface as representing a plane. Then the span of the two "string" vectors is the set of all vectors that lie within the plane. Note that no matter the angle of the strings, if you are doing this correctly it is possible to rest the surface on them.

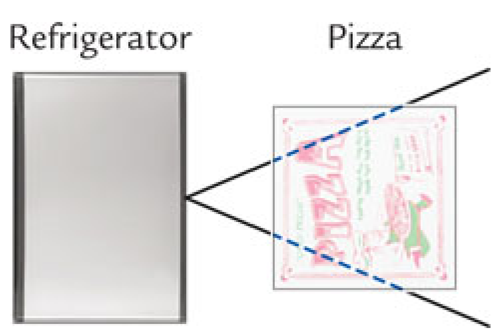


Figure 5 String vectors and pizza box.

Figure 6 shows a plane resting on $\mathbf{u}_1$ and $\mathbf{u}_2$. The span$\{\mathbf{u}_1, \mathbf{u}_2\}$ consists exactly of those vectors that are contained in the plane. Therefore if $\mathbf{u}_3$ is contained in span$\{\mathbf{u}_1, \mathbf{u}_2\}$, then $\mathbf{u}_3$ will lie in the plane, as shown in Figure 7(a). On the other hand, if is not contained in span$\{\mathbf{u}_1, \mathbf{u}_2\}$, then $\mathbf{u}_3$ will be outside the plane, as in Figure 7(b).

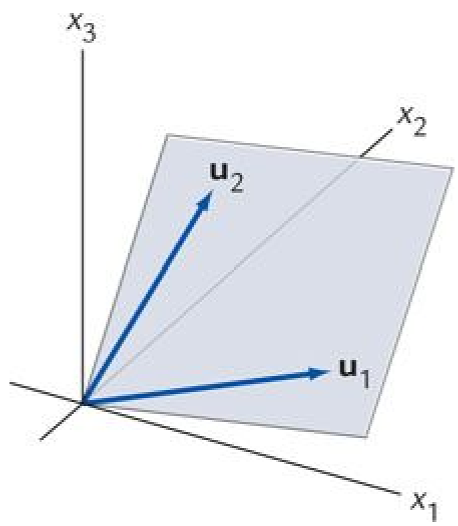


Figure 6 The plane is equal to span$\{\mathbf{u}_1, \mathbf{u}_2\}$.

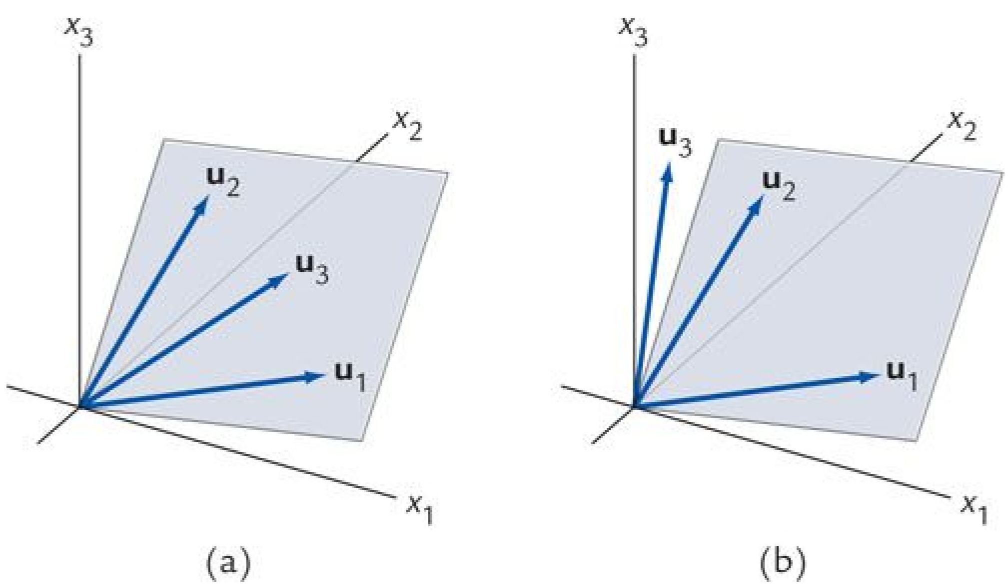


Figure 7 In (a), $\mathbf{u}_3$ is in span$\{\mathbf{u}_1, \mathbf{u}_2\}$. In (b), $\mathbf{u}_3$ is *not* in span$\{\mathbf{u}_1, \mathbf{u}_2\}$.

Three Vectors in $\mathbf{R}^3$

Later in this section we will show that no two vectors can span all of $\mathbf{R}^3$. But three vectors can, as shown in the next example.

Example 3

Show v=[521] is in span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$}, and then show span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} = $\mathbf{R}^3$.

u1=[211],u2=[123],u3=[91-1]

Solution To show that $\mathbf{v}$ is in span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$}, we need to find scalars x_1, x_2, and x_3 such that

x1[211]+x2[123]+x3[91-1]=[521] (3)

This is equivalent to the linear system

2x1+x2+9x3=5x1+2x2+x3=2x1+3x2-x3=1

The augmented matrix and corresponding echelon form are

[21912113-1|521]~[12101-2001|2-1-2] (4)

▶ The row operations used in (4) and (6) are (in order performed):
R1↔R2-2R1+R2→R2-R1+R3→R3R2↔R33R2+R3→R3

This is enough to show that the system has a unique solution. Back substitution gives $x_1 = 14$, $x_2 = -5$, and $x_3 = -2$. Since (3) has a solution, $\mathbf{v}$ is in span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} with

v=14u1-5u2-2u3

To show that span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} = $\mathbf{R}^3$, for any real numbers *a*, *b*, and *c* we need to find scalars x_1, x_2, and x_3 such that

x1[211]+x2[123]+x3[91-1]=[abc] (5)

The augmented matrix is

$$\left[\begin{array}{ccc|c}2&1&9&a\\1&2&1&b\\1&3&-1&c\end{array}\right] \tag{6}$$

In (4) we saw that every row has a pivot position, and none are in the rightmost column. Therefore the same row operations will transform (6) to echelon form, and since (4) had a unique solution so will (5) for any *a*, *b*, and *c*. This is enough to show that span$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \mathbf{R}^3$. For completeness, we determine that the echelon form of (6) is

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

Back substitution gives the solution

$$\begin{aligned}x_1&=-5a+28b-17c\\x_2&=2a-11b+7c\\x_3&=a-5b+3c\end{aligned}$$

In Example 2 and Example 3, the vectors become the columns of an augmented matrix. For instance, the vectors in (5) become the columns in (6),

$$\mathbf{u}_1=\begin{bmatrix}2\\1\\1\end{bmatrix},\quad \mathbf{u}_2=\begin{bmatrix}1\\2\\3\end{bmatrix},\quad \mathbf{u}_3=\begin{bmatrix}9\\1\\-1\end{bmatrix},\quad \mathbf{v}=\begin{bmatrix}a\\b\\c\end{bmatrix} \Rightarrow \left[\begin{array}{ccc|c}2&1&9&a\\1&2&1&b\\1&3&-1&c\end{array}\right]=\left[\begin{array}{ccc|c}\mathbf{u}_1&\mathbf{u}_2&\mathbf{u}_3&\mathbf{v}\end{array}\right]$$

The same is true in general. Given the vector equation

$$x_1\mathbf{u}_1+x_2\mathbf{u}_2+\cdots+x_m\mathbf{u}_m=\mathbf{v}$$

the corresponding augmented matrix is

$$\left[\begin{array}{cccc|c}\mathbf{u}_1&\mathbf{u}_2&\cdots&\mathbf{u}_m&\mathbf{v}\end{array}\right]$$

THEOREM 2.6 ▶

Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ and $\mathbf{v}$ be vectors in $\mathbf{R}^n$. Then $\mathbf{v}$ is an element of span$\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ if and only if the linear system with augmented matrix

[u1u2⋯um|v] (7)

has a solution.

Proof The vector $\mathbf{v}$ is in span$\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ if and only if there exist scalars $x_1, x_2, \dots, x_m$ that satisfy

x1u1+x2u2+⋯+xmum=v

This is true if and only if the corresponding linear system has a solution. As noted above, the linear system has augmented matrix (7), so the proof is complete. ■

When Does a Set Span $\mathbf{R}^n$?

In the examples we have seen, a set of two vectors spanned $\mathbf{R}^2$ and a set of three vectors spanned $\mathbf{R}^3$. This suggests the following question.

Example 4

Is it always true that a set of n vectors will span $\mathbf{R}^n$?

Solution Not always. For example, the span of the vectors

v1=[11],v2=[22]

is a line in $\mathbf{R}^2$ (shown in Figure 8) because $\mathbf{v}_2 = 2\mathbf{v}_1$, so span{$\mathbf{u}_1$, $\mathbf{u}_2$} is not all of $\mathbf{R}^2$.

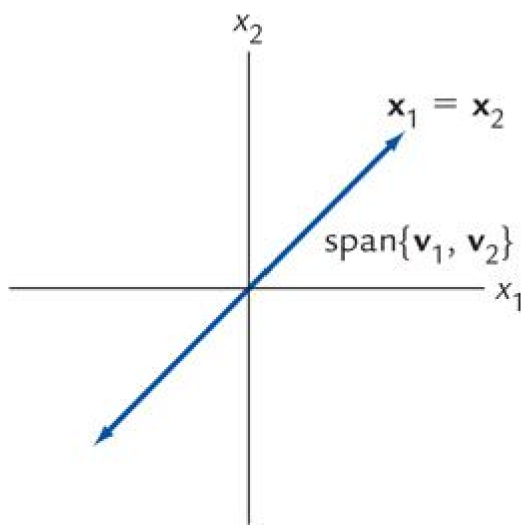


Figure 8 The span of {$\mathbf{v}_1$, $\mathbf{v}_2$} in $\mathbf{R}^2$ in Example 4.

A more subtle example is given by the set of vectors in $\mathbf{R}^3$,

u1=[211],u2=[123],u3=[1-4-7]

It is straightforward to verify that $\mathbf{u}_3$ is equal to the linear combination

u3=2u1-3u2

Thus any vector that is a linear combination of $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ can be expressed as a linear combination of just $\mathbf{u}_1$ and $\mathbf{u}_2$, because

x1u1+x2u2+x3u3=x1u1+x2u2+x3(2u1-3u2) =(x1+2x3)u1+(x2-3x3)u2

Therefore span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} = span{$\mathbf{u}_1$, $\mathbf{u}_2$}, and since we have already shown (in Example 2) that span{$\mathbf{u}_1$, $\mathbf{u}_2$} ≠ $\mathbf{R}^3$, it follows that span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} ≠ $\mathbf{R}^3$.

The above argument serves as a model for proving the next theorem.

THEOREM 2.7 ▸

Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ and $\mathbf{u}$ be vectors in $\mathbf{R}^n$. If $\mathbf{u}$ is in span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$, then

$$\text{span}\{\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}.$$

Proof Let $S_0 = \text{span}\{\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ and $S_1 = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$. We need to show that sets $S_0 = S_1$, which we do by showing that each is a subset of the other. First suppose that a vector $\mathbf{v}$ is in S_1. Then there exist scalars $a_1, \ldots, a_m$ such that

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_m\mathbf{u}_m = 0\mathbf{u} + a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_m\mathbf{u}_m$$

Hence $\mathbf{v}$ is also in S_0, so S_1 is a subset of S_0.

Now suppose that $\mathbf{v}$ is in S_0. Then there exist scalars $b_0, b_1, \ldots, b_m$ such that $\mathbf{v} = b_0\mathbf{u} + b_1\mathbf{u}_1 + \cdots + b_m\mathbf{u}_m$. Since $\mathbf{u}$ is in S_1, there also exist scalars $c_1, \ldots, c_m$ such that $\mathbf{u} = c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m$. Then

$$\mathbf{v} = b_0(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m) + b_1\mathbf{u}_1 + b_1\mathbf{u}_2 + \cdots + b_m\mathbf{u}_m = (b_0c_1 + b_1)\mathbf{u}_1 + (b_0c_2 + b_2)\mathbf{u}_2 + \cdots + (b_0c_m + b_m)\mathbf{u}_m$$

Thus $\mathbf{v}$ is in S_1, so S_0 is a subset of S_1. Since S_0 and S_1 are subsets of each other, it follows that $S_0 = S_1$. ■ ■

Example 5

Find a vector in $\mathbf{R}^4$ that is *not* in span$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, where u1=[1-2-12],u2=[-374-6],u3=[2025]

Solution We start by forming the matrix with our vectors as columns,

[u1 u2 u3]=[1-32-270-1422-65]

Now perform the usual row operations needed to transform the matrix to echelon form, recording each operation along the way. We need only perform enough operations to introduce a row of zeroes on the bottom of the matrix, which must be possible because there are more rows than columns (see Exercise 54 in Section 1.2). The “Forward Operations” shown in the margin yields the echelon form

[1-32-270-1422-65]~[1-32014001000]

▶ Forward Operations:
2R1+R2→R2R1+R3→R3-2R1+R4→R4-R2+R3→R3R3↔R4

The next step is to append a column to the right side of the echelon matrix,

[1-32014001000|0001] (8)

View (8) as an augmented matrix. The bottom row corresponds to the equation 0 = 1, so the associated linear system has no solutions. If we now reverse the row operations used previously (as shown in the margin), the first three columns of the augmented matrix are returned to their original form. This gives us

[1-32014001000|0001]~[1-32-270-1422-65|0010] (9)

▶ Reverse Operations:
R3↔R4R2+R3→R32R1+R4→R4-R1+R3→R3-2R1+R2→R2

Since the system associated with the augmented matrix (8) has no solutions, the system associated with the equivalent augmented matrix (9) also has no solutions. This system corresponds to the vector equation

$$x_1\mathbf{u}_1+x_2\mathbf{u}_2+x_3\mathbf{u}_3=\mathbf{b} \quad \text{where} \quad \mathbf{b}=\begin{bmatrix}0\\0\\1\\0\end{bmatrix}$$

This vector equation cannot have any solutions, so it follows that $\mathbf{b}$ is *not* in span$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

The next theorem draws on ideas from Example 5 to provide a way to determine if a set of vectors spans $\mathbf{R}^n$.

THEOREM 2.8

Suppose that $\mathbf{u}_1, \dots, \mathbf{u}_m$ are in $\mathbf{R}^n$, and let

$$A=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m\end{bmatrix}\sim B$$

where B is in echelon form. Then span$\{\mathbf{u}_1, \dots, \mathbf{u}_m\} = \mathbf{R}^n$ exactly when B has a pivot position in every row.

Proof First suppose B has a pivot position in every row. Consider the augmented matrix

$$\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m & \mathbf{v}\end{bmatrix} \tag{10}$$

where $\mathbf{v}$ is in $\mathbf{R}^n$. Since B has a pivot position in every row, the same sequence of row operations that transformed A to B will transform (10) to echelon form, with no pivot position in the last (rightmost) column. Therefore back substitution will always yield a solution regardless of $\mathbf{v}$, so that span$\{\mathbf{u}_1, \dots, \mathbf{u}_m\} = \mathbf{R}^n$.

Now suppose B does not have a pivot position in every row, which implies the last row of B is zeros. Let $\mathbf{v} = (0, 0, \dots, 0, 1)$ in $\mathbf{R}^n$ and consider the augmented matrix

$$[B \mid v] \tag{11}$$

Since the last row of B is zeros, (11) corresponds to a system with no solutions. Applying the reverse of the transformations used for $A \sim B$ to (11) gives

$$[A \mid v^*]=[u1\ldots um \mid v^*] \tag{12}$$

Thus (12) corresponds to a system with no solutions, so that $\mathbf{v}^*$ is not in span$\{\mathbf{u}_1, \ldots , \mathbf{u}_m\}$. Therefore span$\{\mathbf{u}_1, \ldots , \mathbf{u}_m\} \neq \mathbf{R}^n$.

Example 6

Determine if span$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\} = \mathbf{R}^3$, where

u1=[1-12],u2=[2-12],u3=[-25-10],u4=[3-48]

Solution We populate a matrix with $\mathbf{u}_1, \ldots , \mathbf{u}_4$ as columns then transform to echelon form,

[12-23-1-15-422-108]~[12-23013-10000]

By Theorem 2.8, since the echelon form has a row of zeros, there is no pivot position in the last row and therefore the set does not span $\mathbf{R}^3$.

The following handy result follows from Theorem 2.8. The proof is left as an exercise.

THEOREM 2.9

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbf{R}^n$. If $m < n$, then this set does not span $\mathbf{R}^n$. If $m \geq n$, then the set might span $\mathbf{R}^n$ or it might not. In this case, we cannot say more without additional information about the vectors.

The Equation *Ax* = *b*

By now we are comfortable with translating back and forth between vector equations and linear systems. Here we give new notation that will be used for a variety of purposes, including expressing linear systems in a familiar form.

Let A be the matrix with columns $\mathbf{a}_1=\begin{bmatrix}10\\5\\7\end{bmatrix}$ and $\mathbf{a}_2=\begin{bmatrix}8\\6\\-1\end{bmatrix}$,

$$A=\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}=\begin{bmatrix}10 & 8\\5 & 6\\7 & -1\end{bmatrix}$$

Also let $\mathbf{x}=\begin{bmatrix}x_1\\x_2\end{bmatrix}$. Then the product of the matrix A and the vector $\mathbf{x}$ is defined to be

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

Thus $A\mathbf{x}$ is a linear combination of the columns of A, with the scalars given by the components of $\mathbf{x}$. Next let

$$\mathbf{b}=\begin{bmatrix}18\\31\\3\end{bmatrix}$$

Then $A\mathbf{x} = \mathbf{b}$ is a compact form of the vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$, which in turn is equivalent to the linear system

$$\begin{aligned}10x_1+8x_2&=18\\5x_1+6x_2&=31\\7x_1-x_2&=3\end{aligned}$$

We now have three ways to express a linear system. Below is the general formula for multiplying a matrix by a vector.

DEFINITION 2.10 ▸

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ be vectors in $\mathbf{R}^n$. If

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \tag{13}$$

then $A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m$.

▶ Remember: The product $A\boldsymbol{x}$ only is defined when the number of columns of A equals the number of components (entries) of $\boldsymbol{x}$.

Example 7

Find A, $\mathbf{x}$, and $\mathbf{b}$ so that the equation $A\mathbf{x} = \mathbf{b}$ corresponds to the system of equations

$$\begin{aligned} 4x_1 - 3x_2 + 7x_3 - x_4 &= 13 \\ -x_1 + 2x_2 \phantom{{}+7x_3} + 6x_4 &= -2 \\ x_2 - 3x_3 - 5x_4 &= 29 \end{aligned}$$

Solution Translating the system to the form $A\mathbf{x} = \mathbf{b}$, the matrix A will contain the coefficients of the system, the vector $\mathbf{x}$ has the variables, and the vector $\mathbf{b}$ will contain the constant terms. Thus we have

$$A = \begin{bmatrix} 4 & -3 & 7 & -1 \\ -1 & 2 & 0 & 6 \\ 0 & 1 & -3 & -5 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 13 \\ -2 \\ 29 \end{bmatrix}$$

Example 8

Suppose that

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 7 \\ 5 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -6 \\ 4 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$$

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$. Find the following (if they exist):

(a) $A\mathbf{v}_1$ and $A\mathbf{v}_2$.

(b) The system of equations corresponding to $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{v}_2$.

Solution

(a) In order for $A\mathbf{v}_1$ to exist, the number of columns of A must equal the number of components of $\mathbf{v}_1$. Since this is *not* the case, $A\mathbf{v}_1$ does not exist.
On the other hand, $\mathbf{v}_2$ has two components, so $A\mathbf{v}_2$ exists. We have

$$A\mathbf{v}_2 = -6\mathbf{a}_1 + 4\mathbf{a}_2 = -6\begin{bmatrix}1\\-2\\0\end{bmatrix} + 4\begin{bmatrix}3\\1\\-1\end{bmatrix} = \begin{bmatrix}6\\16\\-4\end{bmatrix}$$

(b) Since $\mathbf{x}$ has two components, $A\mathbf{x}$ exists. Moreover, $A\mathbf{x}$ and $\mathbf{b}$ both have three components, so the equation $A\mathbf{x} = \mathbf{b}$ is defined and corresponds to the system

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

For the second part, $A\mathbf{x}$ exists and has three components, but $\mathbf{v}_2$ has only two components, so $A\mathbf{x} = \mathbf{v}_2$ is not defined.

We close this section with a theorem that ties together several closely related ideas.

THEOREM 2.11

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ and $\mathbf{b}$ be vectors in $\mathbf{R}^n$. Then the following statements are equivalent. That is, if one is true, then so are the others, and if one is false, then so are the others.

(a) $\mathbf{b}$ is in span$\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$.

(b) The vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{b}$ has at least one solution.

(c) The linear system corresponding to $[\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_m \mid \mathbf{b}]$ has at least one solution.

(d) The equation $A\mathbf{x} = \mathbf{b}$, with A and $\mathbf{x}$ given as in (13), has at least one solution.

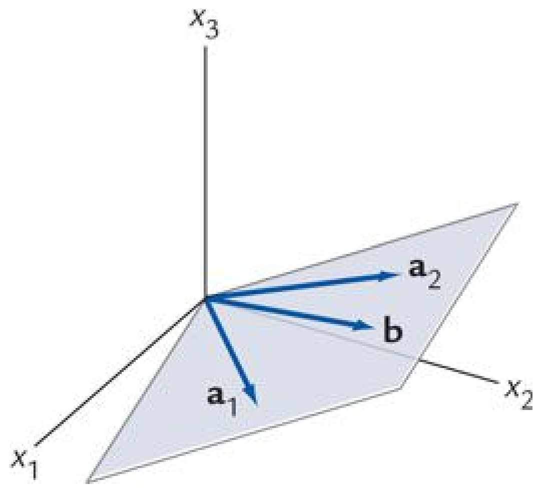


Figure 9 span{$\mathbf{a}_1$, $\mathbf{a}_2$} includes $\mathbf{b}$, so (14) has a solution.

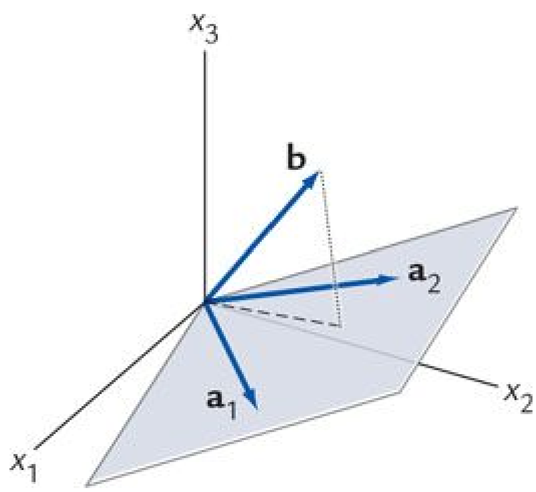


Figure 10 span{$\mathbf{a}_1$, $\mathbf{a}_2$} does not include $\mathbf{b}$, so (14) does not have a solution.

The theorem follows directly from the definitions, so the proof is left as an exercise. Although this result is not hard to arrive at, it is important because it explicitly states the connection between these different formulations of the same basic idea.

As a quick application, note that the vector $\mathbf{b}$ in Figure 9 is in span{$\mathbf{a}_1$, $\mathbf{a}_2$}, so by Theorem 2.11 it follows that

$$x1a1+x2a2=b \tag{14}$$

has at least one solution. On the other hand, if $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{b}$ are as shown in Figure 10, then (14) has no solutions.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. For each set, find three vectors that are in the span of the given vectors.
   (a) u1=[2-3],u2=[41]
   (b) u1=[614],u2=[-23-3]
2. Determine if $\mathbf{b}$ is in span$\{\mathbf{u}_1, \mathbf{u}_2\}$.

   u1=[12-2],u2=[043],b=[-125]

3. Express each linear system in the form $A\mathbf{x} = \mathbf{b}$.
   (a) 7x1-2x2-2x3=6-x1+7x2+4x3=113x1-x2-2x3=1
   (b) 4x1-3x2-x3+5x4=03x1+12x2+6x3 =10
4. Determine if the columns of A span $\mathbf{R}^2$.
   (a) A=[23-1-2]
   (b) A=[411-3]
5. Determine if the columns of A span $\mathbf{R}^3$.
   (a) A=[13-1-1-23025]
   (b) A=[2061-21-141]
6. Determine if each statement is true or false, and justify your answer.
   (a) If a set of vectors spans $\mathbf{R}^3$, then the set also spans $\mathbf{R}^2$.
   (b) If $\{\mathbf{u}_1, \mathbf{u}_2\}$ spans $\mathbf{R}^2$, then so does $\{2\mathbf{u}_1, 3\mathbf{u}_2\}$.
   (c) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ spans $\mathbf{R}^3$ and $A = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$, then $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$ in $\mathbf{R}^3$.
   (d) If $\{\mathbf{u}_1, \mathbf{u}_2\}$ spans $\mathbf{R}^2$, then so does $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

EXERCISES

Exercises 1–6: Find three vectors that are in the span of the given vectors.

1. u1=[26];u2=[915]
2. u1=[-27],u2=[-34]
3. u1=[25-3],u2=[104]
4. u1=[05-2],u2=[126],u3=[-672]
5. u1=[200],u2=[416],u3=[-407]
6. u1=[0130],u2=[-18-52],u3=[12-110]

Exercises 7–12: Determine if **b** is in the span of the other given vectors. If so, express **b** as a linear combination of the other vectors.

7. a1=[35],b=[9-15]
8. a1=[10-15],b=[-3045]
9. a1=[4-210],b=[2-1-5]
10. a1=[-13-1],a2=[-2-36],b=[-692]
11. a1=[-14-3],a2=[28-7],b=[-10-87]
12. a1=[31-2-1],a2=[-4233],b=[01015]

Exercises 13–16: Find A, $\mathbf{x}$, and $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ corresponds to the given linear system.

13. 2x1+8x2-4x3=-10-x1-3x2+5x3=4
14. -2x1+5x2-10x3=4x1-2x2+3x3=-17x1-17x2+34x3=-16
15. x1-x2-3x3-x4=-1-2x1+2x2+6x3+2x4=-1-3x1-3x2+10x3 =5
16. -5x1+9x2=133x1-5x2=-9x1-2x2=-2

Exercises 17–20: Find the vector equation that has the corresponding augmented matrix.

17. [57-21-5-4|92]
18. [4-5-33426-137|012]
19. [4-2-350-573382-1|1262]
20. [4-9241-760|11922]

Exercises 21–24: Determine if the columns of *A* span $\mathbf{R}^2$.

21. A=[15-6-52]

22. A=[4-1226]

23. A=[2106-3-1]

24. A=[105-227]

Exercises 25–28: Determine if the columns of *A* span $\mathbf{R}^3$.

25. A=[3105-2-14-4-3]

26. A=[128-2373-11]

27. A=[21-3514260333]

28. A=[-4-71200385-11-4]

Exercises 29–34: A matrix *A* is given. Determine if the system *A***x** = **b** (where **x** and **b** have the appropriate number of components) has a solution for all choices of **b**.

29. A=[3-442]

30. A=[-9216-14]

31. A=[810-1-32]

32. A=[1-12-23-1105]

33. A=[-3211-1-15-4-3]

34. A=[2-30012-53-9309]

Exercises 35–38: Find a vector of matching dimension that is *not* in the given span.

35. span{[1-2],[-36]}

36. span{[31],[62]}

37. span{[13-2],[2-11]}

38. span{[121],[3-11],[-151]}

Exercises 39–42: Find a vector in $\mathbf{R}^2$ that is *not* in the span of the columns of *A*.

39. A=[1428]

40. A=[-3152-10]

41. A=[2-57-410-14]

42. A=[42-6105-15]

Exercises 43–46: Find a vector in $\mathbf{R}^3$ that is *not* in the span of the columns of *A*.

43. A=[1325-14]

44. A=[341-123]

45. A=[2-5-1124317]

46. A=[12-1-1-30271]

47. Find all values of *h* such that the vectors $\{\mathbf{a}_1, \mathbf{a}_2\}$ span $\mathbf{R}^2$, where

a1=[24],a2=[h6]

48. Find all values of *h* such that the vectors $\{\mathbf{a}_1, \mathbf{a}_2\}$ span $\mathbf{R}^2$, where

a1=[-3h],a2=[5-4]

49. Find all values of *h* such that the vectors $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ span $\mathbf{R}^3$, where

a1=[245],a2=[h810],a3=[126]

50. Find all values of *h* such that the vectors $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ span $\mathbf{R}^3$, where

a1=[-1h7],a2=[4-25],a3=[1-32]

FIND AN EXAMPLE Exercises 51–58: Find an example that meets the given specifications.

51. Four distinct nonzero vectors that span $\mathbf{R}^3$.

52. Four distinct nonzero vectors that span $\mathbf{R}^4$.

53. Four distinct nonzero vectors that do not span $\mathbf{R}^3$.

54. Four distinct nonzero vectors that do not span $\mathbf{R}^4$.

55. Two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in $\mathbf{R}^3$ that span the set of all vectors of the form $\mathbf{v} = (v_1, v_2, 0)$.

56. Three vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ in $\mathbf{R}^4$ that span the set of all vectors of the form $\mathbf{v} = (0, v_2, v_3, v_4)$.

57. Two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in $\mathbf{R}^3$ that span the set of all vectors of the form $\mathbf{v} = (v_1, v_2, v_3)$ where $v_1 + v_2 + v_3 = 0$.

58. Three vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ in $\mathbf{R}^4$ that span the set of all vectors of the form $\mathbf{v} = (v_1, v_2, v_3, v_4)$ where $v_1 + v_2 + v_3 + v_4 = 0$.

TRUE OR FALSE Exercises 59–66: Determine if the statement is true or false, and justify your answer.

59.

(a) If $m < n$, then a set of m vectors cannot span $\mathbf{R}^n$.

(b) If a set of vectors includes $\mathbf{0}$, then it cannot span $\mathbf{R}^n$.

60.

(a) Suppose A is a matrix with n rows and m columns. If $n < m$, then the columns of A span $\mathbf{R}^n$.

(b) Suppose A is a matrix with n rows and m columns. If $m < n$, then the columns of A span $\mathbf{R}^n$.

61.

(a) If A is a matrix with columns that span $\mathbf{R}^n$, then $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

(b) If A is a matrix with columns that span $\mathbf{R}^n$, then $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

62.

(a) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ spans $\mathbf{R}^3$, then so does $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

(b) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ does not span $\mathbf{R}^3$, then neither does $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

63.

(a) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ spans $\mathbf{R}^3$, then so does $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

(b) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ does not span $\mathbf{R}^3$, then neither does $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

64.

(a) If $\mathbf{u}_4$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then

span{u1, u2, u3, u4}=span{u1, u2, u3}.

(b) If $\mathbf{u}_4$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then

span{u1, u2, u3, u4}≠span{u1, u2, u3}.

65.

(a) If $\mathbf{u}_4$ is *not* a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then

span{u1, u2, u3, u4}=span{u1, u2, u3}.

(b) If $\mathbf{u}_4$ is *not* a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then

span{u1, u2, u3, u4}≠span{u1, u2, u3}.

66.

(a) For any vectors $\mathbf{u}_1$, $\mathbf{u}_2$ $\mathbf{u}_3$,

span{u1, u2, u3}=span{0, u1, u2, u3}.

(b) If $\mathbf{u}_1 \neq 0$ is in $\mathbf{R}^2$ and u2=u1=[11], then span$\{\mathbf{u}_1, \mathbf{u}_2\} = \mathbf{R}^2$.

67. Which of the following sets of vectors in $\mathbf{R}^3$ can possibly span $\mathbf{R}^3$? Justify your answer.

(a) $\{\mathbf{u}_1\}$

(b) $\{\mathbf{u}_1, \mathbf{u}_2\}$

(c) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$

(d) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$

68. Which of the following sets of vectors in $\mathbf{R}^3$ cannot possibly span $\mathbf{R}^3$? Justify your answer.

(a) $\{\mathbf{u}_1\}$

(b) $\{\mathbf{u}_1, \mathbf{u}_2\}$

(c) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$

(d) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$

69. Prove that if c is a nonzero scalar, then span$\{\mathbf{u}\}$ = span$\{c\mathbf{u}\}$.

70. Prove that if c_1 and c_2 are nonzero scalars, then

span{u1, u2}=span{c1 u1, c2 u2}.

71. Suppose that S_1 are S_2 are two finite sets of vectors, and that S_1 is a subset of S_2. Prove that the span of S_1 is a subset of the span of S_2.

72. Prove that if span{$\mathbf{u}_1$, $\mathbf{u}_2$} = $\mathbf{R}^2$, then

span{u1+u2, u1-u2}=R2.

73. Prove that if span{$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} = $\mathbf{R}^3$, then

span{u1+u2, u1+u3, u2+u3}=R3.

74. Suppose that {$\mathbf{u}_1$, … , $\mathbf{u}_m$} is a subset of $\mathbf{R}^n$, with $m > n$. Prove that if **b** is in span{$\mathbf{u}_1$, … , $\mathbf{u}_m$}, then there are infinitely many ways to express **b** as a linear combination of {$\mathbf{u}_1$, … , $\mathbf{u}_m$}.

75. Prove Theorem 2.9.

76. Prove Theorem 2.11.

Exercises 77–80: Determine if the claimed equality is true or false.

77. span{[321],[154],[-230]}=R3

78. span{[112],[34-1],[46-6]}=R3

79. span{[4023],[7-467],[13-21],[3202]}=R4

80. span{[3-210],[85-97],[162-3],[2-135]}=R4

2.3 Linear Independence

The myology clinic at a university research hospital helps patients recover muscle mass lost due to illness. After a full evaluation, patients receive exercise training and each is given a nutritional powder that has the exact balance of protein, fat, and carbohydrates required to meet his or her needs. The nutritional powders are created by combining some or all of four powder brands that the clinic keeps in stock. The components for brands A, B, C, and D (in grams per serving) are shown in Table 1.

Stocking all four brands is expensive, as they have a limited shelf life and take up valuable storage space. The clinic would like to eliminate unnecessary brands, but it does not want to sacrifice any flexibility to create specialized combinations. Are all four brands needed, or can they get by with fewer?

We can solve this problem using vectors, but first we need to develop some additional concepts. Recall that in the previous section we noted the set of vectors

$$u1=[211],u2=[123],u3=[1-4-7] \tag{1}$$

are such that the third is a linear combination of the first two, with

$$u3=2u1-3u2 \tag{2}$$

Table 1 **Nutritional Powder Brand Components (grams per serving)**

	Brand			
	A	**B**	**C**	**D**
Protein	16	22	18	18
Fat	2	4	0	2
Carbohydrates	8	4	4	6

Thus, in a sense, $\mathbf{u}_3$ depends on $\mathbf{u}_1$ and $\mathbf{u}_2$. We can also solve (2) for $\mathbf{u}_1$ or $\mathbf{u}_2$,

u1=32u2+12u3oru2=23u1-13u3

so each of the vectors is "dependent" on the others. Rather than separating out one particular vector, we can move all terms to one side of the equation, giving us

2u1-3u2-u3=0

This brings us to the following important definition.

DEFINITION 2.12 ▸

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbf{R}^n$. If the only solution to the vector equation

x1u1+x2u2+⋯+x3u3=0

Linear Independence

is the trivial solution given by $x_1 = x_2 = \cdots = x_m = 0$, then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is **linearly independent**. If there are nontrivial solutions, then the set is **linearly dependent**.

Example 1

Determine if the set

u1=[-14-2-3],u2=[3-1377],u3=[-219-5]

is linearly dependent or linearly independent.

Solution To determine if the set {$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} is linearly dependent or linearly independent, we need to find the solutions of the vector equation

x1u1+x2u2+x3u3=0

- To determine if a set of vectors is linearly dependent or independent, we almost always use the method illustrated in Example 1: Set the linear combination equal to **0** and find the solutions.

This is equivalent to the linear system

-x1+3x2-2x3=04x1-13x2+x3=0-2x1+7x2+9x3=0-3x1+7x2-5x3=0

The corresponding augmented matrix and echelon form are

[-13-24-131-279-37-5|0000]~[-13-20-1-7006000|0000] (3)

- The row operations used in (3) are (in order performed): 4R1+R2→R2-2R1+R3→R3-3R1+R4→R4R2+R3→R3-2R2+R4→R4(-5/2)R3+R4→R4

Back substitution shows that the only solution is the trivial one, $x_1 = x_2 = x_3 = 0$. Hence the set {$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} is linearly independent.

Example 2

Determine if the myology clinic described earlier can eliminate any of the nutritional powder brands with components given in Table 1.

Solution We start by determining if the nutrient vectors for the four brands

a=[1628],b=[2244],c=[1804],d=[1826]

- The row operations in (4) are (in order performed): R1↔R2-8R1+R2→R2-4R1+R3→R3(-6/5)R2+R3→R3(-5/22)R3→R3 are linearly independent. To do so, we must find the solutions to the vector equation $x_1\mathbf{a} + x_2\mathbf{b} + x_3\mathbf{c} + x_4\mathbf{d} = \mathbf{0}$. The augmented matrix of the equivalent linear system and the echelon form are

[1622181824028446|000]~[24020-101820041|000] (4)

Back substitution leads to the general solution

x1=-12s1,x2=-14s1,x3=-14s1,x4=s1

which holds for all choices of s_1. Thus there exist nontrivial solutions, so the set of vectors is linearly dependent. Letting $s_1 = 1$, we have x1=-12, x2=-14, x3=-14, and $x_4 = 1$, which gives us

-12a-14b-14c+d=0⇒d=12a+14b+14c

Thus we obtain a serving of brand D by combining a 12 serving of A, a 14 serving of B, and a 14 serving of C. Hence there is no need to stock brand D.

When working with a new concept, it can be helpful to begin with simple cases, so let's start with the set $\{\mathbf{u}_1\}$. Is this linearly independent? To check, we need to determine the solutions to the equation

x1u1=0 (5)

At first glance, it seems that the only solution is the trivial one $x_1 = 0$, and for most choices of $\mathbf{u}_1$ that is true. Specifically, as long as $\mathbf{u}_1 \neq \mathbf{0}$, then the only solution is $x_1 = 0$ and the set $\{\mathbf{u}_1\}$ is linearly independent. But if it happens that $\mathbf{u}_1 = \mathbf{0}$, then the set is linearly dependent, because now there are nontrivial solutions to (5), such as

3u1=0

In fact, having **0** in *any* set of vectors always guarantees that the set will be linearly dependent.

THEOREM 2.13

Suppose that $\{\mathbf{0}, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is a set of vectors in $\mathbf{R}^n$. Then the set is linearly dependent.

Proof We need to determine if the vector equation

$$x_0\mathbf{0}+x_1\mathbf{u}_1+x_2\mathbf{u}_2+\cdots+x_m\mathbf{u}_m=\mathbf{0}$$

has any nontrivial solutions. Regardless of the values of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$, setting $x_0 = 1$ and $x_1 = x_2 = \cdots = x_m = 0$ gives us an easy (but legitimate) nontrivial solution, so that the set is linearly dependent.

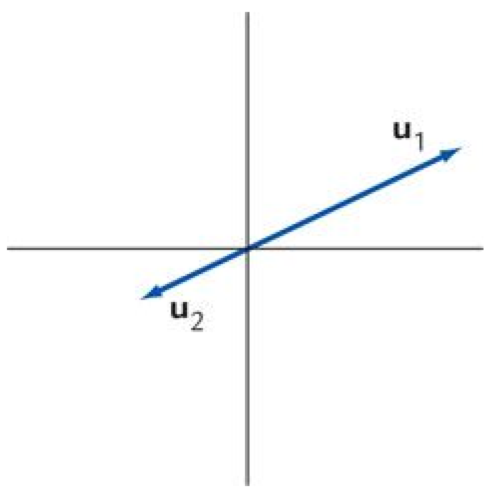


Figure 1 $\mathbf{u}_1$ and are $\mathbf{u}_2$ linearly dependent vectors.

For a set of two vectors $\{\mathbf{u}_1, \mathbf{u}_2\}$, we already know what happens if one of these is **0**. Let's assume that both vectors are nonzero and ask the question: Is this set linearly independent? As usual, we need to determine the nature of the solutions to

$$x_1u_1+x_2u_2=0 \tag{6}$$

If there is a nontrivial solution, then it must be that both x_1 and x_2 are nonzero. (Why?) In this case, we can solve (6) for $\mathbf{u}_1$, giving us

$$u_1=-\frac{x_2}{x_1}u_2$$

Thus we see that the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly dependent if and only if $\mathbf{u}_1$ is a scalar multiple of $\mathbf{u}_2$. Geometrically, the set is linearly dependent if and only if the two vectors point in the same (or opposite) direction (see Figure 1).

When trying to determine if a set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ of three or more vectors is linearly dependent or independent, in general we have to find the solutions to

$$x_1u_1+x_2u_2+\cdots+x_mu_m=0$$

However, there is a special case where virtually no work is required.

THEOREM 2.14 ▶

Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ is a set of vectors in $\mathbf{R}^n$. If $n < m$, then the set is linearly dependent.

In other words, if the number of vectors m exceeds the number of components n, then the set is linearly dependent.

Proof As usual when testing for linear independence, we start with the vector equation

$$x_1u_1+x_2u_2+\cdots+x_mu_m=0 \tag{7}$$

Since $\mathbf{u}_1, \dots, \mathbf{u}_m$ each have n components, this is equivalent to a homogeneous linear system with n equations and m unknowns.

Because $n < m$, this system has more variables than equations, so there are infinitely many solutions (see Exercise 55 in Section 1.2). Therefore (7) has nontrivial solutions, and hence the set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ is linearly dependent. ■ ■

Theorem 2.14 immediately tells us that the myology clinic's set of four nutritional powder brands are not all needed, because we can represent them by four vectors ($\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$) in $\mathbf{R}^3$ (protein, fat, carbohydrates). However, the theorem does not tell us which brand can be eliminated.

Note also that Theorem 2.14 tells us nothing about the case when $n \geq m$. In this instance, it is possible for the set to be linearly dependent, as in (1), where $n = 3$ and $m = 3$. Or the set can be linearly independent, as in Example 1, where $n = 4$ and $m = 3$.

Span and Linear Independence

Span and linear independence both involve sets of vectors and are related. To see the connection, let's return to the above discussion about the set of two nonzero vectors $\{\mathbf{u}_1, \mathbf{u}_2\}$. This set is linearly dependent exactly when $\mathbf{u}_1$ is a multiple of $\mathbf{u}_2$—that is, exactly when $\mathbf{u}_1$ is in span$\{\mathbf{u}_2\}$. This connection between span and linear independence holds more generally.

THEOREM 2.15 ▶

Let $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ be a set of vectors in $\mathbf{R}^n$. Then this set is linearly dependent if and only if one of the vectors in the set is in the span of the other vectors.

Proof First suppose that the set is linearly dependent. Then there exist scalars $c_1, \dots, c_m$, not all zero, such that

$$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m = 0$$

Without loss of generality, assume that $c_1 \neq 0$. Then we can solve for $\mathbf{u}_1$,

$$u_1 = -\frac{c_2}{c_1} u_2 - \cdots - \frac{c_m}{c_1} u_m$$

which shows that $\mathbf{u}_1$ is in span$\{\mathbf{u}_2, \dots, \mathbf{u}_m\}$. This completes one direction of the proof.

Now suppose that one of the vectors in the set is in the span of the remaining vectors—say, $\mathbf{u}_1$ is in span$\{\mathbf{u}_2, \dots, \mathbf{u}_m\}$. Then there exist scalars $c_2, c_3, \dots, c_m$ such that

$$u_1 = c_2 u_2 + \cdots + c_m u_m \tag{8}$$

Moving all terms to the left side in (8), we have

$$u_1 - c_2 u_2 - \cdots - c_m u_m = 0$$

Since the coefficient on $\mathbf{u}_1$ is nonzero, this shows the set is linearly dependent, completing the proof. ■■

Example 3

Give a linearly dependent set of vectors such that one vector is not a linear combination of the others. Explain why this does not contradict Theorem 2.15.

Solution Theorem 2.15 tells us that in a linearly dependent set, at least one vector is a linear combination of the other vectors. However, it does not say that *every* vector is a linear combination of the others, so what we seek does not contradict Theorem 2.15.

Let

u1=[-10-2],u2=[3-22],u3=[5-26],u4=[123]

By Theorem 2.14, this set must be linearly dependent because there are four vectors with three components. Theorem 2.15 says that one of the vectors is a linear combination of the others, and indeed we have $\mathbf{u}_3 = -2\mathbf{u}_1 + \mathbf{u}_2$. On the other hand, the equation

x1u1+x2u2+x3u3=u4 (9)

has corresponding augmented matrix and echelon form

[-1350-2-2-226|123]~[-1350-2-2000|12-3] (10)

▶ The row operations in (10) are (in order performed):
-2R1+R3→R3-2R2+R3→R3

From the echelon form it follows that (9) has no solutions, so that $\mathbf{u}_4$ is not a linear combination of the other vectors in the set.

The next theorem relates the location of pivot positions to span and linear independence.

THEOREM 2.16 ▶

Let $\mathbf{u}_1, \ldots, \mathbf{u}_m$ be in $\mathbf{R}^n$, and suppose

A=[u1 u2 … um]~B

where B is in echelon form. Then

(a) span$\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} = \mathbf{R}^n$ exactly when B has a pivot position in every row.

(b) $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is linearly independent exactly when B has a pivot position in every column.

Proof Part (a) is from Theorem 2.8. For part (b), if B has a pivot position in every column, then the general solution to $A\mathbf{x} = \mathbf{0}$ has a unique solution because there are no free variables. That solution must be the trivial solution, so $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is linearly independent. If a column of B lacks a pivot position, then the general solution to $A\mathbf{x} = \mathbf{0}$ will have free parameters and thus there will be infinitely many solutions, implying $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is linearly dependent. ■ ■

Returning to the vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_4\}$ in Example 3, we see in (10) that every row has a pivot position, so Theorem 2.16(a) implies $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_4\} = \mathbf{R}^3$. Since the third column in (10) does not have a pivot position, the set $\{\mathbf{u}_1, \dots, \mathbf{u}_4\}$ is linearly dependent.

Homogeneous Systems

In Section 2.2, we introduced the notation

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m$$

where $A = [\mathbf{a}_1\ \mathbf{a}_2\ \dots\ \mathbf{a}_m]$ and $\mathbf{x} = (x_1, x_2, \dots, x_m)$, and noted that any linear system can be expressed in the compact form

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

The system $A\mathbf{x} = \mathbf{0}$ is a **homogeneous** linear system, introduced in Section 1.2. There we showed that homogeneous linear systems have either one solution (the trivial solution) or infinitely many solutions.

The next theorem highlights the direct connection between the number of solutions to $A\mathbf{x} = \mathbf{0}$ and whether the columns of A are linearly independent.

THEOREM 2.17 ▶

Let $A = [\mathbf{a}_1\ \mathbf{a}_2\ \dots\ \mathbf{a}_m]$ and $\mathbf{x} = (x_1, x_2, \dots, x_m)$. The set $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ is linearly independent if and only if the homogeneous linear system

Ax=0

has only the trivial solution.

Proof Written as a vector equation, the system $A\mathbf{x} = \mathbf{0}$ has the form

x1a1+x2a2+···+xmam=0 (11)

Thus if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then so does (11) and the columns of A are linearly independent. On the other hand, if $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions, then so does (11) and the columns of A are linearly dependent. ■ ■

Nonhomogeneous, Associated Homogeneous System

If $\mathbf{b} \neq \mathbf{0}$, then the system $A\mathbf{x} = \mathbf{b}$ is **nonhomogeneous**, and the **associated homogeneous system** is $A\mathbf{x} = \mathbf{0}$. There is a close connection between the set of solutions to a nonhomogeneous system $A\mathbf{x} = \mathbf{b}$ and the associated homogeneous system $A\mathbf{x} = \mathbf{0}$, illustrated in the following example.

Example 4

Find the general solution for the linear system

2x1-6x2-x3+8x4=7x1-3x2-x3+6x4=6-x1+3x2-x3+2x4=4 (12)

and the general solution for the associated homogeneous system.

Solution Applying our usual matrix and row reduction methods, we find that the general solution to (12) is

x=[10-50]+s1[3100]+s2[-2041]

The general solution to the associated homogeneous system was found in Example 9 of Section 1.2. It is

x=s1[3100]+s2[-2041]

For both solutions, s_1 and s_2 can be any real numbers.

Comparing the preceding solutions, we see that the only difference is the “constant’’ vector

[10-50]

This type of relationship between general solutions occurs in all such cases. To see why, it is helpful to have the following result showing that the product $A\mathbf{x}$ obeys the distributive law.

THEOREM 2.18

Suppose that $A = [\mathbf{a}_1\ \mathbf{a}_2\ \dots\ \mathbf{a}_m]$, and let $\mathbf{x} = (x_1, x_2, \dots, x_m)$ and $\mathbf{y} = (y_1, y_2, \dots, y_m)$. Then

(a) $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$

(b) $A(\mathbf{x} - \mathbf{y}) = A\mathbf{x} - A\mathbf{y}$

Proof The results follow from the definition of the product and a bit of algebra. Starting with (a), we have

x+y=[x1+y1x2+y2⋮xm+ym]

so that

$$A(\mathbf{x} + \mathbf{y}) = (x_1+y_1)\mathbf{a}_1 + (x_2+y_2)\mathbf{a}_2 + \cdots + (x_m+y_m)\mathbf{a}_m = (x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m) + (y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \cdots + y_m\mathbf{a}_m) = A\mathbf{x} + A\mathbf{y}$$

The proof of (b) is similar and left as an exercise. ■ ■

Particular Solution

Suppose $\mathbf{x}_p$ is a specific solution to $A\mathbf{x} = \mathbf{b}$. We call $\mathbf{x}_p$ a **particular solution** to the system, and it can be thought of as a fixed solution to the system.

THEOREM 2.19 ▶

Let $\mathbf{x}_p$ be a particular solution to

$$A\mathbf{x} = \mathbf{b} \tag{13}$$

Then all solutions $\mathbf{x}_g$ to (13) have the form $\mathbf{x}_g = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_h$ is a solution to the associated homogeneous system $A\mathbf{x} = \mathbf{0}$.

Proof Let $\mathbf{x}_p$ be a particular solution to (13), and suppose that $\mathbf{x}_g$ is any solution to the same system. Let $\mathbf{x}_h = \mathbf{x}_g - \mathbf{x}_p$. Then

$$A\mathbf{x}_h = A(\mathbf{x}_g - \mathbf{x}_p) = A\mathbf{x}_g - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

Thus $\mathbf{x}_h$ is a solution to the associated homogeneous system $A\mathbf{x} = \mathbf{0}$. Solving for $\mathbf{x}_g$, we have

$$\mathbf{x}_g = \mathbf{x}_p + \mathbf{x}_h$$

so that $\mathbf{x}_g$ has the form claimed. ■ ■

Example 5

Find the general solution and solution to the associated homogeneous system for

$$4x1-6x2=-14-6x1+9x2=21 \tag{14}$$

Solution Applying our standard solution procedures yields the general solution to (14)

$$xg=[13]+s[32]$$

which is a line when graphed in $\mathbf{R}^2$. The solutions to the associated homogeneous system are

$$xh=s[32]$$

which is also a line when graphed in $\mathbf{R}^2$. If we let xp=[13], then every value of $\mathbf{x}_g$ can be expressed as the sum of $\mathbf{x}_p$ and one of the homogeneous solutions $\mathbf{x}_h$. Thus the general solution $\mathbf{x}_g$ is a translation by $\mathbf{x}_p$ of the general solution $\mathbf{x}_h$ of the associated homogeneous system. The graphs are shown in Figure 2.

At the end of Section 2.2, Theorem 2.11 linked span with solutions to linear systems. Theorem 2.20 is similar in spirit, this time linking linear independence with solutions to linear systems.

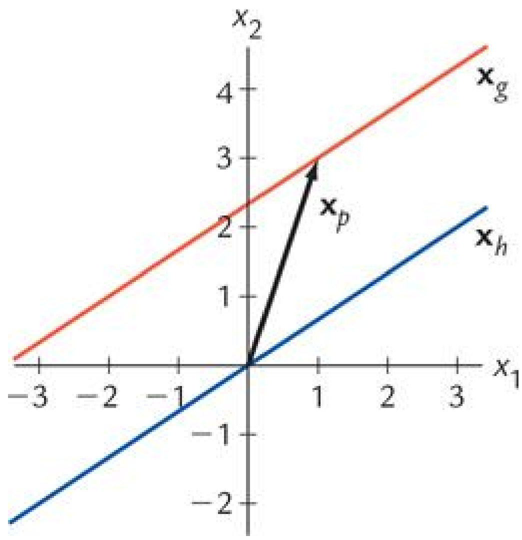


Figure 2 Graphs of $\mathbf{x}_g$, $\mathbf{x}_h$, and $\mathbf{x}_p$ from Example 5.

THEOREM 2.20 ▶

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ and $\mathbf{b}$ be vectors in $\mathbf{R}^n$. Then the following statements are equivalent. That is, if one is true, then so are the others, and if one is false, then so are the others.

(a) The set $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ is linearly independent.

(b) The vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{b}$ has at most one solution for every $\mathbf{b}$.

(c) The linear system corresponding to$[\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_m \mid \mathbf{b}]$ has at most one solution for every $\mathbf{b}$.

(d) The equation $A\mathbf{x} = \mathbf{b}$, with $A = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_m]$, has at most one solution for every $\mathbf{b}$.

Proof The equivalence of (b), (c), and (d) is immediate from the definitions, so we can complete the proof by showing that (a) and (b) are equivalent.

We start by showing that (a) implies (b). Let $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ be linearly independent, and suppose to the contrary that

$$x_1a_1+x_2a_2+\cdots+x_ma_m=b$$

has more than one solution. Then there exist scalars $r_1, \dots, r_m$ and $s_1, \dots, s_m$ such that

$$r_1a_1+r_2a_2+\cdots+r_ma_m=b \quad s_1a_1+s_2a_2+\cdots+s_ma_m=b$$

and so

$$r_1a_1+r_2a_2+\cdots+r_ma_m=s_1a_1+s_2a_2+\cdots+s_ma_m$$

Moving all terms to one side and regrouping yields

$$(r_1-s_1)a_1+(r_2-s_2)a_2+\cdots+(r_m-s_m)a_m=0$$

Since $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ is linearly independent, each coefficient must be 0. Hence $r_1 = s_1, \dots, r_m = s_m$, so there is just one solution to $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{b}$.

Proving that (b) implies (a) is easier. Since $\mathbf{b}$ can be any vector, we can set $\mathbf{b} = \mathbf{0}$. By (b) there is at most one solution to

$$x_1a_1+x_2a_2+\cdots+x_ma_m=0$$

Of course, there is the trivial solution $x_1 = \cdots = x_m = 0$, so this must be the only solution. Hence the set $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ is linearly independent, and (a) follows. ■ ■

The Unifying Theorem – Version 1

Next, we present the first version of the Unifying Theorem, so named because it will serve to unify many of the ideas that we are developing.

THEOREM 2.21 ▶

THE UNIFYING THEOREM – VERSION 1 Let $S = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ be a set of n vectors in $\mathbf{R}^n$, and let $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$. Then the following are equivalent:

(a) S spans $\mathbf{R}^n$.

(b) S is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

▶ In the Unifying Theorem, n is both the number of vectors in S and the number of components in each vector. Thus A has n rows and n columns.

Proof We start by showing that (a) and (b) are equivalent. First suppose that S spans $\mathbf{R}^n$. If S is linearly dependent, then one of $\mathbf{a}_1, \dots, \mathbf{a}_n$—say, $\mathbf{a}_1$—is a linear combination of the others. Then by Theorem 2.7, it follows that

span{a1,… ,an}=span{a2,… ,an}

But this implies that $\mathbf{R}^n = \text{span}\{\mathbf{a}_2, \dots, \mathbf{a}_n\}$, contradicting Theorem 2.9. Hence it must be that S is linearly independent. This shows that (a) implies (b).

To show that (b) implies (a), we assume that S is linearly independent. Now, if S does not span $\mathbf{R}^n$, then there exists a vector that is not a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$. Since S is linearly independent, it follows that the set $\{\mathbf{a}, \mathbf{a}_1, \dots, \mathbf{a}_n\}$ of $n + 1$ vectors is also linearly independent, contradicting Theorem 2.14. Hence S must span $\mathbf{R}^n$. Thus (b) implies (a), and therefore (a) is equivalent to (b).

Now suppose that (a) and (b) are both true. Then by Theorem 2.11, (a) implies that $A\mathbf{x} = \mathbf{b}$ has *at least* one solution for every $\mathbf{b}$ in $\mathbf{R}^n$. On the other hand, from Theorem 2.20 we know that (b) implies that $A\mathbf{x} = \mathbf{b}$ has *at most* one solution for every $\mathbf{b}$ in $\mathbf{R}^n$. This leaves us with only one possibility, that $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $\mathbf{b}$ in $\mathbf{R}^n$, confirming (c) is true.

Finally, suppose that (c) is true. Since $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbf{R}^n$, then in particular $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Appealing to Theorem 2.17, we conclude that S must be linearly independent and hence also spans $\mathbf{R}^n$. ■ ■

Example 6

Suppose that

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 7 \\ -2 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 5 \\ 2 \\ -6 \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$$

Show that the columns of A are linearly independent and span $\mathbf{R}^3$, and that $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbf{R}^3$.

Solution We start with linear independence, so we need to find the solutions to

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{0} \tag{15}$$

The corresponding augmented matrix and echelon form are

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 7 & 0 & 2 & 0 \\ -2 & 1 & -6 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 7 & 4 & 0 & 0 \\ 0 & -2 & 1 & 0 \end{array}\right] \tag{16}$$

From the echelon form it follows that (15) has only the trivial solution, so the columns of A are linearly independent.

Because we have three vectors and each has three components, the other questions follow immediately from the Unifying Theorem. Specifically, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is linearly independent, so the set must also span $\mathbf{R}^3$ and there is exactly one solution to $A\mathbf{x} = \mathbf{b}$ for any $\mathbf{b}$ in $\mathbf{R}^3$.

For one more application of Theorem 2.21, let's return to the nutritional powder problem described at the beginning of the section.

Example 7

Use the Unifying Theorem to show that stocking powder brands A, B, and C is efficient.

Solution We previously determined that the nutrient vectors **a**, **b**, **c**, and **d** are linearly dependent, and we concluded that brand D can be eliminated. For the remaining three brands, it can be verified that

$$x_1\mathbf{a}+x_2\mathbf{b}+x_3\mathbf{c}=\mathbf{0}$$

has only the trivial solution, which tells us that **a**, **b**, and **c** are linearly independent. By Theorem 2.21, we can conclude the following:

- The vectors **a**, **b**, and **c** span all of $\mathbf{R}^3$. Therefore *every* vector in $\mathbf{R}^3$ can be expressed as a linear combination of these three vectors. (Note, though, that some combinations will require negative values of x_1, x_2, and x_3, which is not physically possible when combining powders.)
- Item (c) of the Unifying Theorem tells us that there is exactly one way to combine brands A, B, and C to create any blend with a specific combination of protein, fat, and carbohydrates. Thus stocking brands A, B, and C is efficient, in that there is no redundancy.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine if the given vectors are linearly independent.
   (a) u1=[2-3],u2=[41]

(b) u1=[614],u2=[-23-3]

2. Determine if the columns of A are linearly independent.

   (a) A=[153-4]

   (b) A=[1032-24-372]

3. Determine if the system $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

   (a) A=[142284]

   (b) A=[10-11-1-101-2210]

4. Determine if each statement is true or false, and justify your answer.

   (a) If a set of vectors is linearly independent in $\mathbf{R}^n$, then the set spans $\mathbf{R}^n$.

   (b) If A is a 3 × 3 matrix with columns that span $\mathbf{R}^3$, then the columns are linearly independent.

   (c) If $\mathbf{u}_1 = 4\mathbf{u}_2$, then $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly dependent.

   (d) If the columns of A are linearly dependent, then $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions for every $\mathbf{b}$.

EXERCISES

Exercises 1–6: Determine if the given vectors are linearly independent.

1. u=[3-2],v=[-1-4]
2. u=[6-15],v=[-4-10]
3. u=[71-13],v=[5-32]
4. u=[-40-3],v=[-2-15],w=[-82-19]
5. u=[3-12],v=[041],w=[247]
6. u=[1833],v=[4-25-5],w=[-1201]

Exercises 7–12: Determine if the columns of A are linearly independent.

7. A=[15-6-52]
8. A=[4-1226]

9. A=[10-225-7]
10. A=[1-12-45-5-121]
11. A=[3105-2-14-4-3]
12. A=[-4-710035-1182-4]

Exercises 13–18: A matrix *A* is given. Determine if the homogeneous system *A***x** = **0** (where **x** and **0** have the appropriate number of components) has any nontrivial solutions.

13. A=[-3541]
14. A=[121065]
15. A=[810-1-32]
16. A=[-3211-1-15-4-3]
17. A=[-1314-3-1305]
18. A=[2-30012-53-9309]

Exercises 19–24: Determine by inspection (that is, with only minimal calculations) if the given vectors form a linearly dependent or linearly independent set. Justify your answer.

19. u=[14-6],v=[7-3]
20. u=[21],v=[53]
21. u=[3-1],v=[6-5],w=[14]
22. u=[6-42],v=[3-2-1]
23. u=[1-83],v=[000],w=[-7112]
24. u=[1234],v=[1234],w=[4321]

Exercises 25–28: Determine if one of the given vectors is in the span of the other vectors. (HINT: Check to see if the vectors are linearly dependent, and then appeal to Theorem 2.15.)

25. u=[62-5],v=[170]
26. u=[27-1],v=[116],w=[130]
27. u=[4-13],v=[35-2],w=[-57-7]
28. u=[1784],v=[-1352],w=[31-20]

Exercises 29–32: For each matrix A, determine if $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbf{R}^2$. (HINT: The Unifying Theorem is helpful here.)

29. A=[2-110]

30. A=[41-82]

31. A=[6-9-46]

32. A=[1-227]

Exercises 33–36: For each matrix A, determine if $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbf{R}^3$. (HINT: The Unifying Theorem is helpful here.)

33. A=[2-10101-345]

34. A=[3477-16-202]

35. A=[3-21-410-501]

36. A=[1-3-2011247]

FIND AN EXAMPLE Exercises 37–42: Find an example that meets the given specifications.

37. Three distinct nonzero linearly dependent vectors in $\mathbf{R}^4$.

38. Three linearly independent vectors in $\mathbf{R}^5$.

39. Three distinct nonzero linearly dependent vectors in $\mathbf{R}^2$ that do not span $\mathbf{R}^2$.

40. Three distinct nonzero vectors in $\mathbf{R}^2$ such that any pair is linearly independent.

41. Three distinct nonzero linearly dependent vectors in $\mathbf{R}^3$ such that each vector is in the span of the other two vectors.

42. Four vectors in $\mathbf{R}^3$ such that no vector is a nontrivial linear combination of the other three. (Explain why this does not contradict Theorem 2.15.)

TRUE OR FALSE Exercises 43–52: Determine if the statement is true or false, and justify your answer.

43.

(a) If a set of vectors in $\mathbf{R}^n$ is linearly dependent, then the set must span $\mathbf{R}^n$.

(b) If a set of vectors spans $\mathbf{R}^n$, then the set must be linearly independent.

44.

(a) If $m > n$, then a set of m vectors in $\mathbf{R}^n$ is linearly dependent.

(b) If $m > n$, then a set of m vectors in $\mathbf{R}^n$ must span $\mathbf{R}^n$.

45.

(a) If $A \sim B$ and B has a pivot position in every row, then the columns of A are linearly independent.

(b) If $A \sim B$ and B has a pivot position in every column, then the columns of A are linearly independent.

46.

(a) If $A\mathbf{x} = \mathbf{b}$ has more than one solution for a vector $\mathbf{b}$, then the columns of A must be linearly independent.

(b) If A has linearly dependent columns, then the columns of A must span $\mathbf{R}^n$.

47.

(a) If A is a matrix with more rows than columns, then the columns of A are linearly independent.

(b) If A is a matrix with more columns than rows, then the columns of A are linearly independent.

48.

(a) If A is a matrix with linearly independent columns, then $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

(b) If A is a matrix with linearly independent columns, then $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$.

49.

(a) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly independent, then so is $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

(b) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly dependent, then so is $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

50.

(a) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is linearly independent, then so is $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

(b) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is linearly dependent, then so is $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

51.

(a) If $\mathbf{u}_4$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is linearly independent.

(b) If $\mathbf{u}_4$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is linearly dependent.

52.

(a) If $\mathbf{u}_4$ is *not* a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is linearly independent.

(b) If $\mathbf{u}_4$ is *not* a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is linearly dependent.

53. Which of the following sets of vectors in $\mathbf{R}^3$ could possibly be linearly independent? Justify your answer.

(a) $\{\mathbf{u}_1\}$

(b) $\{\mathbf{u}_1, \mathbf{u}_2\}$

(c) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$

(d) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$

54. Which of the following sets of vectors in $\mathbf{R}^3$ could possibly be linearly independent *and* span $\mathbf{R}^3$? Justify your answer.

(a) $\{\mathbf{u}_1\}$

(b) $\{\mathbf{u}_1, \mathbf{u}_2\}$

(c) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$

(d) $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$

55. Prove that if c_1, c_2, and c_3 are nonzero scalars and $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a linearly independent set of vectors, then so is $\{c_1\mathbf{u}_1, c_2\mathbf{u}_2, c_3\mathbf{u}_3\}$.

56. Prove that if $\mathbf{u}$ and $\mathbf{v}$ are linearly independent vectors, then so are $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$.

57. Prove that if $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a linearly independent set of vectors, then so is $\{\mathbf{u}_1 + \mathbf{u}_2, \mathbf{u}_1 + \mathbf{u}_3, \mathbf{u}_2 + \mathbf{u}_3\}$.

58. Prove that if $U = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is linearly independent, then any nonempty subset of U is also linearly independent.

59. Prove that if a set of vectors is linearly dependent, then adding additional vectors to the set will create a new set that is still linearly dependent.

60. Prove that if $\mathbf{u}$ and $\mathbf{v}$ are linearly independent and the set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a linearly dependent set, then $\mathbf{w}$ is in span$\{\mathbf{u}, \mathbf{v}\}$.

61. Prove that two nonzero vectors **u** and **v** are linearly dependent if and only if **u** = c**v** for some scalar c.

62. Let A be an $n \times m$ matrix that is in echelon form. Prove that the nonzero rows of A, when considered as vectors in $\mathbf{R}^m$, are a linearly independent set.

63. Prove part (b) of Theorem 2.18.

64. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ be a linearly dependent set of nonzero vectors. Prove that some vector in the set can be written as a linear combination of a linearly independent subset of the remaining vectors, with the set of coefficients all nonzero and unique for the given subset. (HINT: Start with Theorem 2.15.)

Exercises 65–70: Determine if the given vectors form a linearly dependent or linearly independent set.

65. [2-35], [3-42], [-117]

66. [-423], [131], [-354]

67. [201-1], [-3256], [670-5], [5-37-3]

68. [35-2-4], [2-43-1], [-4662], [-7226]

69. [2-5-113], [132-21], [-11-20-4], [321-31]

70. [42-152], [231-10], [-32111], [02-132]

Exercises 71–72: Determine if $A\mathbf{x} = \mathbf{b}$ has a unique solution for every **b** in $\mathbf{R}^3$.

71. A=[1-245-3-1-3-7-9],x=[x1x2x3]

72. A=[3-2520-4-271],x=[x1x2x3]

Exercises 73–74: Determine if $A\mathbf{x} = \mathbf{b}$ has a unique solution for every **b** in $\mathbf{R}^4$.

73. A=[25-36-101-152-393-468],x=[x1x2x3x4]

74. A=[5108-24311-382503-18],x=[x1x2x3x4]

SUPPLEMENTARY EXERCISES

Exercises 1–14: Let

u=[1-32] v=[-241] and w=[1-57]

1. Compute **u** + **v** and 3**w**.
2. Compute **v** − **w** and −4**u**.
3. Compute 2**w** + 3**v** and 2**u** − 5**w**.
4. Compute 3**v** + 2**u** and −2**u** + 4**w**.
5. Compute 2**u** + **v** + 3**w** and **u** − 3**v** + 2**w**.
6. Compute **u** − 2**v** + 4**w** and −3**u** + **v** − 2**w**.
7. Express the vector equation x_1**u** + x_2**v** = **w** as a system of linear equations.
8. Express the vector equation x_1**w** + x_2**u** = −2**v** as a system of linear equations.
9. Find three different vectors that are linear combinations of **u** and **v**.
10. Find three different vectors that are linear combinations of **w** and **v**.
11. Determine if **w** is a linear combination of **u** and **v**.
12. Determine if **v** is a linear combination of **w** and **u**.
13. Determine if {**u**, **v**, **w**} is linearly independent.
14. Determine if span{**u**, **v**, **w**} = **R**^{3}.

Exercises 15–16: Express the system of linear equations as a vector equation.

15. 4x1+13x2-x3=-7x1-7x2+4x3=12
16. 3x1-2x2-x3+2x4=0-x1+5x2 +x4=-7-3x1 +10x3-3x4=2

Exercises 17–20: Express the given general solution to a linear system as a linear combination of vectors.

17. x1=-1+2s1x2=3s1x3=s1

18. x1=5-7s1x2=-4x3=s1

19. x1=3-5s1-s2x2=s2x3=-1+8s1x4=s1

20. x1=1+s1+6s3x2=s3x3=6-s1+4s2x4=s2x5=s1

Exercises 21–22: Find the unknowns in the vector equation.

21. 2[-3a]+[-10]-2[b4]=[-25]

22. -[a1-2]+3[3b0]=[1-4c]

Exercises 23–24: Determine if **b** is a linear combination of the other vectors. If so, express **b** as a linear combination.

23. a1=[1-24],a2=[23-1],b=[-1-1110]

24. a1=[1-302],a2=[02-11],a3=[-203-1],b=[-2-453]

Exercises 25–26: Find *A*, **x**, and **b** such that *A***x** = **b** corresponds to the linear system.

25. 2x1+3x2-8x3+x4=56x1-x2+4x3-2x4=9

26. 3x1-x2-7x3=2-4x1+5x2 =-4-8x1+2x2+6x3=3x1+3x2+9x3=7

Exercises 27–28: Determine if **b** is in the span of the other vectors. If so, express **b** as a linear combination.

27. a1=[3-1-2],a2=[145],b=[-157]

28. a1=[1310],a2=[-1234],a3=[220-1],b=[-34-71]

Exercises 29–32: Determine if the given vectors span $\mathbf{R}^2$.

29. a1=[3-1]

30. a1=[6-9],a2=[-23]

31. a1=[12],a2=[-35]

32. a1=[13],a2=[-1-3],a3=[24]

Exercises 33–36: Determine if the given vectors span $\mathbf{R}^3$.

33. a1=[2-17]

34. a1=[16-5],a2=[-253]

35. a1=[125],a2=[-3-54],a3=[4611]

36. a1=[1-31],a2=[-12-2],a3=[1-51],a4=[-22-6]

Exercises 37–44: Determine if the given vectors are linearly independent.

37. a1=[1-5],a2=[-29]

38. a1=[9-6],a2=[-64]

39. a1=[9-6],a2=[-34],a3=[15]

40. a1=[16-2],a2=[-230]

41. a1=[14-5],a2=[-2-810]

42. a1=[1-13],a2=[-234],a3=[2-59]

43. a1=[302],a2=[-23-4],a3=[09-8]

44. a1=[521],a2=[-360],a3=[127],a4=[-54-6]

CHAPTER 3

Matrices

Eric Kruszewski/National Geographic Magazines/Getty Images

Wind power is another form of renewable energy. Shown in the photograph are wind turbines located near Kennewick, Washington. These turbines use the power of the wind to turn kinetic energy into mechanical power, which can then be converted to electrical power. Wind turbines are engineered in different shapes and sizes and can rotate on either a horizontal or vertical axis.

n this chapter we expand our development of matrices. Thus far we have used matrices to solve systems of equations and have defined how to multiply a matrix times a vector. In Section 3.1 we use this multiplication to define an important type of function called a linear transformation and investigate its properties and applications. Section 3.2 and Section 3.3 focus on the algebra of matrices. In Section 3.4 we develop a factorization method that uses matrix multiplication to efficiently find solutions to linear systems. Section 3.5 is about Markov Chains, an application of matrix multiplication that arises in a variety of contexts.

3.1 Linear Transformations

In this section we consider an important class of functions called *linear transformations*. These functions arise in many fields and can be defined naturally in terms of matrix multiplication. The following example gives a sense of how one might encounter linear transformations.

A consumer electronics company makes two different types of smart phones, the j8 and the j8+. The manufacturing cost includes labor, materials, and overhead (facilities, etc.). The company's costs (in dollars) per unit for each type are summarized in Table 1.

Table 1 **Smart Phone Manufacturing Costs**

	j8	j8+
Labor	57	73
Materials	93	101
Overhead	29	34

This table can be organized into two cost vectors,

$$\mathbf{j8}=\begin{bmatrix}57\\93\\29\end{bmatrix},\quad \mathbf{j8+}=\begin{bmatrix}73\\101\\34\end{bmatrix}$$

To determine costs for different manufacturing levels, we compute linear combinations of these vectors. For instance, the cost vector for producing 12 j8's and 19 j8+'s is

$$12\,\mathbf{j8}+19\,\mathbf{j8+}=12\begin{bmatrix}57\\93\\29\end{bmatrix}+19\begin{bmatrix}73\\101\\34\end{bmatrix}=\begin{bmatrix}2071\\3035\\994\end{bmatrix}$$

For this production mix, the company will incur costs of \$2071 for labor, \$3035 for materials, and \$994 for overhead. More generally, let

$$\mathbf{x}=\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

where x_1 and x_2 indicate desired production levels for j8's and j8+'s, respectively. Let T be the function that takes the production vector **x** as input and produces the corresponding cost vector as output. Using the individual cost vectors, we have

T(x)=x1[579329]+x2[7310134]

Now suppose that we define the 3 × 2 matrix

A=[j8j8+]=[5773931012934]

Recalling the formula for multiplying a matrix by a vector (see Definition 2.10 in Section 2.2), we see that we can write the cost function compactly as

T(x)=[5773931012934] [x1x2]=Ax

For example a production level of 10 j8's and 35 j8+'s will have a cost vector

T([1035])=[5773931012934] [1035]=[312544651480]

Linear Transformations

The smart phone cost function is an example of an important class of functions called *linear transformations*. Let

T:Rm→Rn

Domain, Codomain, Image, Range

denote a function T with **domain** $\mathbf{R}^m$ (the input vectors) and **codomain** $\mathbf{R}^n$ (the output vectors). For every vector **u** in $\mathbf{R}^m$, the vector $T(\mathbf{u})$ is called the **image of u under** T. The set of all images of vectors **u** in $\mathbf{R}^m$ under T is called the **range** of T, denoted range(T). The range of T is a subset of the codomain of T.

▶ The range of T is sometimes referred to as the **image** of T.

DEFINITION 3.1 ▶

Linear Transformation

A function T: $\mathbf{R}^m \to \mathbf{R}^n$ is a **linear transformation** if for all vectors **u** and **v** in $\mathbf{R}^m$ and all scalars r we have

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

(b) $T(r\mathbf{u}) = rT(\mathbf{u})$

Example 1

Let T([x1x2])=[3x1-x22x1+5x2]. Find T([2-1]) and show that *T* is a linear transformation.

Solution The domain and range of T are both $\mathbf{R}^2$, so $T : \mathbf{R}^2 \to \mathbf{R}^2$. For the first part we have

T([2-1])=[3(2)-(-1)2(2)+5(-1)]=[7-1]

To show T is a linear transformation requires verifying both parts of Definition 3.1. For part (a), we start by letting u=[u1u2] and v=[v1v2]. Then

T(u+v)=T([u1+v1u2+v2])=[3(u1+v1)-(u2+v2)2(u1+v1)+5(u2+v2)] = [3u1-u22u1+5u2]+[3v1-v22v1+5v2]=T(u)+T(v)

For part (b), suppose r is a scalar. Then

T(ru)=T([ru1ru2])=[3(ru1)-(ru2)2(ru1)+5(ru2)] =[r(3u1-u2)r(2u1+5u2)]=r[3u1-u22u1+5u2]=rT(u)

This verifies that both parts of the definition hold, so T is a linear transformation.

The next example shows that not all functions $T : \mathbf{R}^m \to \mathbf{R}^n$ are linear transformations.

Example 2

Show that $T : \mathbf{R}^3 \to \mathbf{R}^2$ defined by

$$T([x1x2x3])=[x1\text{-}3x2x2x32]$$

is *not* a linear transformation.

Solution We address this question by appealing to Definition 3.1. We only have to find a single specific example where one of the required conditions does not hold. One possibility is to let

$$u=[123]andr=2$$

Then we have

$$T(ru)=T(2[123])=T([246])=[2\text{-}3(4)(4)(6)2]=[\text{-}10144]$$

and

$$rT(u)=2T([123])=2[1\text{-}3(2)(2)(32)]=[\text{-}1036]$$

Since $T(r\mathbf{u}) \neq rT(\mathbf{u})$, it follows that T is not a linear transformation.

Conditions (a) and (b) of Definition 3.1 can be combined into a single condition

$$T(ru+sv)=rT(u)+sT(v) \qquad (1)$$

for all vectors **u** and **v** and all scalars r and s. The proof that these are equivalent is covered in Exercise 58. The following example shows that this is true for a specific case.

Example 3

Let

A=[4-1-2203],u=[-1-3],andv=[25]

Let $r = 2$, $s = -1$, and $T(\mathbf{x}) = A\mathbf{x}$. Verify that (1) holds in this case by separately computing $T(r\mathbf{u} + s\mathbf{v})$ and $rT(\mathbf{u}) + sT(\mathbf{v})$.

Solution We have

ru+sv=2[-1-3]+(-1)[25]=[-4-11]

so that

T(ru+sv)=A(ru+sv)=[4-1-2203] [-4-11]=[-5-14-33]

We also have

rT(u)+sT(v)=2[4-1-2203] [-1-3]+(-1)[4-1-2203] [25] =2[-1-4-9]+(-1)
[3615]=[-5-14-33]

Thus $T(r\mathbf{u} + s\mathbf{v}) = rT(\mathbf{u}) + sT(\mathbf{v})$.

Matrix Dimensions, Square Matrix

Suppose that A is a matrix with n rows and m columns. Then we say that A is an **n × m matrix** and that A has **dimensions** $n \times m$. If $n = m$, then A is a **square matrix**. For instance, if

A=[14-2093019112-1758]andB=[04132078-135986-41]

▶ We say that A is a 3-by-5 matrix and B is a 4-by-4 matrix.

then A is a 3×5 matrix and B is a 4×4 (square) matrix.

THEOREM 3.2 ▸

Let A be an $n \times m$ matrix, and define $T(\mathbf{x}) = A\mathbf{x}$. Then $T : \mathbf{R}^m \to \mathbf{R}^n$ is a linear transformation.

Proof To show that T is a linear transformation, we must verify that the two conditions in Definition 3.1 both hold. Starting with condition (a), given vectors **u** and **v**, we have

$$T(u+v)=A(u+v)=Au+Av=T(u)+T(v) \text{(by Theorem 2.18)}$$

That shows (a) holds. Condition (b) is covered in Exercise 57. Verifying the two conditions completes the proof. ■ ■

Since our smart phone cost function is $T(\mathbf{x}) = A\mathbf{x}$, it follows from Theorem 3.2 that T is a linear transformation. Furthermore, because A is a 3×2 matrix, the domain is $\mathbf{R}^2$ and the codomain is $\mathbf{R}^3$.

It turns out that *every* linear transformation $T : \mathbf{R}^m \to \mathbf{R}^n$ is of the form $T(\mathbf{x}) = A\mathbf{x}$ for some $n \times m$ matrix A. The proof is given in Theorem 3.8.

Example 4

Let

$$A=[1-2430-5], and w=[34] \tag{2}$$

Suppose that $T : \mathbf{R}^3 \to \mathbf{R}^2$ with $T(\mathbf{x}) = A\mathbf{x}$. Determine if **w** is in the range of T.

Solution In order for **w** to be in the range of T, there must exist a solution to $T(\mathbf{x}) = \mathbf{w}$, which is equivalent to the linear system $A\mathbf{x} = \mathbf{w}$. The corresponding augmented matrix and echelon form are

[1-2430-5|34]~[1-2406-17|3-5]

▶ Going forward, we usually will leave out the row operations.

Back substitution yields the general solution

x=[4/3-5/60]+s[5/317/61]

where s can be any real number. Thus $\mathbf{w}$ is the image of infinitely many different vectors and is in the range of T.

THEOREM 3.3 ▶

Let A=[a1a2…am] be an $n \times m$ matrix, and let $T : \mathbf{R}^m \to \mathbf{R}^n$ with $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation. Then

(a) The vector $\mathbf{w}$ is in the range of T if and only if $A\mathbf{x} = \mathbf{w}$ is a consistent linear system.

(b) $\text{range}(T) = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$.

Proof A vector $\mathbf{w}$ is in the range of T if and only if there exists a vector $\mathbf{u}$ such that $T(\mathbf{u}) = \mathbf{w}$. As $T(\mathbf{x}) = A\mathbf{x}$, this is equivalent to $A\mathbf{u} = \mathbf{w}$, which is true if and only if the linear system $A\mathbf{x} = \mathbf{w}$ is consistent.

For part (b), from Theorem 2.11 it follows that $A\mathbf{x} = \mathbf{w}$ is consistent if and only if $\mathbf{w}$ is in the span of the columns of A. Therefore $\text{range}(T) = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. ■ ■

One-to-One and Onto Linear Transformations

Here we consider two special types of linear transformations.

DEFINITION 3.4 ▶

One-to-One, Onto

Let $T : \mathbf{R}^m \to \mathbf{R}^n$ be a linear transformation. Then

(a) T is **one-to-one** if for every vector $\mathbf{w}$ in $\mathbf{R}^n$ there exists *at most* one vector $\mathbf{u}$ in $\mathbf{R}^m$ such that $T(\mathbf{u}) = \mathbf{w}$.

(b) T is **onto** if for every vector $\mathbf{w}$ in $\mathbf{R}^n$ there exists *at least* one vector $\mathbf{u}$ in $\mathbf{R}^m$ such that $T(\mathbf{u}) = \mathbf{w}$.

▶ An alternative for one-to-one is “injective” and an alternative for onto is “surjective.”

Put another way, T is one-to-one if every vector in the domain of T is sent to its “own” unique vector in the range (see Figure 1(a) and 1(c)). T is onto if the range is equal to the codomain (see Figure 1(a) and 1(b)).

An equivalent formulation (see Exercise 59) for the definition of one-to-one is given below.

DEFINITION ▶

(ALTERNATE) A linear transformation T is one-to-one if $T(\mathbf{u}) = T(\mathbf{v})$ implies that $\mathbf{u} = \mathbf{v}$.

This version of one-to-one often is more convenient for proofs. It is used in the proof of the next theorem, which provides an easy way to determine if a linear transformation is one-to-one.

THEOREM 3.5 ▶

Let T be a linear transformation. Then T is one-to-one if and only if the only solution to $T(\mathbf{x}) = \mathbf{0}$ is the trivial solution $\mathbf{x} = \mathbf{0}$.

Proof First suppose that T is one-to-one. Then there is at most one solution to $T(\mathbf{x}) = \mathbf{0}$. Moreover, since T is a linear transformation, it follows that $T(\mathbf{0}) = \mathbf{0}$ (see Exercise 55). Therefore the only solution to $T(\mathbf{x}) = \mathbf{0}$ is the trivial solution $\mathbf{x} = \mathbf{0}$.

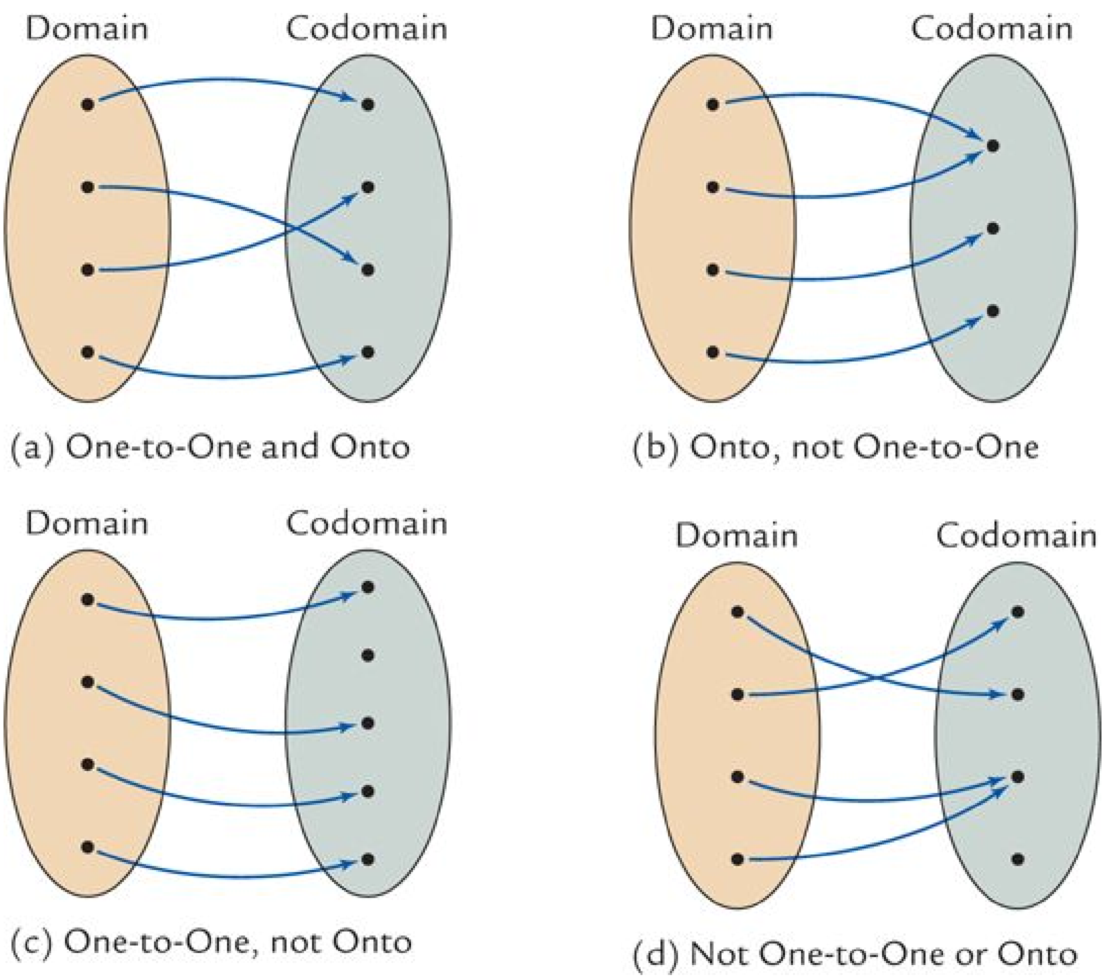


Figure 1 Graphical depiction of various combinations of one-to-one and onto.

Now suppose that the only solution to $T(\mathbf{x}) = \mathbf{0}$ is the trivial solution. If $T(\mathbf{u}) = T(\mathbf{v})$, then

$$T(u)-T(v)=0 \Rightarrow T(u-v)=0 (T \text{ is a linear transformation})$$

Since $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution, it follows that $\mathbf{u} - \mathbf{v} = \mathbf{0}$ and thus $\mathbf{u} = \mathbf{v}$. Therefore T is one-to-one.

Example 5

Let A be as given in Example 3. Determine if $T(\mathbf{x}) = A\mathbf{x}$ one-to-one.

Solution By Theorem 3.5, we need only find the solutions to $T(\mathbf{x}) = \mathbf{0}$, which is equivalent to solving $A\mathbf{x} = \mathbf{0}$. Populating the augmented matrix and reducing to echelon form gives

[4-1-2203|000]~[-220300|000]

From the echelon form we can see that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Thus $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution and thus T is one-to-one.

The next two theorems follow from results on span and linear independence developed in Section 2.2 and Section 2.3.

THEOREM 3.6

Let A be an $n \times m$ matrix and define $T : \mathbf{R}^m \to \mathbf{R}^n$ by $T(\mathbf{x}) = A\mathbf{x}$. Then

(a) T is one-to-one if and only if the columns of A are linearly independent.

(b) If $A \sim B$ and B is in echelon form, then T is one-to-one if and only if B has a pivot position in every column.

(c) If $n < m$, then T is *not* one-to-one.

This proof is completed by applying earlier results, and illustrates the interconnections in linear algebra.

Proof To prove part (a), note that by Theorem 3.5 T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution. By Theorem 2.17, $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution if and only if the columns of A are linearly independent.

Part (b) follows from Theorem 2.16 and part (a).

For part (c), if A has more columns than rows, then by Theorem 2.14 the columns are linearly dependent. Thus by part (a), T is not one-to-one. ■ ■

Returning to Example 5, we see that the two columns of A are linearly independent. (Why?) Therefore we can also conclude from Theorem 3.6 that the linear transformation T is one-to-one.

The next theorem is the counterpart to Theorem 3.6 that shows when a linear transformation is onto.

THEOREM 3.7 ▶

Let A be an $n \times m$ matrix and define $T : \mathbf{R}^m \to \mathbf{R}^n$ by $T(\mathbf{x}) = A\mathbf{x}$. Then

(a) T is onto if and only if the columns of A span the codomain R^n.

(b) If $A \sim B$ and B is in echelon form, then T is onto if and only if B has a pivot position in every row.

(c) If $n > m$, then T is *not* onto.

Proof For part (a), if T is onto then range(T) = $\mathbf{R}^n$. By Theorem 3.3, range(T) equals the span of the columns of A. Hence T is onto if and only if the columns of A span $\mathbf{R}^n$.

Part (b) follows from Theorem 2.16 and part (a).

For part (c), by Theorem 2.9 if $n > m$ then the columns of A cannot span $\mathbf{R}^n$. Thus, by part (a), T is not onto. ■ ■

Example 6

Suppose that A is the matrix given in Example 3. Determine if $T(\mathbf{x}) = A\mathbf{x}$ is onto.

Solution Since A is a 3×2 matrix, by Theorem 3.7(b) it follows that T is not onto.

Example 7

Suppose that

$$A=[211120130]$$

Determine if the linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is onto.

Solution In Section 2.2 we showed that the set of vectors

$$\{[211], [123], [100]\}$$

spans $\mathbf{R}^3$. Thus, by Theorem 3.7(a), T is onto.

In Theorem 3.2, we showed that a function $T : \mathbf{R}^m \to \mathbf{R}^n$ of the form $T(\mathbf{x}) = A\mathbf{x}$ must be a linear transformation. The next theorem combines Theorem 3.2 with its converse.

THEOREM 3.8 ▶

Let $T : \mathbf{R}^m \to \mathbf{R}^n$. Then T is a linear transformation if and only if $T(\mathbf{x}) = A\mathbf{x}$ for some $n \times m$ matrix A.

Proof On direction of this theorem is proved in Theorem 3.2. For the other direction, suppose that *T* is a linear transformation. Let

e1=[100⋮0], e2=[010⋮0], e3=[001⋮0],… ,em=[000⋮1] (3)

be vectors in $\mathbf{R}^m$, and then let *A* be the $n \times m$ matrix with columns $T(\mathbf{e}_1)$, $T(\mathbf{e}_2)$, … , $T(\mathbf{e}_m)$,

A=[T(e1)T(e2)⋯T(em)]

Note that any vector $\mathbf{x}$ in $\mathbf{R}^m$ can be written as a linear combination of $\mathbf{e}_1$, $\mathbf{e}_2$, … , $\mathbf{e}_m$, by

x=[x1x2x3⋮xm]=x1[100⋮0]+⋯+xm[000⋮1]=x1e1+x2e2+⋯+xmem

From the properties of linear transformations, we have

T(x)=T(x1e1+x2e2+⋯+xmem) =x1T(e1)+x2T(e2)+⋯+xmT(em) =Ax

Thus *T* has the required form and the proof is complete. ■ ■

Example 8

Suppose that $T : \mathbf{R}^3 \to \mathbf{R}^4$ is defined by

T([x1x2x3])=[2x1+x3 -x1+2x2x1-3x2+5x3 4x2]

Show that *T* is a linear transformation.

Solution We could solve this by directly appealing to Definition 3.1. But instead, let's apply Theorem 3.8 by finding the matrix *A* such that $T(\mathbf{x}) = A\mathbf{x}$. We start by noting that

T([x1x2x3])=[2x1+x3 -x1+2x2x1-3x2+5x3 4x2]=[2x2+0x2+1x3-1x1+2x2+0x31x1-3x2+5x30x1+4x2+0x3]=[201-1201-35040] [x1x2x3]

Thus, if

$$A=[201-1201-35040]$$

then $T(\mathbf{x}) = A\mathbf{x}$. Hence T is a linear transformation by Theorem 3.8.

The Unifying Theorem, Version 2

Incorporating the preceding work, let us add two new conditions to the Unifying Theorem, Version 1 (Theorem 2.21) that we proved in Section 2.3.

THEOREM 3.9

(THE UNIFYING THEOREM, VERSION 2) Let $\mathcal{S} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ be a set of n vectors in $\mathbf{R}^n$, let $A=[a1\cdots an]$, and let $T: \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent:

(a) $\mathcal{S}$ spans $\mathbf{R}^n$.

(b) $\mathcal{S}$ is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

(d) T is onto.

(e) T is one-to-one.

Proof From the Unifying Theorem, Version 1, we know that (a), (b), and (c) are equivalent. It follows from Theorem 3.7 that (a) and (d) are equivalent, and it follows from Theorem 3.6 that (b) and (e) are equivalent. Thus all five conditions are equivalent to each other.

Geometry of Linear Transformations

One reason for the name of linear transformations is because they frequently transform lines in the domain to lines in the range. To see why, recall that (see Exercise 64) the line segment from **u** to **v** can be parameterized by

$$(1+-s)u+sv,0≤s≤1$$

Applying a linear transformation T to this, we find that

$$T[(1-s)u+sv]=(1-s)T(u)+sT(v),0≤s≤1$$

which is the parameterization of a line in the range of T. Figure 2 illustrates this for two different linear transformations, T_1: $\mathbf{R}^2 \to \mathbf{R}^2$ and T_2: $\mathbf{R}^2 \to \mathbf{R}^2$.

What effect do linear transformations have on regions? Figure 3 shows the result of applying three different linear transformations to the unit square $\mathcal{S}$ in the first quadrant. The boundaries of $\mathcal{S}$ are mapped to the boundaries of the image. When the linear transformation is one-to-one, the unit square $\mathcal{S}$ is mapped to a parallelogram.

Linear transformations are used extensively in computer graphics. Figure 4 shows the results of linear transformations that reflect, rotate, and shear the unit square $\mathcal{S}$ in Figure 3.

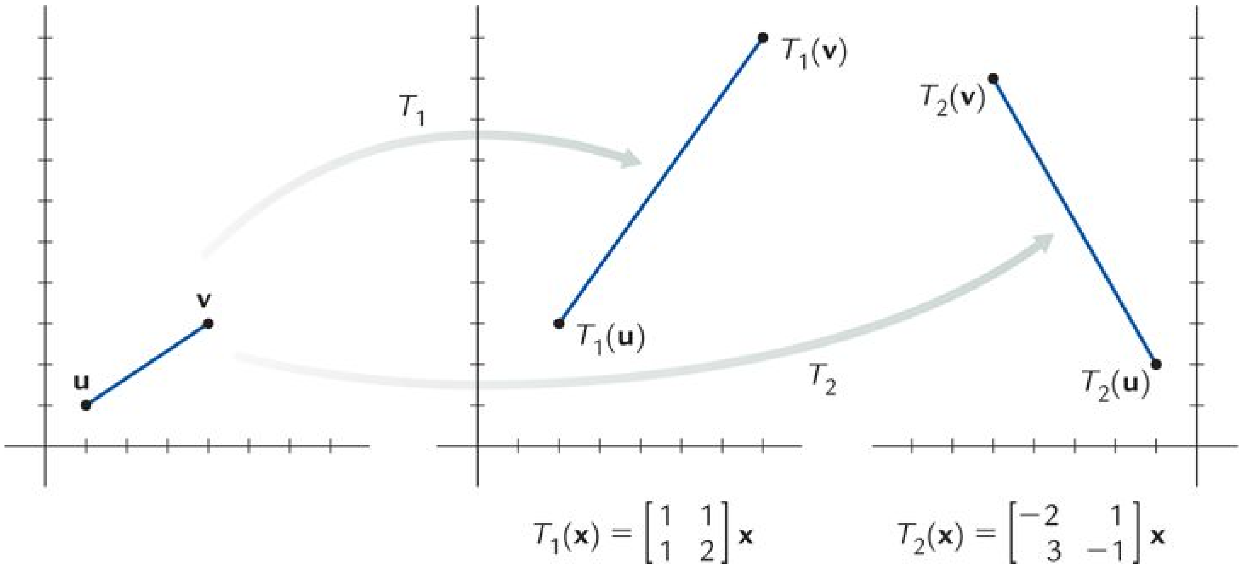


Figure 2 Image of the line segment between **u** and **v** under linear transformations $T_1(\mathbf{x})$ and $T_2(\mathbf{x})$.

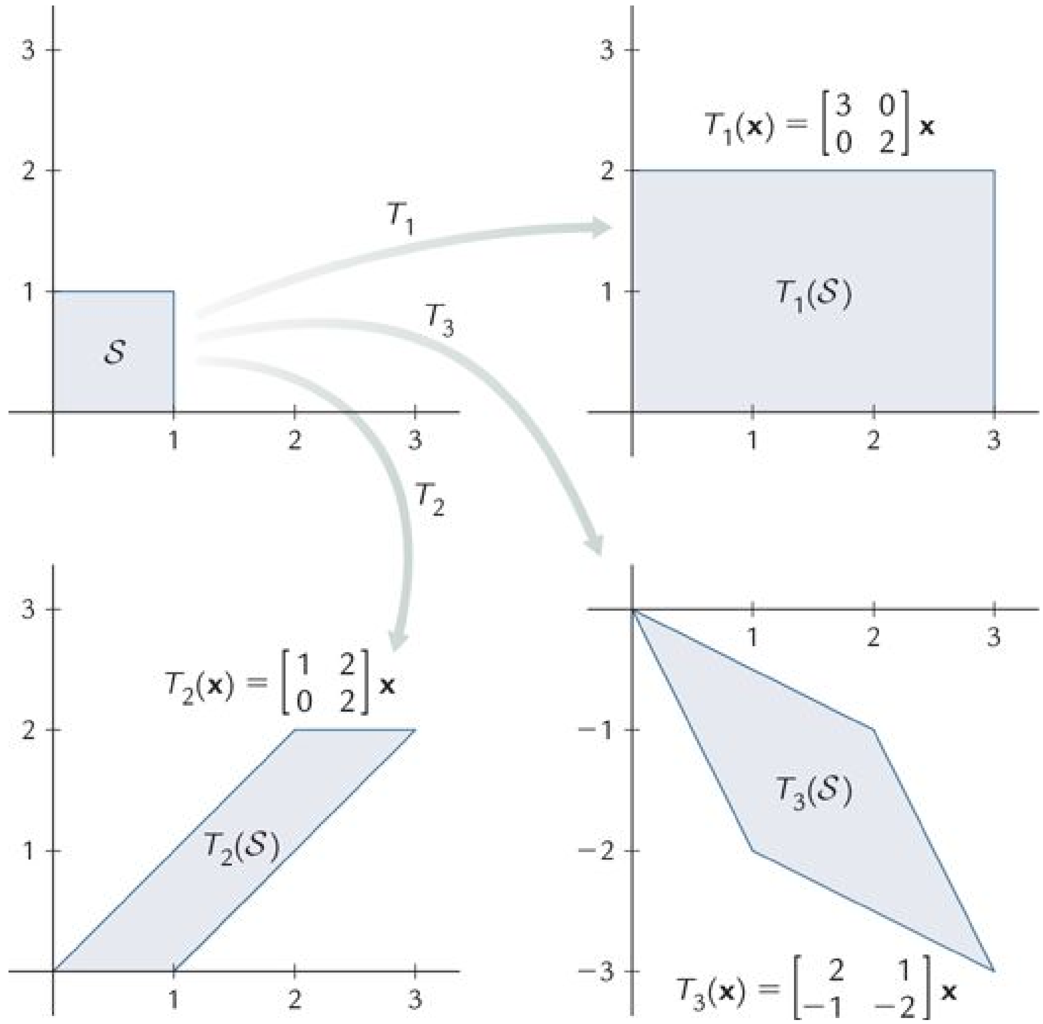


Figure 3 The image of the unit square $\mathcal{S}$ (*upper left*) under three different linear transformations.

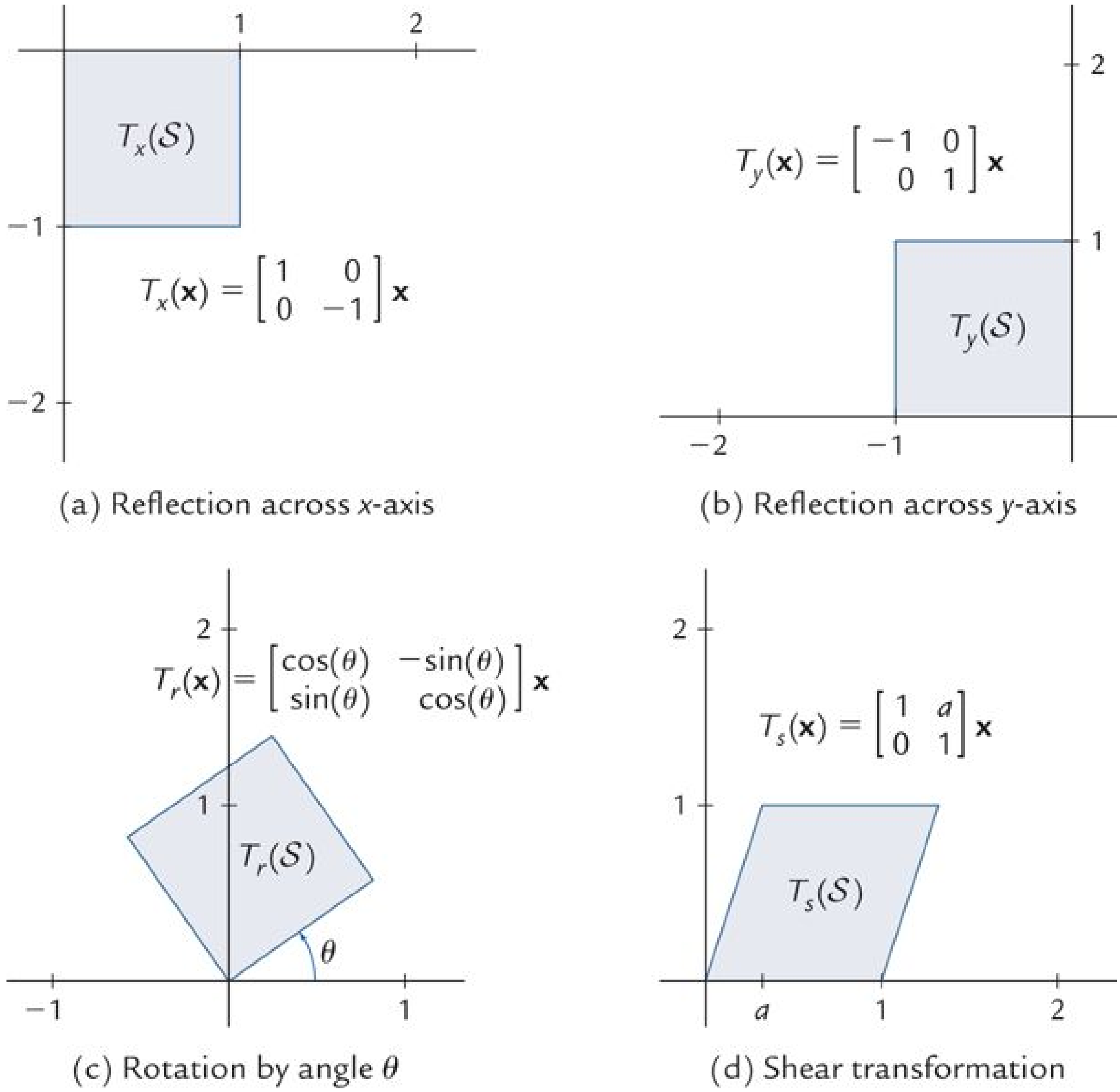


Figure 4 The image of the unit square $\mathcal{S}$ under reflection, rotation, and shear linear transformations used in computer graphics.

Figure 5 Karl Jacobi.
Courtesy of the Smithsonian Libraries, Washington, D.C.

Figure 6 The linear transformations from Figure 4 applied to an image of Karl Jacobi.
Courtesy of the Smithsonian Libraries, Washington, D.C.

In Section 1.4 we encountered Karl Jacobi, whose picture is shown in Figure 5. Reflection, rotation, and shear transformations applied to that picture are shown in Figure 6.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Suppose that T([x1x2])=[3x1+2x2-x1+x2-4x1-3x2].
   (a) Find T([2-1]).
   (b) Find A such that $T(\mathbf{x}) = A\mathbf{x}$.
   (c) Determine if T is one-to-one and if T is onto.
2. Suppose T(u1)=[23] and T(u2)=[-41]. Find the following:
   (a) $T(\mathbf{u}_1 + \mathbf{u}_2)$
   (b) $T(3\mathbf{u}_1)$
   (c) $T(2\mathbf{u}_1 - \mathbf{u}_2)$
3. Show that T([x1x2])=[x1+2x2x1-2] is not a linear transformation.
4. Suppose $T(\mathbf{x}) = A\mathbf{x}$ where A=[2-123]. Sketch the graph of the image under T of the unit square in the first quadrant in $\mathbf{R}^2$.
5. Determine if each statement is true or false, and justify your answer.
   (a) T([x1x2])=[x1-2x23x1+x2] is a linear transformation.
   (b) If $T(\mathbf{x}) = A\mathbf{x}$ where A is a 3×4 matrix, then T cannot be onto.
   (c) Linear transformations are functions.
   (d) T is onto if and only if $T(\mathbf{x}) = \mathbf{0}$ only when $\mathbf{x} = \mathbf{0}$.

EXERCISES

Exercises 1–4: Let $T(\mathbf{x}) = A\mathbf{x}$ for the given matrix A, and find $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$ for the given $\mathbf{u}_1$ and $\mathbf{u}_2$.

1. A=[21-35],u1=[-4-2],u2=[1-6]
2. A=[102-433],u1=[12],u2=[-50]
3. A=[0-4231-2],u1=[321],u2=[4-5-2]
4. A=[-25-20-1-20-1-1],u1=[07-2],u2=[35-1]

Exercises 5–8: Determine if the given vector is in the range of $T(\mathbf{x}) = A\mathbf{x}$, where

A=[1-20321]

5. y=[-36]

6. y=[1-4]

7. y=[27]

8. y=[45]

9. Suppose that a linear transformation T satisfies

T(u1)=[21],T(u2)=[-32]

Find $T(-2\mathbf{u}_1 + 3\mathbf{u}_2)$.

10. Suppose that a linear transformation T satisfies

T(u1)=[3-1-2],T(u2)=[114]

Find $T(3\mathbf{u}_1 - 2\mathbf{u}_2)$.

11. Suppose that a linear transformation T satisfies

T(u1)=[-30],T(u2)=[2-1],T(u3)=[05]

Find $T(-\mathbf{u}_1 + 4\mathbf{u}_2 - 3\mathbf{u}_3)$.

12. Suppose that a linear transformation T satisfies

T(u1)=[3-1-2],T(u2)=[114],T(u3)=[600]

Find $T(\mathbf{u}_1 + 4\mathbf{u}_2 - 2\mathbf{u}_3)$.

Exercises 13–20: Determine if T is a linear transformation. If so, identify the matrix A such that $T(\mathbf{x}) = A\mathbf{x}$. If not, explain why not.

13. $T(x_1, x_2) = (3x_1 + x_2, -2x_1 + 4x_2)$

14. $T(x_1, x_2) = (x_1 - x_2, x_1x_2)$

15. $T(x_1, x_2, x_3) = (2 \cos(x_2), 3 \sin (x_3), x_1)$

16. $T(x_1, x_2, x_3) = (-5x_2, 7x_3)$

17. $T(x_1, x_2, x_3) = (-4x_1 + x_3, 6x_1 + 5x_2)$

18. $T(x_1, x_2, x_3) = (-x_1 + 3x_2 + x_3, 2x_1 + 7x_2 + 4, 3x_3)$

19. $T(x_1, x_2) = (x_2 \sin (\pi/4), x_1 \ln(2))$

20. $T(x_1, x_2) = (3x_2, -x_1 + 5|x_2|, 2x_1)$

Exercises 21–28: Let $T(\mathbf{x}) = A\mathbf{x}$ for the matrix A. Determine if T is one-to-one and if T is onto.

21. A=[1-3-25]
22. A=[3296]
23. A=[54-23-10]
24. A=[-1324-12-8]
25. A=[1-2-352-7]
26. A=[2-45-10-48]
27. A=[2843231145]
28. A=[12-537-8-2-46]

Exercises 29–32: Suppose that $T(\mathbf{x}) = A\mathbf{x}$ for the matrix A. Sketch a graph of the image under T of the unit square in the first quadrant of $\mathbf{R}^2$.

29. A=[3003]
30. A=[-2004]
31. A=[1-231]
32. A=[-316-2]

FIND AN EXAMPLE Exercises 33–38: Find an example that meets the given specifications.

33. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ such that

T([10])=[23]

34. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^3$ such that

T([01])=[145]

35. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ such that

T([31])=[70]

36. A linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ such that

T([31-2])=[7-1]

37. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ such that

T([21])=[07]andT([13])=[-56]

38. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ such that

T([-12])=[-18]andT([2-3])=[2-13]

TRUE OR FALSE Exercises 39–48: Determine if the statement is true or false, and justify your answer.

39.

(a) The codomain of a linear transformation is a subset of the range.

(b) All linear transformations $T : \mathbf{R}^n \to \mathbf{R}^m$ have the form $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A.

40.

(a) The range of a linear transformation must be a subset of the domain.

(b) If T is onto, then range(T) is equal to the codomain.

41.

(a) If T is a linear transformation and $\mathbf{v}$ is in range(T), then there is at least one $\mathbf{u}$ in the domain such that $T(\mathbf{u}) = \mathbf{v}$.

(b) If S and T are onto linear transformations and $W(\mathbf{x}) = S(T(\mathbf{x}))$ is defined, then W is onto.

42.

(a) If $T(\mathbf{x})$ is not a linear transformation, then $T(r\mathbf{x}) \neq rT(\mathbf{x})$ for all r and $\mathbf{x}$.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ is one-to-one, then the columns of A span the codomain of T.

43.

(a) The function $T(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ is a linear transformation only when $\mathbf{b} = \mathbf{0}$.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ is onto, then the columns of A span the codomain of T.

44.

(a) If $T : \mathbf{R}^2 \to \mathbf{R}^2$ is a linear transformation, then the image under T of the unit square in the first quadrant will be a parallelogram.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ is one-to-one, then the columns of A are linearly independent.

45.

(a) If $T_1(\mathbf{x})$ and $T_2(\mathbf{x})$ are one-to-one linear transformations from $\mathbf{R}^n$ to $\mathbf{R}^m$, then so is $W(\mathbf{x}) = T_1(\mathbf{x}) + T_2(\mathbf{x})$.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ is one-to-one and onto, then A must be a square matrix.

46.

(a) If $T_1(\mathbf{x})$ and $T_2(\mathbf{x})$ are onto linear transformations from $\mathbf{R}^n$ to $\mathbf{R}^m$, then so is $W(\mathbf{x}) = T_1(\mathbf{x}) + T_2(\mathbf{x})$.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ is not one-to-one, then the system $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$.

47.

(a) If a linear transformation $T : \mathbf{R}^4 \to \mathbf{R}^4$ is one-to-one, then $T(\mathbf{x}) = \mathbf{0}$ has nontrivial solutions.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ and an echelon form of A has pivot positions in every row, then T is onto.

48.

(a) If a linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ is one-to-one, then T also must be onto.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ and an echelon form of A has pivot positions in every column, then T is onto.

49. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ is called a *dilation* if $T(\mathbf{x}) = r\mathbf{x}$ for $r > 1$. (It is called a *contraction* if $0 < r < 1$.)

(a) Find the matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

(b) Let $r = 2$, and then sketch the graphs of u=[2-1] and $T(\mathbf{u})$.

50. Suppose that $T : \mathbf{R}^3 \to \mathbf{R}^2$ is given by

T([x1x2x3])=[x1x2]

The T is called a *projection transformation* because it projects vectors in $\mathbf{R}^3$ onto $\mathbf{R}^2$.

(a) Prove that T is a linear transformation.

(b) Find the matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

(c) Describe the set of vectors in $\mathbf{R}^3$ such that $T(\mathbf{x}) = \mathbf{0}$.

51. Suppose that $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ are vectors in $\mathbf{R}^n$. Then the *dot product* of $\mathbf{x}$ and $\mathbf{y}$ is given by

x·y=x1y1+⋯+xnyn

Now let **u** be a fixed vector in $\mathbf{R}^n$, and define $T(\mathbf{x}) = \mathbf{u} \cdot \mathbf{x}$. Show that T is a linear transformation.

52. Suppose that $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ are vectors in $\mathbf{R}^3$. Then the *cross product* of **x** and **y** is given by

x×y=[x2y3-x3y2x3y1-x1y3x1y2-x2y1]

Now let **u** be a fixed vector in $\mathbf{R}^n$, and define $T(\mathbf{x}) = \mathbf{u} \times \mathbf{x}$. Show that T is a linear transformation.

53. Suppose that $T : \mathbf{R}^3 \to \mathbf{R}^2$ is a linear transformation. Prove that T is not one-to-one.

54. Suppose that $T : \mathbf{R}^2 \to \mathbf{R}^4$ is a linear transformation. Prove that T is not onto.

55. Suppose that T is a linear transformation. Show that $T(\mathbf{0}) = \mathbf{0}$.

56. Suppose that $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation and that there exists $\mathbf{u} \neq \mathbf{0}$ such that $T(\mathbf{u}) = \mathbf{0}$. Show that the columns of A must be linearly dependent.

57. Suppose that $T(\mathbf{x}) = A\mathbf{x}$. Show that

T(ru)=rT(u)

for all scalars r and all vectors **u**.

58. Suppose that $T : \mathbf{R}^m \to \mathbf{R}^n$.

(a) Show that if T is a linear transformation, then

T(rx+sy)=rT(x)+sT(y)

for all scalars r and s and all vectors **x** and **y**.

(b) Now show the converse: If

T(rx+sy)=rT(x)+sT(y)

for all scalars r and s and all vectors **x** and **y**, then T is a linear transformation.

59. Prove that a linear transformation T is one-to-one if and only if $T(\mathbf{u}) = T(\mathbf{v})$ implies that $\mathbf{u} = \mathbf{v}$.

60. Suppose that T is a linear transformation and that $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly dependent. Prove that $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$ are also linearly dependent.

61. Suppose that T is a linear transformation and that $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$ are linearly independent. Prove that $\mathbf{u}_1$ and $\mathbf{u}_2$ must be linearly independent.

62. Suppose that T is a linear transformation and that $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent. Show that $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$ need not be linearly independent.

63. Suppose that $\mathbf{y}$ is in the range of a linear transformation T and that there exists a vector $\mathbf{u} \neq \mathbf{0}$ such that $T(\mathbf{u}) = \mathbf{0}$. Show that there are infinitely many solutions $\mathbf{x}$ to $T(\mathbf{x}) = \mathbf{y}$. (HINT: First, explain why $T(r\mathbf{u}) = \mathbf{0}$ for any scalar r, and then show that $T(\mathbf{x} + r\mathbf{u}) = \mathbf{y}$ when $T(\mathbf{x}) = \mathbf{y}$.)

64. Let $\mathbf{u}$ and $\mathbf{v}$ be two distinct vectors in $\mathbf{R}^2$. Show that the set of points on the line segment connecting $\mathbf{u}$ and $\mathbf{v}$ is the same as the set of points

$$(1-s)\mathbf{u}+s\mathbf{v}, \quad 0 \leq s \leq 1$$

65. Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be a linear transformation with $T(\mathbf{x}) = A\mathbf{x}$. Prove that the image of the unit square in the first quadrant is a parallelogram if the columns of A are linearly independent, and a line segment (possibly of zero length) if the columns of A are linearly dependent.

66. In graph theory, an *adjacency matrix* A has an entry of 1 at a_{ij} if there is an edge connecting node i with node j, and a zero otherwise. (Such matrices come up in network analysis.) Suppose that a graph with five nodes has adjacency matrix

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

Let $T : \mathbf{R}^5 \to \mathbf{R}^5$ be given by $T(\mathbf{x}) = A\mathbf{x}$.

(a) Describe how to use $T(\mathbf{x})$ to determine the number of edges connected to node j.

(b) How can one use $T(\mathbf{x})$ to help determine the total number of graph edges?

67. **Calculus Required** Suppose that for each polynomial of degree 2 or less, we identify the coefficients with a vector in $\mathbf{R}^3$ by

$$ax^2+bx+c \leftrightarrow \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

(a) Show that addition of polynomials corresponds to vector addition and that multiplication of a polynomial by a constant corresponds to scalar multiplication of a vector.

(b) Let $T : \mathbf{R}^3 \to \mathbf{R}^3$ be the function that takes a polynomial vector as input and produces the vector of the derivative as output. Prove that T is a linear transformation.

(c) Find the matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

(d) Is T one-to-one? Onto? Give a proof or counter-example for each.

68. **Calculus Required** Complete Exercise 67 for polynomials of degree 3 or less identified with vectors in $\mathbf{R}^4$.

Calculus Required Exercises 69–70: In Chapter 9 we extend the concept of a linear transformation by observing that the two conditions given in Definition 3.1 exist for other types of mathematical operations. Here we provide a sneak preview. Assume that $f(x)$ and $g(x)$ are in the set of functions $C^\infty(\mathbf{R})$ that have infinitely many derivatives on $\mathbf{R}$ and that r is a real number.

69. Let $T : C^\infty(\mathbf{R}) \to C^\infty(\mathbf{R})$ be defined by $T(f(x)) = f'(x)$.

(a) Evaluate $T(x^2 + \sin(x))$.

(b) Prove that T satisfies conditions analogous to those given in Definition 3.1:

(i) $T(f(x) + g(x)) = T(f(x)) + T(g(x))$

(ii) $T(rf(x)) = rT(f(x))$

70. Let $T : C^\infty(\mathbf{R}) \to \mathbf{R}$ be defined by

$$T(f(x))=\int_0^1 f(x)\,dx$$

(a) Evaluate $T(4x^3 - 6x^2 + 1)$.

(b) Prove that T satisfies conditions analogous to those given in Definition 3.1:

(i) $T(f(x) + g(x)) = T(f(x)) + T(g(x))$

(ii) $T(rf(x)) = rT(f(x))$

Exercises 71–74: Refer to the smart phone scenario given at the beginning of the section. Use the linear transformation T to determine the cost vector that results from producing the specified number of j8’s and j8+’s.

71. 5 j8’s, 6 j8+’s.
72. 6 j8’s, 10 j8+’s.
73. 8 j8’s, 16 j8+’s.
74. 12 j8’s, 20 j8+’s.

Exercises 75–80: Let $T(\mathbf{x}) = A\mathbf{x}$ for the given matrix A. Determine if T is one-to-one and if T is onto.

75. A=[42-5267-20-4103-57-1]
76. A=[4-25215144-58-16-2-32]
77. A=[2-1403-3111-1830-214]
78. A=[3205012-3-2-1314-23-1]
79. A=[2-351603-2-4211823-4-125-3]
80. A=[43-29-101-130-242-4335-703]

3.2 Matrix Algebra

In this section, we develop the algebra of matrices. This algebraic structure has many things in common with the algebra of the real numbers, but there are also some important differences.

To get us started, consider a hypothetical natural foods store that sells organic chicken eggs that come from two suppliers, *The Happy Coop* and *Eggspeditious*. Each provides both white and brown eggs by the dozen in medium, large, and extra large sizes. The store's current inventory of 12-egg cartons is given in Table 1.

Table 1 **Egg Inventories at a Natural Foods Store**

	The Happy Coop			**Eggspeditious**	
	White	**Brown**		**White**	**Brown**
Medium	5	3	Medium	8	5
Large	11	6	Large	3	6
XLarge	4	6	XLarge	8	10

This information can be transferred into a pair of matrices

$$H=\begin{bmatrix} 5 & 3 \\ 11 & 6 \\ 4 & 6 \end{bmatrix} \text{ and } E=\begin{bmatrix} 8 & 5 \\ 3 & 6 \\ 8 & 10 \end{bmatrix}$$

If we want to know the total number of egg cartons of each type in stock, we just add the corresponding terms in each matrix, giving

$$\begin{bmatrix} (5+8) & (3+5) \\ (11+3) & (6+6) \\ (4+8) & (6+10) \end{bmatrix}=\begin{bmatrix} 13 & 8 \\ 14 & 12 \\ 12 & 16 \end{bmatrix}$$

Thus, for instance, there are 13 cartons of medium white eggs in stock. If instead we want to find the total number of each type of Eggspeditious eggs, we multiply each term in E by 12:

$$\begin{bmatrix} (12\cdot 8) & (12\cdot 5) \\ (12\cdot 3) & (12\cdot 6) \\ (12\cdot 8) & (12\cdot 10) \end{bmatrix}=\begin{bmatrix} 96 & 60 \\ 36 & 72 \\ 96 & 120 \end{bmatrix}$$

Equal Matrices

These computations illustrate addition and scalar multiplication of matrices, and are analogous to those for vectors. As these operations suggest, two matrices are **equal** if they have the same dimensions and if their corresponding entries are equal.

This example serves as a model for a formal definition of addition and scalar multiplication.

DEFINITION 3.10 ▶

Let c be a scalar, and let

A=[a11a12…a1ma21a22…a2m⁝⁝ ⁝an1an2…anm]andB=[b11b12…b1mb21b22…b2m⁝⁝ ⁝bn1bn2…bnm]

be $n \times m$ matrices. Then addition and scalar multiplication of matrices are defined as follows:

Addition, Scalar Multiplication of Matrices

(a) **Addition**: A+B=[(a11+b11)(a12+b12)…(a1m+b1m)(a21+b21)(a22+b22)…(a2m+b2m)⁝⁝ ⁝(an1+bn1)(an2+bn2)…(anm+anm)]

(b) **Scalar Multiplication**: cA=[ca11ca12…ca1mca21ca22…ca2m⁝⁝ ⁝can1can2…canm]

Example 1

Let

A=[4-12-370]andB=[3-15002]

Find $3A$ and $A - 2B$.

Solution We have

3A=3[4-12-370]=[3(4)3(-1)3(2)3(-3)3(7)3(0)]=[12-36-9210]

and

A-2B=[4-12-370] -2[3-15002]=[4-12-370]-[6-210004]=[-21-8-37-4]

Suppose that r, s, and t are real numbers. Recall the following laws for arithmetic:

(a) $r + s = s + r$ (Commutative)
(b) $(r + s) + t = r + (s + t)$ (Associative)
(c) $r(s + t) = rs + rt$ (Distributive)
(d) $r + 0 = r$ (Additive Identity)

Similar laws hold for addition and scalar multiplication of matrices.

THEOREM 3.11

Let s and t be scalars, A, B, and C be matrices of dimension $n \times m$, and 0_{nm} be the $n \times m$ matrix with all zero entries. Then

(a) $A + B = B + A$
(b) $s(A + B) = sA + sB$
(c) $(s + t)A = sA + tA$
(d) $(A + B) + C = A + (B + C)$
(e) $(st) A = s(tA)$
(f) $A + 0_{nm} = A$

Proof The proof of each part follows from the analogous laws for the real numbers. We prove part (e) here and leave the rest as exercises.

Let

$$A=[a11a12\ldots a1ma21a22\ldots a2m\vdots\vdots\ \vdots an1an2\ldots anm]$$

Then we have

$$(st)A=(st)[a11a12\ldots a1ma21a22\ldots a2m\vdots\vdots\ \vdots an1an2\ldots anm]= [sta11sta12\ldots sta1msta21sta22\ldots sta2m\vdots\vdots\ \vdots stan1stan2\ldots stanm]$$

On the other hand,

$$s(tA)=s\ [ta11ta12\ldots ta1mta21ta22\ldots ta2m\vdots\vdots\ \vdots tan1tan2\ldots tanm]= [sta11sta12\ldots sta1msta21sta22\ldots sta2m\vdots\vdots\ \vdots stan1stan2\ldots stanm]$$

The two are the same, so $(st)A = s(tA)$.

Matrix Multiplication

Matrix multiplication is defined by extending the definition for multiplying a matrix times a vector,

$$Ax=[a1a2\cdots ak]\ [x1x2\vdots xk]=x1a1+x2a2+\cdots+xkxk$$

The product $A\mathbf{x}$ is defined as long as the number of columns of A is equal to the number of components of $\mathbf{x}$. Now let

$$B=[b1b2\cdots bm]$$

where each $\mathbf{b}_1, \ldots , \mathbf{b}_m$ has k components. Then

$$Ab1,Ab2,\ldots ,Abk \qquad (1)$$

are all defined, with each having the same number of components as A has rows. The vectors in (1) will be the columns of the product AB.

DEFINITION 3.12 ▶

Matrix Multiplication

Let A be an $n \times k$ matrix and $B=[\mathbf{b}_1\ \mathbf{b}_2 \cdots \mathbf{b}_m]$ a $k \times m$ matrix. We define the product

$$AB=[A\mathbf{b}_1\ A\mathbf{b}_2 \cdots A\mathbf{b}_m]$$

which is an $n \times m$ matrix.

▶ For AB to exist, the number of columns of A must equal the number of rows of B.

Example 2

Let

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

Find (if they exist) AB and BA.

Solution Since A is a 2×2 matrix and B is a 2×3 matrix, it follows that AB exists and is a 2×3 matrix, with columns

$$A\mathbf{b}_1=-1\begin{bmatrix} 3 \\ -2 \end{bmatrix}+4\begin{bmatrix} 1 \\ 0 \end{bmatrix}=\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad A\mathbf{b}_2=0\begin{bmatrix} 3 \\ -2 \end{bmatrix}-3\begin{bmatrix} 1 \\ 0 \end{bmatrix}=\begin{bmatrix} -3 \\ 0 \end{bmatrix} \quad A\mathbf{b}_3=2\begin{bmatrix} 3 \\ -2 \end{bmatrix}-1\begin{bmatrix} 1 \\ 0 \end{bmatrix}=\begin{bmatrix} 5 \\ -4 \end{bmatrix}$$

Thus

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

Turning to BA, since B is 2×3 and A is 2×2, B has three columns but A has two rows. These do not match, so the product BA is not defined.

The above example shows that even when AB is defined, BA might not be. Example 3 shows that even if both AB and BA are defined, they might not be equal.

Example 3

Let

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

Find (if they exist) AB and BA.

Solution Since A and B both are 2×2 matrices, AB and BA are defined. Starting with AB, we have

$$A\mathbf{b}_1=4\begin{bmatrix} 2 \\ 1 \end{bmatrix}-1\begin{bmatrix} -1 \\ 3 \end{bmatrix}=\begin{bmatrix} 9 \\ 1 \end{bmatrix} \quad A\mathbf{b}_2=-2\begin{bmatrix} 2 \\ 1 \end{bmatrix}+1\begin{bmatrix} -1 \\ 3 \end{bmatrix}=\begin{bmatrix} -5 \\ 1 \end{bmatrix}$$

so that

$$AB=\begin{bmatrix} 9 & -5 \\ 1 & 1 \end{bmatrix}$$

Turning to BA, we compute

$$B\mathbf{a}_1=2\begin{bmatrix} 4 \\ -1 \end{bmatrix}+1\begin{bmatrix} -2 \\ 1 \end{bmatrix}=\begin{bmatrix} 6 \\ -1 \end{bmatrix} \quad B\mathbf{a}_2=-1\begin{bmatrix} 4 \\ -1 \end{bmatrix}+3\begin{bmatrix} -2 \\ 1 \end{bmatrix}=\begin{bmatrix} -10 \\ 4 \end{bmatrix}$$

▶ Example 3 shows that matrix multiplication is not commutative.

Therefore

$$BA=\begin{bmatrix} 6 & -10 \\ -1 & 4 \end{bmatrix}$$

Thus we see that $AB \neq BA$.

In some applications, we need only a single entry of the matrix product AB. Assume that A is an $n \times k$ matrix and B is a $k \times m$ matrix. Then $C = AB$ is defined, with jth column

$$c_j = A\mathbf{b}_j = b_{1j}\mathbf{a}_1 + b_{2j}\mathbf{a}_2 + \cdots + b_{kj}\mathbf{a}_k$$

for $j = 1, \ldots, m$. Selecting the ith gives component from each of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ the formula for individual entries of C,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} \qquad (2)$$

Note that we are multiplying each term from row i of A times the corresponding term in column j of B and then adding the resulting products. (This is called the *dot product*, which we will study later.) Figure 1 shows a graphical depiction of computing an entry.

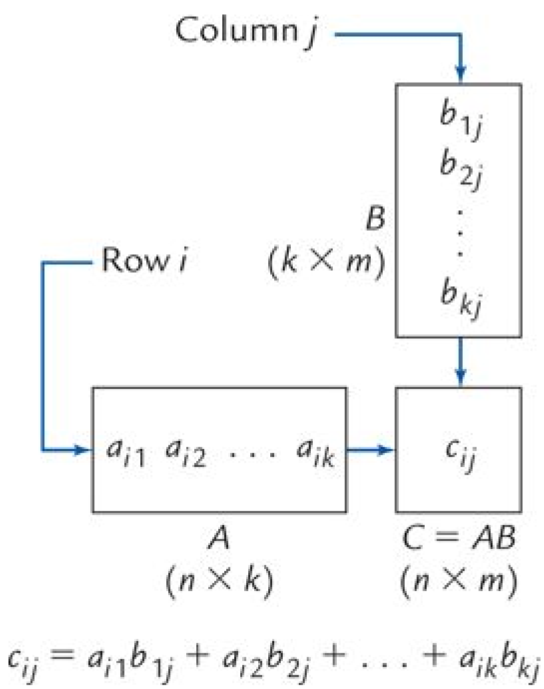


Figure 1 Computing an entry in matrix multiplication.

Example 4

Use (2) to compute the entries of $C = AB$, where

A=[2-30-114]and[23-424-25-3]

Solution Let’s start with c_{11}. To compute this, we need the entries in row 1 of A and column 1 of B. Setting $i = 1$ and $j = 1$, our formula 2 gives us

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} = (2)(2) + (-3)(4) = -8$$

All entries of C are found in a similar manner, with

$c_{11} = (2)(2) + (-3)(4) = -8$	$c_{13} = (2)(-4) + (-3)(5) = -23$
$c_{21} = (0)(2) + (-1)(4) = -4$	$c_{23} = (0)(-4) + (-1)(5) = -5$
$c_{31} = (1)(2) + (4)(4) = 18$	$c_{33} = (1)(-4) + (4)(5) = 16$
$c_{12} = (2)(3) + (-3)(-2) = 12$	$c_{14} = (2)(2) + (-3)(-3) = 13$
$c_{22} = (0)(3) + (-1)(-2) = 2$	$c_{24} = (0)(2) + (-1)(-3) = 3$
$c_{32} = (1)(3) + (4)(-2) = -5$	$c_{34} = (1)(2) + (4)(-3) = -10$

This gives the product

$$AB = C = \begin{bmatrix} -8 & 12 & -23 & 13 \\ -4 & 2 & -5 & 3 \\ 18 & -5 & 16 & -10 \end{bmatrix}$$

Some find that the method in Example 4 is easier for computing products by hand than that shown in Example 2 and Example 3. Both methods are perfectly fine, so use whichever you prefer.

The Identity Matrix

Given an $n \times m$ matrix $A = [\mathbf{a}_1 \cdots \mathbf{a}_m]$, we can see that the zero matrix 0_{nm} satisfies

$$A + 0_{nm} = A$$

Additive Identity

For this reason, the matrix 0_{nm} is called the **additive identity**. It is less apparent which matrix plays the role of the multiplicative identity—that is, the matrix I such that

$$A = AI$$

for all $n \times m$ matrices A. But we can deduce I. For AI to be defined and equal to A, I must be an $m \times m$ matrix, so let $I_m = [\mathbf{i}_1\ \mathbf{i}_2 \cdots \mathbf{i}_m]$. The first columns of A and AI must be the same, so that

$$\mathbf{a}_1 = A\mathbf{i}_1 = i_{11}\mathbf{a}_1 + i_{21}\mathbf{a}_2 + \cdots + i_{m1}\mathbf{a}_m$$

The only way that this will hold for every possible matrix A is if

$$\mathbf{i}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Taking each of the other columns in turn, similar arguments show that

$$\mathbf{i}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{i}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots \mathbf{i}_m = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Note that $\mathbf{i}_1 = \mathbf{e}_1, \ldots, \mathbf{i}_m = \mathbf{e}_m$, where $\mathbf{e}_1, \ldots, \mathbf{e}_m$ are defined in (3) of Section 3.1. Thus we find that if I_m is the $m \times m$ matrix

$$I_m = [\mathbf{e}_1\ \mathbf{e}_2 \cdots \mathbf{e}_m] = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

then $A = AI_m$ for all $n \times m$ matrices A. Similarly, we have

$$A = I_n A \tag{3}$$

Identity Matrix

for all $n \times m$ matrices A. (Verification of (3) is left as an exercise.) The matrices I_n for $n = 1, 2, \ldots$ are multiplicative identities. For brevity they are referred to as **identity matrices**.

Properties of Matrix Algebra

As we have already seen, addition and scalar multiplication of matrices share many of the same properties as addition and multiplication of real numbers. However, not all properties of multiplication of real numbers carry over to multiplication of matrices. Some of those that do are given in Theorem 3.13.

THEOREM 3.13 ▶

Let s be a scalar, and let A, B, and C be matrices. Then each of the following holds in the cases where the indicated operations are defined:

(a) $A(BC) = (AB)C$

(b) $A(B + C) = AB + AC$

(c) $(A + B)C = AC + BC$

(d) $s(AB) = (sA)B = A(sB)$

(e) $AI = A$

(f) $IA = A$

Here I denotes an identity matrix of appropriate dimension.

▶ $D = [d_{ij}]$ is shorthand notation for

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & d_{2m} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nm} \end{bmatrix}$$

Proof We have already supplied a proof for part (e). Here we prove part (c) and leave the rest as exercises.

Suppose that $A = [a_{ij}]$ and $B = [b_{ij}]$ are $n \times m$ matrices and that $C = [c_{ij}]$ is an $m \times k$ matrix. Let $F = [f_{ij}] = (A + B)C$ and $G = [g_{ij}] = AC + BC$. It is straightforward to verify that F and G both have dimension $n \times k$. Using the formula for calculating product entries given in (2), we find that

$$\begin{aligned} f_{ij} &= (a_{i1}+b_{i1})c_{1j} + (a_{i2}+b_{i2})c_{2j} + \cdots + (a_{im}+b_{im})c_{mj} \\ &= (a_{i1}c_{1j} + a_{i2}c_{2j} + \cdots + a_{im}c_{mj}) + (b_{i1}c_{1j} + b_{i2}c_{2j} + \cdots + b_{im}c_{mj}) \\ &= g_{ij} \end{aligned}$$

which shows that $(A + B)C = AC + BC$ as claimed. ■ ■

Example 5

Let

A=[2-3-15],B=[074-2],andC=[-3-40-1]

Verify that $A(BC) = (AB)C$ and $A(B + C) = AB + AC$.

Solution We have

A(BC)=[2-3-15] ([074-2] [-3-40-1]) =[2-3-15] [0-7-12-14]=[3628-60-63]

and

(AB)C=([2-3-15] [074-2]) [-3-40-1] =[-122020-17] [0-7-12-14]=[3628-60-63]

so that $A(BC) = (AB)C$. We also have

A(B+C)=[2-3-15] ([074-2]+[-3-40-1]) =[2-3-15] [-334-3]=[-181523-18]

and

AB+AC=[2-3-15] [074-2]+[2-3-15] [-3-40-1] =[-122020-17] [-6-53-1]=[-181523-18]

Thus $A(B + C) = AB + AC$.

Although many of the rules for the algebra of matrices are the same as the rules for the algebra of real numbers, there are important differences. For instance, if a and b are any real numbers, then $ab = ba$. The same is not true of matrices, where multiplication is not generally commutative (see Example 3). This and two other properties of real numbers that do not carry over to matrices are given in the next theorem.

THEOREM 3.14

Let A, B, and C be nonzero matrices.

(a) It is possible that $AB \neq BA$.

(b) $AB = 0$ does not imply that $A = 0$ or $B = 0$.

(c) $AC = BC$ does not imply that $A = B$ or $C = 0$.

Here 0 represents the zero matrix of appropriate dimension.

Proof Part (a) follows from Example 3. For part (b), let

$$A=[1236] \text{and} B=[-462-3] \Rightarrow AB=[0000]$$

Thus $AB = 0_{22}$ even though A and B are nonzero matrices. For (c), let

$$A=[-3311-3], B=[-1231], C=[1326]$$

Although $A \neq B$ and C is a nonzero matrix, we have

$$AC=[39515] \text{and} BC=[39515] \Rightarrow AC=BC$$

Because of the results in Theorem 3.14, we must take care when performing algebra with matrices.

Transpose of a Matrix

Transpose

The **transpose** of a matrix A is denoted by A^T and results from interchanging the rows and columns of A. For example,

$$A=[123456789101112] \Rightarrow AT=[159261037114812]$$

Focusing on individual entries, the entry in row i and column j of A becomes the entry in row j and column i of A^T.

A few properties of matrix transposes are given in the next theorem.

THEOREM 3.15

Let A and B be $n \times m$ matrices, C an $m \times k$ matrix, and s a scalar. Then

(a) $(A + B)^T = A^T + B^T$

(b) $(sA)^T = sA^T$

(c) $(AC)^T = C^T A^T$

Theorem 3.15 (c) says that the transpose of a product is the product of transposes but with the order reversed.

The proofs of parts (a) and (b) are straightforward and left as exercises. A general proof of part (c) is not difficult but is notationally messy and is omitted. The next example illustrates Theorem 3.15 (c) for a specific pair of matrices.

Example 6

Verify that $(AC)^T = C^T A^T$ for

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

Solution We have

$$AC = \begin{bmatrix} 7 & -4 \\ 14 & -10 \end{bmatrix} \Rightarrow (AC)^T = \begin{bmatrix} 7 & 14 \\ -4 & -10 \end{bmatrix}$$

On the other hand,

$$C^T A^T = \begin{bmatrix} 5 & -1 & 0 \\ 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -2 & 1 \\ 0 & -4 \end{bmatrix} = \begin{bmatrix} 7 & 14 \\ -4 & -10 \end{bmatrix}$$

Thus $(AC)^T = C^T A^T$.

Symmetric Matrix

A special class of square matrices are those such that $A = A^T$. Matrices with this property are said to be **symmetric**. For instance,

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

so that A is a symmetric matrix.

Composition of Linear Transformations

Our definition for matrix multiplication is consistent with composition of linear transformations. Let $S : \mathbf{R}^m \to \mathbf{R}^k$ and $T : \mathbf{R}^k \to \mathbf{R}^n$ be linear transformations. By Theorem 3.8, there exists a $k \times m$ matrix $B = [\mathbf{b}_1 \cdots \mathbf{b}_m]$ and an $n \times k$ matrix A such that

$$S(\mathbf{x}) = B\mathbf{x} \quad \text{and} \quad T(\mathbf{y}) = A\mathbf{y}$$

where $\mathbf{x}$ is in $\mathbf{R}^m$ and $\mathbf{y}$ is in $\mathbf{R}^k$. Now let

$$W(\mathbf{x}) = T(S(\mathbf{x}))$$

be the composition of T with S (see Figure 2). Then $W : \mathbf{R}^m \to \mathbf{R}^n$ and

$$
\begin{aligned}
W(\mathbf{x}) &= T(S(\mathbf{x})) = T(B\mathbf{x}) \\
&= T\left(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_m\mathbf{b}_m\right) \\
&= x_1T(\mathbf{b}_1) + x_2T(\mathbf{b}_2) + \cdots + x_mT(\mathbf{b}_m) \\
&= x_1A\mathbf{b}_1 + x_2A\mathbf{b}_2 + \cdots + x_mA\mathbf{b}_m \\
&= \underbrace{\begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_m \end{bmatrix}}_{n \times m \text{ matrix}} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \\
&= AB\mathbf{x}
\end{aligned}
$$

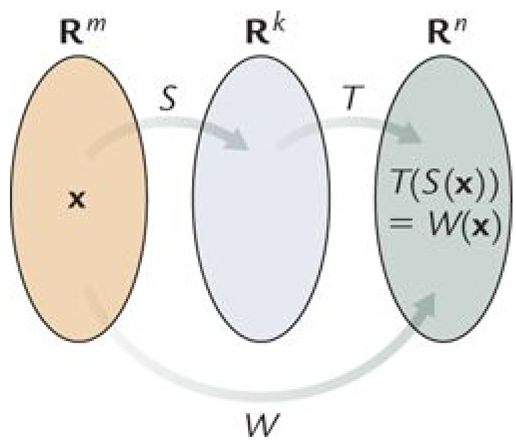


Figure 2 The composition $W(\mathbf{x}) = T(S(\mathbf{x}))$.

Since $W(\mathbf{x}) = AB\mathbf{x}$, then W is a linear transformation by Theorem 3.8. This proves the next theorem.

THEOREM 3.16 ▶

Let $S(\mathbf{x}) = B\mathbf{x}$ and $T(\mathbf{y}) = A\mathbf{y}$ be linear transformations, where A is $n \times k$ and B is $k \times m$. Then $W(\mathbf{x}) = T(S(\mathbf{x}))$ is a linear transformation with $W(\mathbf{x}) = AB\mathbf{x}$.

Example 7

Find the matrix C such that $W(\mathbf{x}) = C\mathbf{x}$ for the linear transformation $W(\mathbf{x}) = T(S(\mathbf{x}))$, where

$$S(\mathbf{x}) = \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 4x_2 \\ 3x_1 + x_2 \end{bmatrix} \quad \text{and} \quad T(\mathbf{x}) = \begin{bmatrix} x_1 + x_2 - 2x_3 \\ 2x_1 - x_2 + 3x_3 \end{bmatrix}$$

Solution Based on the definitions for S and T, we have $S(\mathbf{x}) = B\mathbf{x}$ and $T(\mathbf{x}) = A\mathbf{x}$, where

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

Since $W(\mathbf{x}) = T(S(\mathbf{x}))$, we have $W(\mathbf{x}) = C\mathbf{x}$ for

$$C = AB = \begin{bmatrix} 1 & 1 & -2 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 4 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} -5 & 1 \\ 14 & -3 \end{bmatrix}$$

Powers of a Matrix

Suppose that A is the 2×2 matrix

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

Then we let A^2 denote $A \cdot A$, so that

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

Similarly, $A^3 = A \cdot A \cdot A$. By the associative law, we have $A \cdot (A \cdot A) = (A \cdot A) \cdot A$, so that we can interpret $A^3 = A \cdot A^2$ or $A^3 = A^2 \cdot A$—either way, we get the same result. For our matrix, we have

$$A^3 = A \cdot A^2 = \begin{bmatrix} 2 & 2 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ -3 & -5 \end{bmatrix} = \begin{bmatrix} -10 & -6 \\ 9 & -1 \end{bmatrix}$$

In general, if A is an $n \times n$ matrix, then

$$A^k = \underbrace{A \cdot A \cdot \cdots A}_{k \text{ factors}}$$

As with A^3, the associative law ensures that we get the same result regardless of how we organize the products.

Example 8

In a small town there are 10,000 homes. When it comes to television viewing, the residents have three choices: they can subscribe to cable, they can pay for satellite service, or they watch no TV. (The town is sufficiently remote so that an antenna does not work.) In any given year, 80% of the cable customers stick with cable, 10% switch to satellite, and 10% quit watching TV. Over the same time period, 90% of satellite viewers continue with satellite service, 5% switch to cable, and 5% quit watching TV. And of those people who start the year not watching TV, 85% continue not watching, 5% subscribe to cable, and 10% get satellite service (see Figure 3). If the current distribution is 6000 homes with cable, 2500 with satellite service, and 1500 with no TV, how many of each type will there be a year from now? How about two years from now? Three years from now?

Solution The information given is summarized in Table 2. Reading down each column, we see the percentage of viewers in a given group that switches to one of the other groups. At the start of the year, 6000 homes have cable, 2500 have satellite service, and 1500 have no TV. From our table we see that at the end of the year, the number of homes with cable is

$$0.80(6000)+0.05(2500)+0.05(1500)=5000 \tag{4}$$

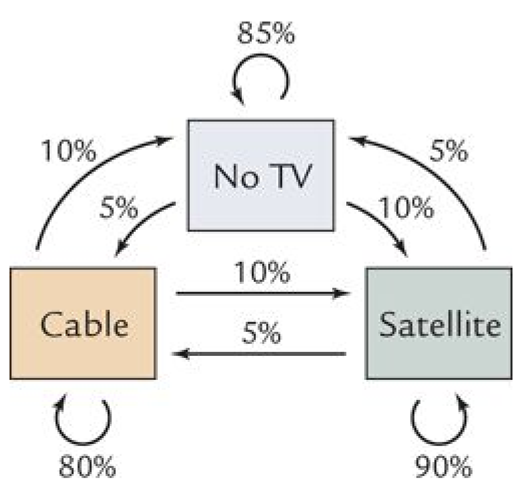


Figure 3 Customer transition percentages between cable, satellite, and no television.

Table 2 Rates of Customer Transitions

		Start of Year		
		Cable	**Satellite**	**No TV**
End of Year	Cable	80%	5%	5%
	Satellite	10%	90%	10%
	No TV	10%	5%	85%

Similar calculations can be performed to determine the number of satellite customers and the number with no TV at year's end. We can simplify these calculations by letting A be the matrix formed from the values in our table (converted from percentages to proportions) and **x** be the vector containing the initial number of people in each category,

$$A=\begin{bmatrix}0.80 & 0.05 & 0.05\\ 0.10 & 0.90 & 0.10\\ 0.10 & 0.05 & 0.85\end{bmatrix} \text{ and } \mathbf{x}=\begin{bmatrix}6000\\ 2500\\ 1500\end{bmatrix}$$

Note that the top entry of $A\mathbf{x}$ is the same as the left side of (4) and that in general $A\mathbf{x}$ gives the number of people in each category after a year has passed. We have

$$A\mathbf{x}=\begin{bmatrix}0.80 & 0.05 & 0.05\\ 0.10 & 0.90 & 0.10\\ 0.10 & 0.05 & 0.85\end{bmatrix}\begin{bmatrix}6000\\ 2500\\ 1500\end{bmatrix}=\begin{bmatrix}5000\\ 3000\\ 2000\end{bmatrix}$$

which shows that after one year the town will have 5000 cable subscribers, 3000 receiving satellite service, and 2000 with no TV. If the proportion of homes switching among categories remains unchanged, after two years the number of households in each group will be

$$A(A\mathbf{x})=A^2\mathbf{x}=\begin{bmatrix}0.65 & 0.0875 & 0.0875\\ 0.18 & 0.82 & 0.18\\ 0.17 & 0.0925 & 0.7325\end{bmatrix}\begin{bmatrix}6000\\ 2500\\ 1500\end{bmatrix}=\begin{bmatrix}4250\\ 3400\\ 2350\end{bmatrix}$$

Similarly, after three years, the number of homes in each group is

$$A^3\mathbf{x} = \begin{bmatrix} 0.655375 & 0.115625 & 0.115625 \\ 0.244 & 0.756 & 0.244 \\ 0.2185 & 0.128375 & 0.640375 \end{bmatrix} \begin{bmatrix} 6000 \\ 2500 \\ 1500 \end{bmatrix} \approx \begin{bmatrix} 3688 \\ 3720 \\ 2593 \end{bmatrix}$$

More generally, the number of households in each category after n years is given by $A^n\mathbf{x}$. In Section 3.5 we see what happens as $n \to \infty$.

▶ Here we have rounded the entries of $A^3\mathbf{x}$ to the nearest integer.

Diagonal Matrix

There are two types of matrices that retain their form when raised to powers. The first of these is the **diagonal matrix**, which has the form

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix} \qquad (5)$$

The **diagonal** of A consists of the entries $a_{11}, \ldots, a_{nn}$, each of which can be zero or nonzero. For instance,

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

are both diagonal matrices.

Example 9

Compute A^2 and A^3 for the diagonal matrix

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

Solution We have

$$A^2=\begin{bmatrix}2&0&0\\0&-3&0\\0&0&5\end{bmatrix}\begin{bmatrix}2&0&0\\0&-3&0\\0&0&5\end{bmatrix}=\begin{bmatrix}2^2&0&0\\0&(-3)^2&0\\0&0&5^2\end{bmatrix}=\begin{bmatrix}4&0&0\\0&9&0\\0&0&25\end{bmatrix}$$

and

$$A^3=A^2\cdot A=\begin{bmatrix}2^2&0&0\\0&(-3)^2&0\\0&0&5^2\end{bmatrix}\begin{bmatrix}2&0&0\\0&-3&0\\0&0&5\end{bmatrix}=\begin{bmatrix}2^3&0&0\\0&(-3)^3&0\\0&0&5^3\end{bmatrix}=\begin{bmatrix}8&0&0\\0&-27&0\\0&0&125\end{bmatrix}$$

In Example 9 the powers of the diagonal matrix A are just the powers of the diagonal entries. This is true for any diagonal matrix.

THEOREM 3.17

If A is the diagonal matrix in (5), then for each integer $k \geq 1$,

$$A^k=\begin{bmatrix}a_{11}^k&0&0&\cdots&0\\0&a_{22}^k&0&\cdots&0\\0&0&a_{33}^k&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&a_{nn}^k\end{bmatrix}$$

Proof We use induction for the proof. First, if $k = 1$ then $A^k = A$ so that A^k clearly has the form shown. Next is the induction hypothesis, which states that the theorem is true for exponent $k - 1$, so that A^{k-1} is diagonal and given by

$$A^{k-1}=\begin{bmatrix}a_{11}^{k-1}&0&0&\cdots&0\\0&a_{22}^{k-1}&0&\cdots&0\\0&0&a_{33}^{k-1}&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&a_{nn}^{k-1}\end{bmatrix}$$

Therefore

$$A^k=A^{k-1}\cdot A=\begin{bmatrix}a_{11}^{k-1}&0&0&\cdots&0\\0&a_{22}^{k-1}&0&\cdots&0\\0&0&a_{33}^{k-1}&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&a_{nn}^{k-1}\end{bmatrix}\begin{bmatrix}a_{11}&0&0&\cdots&0\\0&a_{22}&0&\cdots&0\\0&0&a_{33}&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&a_{nn}\end{bmatrix}=\begin{bmatrix}a_{11}^k&0&0&\cdots&0\\0&a_{22}^k&0&\cdots&0\\0&0&a_{33}^k&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&a_{nn}^k\end{bmatrix}$$

Hence A^k has the claimed form. ■■

Upper Triangular Matrix

A second class of matrices whose form is unchanged when raised to a power are triangular matrices. An $n \times n$ matrix A is **upper triangular** if it has the form

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Lower Triangular Matrix

That is, A is upper triangular if the entries below the diagonal are all zero. Similarly, an $n \times n$ matrix A is **lower triangular** if the terms above the diagonal are all zero,

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

▶ As with diagonal matrices, triangular matrices can have entries equal to zero along the diagonal. Note that this is different than triangular linear systems, where the leading (diagonal) coefficients must be nonzero.

A matrix is **triangular** if it is either upper or lower triangular. (Diagonal matrices are both.) For example, the matrix

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

is upper triangular, as are the matrix powers

$$A^2 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 4 & 4 \\ 0 & 0 & 9 \end{bmatrix} \quad \text{and} \quad A^3 = \begin{bmatrix} -1 & 0 & 7 \\ 0 & -8 & 28 \\ 0 & 0 & 27 \end{bmatrix}$$

In fact, powers of upper (or lower) triangular matrices are also upper (or lower) triangular.

THEOREM 3.18 ▶

Let A be an $n \times n$ upper (lower) triangular matrix and $k \geq 1$ an integer. Then A^k is also an upper (lower) triangular.

The proof makes use of the fact that the product of two upper (lower) triangular matrices is again an upper (lower) triangular matrix (see Exercises 71–72), and is left as an exercise.

Elementary Matrices

Consider the matrix A and the indicated elementary row operation to transform to A_1.

A=[31-126-201451-1]2R1+R2→R2~[31-12120-25451-1]=A1

Now suppose that we perform the same row operation on

I3=[100010001]2R1+R2→R2~[100210001]=E

Forming the product EA gives us

EA=[100210001] [31-126-201451-1]=[31-12120-25451-1]=A1

It turns out that we can perform *any* row operation by using the same approach. Given an $n \times m$ matrix A and a desired row operation, all we do is start with I_n, perform the row operation on to produce a new matrix E, and then compute EA.

DEFINITION 3.19 ▶

Elementary Matrix

If we perform a single elementary row operation on an identity matrix I_n, then the result is called an **elementary matrix**.

Examples of row operations on a matrix A and the corresponding 3×3 elementary matrices are shown below.

Row Operation	Elementary Matrix
$-4R_1 + R_3 \to R_3$	[100010-401]

Row Operation	Elementary Matrix
$5R_2 \rightarrow R_2$	[100050001]
R1↔R2	[010100001]

Example 10

Use elementary matrices to perform the sequence of row operations R1↔R2, $3R_1 + R_2 \rightarrow R_2$, and $-2R_2 \rightarrow R_2$ on A. Then find a single matrix that will do the same thing.

A=[-3-2310-2]

Solution Since A has two rows, the required elementary matrices are 2 × 2.

R1↔R2⇒E1=[0110]3R1+R2→R2⇒E2=[1031]-2R2→R2⇒E3=[100-2]

The sequence of row operations is then

A1=E1A=[0110] [-3-2310-2]=[10-2-3-23]⇒A2=E2A1=[1031] [10-2-3-23]=[10-20-2-3]⇒A3 =E3A2=[100-2] [10-20-2-3]=[10-2046]

It can be verified that the sequence of row operations will result in A_3. The multiplications by elementary matrices was also sequential, in reverse order of the operations,

A3=E3A2=E3(E2A1)=(E3E2)A1=(E3E2)(E1A)=(E3E2E1)A

The single matrix

B=E3E2E1=[100-2][1031][0110]=[01-2-6]

will produce the same sequence of row operations,

BA=[01-2-6] [-3-2310-2]=[10-2046]=A3

Partitioned Matrices

Some applications require working with really, really big matrices. (Think tens of millions of entries.) In such situations, we can divide the matrices into smaller submatrices that are more manageable.

- The material on partitioned matrices is optional.

- For instance, the Google Page Rank search algorithm uses a matrix with several billion rows and columns.

Let's start with a concrete example, say

A=[20−317−142046−118−3027−3320−6901−185−146978]

Partitioned Matrix, Blocks

Below A is shown **partitioned** into six different submatrices (called **blocks**),

A=[A11A12A21A22A31A32]

where

A11=[20-3-1426-11] A12=[17048-3]A21=[02720-6] A22=[-3390]A31=[1-18469] A32=[5-178]

Matrices can be partitioned in any manner desired. The advantage of working with partitioned matrices is that we can do arithmetic on a few blocks at a time to make better use of computer memory. In addition, if numerous processors are available, computations can be distributed across them and simultaneously performed in parallel.

Of the arithmetic operations that can be performed on partitioned matrices, addition and scalar multiplication are the easiest to understand. Suppose that A and B are $n \times m$ matrices, partitioned into blocks as shown:

A=[A11A12A13A21A22A23]n2n1 m1m2m3B=
[B11B12B13B21B22B23]n2n1 m1m2m3

The notation around each matrix indicates the block dimensions. For example, A_{21} and B_{21} are both $n_2 \times m_1$ submatrices. Since the corresponding blocks have the same dimensions, they can be added together in the usual manner, yielding

A+B=[(A11+B11)(A12+B12)(A13+B13)(A21+B21)(A22+B22)
(A23+B23)]

The formula for scalar multiplication of partitioned matrices is quite natural,

rA=[rA11rA12rA13rA21rA22rA23]

Example 11

Suppose that A and B are two matrices partitioned into blocks,

A=[1−20344211−27−350−1]   and   B= [6241037−12245016]

Use addition and scalar multiplication of partitioned matrices to find $A + B$ and $-4B$.

Solution We have

A11+B11=[1-242]+[6237]=[7079]A12+B12=[03411-2]+[410-122]=
[444030]A21+B21=[7-3]+[45]=[112]A22+B22=[50-1]+[016]=[515]

Pulling the block sums back together gives

A+B=[(A11+B11)(A12+B12)(A21+B21)
(A22+B22)]= [7044479030112515]

To find $-4B$, we need

-4B11=-4[6237]=[-24-8-12-28]-4B12=-4[410-122]=[-16-404-8-8]-4B21=-4[45]=[-16-20]-4B22=-4[016]=[0-4-24]

Putting everything back together yields

−4B=[−4B11−4B12−4B21−4B22]=[−24−8−16−40−12−284−8−8−16−200−4−24]

Multiplication of partitioned matrices nicely mimics the usual multiplication of matrices. Suppose that A is an $n \times k$ matrix and B is a $k \times m$ matrix, so that AB is defined. Partition A and B as shown:

A=[A11A12⋯A1iA21A22⋯A2i⋮⋮⋯⋮Aj1Aj2⋯Aji]n1n2⋮nj k1k2⋯kiB=[B11B12⋯B1lB21B22⋯B2l⋮⋮⋯⋮Bi1Bi2⋯Bil]k1k2⋮ki m1m2⋯ml

The only requirement when setting the size of the partitions is that the column sizes for A must match the row sizes for B. (Here both are k_1, k_2, … , k_i.) We can compute AB using the blocks in exactly the same manner as we do with regular matrix multiplication. For instance, the upper left block of AB is given by

A11B11+A12B21+⋯+A1iBi1

Note that each product in this sum is an $n_1 \times m_1$ matrix, so that the upper left block of AB is also an $n_1 \times m_1$ matrix.

Example 12

Suppose that

A=[3−1240021−3123−40−41602−2]=[A11A12A21A22]B=[312402−1702410−1−1]=[B11B12B13B21B22B23]

Find AB using block multiplication with the given partitions.

Solution First, the block multiplication yields

AB=[A11A12A21A22] [B11B12B13B21B22B23] = [(A11B11+A12B21)(A11B12+A12B22)(A11B13+A12B23)(A21B11+A22B21)(A21B12+A22B22)(A21B13+A22B23)]

For the upper left block we need the products

A11B11=[3-10223] [34]=[5818]andA12B21=[2401-31-40-4] [-120]=[6-74]

Thus the upper left block of *AB* is given by

A11B11+A12B21=[5818]+[6-74]=[11122]

Similar computations give us

A11B12+A12B22=[33-4-30],A11B13+A12B23=[749],A12B11+A22B21=[31],A21B12+A22B22=[11],A21B13+A22B23=[24]

Finally, we pull everything together to arrive at

AB=[113371−4422−309311124]

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Perform the indicated computations when possible using the given matrices.

A=[253-4],B=[314-5],C=[21540-1],D=[2-2-3031]

(a) $A + B$; AC

(b) $B - 3I_2$; DB

(c) CB; A^2

(d) $C^T - D$; $BA + DC$

2. Find all values of a such that $A^2 = A$, where A=[a2-11].

3. Let T_1 and T_2 be linear transformations given by

T1([x1x2])=[x1-2x2-x1+3x2]T2([x1x2])=[-x1+3x2x1-x2]

Find the matrix A such that

(a) $T_1(T_2(\mathbf{x})) = A\mathbf{x}$

(b) $T_2(T_2(\mathbf{x})) = A\mathbf{x}$

4. Describe the row operation that corresponds to the elementary matrix E.

(a) E=[10-21]

(b) E=[1004]

(c) E=[100310001]

(d) E=[001010100]

5. Determine if each statement is true or false, and justify your answer.

(a) If A is an $n \times m$ matrix, then A^5 is defined if and only if $n = m$.

(b) If A and B have the same dimensions, then $(2A - 3B)^T = 2A^T - 3B^T$.

(c) If A is an upper triangular 3×3 matrix and B is a lower triangular 3×3 matrix, then AB is a diagonal matrix.

(d) The product of two elementary matrices is another elementary matrix.

EXERCISES

Exercises 1–6: Perform the indicated computations when possible, using the matrices given below. If a computation is not possible, explain why.

A=[-312-1],B=[04-25],C=[50-1433]D=[10-3-25-1],E=[14-5-21-3026]

1. (a) $A + B$, (b) $AB + I_2$, (c) $A + C$
2. (a) AC, (b) $C + D^T$, (c) $CB + I_2$
3. (a) $(AB)^T$, (b) CE, (c) $(A - B)D$
4. (a) A^3, (b) BC^T, (c) $EC + I_3$
5. (a) $(C + E)B$, (b) $B(C^T + D)$, (c) $E + CD$
6. (a) $AD - C^T$, (b) $AB - DC$, (c) $DE + CB$

Exercises 7–10: Find the missing values in the given matrix equation.

7. [2a3-2] [b-3-12]=[3-85c]
8. [14a7] [2-1b3]=[6d11c]
9. [a3-23-24] [2-10bc1]=[4d-6-5]
10. [1a0-25b] [3cd-212]=[-337e-2-4f-21]
11. Find all values of a such that $A^2 = A$ for

    A=[5-10a-4]

12. Find all values of a such that $A^3 = 2A$ for

    A=[-22-1a]

13. Let T_1 and T_2 be linear transformations given by

    T1([x1x2])=[3x1+5x2-2x1+7x2]T2([x1x2])=[-2x1+9x25x2]

    Find the matrix A such that
    - (a) $T_1(T_2(\mathbf{x})) = A\mathbf{x}$
    - (b) $T_2(T_1(\mathbf{x})) = A\mathbf{x}$
    - (c) $T_1(T_1(\mathbf{x})) = A\mathbf{x}$
    - (d) $T_2(T_2(\mathbf{x})) = A\mathbf{x}$
14. Let T_1 and T_2 be linear transformations given by

    T1([x1x2])=[-2x1+3x2x1+6x2]T2([x1x2])=[4x1-5x2x1+5x2]

    Find the matrix A such that

(a) $T_1(T_2(\mathbf{x})) = A\mathbf{x}$

(b) $T_2(T_1(\mathbf{x})) = A\mathbf{x}$

(c) $T_1(T_1(\mathbf{x})) = A\mathbf{x}$

(d) $T_2(T_2(\mathbf{x})) = A\mathbf{x}$

Exercises 15–18: Expand each of the given matrix expressions and combine as many terms as possible. Assume that all matrices are $n \times n$.

15. $(A + I)(A - I)$

16. $(A + I)(A^2 + A)$

17. $(A + B^2)(BA - A)$

18. $A(A + B) + B(B - A)$

Exercises 19–22: The given matrix equation is false in general. Explain why, and give a 2×2 example showing the equations are false. Assume that all matrices are $n \times n$.

19. $(A + B)^2 = A^2 + 2AB + B^2$

20. $(A - B)^2 = A^2 - 2AB + B^2$

21. $A^2 - B^2 = (A - B)(A + B)$

22. $A^3 + B^3 = (A + B)(A^2 - AB + B^2)$

23. Suppose that A has four rows and B has five columns. If AB can be defined, what are its dimensions?

24. Suppose that A has four rows and B has five columns. If BA can be defined, what are its dimensions?

Exercises 25–30: Assume that A is a matrix with three rows. Find the elementary matrix E such that EA gives the matrix resulting from A after the given row operation is performed.

25. $4R_1 \to R_1$

26. $-3R_2 \to R_2$

27. R2↔R1

28. R3↔R2

29. $2R_1 + R_3 \to R_3$

30. $-4R_3 + R_2 \rightarrow R_2$

Exercises 31–36: Assume that *A* is a matrix with three rows. Find the matrix *B* such that *BA* gives the matrix resulting from *A* after the given row operations are performed.

31. -2R1+R2→R2,5R3→R3

32. -6R2+R3→R3,R1↔R3

33. R2↔R1,3R1+R2→R2

34. -2R1→R1,7R2+R3→R3

35. -3R1→R1,R1↔R2,4R1+R2→R2

36. -3R1+R2→R2,2R1+R3↔R3,-R2+R3→R3

Exercises 37–40: Let

A=[1-2-13-2014-12-200121], B=[20-11-31210-1-2322-1-2]

37. Partition *A* and *B* into four 2 × 2 blocks, and when possible use the partitions to compute the following:

(a) $A - B$

(b) AB

(c) BA

38. Partition *A* and *B* into four blocks, with the upper left of each a 3 × 3 matrix, and when possible use the partitions to compute the following:

(a) $A + B$

(b) AB

(c) BA

39. Partition *A* and *B* into four blocks, with the lower left of each a 3 × 3 matrix, and when possible use the partitions to compute the following:

(a) $B - A$

(b) AB

(c) $BA + A$

40. Partition *A* and *B* into four blocks, with the lower right of each a 3 × 3 matrix, and when possible use the partitions to compute the following:

(a) $A + B$

(b) AB

(c) BA

41. Suppose that A is a 3×3 matrix. Find a 3×3 matrix E such that the product EA is equal to A with the

(a) first and second rows interchanged.

(b) first and third rows interchanged.

(c) second row multiplied by -2.

42. Suppose that A is a 4×3 matrix. Find a 4×4 matrix E such that the product EA is equal to A with the

(a) first and fourth rows interchanged.

(b) second and third rows interchanged.

(c) third row multiplied by -2.

FIND AN EXAMPLE Exercises 43–50: Find an example that meets the given specifications.

43. 3×3 matrices A and B such that $AB \neq BA$.

44. 3×3 matrices A and B such that $AB = BA$.

45. 2×2 nonzero matrices A and B such that $AB = 0_{22}$.

46. 3×3 nonzero matrices A and B such that $AB = 0_{33}$.

47. 2×2 matrices A and B that have *no* zero entries and yet $AB = 0_{22}$.

48. 3×3 matrices A and B that have *no* zero entries and yet $AB = 0_{33}$.

49. 2×2 matrices A, B, and C that are nonzero, where $A \neq B$ but $AC = BC$.

50. 3×3 matrices A, B, and C that are nonzero, where $A \neq B$ but $AC = BC$.

TRUE OR FALSE Exercises 51–62: Determine if the statement is true or false, and justify your answer. You may assume that A, B, and C are $n \times n$ matrices.

51.

(a) If A and B are nonzero (that is, not equal to 0_{nn}), then so is $A + B$.

(b) $(A + B^T)^T = A^T + B$

52.

(a) If A and B are diagonal matrices, then so is $A - B$.

(b) If A is upper triangular, then so is AB.

53.

(a) If A is upper triangular, then A^T is lower triangular.

(b) If $AB = 0$, then either $A = 0$ or $B = 0$.

54.

(a) $AB \neq BA$ for all matrices A and B.

(b) $AA^T = I_n$ for all matrices A.

55.

(a) $C + I_n = C$ for all matrices C.

(b) If $S : \mathbf{R}^n \to \mathbf{R}^n$ is a linear transformation, then so is $S(S(\mathbf{x}))$.

56.

(a) If A is symmetric, then so is $A + I_n$.

(b) If B is a diagonal matrix, then $B = B^T$.

57.

(a) $(ABC)^T = C^T B^T A^T$

(b) If A is upper triangular, then $A^T A$ is diagonal.

58.

(a) If $AB = BA$, then either $A = I_n$ or $B = I_n$.

(b) $ABAB = A^2B^2$.

59.

(a) $(AB + C)^T = C^T + B^T A^T$

(b) If A is diagonal, then $AB = BA$.

60.

(a) All elementary matrices corresponding to interchanging two rows are diagonal.

(b) All elementary matrices corresponding to multiplying a row times a constant are diagonal.

61.

(a) Suppose that E_1 and E_2 are two elementary matrices. Then $E_1E_2 = E_2E_1$.

(b) If E is an $n \times n$ elementary matrix corresponding to interchanging two rows, then $E^2 = I_n$.

62.

(a) $(AB)^2 = A^2B^2$ for all matrices A and B.

(b) $(I_n + I_n)^3 = 3I_n$

63. Prove the remaining unproven parts of Theorem 3.11.

(a) $A + B = B + A$

(b) $s(A + B) = sA + sB$

(c) $(s + t)A = sA + tA$

(d) $(A + B) + C = A + (B + C)$

(e) $A + 0_{nm} = A$

64. Prove the remaining unproven parts of Theorem 3.13.

(a) $A(BC) = (AB)C$

(b) $A(B + C) = AB + AC$

(c) $s(AB) = (sA)B = A(sB)$

(d) $IA = A$

65. Prove the remaining unproven parts of Theorem 3.15.

(a) $(A + B)^T = A^T + B^T$

(b) $(sA)^T = sA^T$

66. Verify Equation (3): If A is an $n \times m$ matrix and I_n is the $n \times n$ identity matrix, then $A = I_nA$.

67. Show that if A and B are symmetric matrices and $AB = BA$, then AB is also a symmetric matrix.

68. Let A and D be $n \times n$ matrices, and suppose that the only nonzero terms of D are along the diagonal. Must $AD = DA$? If so, prove it. If not, give a counter-example.

69. Let A be an $n \times m$ matrix.

(a) What are the dimensions of $A^T A$?

(b) Show that $A^T A$ is symmetric.

70. Suppose that A and B are both $n \times n$ diagonal matrices. Prove that AB is also an $n \times n$ diagonal matrix. (HINT: The formula given in (2) can be helpful here.)

71. Suppose that A and B are both $n \times n$ upper triangular matrices. Prove that AB is also an $n \times n$ upper triangular matrix. (HINT: The formula given in (2) can be helpful here.)

72. Suppose that A and B are both $n \times n$ lower triangular matrices. Prove that AB is also an $n \times n$ lower triangular matrix. (HINT: The formula given in (2) can be helpful here.)

73. Prove Theorem 3.18: If A is an upper (lower) triangular matrix and $k \geq 1$ is an integer, then A^k is also an upper (lower) triangular matrix.

74. If A is a square matrix, show that $A + A^T$ is symmetric.

75. A square matrix A is **skew symmetric** if $A^T = -A$.

(a) Find a 3×3 skew symmetric matrix.

(b) Show that the same numbers must be on the diagonal of all skew symmetric matrices.

76. A square matrix A is **idempotent** if $A^2 = A$.

(a) Find a 2×2 matrix, not equal to 0_{22} or I, that is idempotent.

(b) Show that if A is idempotent, then so is $I - A$.

77. If A is a square matrix, show that $(A^T)^T = A$.

78. The **trace** of a square matrix A is the sum of the diagonal terms of A and is denoted by $\mathrm{tr}(A)$.

(a) Find a 3×3 matrix A with nonzero entries such that $\mathrm{tr}(A) = 0$.

(b) If A and B are both $n \times n$ matrices, show that $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$.

(c) Show that $\mathrm{tr}(A) = \mathrm{tr}(A^T)$.

(d) Select two nonzero 2×2 matrices A and B of your choosing, and check if $\mathrm{tr}(AB) = \mathrm{tr}(A)\mathrm{tr}(B)$.

79. In Example 8, suppose that the current distribution is 8000 homes with cable, 1500 homes with satellite, and 500 homes with no TV. Find the distribution one year, two years, three years, and four years from now.

80. In Example 8, suppose that the current distribution is 5000 homes with cable, 3000 homes with satellite, and 2000 homes with no TV. Find the distribution one year, two years, three years, and four years from now.

81. In an office complex of 1000 employees, on any given day some are at work and the rest are absent. It is known that if an employee is at work today, there is an 85% chance that she will be at work tomorrow, and if the employee is absent today, there is a 60% chance that she will be absent tomorrow. Suppose that today there are 760 employees at work. Predict the number that will be at work tomorrow, the following day, and the day after that.

82. The star quarterback of a university football team has decided to return for one more season. He tells one person, who in turn tells someone else, and so on, with each person talking to someone who has not heard the news. At each step in this chain, if the message heard is “yes” (he is returning), then there is a 10% chance it will be changed to “no,” and if the message heard is “no,” then there is a 15% chance that it will be changed to “yes.” Determine the probability that the fourth person in the chain hears the correct news.

Exercises 83–88: Perform the indicated computations when possible, using the matrices given below. If a computation is not possible, explain why.

A=[2-104033-168115-312],B=[-62-31-520303-1485-20]C=
[20111512436240873332],D=[52002513071436921471]

83. (a) $A + B$, (b) $BA - I_4$, (c) $D + C$

84. (a) AC, (b) $C^T - D^T$, (c) $CB + I_2$

85. (a) AB, (b) CD, (c) $(A - B)C^T$

86. (a) B^4, (b) BC^T, (c) $D + I_4$

87. (a) $(C + A)B$, (b) $C(C^T + D)$, (c) $A + CD$

88. (a) $AB - D^T$, (b) $AB - DC$, (c) $D + CB$

3.3 Inverses

In Section 3.1, we defined the linear transformation and developed the properties of this type of function. In this section we consider the problem of “reversing” a linear transformation. An application of this can be found in encoding messages. The history of secret codes is long, going back at least as far as Julius Caesar. Here we give a brief description of an encoding method that uses linear transformations.

We start by setting numerical equivalencies between letters and numbers,

$$a=1, b=2, c=3, \ldots z=26 \tag{1}$$

▶ We could also have numerical equivalencies for spaces, punctuation, and uppercase letters, but these are not needed here.

One way to encode messages is to map each number 1, 2, … , 26 to some other number, such as

$$1 \to 4, 2 \to 5, 3 \to 6, \ldots 26 \to 29$$

▶ This particular shift cipher is called the *Caesar cipher* because it is said to have been invented by Julius Caesar.

This is called a *shift cipher* and gives encodings such as

$$\text{“linear”} = \underbrace{\{12, 9, 14, 5, 1, 18\}}_{\text{numerical equiv.}} \to \underbrace{\{15, 12, 17, 8, 4, 21\}}_{\text{encoded message}}$$

We could convert the encoded message back to letters before transmitting the message, but that is not necessary—the string of encoded numbers can be sent. This type of encoding system is easy to implement but not very secure. In particular, it is vulnerable to *frequency analysis*, which involves breaking the code by matching up the numbers that occur most frequently in the encoded message with the letters that occur most frequently in the language of the original message (Table 1).

Table 1 **Relative Frequency of Letters in English**

Letter	Frequency
e	12.70%
t	9.06%
a	8.17%
o	7.51%
i	6.97%
n	6.75%
s	6.33%
h	6.09%
r	5.99%
d	4.53%

One way to deter frequency analysis is by encoding letters in groups called *blocks*. Although there are only 26 letters in English, there are 26^3 = 17,576 possible blocks of 3 letters, making frequency analysis more difficult. We start by converting each letter of a block using the equivalences in (1) and placing them in a vector in $\mathbf{R}^3$. For example, we would have

“the”=[2085] and “dog”=[4157]

We then apply a linear transformation to encode the vector. For instance, we could use $T(\mathbf{x}) = A\mathbf{x}$, where

A=[132-2-72265]

For example, we encode “linear” by splitting it into blocks and then applying T to both blocks,

“linear”={[12914],⏟“lin”[5118]}⏟“ear”→{T([12914]),T([5118])}=
{[67−59148],[4419106]}⏟encoded message

Of course, it does no good to encode a message if it cannot be decoded. To see how we decode, suppose that we have received the encoded block

y=[109-95241]

To decode, we need to find the vector **x** such that $T(\mathbf{x}) = \mathbf{y}$, or equivalently, $A\mathbf{x} = \mathbf{y}$. Expressed as a linear system, we have

x1+3x2+2x3=109-2x1-7x2+2x3=-952x1+6x2+5x3=241 (2)

Our usual solution methods can be used to show that $x_1 = 18$, $x_2 = 15$, and $x_3 = 23$, so that

x = [181523] = “row”

In solving for x_1, x_2, and x_3, we have computed the *inverse* of T for the vector **y**. The notation for the inverse function is T^{-1}, and we have the relationship

T(x)=y⇔T-1(y)=x

A typical encoded message will consist of many blocks. Instead of repeatedly solving (2) with only changes to the right-hand numbers, it is more efficient to find a formula for $T^{-1}(\mathbf{y})$ for a generic vector $\mathbf{y} = (y_1, y_2, y_3)$. To find T^{-1}, we need to solve

x1+3x2+2x3=y1-2x1-7x2+2x3=y22x1+6x2+5x3=y3 (3)

for x_1, x_2, and x_3 in terms of y_1, y_2, and y_3. Transferring (3) to an augmented matrix and transforming to reduced echelon form yields the solution

x1=47y1+3y2-20y3x2=-14y1-y2+6y3x3=-2y1 +y3

Thus if we set

B=[473-20-14-16-201]

then we have $T^{-1}(\mathbf{y}) = B\mathbf{y}$. Therefore $T(\mathbf{x})$ and $T^{-1}(\mathbf{y})$ are both linear transformations. As a quick application of $T^{-1}(\mathbf{y}) = B\mathbf{y}$, we have

{[85−132179],[74−4173]}⏟
encoded message →{T−1([85−132179]),T−1([74−4173])}={[19169],⏟"spi"[6625]}⏟“ffy”=“spiffy”

Larger blocks and corresponding encoding matrices can be used if needed. More can be learned about encoding messages in *Cryptological Mathematics* by Robert Lewand (Mathematical Association of America Textbooks).

Inverse Linear Transformations

A linear transformation $T : \mathbf{R}^m \to \mathbf{R}^n$ that is one-to-one and onto pairs up each vector $\mathbf{x}$ in $\mathbf{R}^m$ with a unique vector $\mathbf{y} = T(\mathbf{x})$ in $\mathbf{R}^n$. The *inverse* T^{-1} creates the same pairing but in reverse, so that $T^{-1}(\mathbf{y}) = \mathbf{x}$ (see Figure 1). Hence we can think of T^{-1} as reversing T. Here we put the notion of inverse on a firmer footing with a definition of an inverse linear transformation.

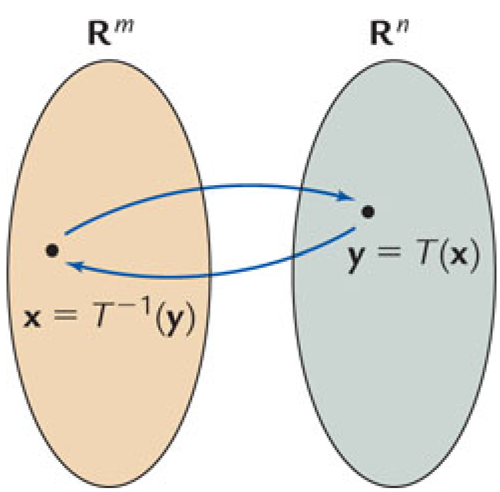


Figure 1 The relationship between T and T^{-1}.

DEFINITION 3.20

Inverse, Invertible

A linear transformation $T : \mathbf{R}^m \to \mathbf{R}^n$ is **invertible** if T is one-to-one and onto. When T is invertible, the **inverse** function $T^{-1} : \mathbf{R}^n \to \mathbf{R}^m$ is defined by

$$T^{-1}(\mathbf{y}) = \mathbf{x} \text{ if only if } T(\mathbf{x}) = \mathbf{y}$$

If T is an invertible linear transformation, then the inverse T^{-1} is unique and satisfies

$$T(T^{-1}(\mathbf{y})) = \mathbf{y} \text{ and } T^{-1}(T(\mathbf{x})) = \mathbf{x}$$

In our secret-code example, the encoding linear transformation mapped vectors in $\mathbf{R}^3$ to other vectors in $\mathbf{R}^3$, and we observed that the inverse function for decoding was another linear transformation. The next theorem gives a required condition for a linear transformation to be invertible and tells us that the inverse of a linear transformation is also a linear transformation.

THEOREM 3.21

Let $T : \mathbf{R}^m \to \mathbf{R}^n$ be a linear transformation. Then

(a) The only way T has an inverse is if $m = n$.

(b) If T is invertible, then T^{-1} is also a linear transformation.

Proof For (a), note that Theorem 3.6 in Section 3.1 tells us that the only way T can be one-to-one is if $n \geq m$. Moreover, Theorem 3.7 in Section 3.1 tells us the only way that T can be onto is if $n \geq m$.

Thus the only way that T can be one-to-one and onto, and hence invertible, is if $m = n$.

For (b), let $\mathbf{y}_1$ and $\mathbf{y}_2$ be vectors in $\mathbf{R}^n$. If T is invertible, then T is onto, so there exist vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbf{R}^n$ such that $T(\mathbf{x}_1) = \mathbf{y}_1$ and $T(\mathbf{x}_2) = \mathbf{y}_2$. Hence

$$\begin{aligned} T^{-1}(\mathbf{y}_1+\mathbf{y}_2) &= T^{-1}(T(\mathbf{x}_1)+T(\mathbf{x}_2)) \\ &= T^{-1}(T(\mathbf{x}_1+\mathbf{x}_2)) && (T \text{ is a linear transformation}) \\ &= \mathbf{x}_1+\mathbf{x}_2 \\ &= T^{-1}(\mathbf{y}_1)+T^{-1}(\mathbf{y}_2) \end{aligned}$$

A similar argument can be used to show that $T^{-1}(r\mathbf{y}) = rT^{-1}(\mathbf{y})$ (see Exercise 69). Therefore T^{-1} is a linear transformation. ■ ■

Note that having $m = n$ does not guarantee that T will have an inverse. For instance, if

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

and $T(\mathbf{x}) = A\mathbf{x}$, then $T : \mathbf{R}^2 \to \mathbf{R}^2$ but is not invertible, because T is neither one-to-one nor onto.

If $T : \mathbf{R}^n \to \mathbf{R}^n$ is an invertible linear transformation, then $T^{-1} : \mathbf{R}^n \to \mathbf{R}^n$ is also a linear transformation. Hence there exist $n \times n$ matrices A and B such that $T(\mathbf{x}) = A\mathbf{x}$ and $T^{-1}(\mathbf{x}) = B\mathbf{x}$. Furthermore, for each $\mathbf{x}$ in $\mathbf{R}^n$ we have

$$\mathbf{x} = T(T^{-1}(\mathbf{x})) = T(B\mathbf{x}) = AB\mathbf{x}$$

Since $AB\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbf{R}^n$, it follows that $AB = I_n$. We use this relationship to characterize what it means for a matrix to be invertible.

DEFINITION 3.22 ▶

Invertible Matrix

An $n \times n$ matrix A is **invertible** if there exists an $n \times n$ matrix B such that $AB = I_n$.

A matrix A is invertible precisely when the associated linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is invertible.

Example 1

Let

A=[12-125-1120]andB=[2-23-11-1-101]

Prove that A is invertible by showing $AB = I_3$.

Solution We have

AB=[12-125-1120][2-23-11-1-101]=[100010001]=I3

so A is invertible.

In Section 3.2 we showed that sometimes $AB \neq BA$ because matrix multiplication is generally not commutative. However, somewhat surprisingly, if A and B are square with $AB = I_n$, then $BA = AB = I_n$.

THEOREM 3.23

Suppose that A is an invertible matrix with $AB = I_n$. Then $BA = I_n$, and the matrix B such that $AB = BA = I_n$ is unique.

Proof Let be in $\mathbf{R}^n$. Since $AB = I_n$, we have

$$AB(A\mathbf{x}) = I_n(A\mathbf{x}) = A\mathbf{x} \Rightarrow A(B\,A\mathbf{x}) = A\mathbf{x} \Rightarrow A(B\,A\mathbf{x} - \mathbf{x}) = \mathbf{0}$$

Since A is invertible, $A\mathbf{y} = \mathbf{0}$ has only the trivial solution because the corresponding linear transformation is one-to-one. Hence $A(B\,A\mathbf{x} - \mathbf{x}) = \mathbf{0}$ implies that

$$B\,A\mathbf{x} - \mathbf{x} = \mathbf{0} \Rightarrow B\,A\mathbf{x} = \mathbf{x}$$

Since $B\,A\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbf{R}^n$, we may conclude that $BA = I_n$.

To show that B is unique, suppose that there is another $n \times n$ matrix C such that $AB = AC = I_n$. Then

$$B(AB) = B(AC) \Rightarrow (BA)B = (BA)C \Rightarrow B = C$$

because $BA = I_n$. Since $B = C$, it follows that B is unique. ■■

As a quick application of Theorem 3.23, for the matrices A and B in Example 1 we have

$$BA = \begin{bmatrix} 2 & -2 & 3 \\ -1 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 2 & 5 & -1 \\ 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

Since for an invertible matrix A there is exactly one matrix B such that $AB = BA = I_n$, the next definition makes sense.

DEFINITION 3.24 ▶

Inverse Matrix

If an $n \times n$ matrix A is invertible, then A^{-1} is called the **inverse** of A and denotes the unique $n \times n$ matrix such that $AA^{-1} = A^{-1}A = I_n$.

Nonsingular, Singular

A square matrix A that is invertible is also called **nonsingular**. If A does not have an inverse, it is **singular**. Definition 3.24 is symmetric in that if A^{-1} is the inverse of A, then A is the inverse of A^{-1}.

Our next theorem gives several important properties of invertible matrices. Note the distinction between (c) and (d) of Theorem 3.25 and those given in Theorem 3.14 in Section 3.2.

THEOREM 3.25 ▸

Let A and B be invertible $n \times n$ matrices and C and D be $n \times m$ matrices. Then

(a) A^{-1} is invertible, with $(A^{-1})^{-1} = A$.

(b) AB is invertible, with $(AB)^{-1} = B^{-1}A{-1}$.

(c) If $AC = AD$ then $C = D$.

(d) If $AC = 0_{nm}$, then $C = 0_{nm}$.

Proof We prove (a) and (b) here and leave (c) and (d) as an exercise. For (a), since A is invertible we know that $A^{-1}A = I_n$. This implies that A is the inverse of A^{-1}—that is, $A = (A^{-1})^{-1}$.

For (b), note that

$$(AB)(B^{-1}A^{-1})=A(BB^{-1})A^{-1}=AI_nA^{-1}=AA^{-1}=I_n$$

Hence AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. ■ ■

Finding A^{-1}

We now develop a method for computing A^{-1}. Let's start by supposing that A and B are $n \times n$ matrices with $AB = I_n$, where

$$A=[\mathbf{a}_1 \cdots \mathbf{a}_n],\quad B=[\mathbf{b}_1 \cdots \mathbf{b}_n],\quad \text{and} \quad I_n=[\mathbf{e}_1 \cdots \mathbf{e}_n]$$

(Later we will set $B = A^{-1}$, but for now using B simplifies notation.) If $AB = I_n$, then, taking multiplication one column at a time, we have

Ab1=e1,Ab2=e2,… ,Abn=en

Thus $\mathbf{b}_1$ is a solution to the linear system $A\mathbf{x} = \mathbf{e}_1$, $\mathbf{b}_2$ is a solution to the linear system $A\mathbf{x} = \mathbf{b}_2$, and so on. We could solve these systems one at a time by transforming to reduced echelon form each of the augmented matrices

[a1⋯an|e1],[a1⋯an|e2],⋯[a1⋯an|en]

However, since we do the same row operations for each matrix, we can save ourselves some work by setting up one large augmented matrix

[a1⋯an|e1⋯en]

and go through the row operations once to transform the left half from A to I_n. If this can be done, then the right half will be transformed from to I_n to $B = A^{-1}$. In brief, we want

[A|In]transformed to[In|A-1]

Let’s look at an example.

Example 2

Find the inverse of A=[1325].

Solution We begin by setting up the augmented matrix

[A|I2]=[1325|1001]

Now we use our usual row operations to transform the left half to I_2,

[1325|1001]-2R1R2→R2~[130-1|10-21] -R2→R2~[1301|102-1] -3R2+R1→R1~[1001|-532-1]

Thus we find that A-1=[-532-1]. Let's test it out:

AA-1=[1325][-532-1]=[1001]

and

A-1A=[-532-1][1325]=[1001]

For a given $n \times n$ matrix A, if A is invertible, then this procedure will find A^{-1}. If A is not invertible, then the reduced row echelon form of $[A \mid I_n]$ will not have I_n on the left side.

Example 3

Let $T(\mathbf{x}) = A\mathbf{x}$, where A=[1-21-37-62-30]. Find T^{-1}, if it exists.

Solution T^{-1} exists if and only if A^{-1} exists. To find A^{-1}, we set up the augmented matrix $[A \mid I_3]$ and then use row operations to transform to $[I_3 \mid A^{-1}]$:

[1-21-37-62-30|100010001]3R1+R2→R2-2R1+R3→R3~[1-2101-301-2|100310-201] -R2+R3→R3~[1-2101-3001|100310-5-11] 3R3+R2→R2-R3+R1→R1~[1-20010001|61-1-12-23-5-11] 2R2+R1→R1~[100010001|-18-35-12-23-5-11]

Thus we have $T^{-1}(\mathbf{x}) = A^{-1}\mathbf{x}$, where

A-1=[-18-35-12-23-5-11]

We can check our work by computing

AA-1=[1-21-37-62-30][-18-35-12-23-5-11]=[100010001]=I3

and

A-1A=[-18-35-12-23-5-11][1-21-37-62-30]=[100010001]=I3

The next example shows what happens when there is no inverse.

Example 4

Find the inverse, if it exists, of A=[11-221-3-3-14].

Solution We set up the augmented matrix $[A \mid I_3]$, and using row operations we get

[11-221-3-3-14|100010001]-2R1+R2→R23R1+R3→R3~[11-20-1102-2|100-210301] 2R2+R3→R3~[01-20-11000|100-210-121]

We can stop there. The left half of the bottom row consists entirely of zeros, so there is no way to use row operations to transform the left half of the augmented matrix to I_3. Thus A has no inverse.

The next theorem highlights how invertibility is related to the solutions of linear systems.

THEOREM 3.26 ▸

Let A be an $n \times n$ matrix. Then the following are equivalent:

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$, given by $\mathbf{x} = A^{-1}\mathbf{b}$.

(c) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Proof A is invertible if and only if $T(\mathbf{x}) = A\mathbf{x}$ is an invertible linear transformation, which in turn is true if and only if T is one-to-one and onto. This implies that (a) and (b) are equivalent. By setting $\mathbf{b} = \mathbf{0}$, we see that (b) implies (c), so all that remains to complete the proof is to show that (c) implies (a).

If $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then by Theorem 3.5 we know that T is one-to-one. Therefore, since A is $n \times n$, by Theorem 3.9 in Section 3.1, T also must be onto. Since T is one-to-one and onto, we can conclude that T is invertible, and so A is invertible. Therefore (c) implies (a).

Finally, we note that if A is invertible, then we have

$$A\mathbf{x} = \mathbf{b} \Rightarrow A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \Rightarrow I_n\mathbf{x} = A^{-1}\mathbf{b} \Rightarrow \mathbf{x} = A^{-1}\mathbf{b}$$

Example 5

Find the unique solution to the linear system

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

Solution Start by setting

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

Then our system is equivalent to $A\mathbf{x} = \mathbf{b}$, so that by Theorem 3.26, if A is invertible, then the solution is given by $\mathbf{x} = A^{-1}\mathbf{b}$. Happily, in Example 2 we found A^{-1}, which we use here to give us

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -5 & 3 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 4 \\ -3 \end{bmatrix} = \begin{bmatrix} -29 \\ 11 \end{bmatrix}$$

A Quick Formula

Typically, we use the row reduction method to compute inverses for matrices. However, in the case of a 2 × 2 matrix

$$A=[abcd]$$

there exists a quick formula. It can be shown that A has an inverse exactly when $ad - bc \neq 0$, and that the inverse is

$$A-1=1ad-ac[d-b-ca]$$

Instead of using an augmented matrix and row operations to verify the formula, we check it by multiplying,

$$A-1A=1ad-ac[d-b-ca] [abcd]=1ad-ac[ad-c00ad-bc]=[1001]=I2$$

Since $A^{-1}A = I_2$ and the inverse is unique, it must be that the formula for A^{-1} is correct.

Example 6

Use the Quick Formula to find A^{-1} for

$$A=[2715]$$

Solution From the Quick Formula, we have

$$A-1=1(2)(5)-(7)(1)[5-7-12]=[53-73-1323]$$

▶ This updates the Unifying Theorem, Version 2, from Section 3.1.

The Unifying Theorem, Version 3

Summarizing the results we have proved about invertible matrices, we add one more important condition to the Unifying Theorem.

THEOREM 3.27

(THE UNIFYING THEOREM, VERSION 3) Let $\mathcal{S} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ be a set of n vectors in $\mathbf{R}^n$, let $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, and let $T : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent:

(a) $\mathcal{S}$ spans $\mathbf{R}^n$.

(b) $\mathcal{S}$ is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

(d) T is onto.

(e) T is one-to-one.

(f) A is invertible.

Proof From the Unifying Theorem, Version 2, we know that (a) through (e) are equivalent. Moreover, from Theorem 3.26, we know that (c) and (f) are equivalent. Thus we can conclude that all six conditions are equivalent. ■ ■

The material on partitioned matrices is optional.

Partitioned Matrices

For matrices in certain special forms, it is possible to use matrix partitions to efficiently compute the inverse. For instance, suppose that

$$A = \begin{bmatrix} A_{11} & 0_{12} \\ 0_{21} & A_{22} \end{bmatrix}$$

where A_{11} is $n_1 \times n_1$, A_{22} is $n_2 \times n_2$, both A_{11} and A_{22} are invertible, and 0_{12} and 0_{21} represent matrices with all zero entries of dimension $n_1 \times n_2$ and $n_2 \times n_1$, respectively. A is an example of a **block diagonal** matrix.

To find the inverse of A, let

$$B=\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where the blocks of B have the appropriate dimensions so that AB is defined and can be computed using block multiplication. Now we assume that $AB = I$, and from this we determine the form required of B that will give us a formula for A^{-1}.

To start, write

$$I=\begin{bmatrix} I_{n_1} & 0_{12} \\ 0_{21} & I_{n_2} \end{bmatrix}$$

where I_{n_1} and I_{n_2} are the $n_1 \times n_1$ and $n_2 \times n_2$ identity matrices, respectively, and 0_{12} and 0_{21} are defined as above. If $AB = I$, then from our block multiplication formulas we must have

$$\begin{aligned} A_{11}B_{11}+0_{12}B_{21} &= I_{n_1} \\ A_{11}B_{12}+0_{12}B_{22} &= 0_{12} \\ 0_{21}B_{11}+A_{22}B_{21} &= 0_{21} \\ 0_{21}B_{12}+A_{22}B_{22} &= I_{n_2} \end{aligned}$$

Focusing on the first and fourth equations, we see that these imply that $B_{11}=A_{11}^{-1}$ and $B_{22}=A_{22}^{-1}$. Next, the second equation reduces to $A_{11}B_{12} = 0_{12}$, and since A_{11} is invertible it follows that $B_{12} = 0_{12}$. A similar argument can be used to show that $B_{21} = 0_{21}$, and so we can conclude that

$$A^{-1}=\begin{bmatrix} A_{11}^{-1} & 0_{12} \\ 0_{21} & A_{22}^{-1} \end{bmatrix} \tag{4}$$

We can stretch this method further to find an inverse if A has the form

$$A=\begin{bmatrix} A_{11} & 0_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

which is a **block lower triangular** matrix. Defining *B* as we did previously (with the assumptions required for *A* so that the inverse exists) and computing the product *AB* using block multiplication, we arrive at the equations

$$\begin{aligned} A_{11}B_{11} + 0_{12}B_{21} &= I_{n_1} \\ A_{11}B_{12} + 0_{12}B_{22} &= 0_{12} \\ A_{21}B_{11} + A_{22}B_{21} &= 0_{21} \\ A_{21}B_{12} + A_{22}B_{22} &= I_{n_2} \end{aligned}$$

Using the same line of reasoning as above, we find that $B_{11} = A_{11}^{-1}$, $B_{12} = 0_{12}$, $B_{21} = -A_{22}^{-1}A_{21}A_{11}^{-1}$, and $B_{22} = A_{22}^{-1}$. That is,

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix} \tag{5}$$

Example 7

Find the inverse for $A = \begin{bmatrix} 2 & 3 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 1 & -1 & -2 & 0 & 0 \\ 2 & -4 & 0 & 9 & 4 \\ 0 & 3 & 0 & 7 & 3 \end{bmatrix}$.

Solution All of the zeros in the upper right of *A* suggest that we partition *A* into the blocks

$$A_{11} = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix} \quad A_{12} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad A_{21} = \begin{bmatrix} 1 & -1 \\ 2 & -4 \\ 0 & 3 \end{bmatrix} \quad A_{22} \begin{bmatrix} -2 & 0 & 0 \\ 0 & 9 & 4 \\ 0 & 7 & 3 \end{bmatrix}$$

Since *A* is block lower triangular, we can apply the formula given in (5). To do so, we need A_{11}^{-1} and A_{22}^{-1}. As A_{11} is 2 × 2, we can apply our quick inverse formula to find

$$A_{11}^{-1} = \frac{1}{8-9}\begin{bmatrix} 4 & -3 \\ -3 & 2 \end{bmatrix} = \begin{bmatrix} -4 & 3 \\ 3 & -2 \end{bmatrix}$$

The block A_{22} is itself block diagonal, so we can apply 4 to find A_{22}^{-1}. The upper-left entry has inverse −1/2, and the quick inverse formula can be used to find the inverse for the lower-right sub-block. Combining these gives us

$$A_{22}^{-1} = \begin{bmatrix} -1/2 & 0 & 0 \\ 0 & -3 & 4 \\ 0 & 7 & -9 \end{bmatrix}$$

Finally, we compute

-A22-1A21A11-1=-[-1/2000-3407-9] [1-12-403] [-433-2]=[-7/25/2-9666221-152]

Combining all these ingredients together as specified in (5), we arrive at

A-1=[-430003-3000-7/25/2-1/200-96660-34221-15207-9]

The formulas illustrated here can readily be extended to other, more complex block matrices.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Use the Quick Formula to find the inverse of A=[21115].
2. Use the results from the previous problem to find the solution to

   2x1+11x2=5x1+5x2=2

3. Find the inverse of the linear transformation

   T([x1x2])=[3x1+2x25x1+3x2].

4. If it exists, find the inverse of the matrix

   A=[1-11210-20-1].

5. Determine if each statement is true or false, and justify your answer.
   - (a) If A is invertible and $\mathbf{x}$ is nonzero, then $A\mathbf{x}$ is nonzero.
   - (b) If A and B are invertible $n \times n$ matrices, then so is AB.
   - (c) If A is an upper triangular 3×3 matrix with nonzero entries on the diagonal, then A has an inverse.

(d) If A is singular, then the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions.

EXERCISES

Exercises 1–4: Use the Quick Formula to find A^{-1} if it exists.

1. A=[7321]

2. A=[5-2-43]

3. A=[2-5-410]

4. A=[-625-1]

Exercises 5–16: Use an augmented matrix and row operations to find A^{-1} if it exists.

5. A=[1429]

6. A=[41313]

7. A=[101010111]

8. A=[011010101]

9. A=[12-1013001]

10. A=[12-1-4-77-1-15]

11. A=[1-312-54-23-8]

12. A=[3-191-142-210]

13. A=[0010100000010100]

14. A=[100-201-100-230200-3]

15. A=[131-401-2200110001]

16. A=[1-31-22-54-2-39-254-124-7]

17. Use the answer to Exercise 6 to find the solutions to the linear system

4x1+13x2=-3x1+3x2=2

18. Use the answer to Exercise 10 to find the solutions to the linear system x1+2x-x3=-2-4x1-7x2+7x3=1-x1-x2+5x3=-1

19. Use the answer to Exercise 12 to find the solutions to the linear system

3x1-x2+9x3=4x1-x2+4x3=-12x1-2x2+10x3=3

20. Use the answer to Exercise 14 to find the solutions to the linear system

x1 -2x4=-1 x2-x3 =-2 -2x2+3x3 =22x1 -3x4=-1

Exercises 21–26: Determine if the given linear transformation T is invertible, and if so find T^{-1}. (HINT: Start by finding the matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.)

21. T([x1x2])=[4x1+3x23x1+2x2]

22. T([x1x2])=[2x1-5x2-x1+4x2x1+x2]

23. T([x1x2])=[x1-5x2-2x1+10x2]

24. T([x1x2x3])=[x1+x3x2+x3x1-x2+x3]

25. T([x1x2x3])=[x1+2x2-x3x1+x2-x3]

26. T([x1x2x3])=[x1+x2-x3x2-x3x1-x2+x3]

27. Let T_1 and T_2 be linear transformations given by

T1([x1x2])=[2x1+x2x1+x2]T([x1x2])=[3x1+2x2x1+x2]

Find the matrix A such that

(a) T1-1(T2(x))=Ax

(b) T1(T2-1(x))=Ax

(c) T2-1(T1(x))=Ax

(d) T2(T1-1(x))=Ax

28. Let T_1 and T_2 be linear transformations given by

T1([x1x2])=[3x1+5x24x1+7x2]T2([x1x2])=[2x1+9x2x1+5x2]

Find the matrix A such that

(a) T1-1(T2(x))=Ax

(b) T1(T2-1(x))=Ax

(c) T2-1(T1(x))=Ax

(d) T2(T1-1(x))=Ax

Exercises 29–34: Use an appropriate partitioning of the matrix A to find A^{-1}.

29. A=[100027014]

30. A=[520210001]

31. A=[2500380000140013]

32. A=[13003800-12254313]

33. A=[1300038000-1212-2430101-2001]

34. A=[720004100013100-2301-25-203-5]

FIND AN EXAMPLE Exercises 35–42: Find an example that meets the given specifications.

35. A diagonal 3×3 invertible matrix A.

36. A singular 3×3 matrix A that has no zero entries.

37. 2×2 matrices A and B such that $AB = 3I_2$.

38. 3×3 matrices A and B such that $BA = -2I_3$.

39. A 2×3 matrix A and a 3×2 matrix B such that $AB = I_2$ but $BA \neq I_3$.

40. A 3×4 matrix A and a 4×3 matrix B such that $AB = I_3$ but $BA \neq I_4$.

41. A lower triangular 4×4 matrix that is singular.

42. A 2×4 matrix B such that BB^T is nonsingular.

TRUE OR FALSE Exercises 43–50: Determine if the statement is true or false, and justify your answer.

43.

(a) If A is an invertible $n \times n$ matrix, then the number of solutions to $A\mathbf{x} = \mathbf{b}$ depends on the vector $\mathbf{b}$ in $\mathbf{R}^n$.

(b) A must be a square matrix to be invertible.

44.

(a) If an $n \times n$ matrix A is equivalent to I_n, then A^{-1} is also equivalent to I_n.

(b) If an $n \times n$ matrix A is singular, then the columns of A must be linearly independent.

45.

(a) The Caesar cipher encoding system is an example of a linear transformation.

(b) If B is invertible, then $(B^T)^{-1} = (B^{-1})^T$.

46.

(a) If the columns of an $n \times n$ matrix A span $\mathbf{R}^n$, then A is singular.

(b) If B is invertible, then AB-1=AB.

47.

(a) If A and B are invertible $n \times n$ matrices, then the inverse of AB is $B^{-1}A^{-1}$.

(b) A is invertible if and only if any echelon form of A has a pivot position in every row.

48.

(a) If A and B are invertible $n \times n$ matrices, then the inverse of $A + B$ is $A^{-1} + B^{-1}$.

(b) If $n \times n$ matrix B has rows that span $\mathbf{R}^n$, then B is invertible.

49.

(a) If A is invertible, then $(A^{-1})^{-1} = A$.

(b) The composition of two invertible linear transformations is an invertible linear transformation.

50.

(a) If $AB = 2I_3$, then $BA = 2I_3$.

(b) If A has linearly dependent rows, then A is invertible.

Exercises 51–54: Solve for the matrix X. Assume that all matrices are $n \times n$ and invertible as needed.

51. $AX = B$

52. $BX = A + CX$

53. $B(X + A)^{-1} = C$

54. $AX(D + BX)^{-1} = C$

55. Find all 2×2 matrices A such that $A^{-1} = A$.

56. Suppose that A is a square matrix with two equal rows. Is A invertible? Justify your answer.

57. Suppose that A is a square matrix with two equal columns. Is A invertible? Justify your answer.

58. Let A=[a00d] be a 2 × 2 diagonal matrix. For what values of a and d will A be invertible?

59. For what values of c will A=[11cc2] be invertible?

60. Suppose that A-1=[4-6214]. Find $(2A)^{-1}$.

61. Let A be an $n \times n$ matrix and $\mathbf{b}$ be in $\mathbf{R}^n$. If $A\mathbf{x} = \mathbf{b}$ has a unique solution, show that A must be invertible.

62. Suppose that $A = PDP^{-1}$, where all matrices are square. Find an expression for each of A^2 and A^3, and then give a general formula for A^n.

63. Suppose that A, B, and C are $n \times n$ invertible matrices. Solve $AC = CB$ for B.

64. Suppose that A is an invertible $n \times n$ matrix and that X and B are $n \times m$ matrices. If $AX = B$, prove that $X = A^{-1}B$.

65. Suppose that A is an invertible $m \times m$ matrix and that B and C are $n \times m$ matrices. If $(B - C)A = 0_{nm}$, prove that $B = C$.

66. Let A and B be $n \times n$ matrices, and suppose that B and AB are both invertible. Prove that A is also invertible.

67. Let A and B be $n \times n$ matrices. Prove that if B is singular, then so is AB.

68. Let A and B be $n \times n$ matrices. Prove that if A is singular, then so is AB.

69. Complete the proof of Theorem 3.21: If T is an invertible linear transformation, then $T^{-1}(r\mathbf{x}) = rT^{-1}(\mathbf{x})$.

70. Complete the proof of Theorem 3.25: Let A be an invertible $n \times n$ matrix and C and D be $n \times m$ matrices.

(a) If $AC = AD$, then $C = D$.

(b) If $AC = 0_{nm}$, then $C = 0_{nm}$.

Exercises 71–74: Refer to the smart phone scenario given at the beginning of the Section 3.1, and suppose that $\mathbf{j}_9$ corresponds to a new smart phone that costs (per phone) \$81 for labor, \$113 for materials, and \$38 for overhead. Suppose T is the linear transformation that takes as input a vector of unit counts for j8's, j8+'s, and j9's (in that order), and produces for output a vector of total labor, material, and overhead (again in order). Find a formula for T, then determine T^{-1} and use it to find the production level for each type of phone that will result in the given costs.

71. Labor = \$2150, Materials = \$3114, and Overhead = \$1027.

72. Labor = \$2152, Materials = \$3228, and Overhead = \$1047.

73. Labor = \$2946, Materials = \$4254, and Overhead = \$1404.

74. Labor = \$5062, Materials = \$7302, and Overhead = \$2413.

Exercises 75–78: Suppose that you are in the garden supply business. Naturally, one of the things that you sell is fertilizer. You have three brands available: Vigoro and Parker's as introduced in Section 2.1, and a third brand, Bleyer's SuperRich. The amount of nitrogen, phosphoric acid, and potash per 100 pounds for each brand is given by the nutrient vectors

V=[2934]p=[18256]b=[50199]VigoroParker'sBleyer's

Determine the linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ that takes a vector of brand amounts (in hundreds of pounds) as input and gives the nutrient vector as output. Then find a formula for T^{-1}, and use it to determine the amount of Vigoro, Parker's, and Bleyer's required to produce the specified nutrient mix.

75. 409 pounds of nitrogen, 204 pounds of phosphoric acid, and 81 pounds of potash

76. 439 pounds of nitrogen, 147 pounds of phosphoric acid, and 76 pounds of potash

77. 1092 pounds of nitrogen, 589 pounds of phosphoric acid, and 223 pounds of potash

78. 744 pounds of nitrogen, 428 pounds of phosphoric acid, and 156 pounds of potash

Exercises 79–82: The given set of vectors are the encoded version of a short message. Decode the message, given that the encoding matrix is

A=[1-322-79-414-17]

79. {[41161-306],[779-142]}

80. {[-30-4596],[122-39]}

81. {[775-136],[-25-3779],[47158-303]}

82. {[44157-300],[-46-405211],[2388-167],[40152-289]}

Exercises 83–86: Find A^{-1} if it exists.

83. A=[31-202251-30-224123]

84. A=[52-102-31421-3235-2-4]

85. A=[51212-31110231015-1-1-1300321]

86. A=[22904955212300595070499 51]

3.4 LU Factorization

In this section we revisit the problem of finding solutions to a system of linear equations, but we develop a new approach that can be more efficient in certain situations.

▶ This section is optional. However, LU factorizations are revisited in optional Section 11.2.

Figure 1 shows a diagram of three one-way streets that intersect. The arrows indicate the direction of traffic flow, and the numbers shown give the number of cars per minute passing along that stretch of road at a particular time. Our goal is to use this information to find x_1, x_2, and x_3.

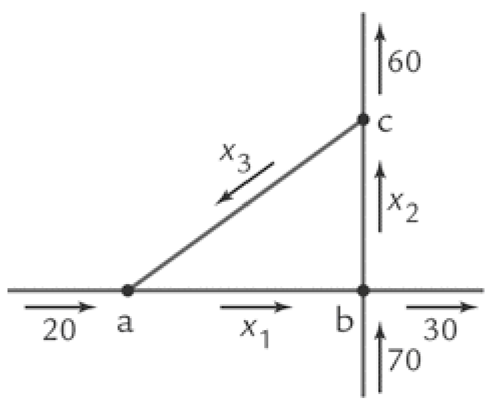


Figure 1 Traffic flow rates. (See Section 1.3 for additional discussion about this type of problem.)

The number of cars entering and exiting an intersection must be equal, so that for our three intersections a, b, and c we have the following equations:

$$a: 20+x_3=x_1 \quad b: 70+x_1=30+x_2 \quad c: x_2=60+x_3$$

In addition, there is a safety metering system in place that constrains $x_1 + x_2 = 160$. Combining these four equations into a linear system gives us

x1+x2 =160x1 -x3=20x1-x2 =-40 x2-x3=60 (1)

Example 1

Find the solution to the linear system 1.

Solution This system can be expressed in matrix form $A\mathbf{x} = \mathbf{b}$, where

A=[11010-11-1001-1]andb=[16020-4060]

Here we use a new approach to find the solutions to this system. It can be shown (we will see how later) that

A=[1000110012100-1-11] [1100-1-1002000]

Denoting the left matrix L (for Lower triangular) and the right by U (for Upper triangular, which it nearly is), we have $A = LU$. Thus the system can be written

LUx=b

which looks more complicated but actually allows us to break the problem into two systems that can both be solved with back substitution.

The first step is to denote $\mathbf{y} = U\mathbf{x}$, so that our system can be expressed as $L\mathbf{y} = \mathbf{b}$, or

[1000110012100-1-11] [y1y2y3y4]=[16020-4060]

This is equivalent to the system

y1 =160y1+y2 =20y1+2y2+y3 =-40 -y2-y3+y4=60

To find the solution, we modify our usual back substitution method by starting at the top (the details are left to the reader) to find

$$y=[160-140800]$$

Now that we know **y**, we can return to the system **y** = *U***x**, or

$$[1100-1-1002000]\ [x1x2x3]=[160-140800]$$

This is equivalent to

$$x1+x2\ \ =160\ -x2-x3=-140\quad 2x3=80$$

This system can be solved using standard back substitution, which gives us x_1 = 60, x_2 = 100, and x_3 = 40.

Having *L* and *U* made it easier to find the solution, but usually *L* and *U* are not known. Below we present the method for finding *L* and *U*, but before we get to that, we pause to discuss whether doing so is worth the trouble.

The number of computations required to find *L* and *U* from *A*, then perform the two back substitutions, is about the same as simply reducing *A* to echelon form and back substituting. Therefore for a single system *A***x** = **b** there is no benefit to finding *L* and *U* first. However, some applications require the solutions to many systems that all have the same coefficient matrix,

$$Ax=b1,Ax=b2,Ax=b3,\cdots$$

For instance, the traffic rates in our example are likely to change from minute to minute, each time generating new systems with the same coefficient matrix. In such situations, once we have *L* and *U* we can use them over and over. Thus, from the second system on, all we have to do are the two back substitutions, which is much faster than solving each individual system from scratch.

Finding *L* and *U*

LU factorization

If $A = LU$, where U is in echelon form and L is lower triangular with 1's on the diagonal (making L *unit lower triangular*), then the product is called an **LU factorization** of A. In the following examples, we show how to find an LU factorization.

Example 2

Find an LU factorization for

A=[-31262-595-6]

Solution We obtain U by transforming A to echelon form, and build up L one column at a time as we transform A.

Step 1a: Take the first column of A, divide each entry by the pivot (−3), and use the resulting values to form the first column of L.

▶ The • symbol represents a matrix entry that has not yet been determined.

A=[-31262-595-6]⇒L=[1••-2••-3••]

Step 1b: Perform row operations on A as usual to introduce zeros down the first column of A.

Note: Do not scale rows by constants (i.e., dividing the first row by −3). Doing so will give incorrect results.

▶ It is possible to modify this procedure to allow row scaling, but it makes the algorithm more complicated. Since scaling is not necessary to reduce a matrix to echelon form, we leave it out.

A=[-31262-595-6]2R1+R2→R23R1+R3→R3~[-31204-1080]=A1

Step 2a: Take the second column of A_1, starting from the pivot entry (4) down (the entries are shown in boldface), and divide each entry by the pivot. Use the resulting values to form the lower portion of the second column of L.

$$A_1 = \begin{bmatrix} -3 & 1 & 2 \\ 0 & 4 & -1 \\ 0 & 8 & 0 \end{bmatrix} \Rightarrow L = \begin{bmatrix} 1 & \bullet & \bullet \\ -2 & 1 & \bullet \\ -3 & 2 & \bullet \end{bmatrix}$$

Step 2b: Perform row operations on A_1 to introduce zeros down the second column of A_1.

$$A_1 = \begin{bmatrix} -3 & 1 & 2 \\ 0 & 4 & -1 \\ 0 & 8 & 0 \end{bmatrix} \begin{matrix} \\ \\ -2R_2 + R_3 \rightarrow R_3 \end{matrix} \sim \begin{bmatrix} -3 & 1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{bmatrix} = A_2$$

Step 3: Now that the original matrix is in echelon form (indeed, it is upper triangular), two small items remain. First, set U equal to A_2, our echelon form of A. Second, finish filling in L. Since L must be unit lower triangular, we add a 1 in the lower right corner and fill in the remaining entries with 0's. Thus we end up with

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 2 & 1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} -3 & 1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{bmatrix}$$

as seen in the previous example. Standard matrix multiplication can be used to verify that $A = LU$.

Not all matrices have an LU factorization. The following theorem gives a condition that will insure such a factorization exists.

THEOREM 3.28 ▸

Let A be an $n \times m$ matrix. If A can be transformed to echelon form without interchanging rows, then there exists an $n \times n$ unit lower triangular matrix L and an $n \times m$ echelon matrix U such that $A = LU$.

Note that if *U* is a square matrix, then it is an echelon matrix and is also upper triangular. Later in this section, we explain why the algorithm for finding an LU factorization works and justify Theorem 3.28. For now let’s consider additional examples.

Example 3

Find an LU factorization for

A=[2104-4-35-1064-8172-329-9]

Solution We proceed in the same manner as in the previous example.

Step 1a: Divide by the pivot to form the first column of *L*.

A=[2104-4-35-1064-8172-329-9]⇒L=[1•••-2•••3•••1•••]

Step 1b: Perform row operations on *A* to introduce zeros down the first column.

A=[2104-4-35-1064-8172-329-9]2R1+R2→R2-3R1+R3→R3-R1+R4→R4~[21040-15-201-850-429-13]=A1

Step 2a: Divide by the pivot to form the second column of *L*.

A1=[21040-15-201-850-429-13]⇒L=[1•••-21••3-1••14••]

Step 2b: Perform row operations on A_1 to introduce zeros down the second column.

A1=[21040-15-201-850-429-13]R2+R3→R3-4R2+R4→R4~[21040-15-201-33009-5]=A2

Step 3a: Divide by the pivot to form the third column of *L*.

A2=[21040-15-200-32009-5]⇒L=[1•••-21••3-11•14-3•]

Step 3b: Perform row operations on A_2 to introduce zeros down the third column.

A2=[21040-15-200-33009-5]3R3+R4→R4~[21040-15-200-330004]=A3

Step 4: Finish up by placing in L a 1 in the bottom right entry and 0's elsewhere and setting $U = A_3$.

L=[1000-21003-11014-31]andU=[21040-15-200-330004]

Thus L is unit lower triangular, U is in echelon form (and upper triangular), and it is easily verified that $A = LU$.

Example 4

Find an LU factorization for

A=[21-1-4-256211]

Solution Since A has three rows, L will be a 3×3 matrix.
Step 1a: Divide by the pivot to form the first column of L:

A=[21-1-4-256211]⇒L=[1••-2••3••]

Step 1b: Perform row operations on A to introduce zeros down the first column:

A=[21-1-4-256211]2R1+R2→R2-3R1+R3→R3~[21-10030-114]=A1

At this point we can see from the second column of A_1 that there is no way to transform A to echelon form without interchanging the second and third rows. Our algorithm will not work, so we stop. There is no LU factorization.

The matrix A need not be square to have an LU factorization. In general, if A is $n \times m$, then L will be $n \times n$ and U will be $n \times m$. The next two examples consider nonsquare matrices.

Example 5

Find an LU factorization for

A=[4-3-152-16122-17-78-6-122210]

Solution Since A has three rows, L will be a 3×3 matrix. The solution method is similar to previous examples.
Step 1a: Divide by the pivot to form the first column of L.

A=[4-3-152-16122-17-78-6-122210]⇒L=[1••-4••2••]

Step 1b: Perform row operations on A to introduce zeros down the first column.

A=[4-3-152-16122-17-78-6-122210]4R1+R2→R2-
2R1+R3→R3~[4-3-15200-23100-10126]=A1

Step 2a: Since the second and third entries in the second column of A_1 are both zero, we bypass it and move to the third column. Divide by the pivot of the third column to form the next column of L.

A1=[4-3-15200-23100-10126]⇒L=[1••-41•25•]

Step 2b: Perform row operations on A_1 to introduce zeros down the third column.

A1=[4-3-15200-23100-10126]-5R2+R3→R3~[4-3-15200-231000-
31]=A2

Step 3: Our matrix is now in echelon form. We place a 1 in the lower-right position of L and 0's elsewhere and set $U = A_2$, giving us

L=[100-410251],U=[4-3-15200-231000-31]

Example 6

Find an LU factorization for

A=[3-149-51515-110-62-4-3-310]

Solution Since A has five rows, L will be a 5 × 5 matrix.
Step 1a: Divide by the pivot to form the first column of L.

A=[3-149-51515-110-62-4-3-310]⇒L=[1••••3••••5••••-2••••-1••••]

Step 1b: Perform row operations on A to introduce zeros down the first column.

A=[3-149-51515-110-62-4-3-310]-3R1+R2→R2-5R1+R3→R32R1+R4→R4R1+R5→R5~[3-140-2304-100040-414]=A1

Step 2a: Divide by the pivot to form the second column of L.

A1=[3-140-2304-100040-414]⇒L=[1••••31•••5-2•••-20•••-12•••]

Step 2b: Perform row operations on A_1 to introduce zeros down the second column.

A1=[3-140-2304-100040-414]2R2+R3→R3-2R2+R5→R5~[3-140-2300-4004008]=A2

Step 3a: Divide by the pivot to form the third column of L.

A2=[3-140-2300-4004008]⇒L=[1••••31•••5-21••-20-1••-12-2••]

Step 3b: Perform row operations on A_2 to introduce zeros down the third column.

A2=[3-140-2300-4004008]R3+R4→R42R3+R5→R5~[3-140-2300-4000000]=A3

Step 4: Since A_3 is in echelon form, we can set $U = A_3$. However, we still need the last two columns of L. Since the bottom two rows of U are all zeros, a moment's thought reveals that the product LU will be the same no matter which entries are in the last two columns of L. (See Exercise 50.) Because we need L to be unit lower triangular and in applications we do back substitution, the smart way to fill out the last two columns of L is with the last two columns of I_5 Thus we end up with

L=[10000310005-2100-20-110-12-201],U=[3-340-2300-4000000]

Regarding the LU factorization algorithm:

- The number of zero rows in U at the end of the algorithm is equal to the number of columns from I (taken from the right side) required to fill out the remainder of L (see Example 6).
- There can be more than one LU factorization for a given matrix.
- As seen in Example 4, if at any point in the algorithm a row interchange is required, stop. The matrix does not have an LU factorization.

LDU Factorization

A variant of LU factorization is called *LDU factorization*. Starting with a matrix A, now the goal is to write $A = LDU$, where L and U are as in LU factorization (except that U has 1's in the pivot positions) and D is a diagonal matrix with the same dimensions as L. To find the LDU factorization, we follow the same procedure as in finding the LU factorization, and then at the end form D and modify U.

Example 7

Find an LDU factorization for

A=[280418-4-2-2-13]

Solution We start by finding an LU factorization. Since *A* has three rows, *L* will be a 3 × 3 matrix.
Step 1a: Take the first column of *A*, divide each entry by the pivot, and use the resulting values to form the first column of *L*.

A=[280418-4-2-2-13]⇒L=[1••2••-1••]

Step 1b: Perform row operations on *A* as usual to introduce zeros down the first column.

A=[280418-4-2-2-13]-2R1+R2→R2R1+R3→R3~[28002-406-13]=A1

Step 2a: Take the second column of A_1, starting from the pivot entry (2), and divide each entry by the pivot. Use the resulting values to form the second column of *L*.

A1=[28002-406-13]⇒L=[1••21•-13•]

Step 2b: Perform row operations on A_1 to introduce zeros down the second column of A_1.

A1=[28002-406-13]-3R2+R3→R3~[28002-400-1]=A2

Step 3: The matrix A_2 is in echelon form. We have

L=[100210-131]andU=[28002-400-1]

Step 4a: We now have *A* = *LU*. The diagonal matrix *D* has entries taken from the pivots of *U*.

U=[28002-400-1]⇒D=[20002000-1]

Step 4b: We modify *U* by dividing each row by the row pivot, leaving 1’s in the pivot positions.

(old) U=[28002-400-1]⇒(new) U=[14001-2001]

It is not hard to check that

[28002-400-1]=[20002000-1] [14001-2001]

Setting *U* = (new) *U*, we have *A* = *LDU*, where

L=[100210-131],D=[20002000-1],U=[14001-2000]

Example 8

Find an LDU factorization for the matrix in Example 5,

A=[4-3-152-16122-17-78-6-122210]

Solution In Example 5 we found that *A* = *LU*, where

L=[100-410251],U=[4-3-15200-231000-31]

All that is left is to form the 3 × 3 matrix *D* and modify *U*.
Step 1a: Form the 3 × 3 diagonal matrix *D*, with the diagonal terms coming from the values in the pivot positions of *U* (in boldface).

U=[4-3-15200-231000-31]⇒D=[4000-2000-3]

Step 1b: Modify U by dividing each row by its pivot, leaving 1’s in the pivot positions.

(old) U=[4-3-15200-231000-31]⇒U=[1-3/4-1/45/41/2001-3/2-1/20001-1/3]

By forming *D* and modifying *U* in this manner, it follows that the product *DU* is equal to the matrix *U* found earlier. Therefore we have *A* = *LDU*, where

L=[100-410251],D=[4000-2000-3],U=[1-3/4-1/45/41/2001-3/2-1/20001-1/3]

Example 9

Find an LDU factorization for the matrix in Example 6,

A=[3-149-51515-110-62-4-3-310]

Solution In Example 6 we found that *A* = *LU*, where

L=[10000310005-2100-20-110-12-201],U=[3-140-2300-4000000]

The last steps are to form the 5 × 5 matrix *D* and modify *U*.

Step 1a: To find the terms of *D*, we start as in Example 8, setting the diagonal terms in the first three columns equal to the pivots of *U*. As with *L*, we have *A* = *LDU* regardless of which entries are in the last two diagonal positions of *D*. In applications it is convenient to have *D* invertible, so we put 1’s in the last two diagonal positions.

D=[300000-200000-4000001000001]

Step 1b: As in the previous example, we modify *U* by dividing each nonzero row by its pivot. Doing so, we have *A* = *LDU*, where

L=[10000310005-2100-20-110-12-201],D=[300000-200000-4000001000001],U=[0-1/34/301-3/2001000000]

Why Does the LU Algorithm Work?

As illustrated in our examples, a central part of the LU factorization algorithm is using elementary row operations to reduce the original matrix to echelon form.

For example, for the following matrix A, the first step toward echelon form in the LU factorization algorithm is

A=[31-126-201401-1]-2R1+R2→R2∼[31-120-42-3401-1]=A1

We can perform the same operation by left-multiplying by the corresponding elementary matrix

E1A=[100-210001] [31-126-201401-1]=[31-120-42-3401-1]=A1

The LU factorization algorithm involves a number of row operations to transform the original matrix A to echelon form, producing the matrix U. Let $E_1, E_2, \ldots, E_k$ denote the elementary matrices corresponding to the row operations, in the order performed. Then we have

(Ek⋯E2E1)A=U (1)

In our algorithm, the only type of row operation that we perform involves adding a multiple of one row to a row below. The elementary matrix corresponding to such a row operation will be unit lower triangular. Thus each of $E_1, E_2, \ldots, E_k$ is unit lower triangular, and it follows that

(Ek⋯E2E1)and(Ek⋯E2E1)-1

also are both unit lower triangular (see Exercise 52).

Returning to (2) and solving for A, we have

A=(Ek⋯E2E1)-1U

Thus, if we let $L = E_k \cdots E_2E_1)^{-1}$, then $A = LU$ where L is unit lower triangular and U is in echelon form. This shows that an LU factorization is possible. To see why the construction of L in the algorithm works, note that

$$(E_k \cdots E_2 E_1)L = I$$

Comparing to (2) we see that the same sequence of row operations that transform A into U will also transform L into I. This provides the rationale for the construction method for L—we set the columns of L so that the row operations that transform A to U will also transform L to I.

COMPUTATIONAL COMMENTS

- For many systems, interchanging rows is necessary when reducing to echelon form, especially when partial pivoting is being used to minimize round-off error. It is possible to modify the LU factorization algorithm given here to accommodate row swaps, although one ends up with a matrix L that is *permuted lower triangular*, meaning that the rows can be reorganized to form a lower triangular matrix.
- When presented with the problem of solving many systems of the form

  $$A\mathbf{x} = \mathbf{b}_1, A\mathbf{x} = \mathbf{b}_2, A\mathbf{x} = \mathbf{b}_3, \cdots$$

  where A is a square invertible matrix, it is tempting to compute A^{-1} and then use this to find the solution to each system. However, for an $n \times n$ matrix A, it takes about $2n^3/3$ flops to find L and U as opposed to $2n^3$ flops to find A^{-1}. In addition, if A is a sparse matrix (that is, most of the entries are 0), then L and U typically will be sparse, while A^{-1} will not. In this instance, using the LU factorization usually will be more efficient than using A^{-1}.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Use the given LU factorization to find all solutions to $A\mathbf{x} = \mathbf{b}$.
   (a) A=[1031] [-2501],b=[13]
   (b) A=[100310-211] [31-102-2022],b=[-121]
2. Find an LU factorization for each matrix.
   (a) A=[21-1440-2160]
   (b) A=[213-1010223242122]
   (c) A=[210-2223-1-4078]
   (d) A=[212-2-2-1427209]
3. Find an LDU factorization for each matrix in the previous exercise.

EXERCISES

Exercises 1–4: A few terms are missing from the given LU factorization. Find them.

1. [10-71] [22-301-4]=[2a-3b-1317]
2. [100-410-2-11] [-203011002]=[-2a381bc-1-5]
3. [100310a21] [520b00]=[5c1592014]
4. [10a1] [423b0231]=[4c3186d3]

Exercises 5–12: Use the given LU factorization to find all solutions to $A\mathbf{x} = \mathbf{b}$.

5. A=[10-21] [2-203],b=[22]
6. A=[1031] [140-2],b=[-7-17]
7. A=[100-1102-21] [2-1301200-2],b=[40-4]
8. A=[100-210131] [1-2003-100-2],b=[-4115]

9. A=[100210-341] [1-20100],b=[014]
10. A=[1031] [1-120-2-1],b=[213]
11. A=[1000-210003102-101] [1-20-10113001-10001],b=[00-10]
12. A=[10002100-1310-3001] [131012001000],b=[-1-3-23]

Exercises 13–24: Find an LU factorization for *A*.

13. A=[1-4-29]
14. A=[2-3-610]
15. A=[-2-11-6042-2-1]
16. A=[-321-6230-86]
17. A=[-10-12132-2-2-9-33-19-25]
18. A=[-321402036-6-1-6-62-1-9]
19. A=[-12134-7-7-17-26-3-2]
20. A=[-20-11341-12-5-23-8145]
21. A=[11010-11-1001-1]
22. A=[-1244-6-17-32152-4-9]
23. A=[-213208-411220-10-427]
24. A=[132-1-5-10-6-3157-393]

Exercises 25–30: Find an LDU factorization for *A* given in the referenced exercise.

25. Exercise 5.
26. Exercise 8.
27. Exercise 10.
28. Exercise 13.
29. Exercise 15.
30. Exercise 19.

Exercises 31–36: If *L* and *U* are invertible, then $(LU)^{-1} = U^{-1}L^{-1}$. Find A^{-1} from the LU factorization for *A* given in the referenced exercise.

31. Exercise 5.

32. Exercise 8.

33. Exercise 11.

34. Exercise 13.

35. Exercise 16.

36. Exercise 17.

FIND AN EXAMPLE Exercises 37–42: Find an example that meets the given specifications.

37. A matrix A that has an LU factorization where L is 4×4 and U is 4×3.

38. A matrix A that has an LU factorization where L is 3×3 and U is 3×6.

39. A 2×2 matrix A that has an LU factorization where U is diagonal.

40. A 3×2 matrix A that has an LU factorization where L is diagonal.

41. A 4×4 matrix A that has an LU factorization where L and U are both diagonal.

42. A 3×4 matrix A that has an LDU factorization where $D = I_2$.

TRUE OR FALSE Exercises 43–48: Determine if the statement is true or false, and justify your answer.

43.

(a) The dimensions of L and U are the same in an LU factorization.

(b) The L matrix is square in an LU factorization.

44.

(a) A matrix has an LU factorization provided that it can be transformed to echelon form without the use of row interchanges.

(b) The U matrix in an LU factorization has 1's along the diagonal.

45.

(a) If an $n \times n$ matrix A is lower triangular, then A has an LU factorization.

(b) If A is 4×3, then so is L in an LU factorization of A.

46.

(a) If A is a nonsingular $n \times n$ matrix, then A has an LU factorization.

(b) Every elementary matrix is nonsingular.

47.

(a) The LU factorization for a given matrix is unique.

(b) The L matrix for the LU factorization of a lower triangular matrix is the identity.

48.

(a) The L matrix in an LU factorization satisfies $L^2 = I_n$.

(b) The U matrix in an LU factorization is nonsingular.

49. Let U be an $n \times n$ matrix that is in echelon form, and let D be an $n \times n$ diagonal matrix. Prove that the product DU is equal to the matrix U with each row multiplied by the corresponding diagonal entry of D. (That is, the first row of U multiplied by d_{11}, the second row of U multiplied by d_{22}, and so on.)

50. Let U be an $n \times m$ matrix in echelon form, with the bottom k rows of U having all zero entries. Suppose that L_1 and L_2 are both $n \times n$ matrices and that the leftmost $n - k$ columns of both are identical. Prove that $L_1U = L_2U$.

51. Prove or disprove: If A is an $n \times n$ upper triangular matrix, then we can have $L = I_n$ in an LU factorization of A.

52. Let $L_1, L_2, \dots, L_k$ be unit lower triangular matrices.

(a) Prove that L_1L_2 is unit lower triangular.

(b) Use induction to prove that $L_k \cdots L_2L_1$ is unit lower triangular.

(c) Prove that Li-1 is unit lower triangular for $i = 1, \dots, k$.

(d) Prove that $(L_k \cdots L_2L_1)^{-1}$ is unit lower triangular.

53. In graph theory, an *adjacency matrix A* has an entry of 1 at a_{ij} if there is an edge connecting node i with node j, and a zero otherwise. (Such matrices arise in network analysis.) Figure 2 shows a *cyclic graph* with six nodes.

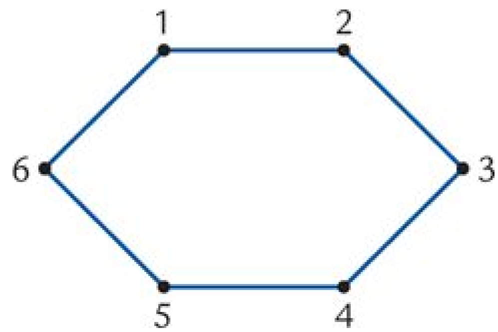


Figure 2 A cyclic graph with six nodes.

The adjacency matrix for this graph is

$$A=[010001101000010100001010000101100010]$$

If possible find an LU factorization for the matrix *A*.

54. *Band matrices* arise in numerous applications, such as finite difference problems in engineering. Such matrices have their nonzero entries clustered along the diagonal, as with

$$A=[12 \quad 311 \quad 221 \quad 121 \quad 213 \quad 141 \quad 32]$$

The missing entries are zero. Find an LU factorization for the matrix *A*.

Exercises 55–56: Find an LU factorization for *A* if one exists. Note that some computer algorithms do not compute LU factorizations in the same manner as presented here, so use caution.

55. $A=[1020\text{-}4251\text{-}145221020\text{-}92153\text{-}4\text{-}3\text{-}7]$

56. $A=[\text{-}15\text{-}321\text{-}103295\text{-}1354\text{-}26]$

3.5 Markov Chains

▶ This section is optional and can be omitted without loss of continuity.

Sequences of vectors arise naturally in certain applications. In those considered here, each vector in the sequence is found by multiplying a special matrix times the preceding vector. To illustrate this, we begin by recalling the following scenario, first introduced in Example 8, Section 3.2.

Example 1

In a small town there are 10,000 homes. When it comes to television viewing, the residents have three choices: they can subscribe to cable, they can pay for satellite service, or they watch no TV. (The town is sufficiently remote so that an antenna does not work.) In any given year, 80% of the cable customers stick with cable, 10% switch to satellite, and 10% get totally disgusted and quit watching TV. Over the same time period, 90% of satellite viewers continue with satellite service, 5% switch to cable, and 5% quit watching TV. And of those people who start the year not watching TV, 85% continue not watching, 5% subscribe to cable, and 10% get satellite service (see Figure 1). If the current distribution is 6000 homes with cable, 2500 with satellite service, and 1500 with no TV, how many of each type will there be a year from now? How about two years from now? Three years from now?

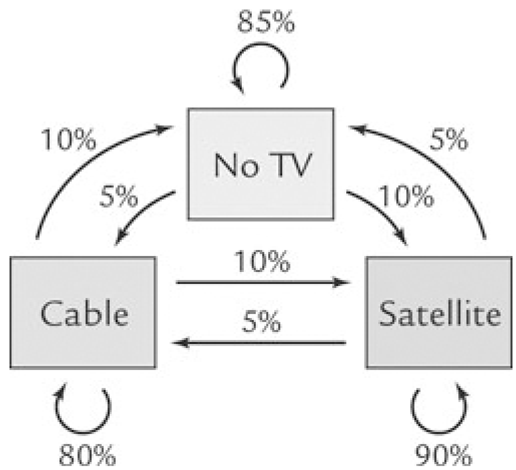


Figure 1 Customer transition percentages among cable, satellite, and no television.

To answer the questions posed, we started by forming the matrix

$$A=\begin{bmatrix}0.80 & 0.05 & 0.05\\ 0.10 & 0.90 & 0.10\\ 0.10 & 0.05 & 0.85\end{bmatrix}$$

Stochastic Matrix, Doubly Stochastic Matrix

A square matrix like *A* that has nonnegative entries and columns that each add to 1 is called a **stochastic matrix**. If a stochastic matrix also has rows that add to 1, then it is a **doubly stochastic matrix**.

Previously, we formed the vector **x** giving the initial number of homes with each of satellite, cable, and no television. This time, we start with the vector that gives the initial *proportion* of homes of each type,

$$x_0=\begin{bmatrix}0.60\\ 0.25\\ 0.15\end{bmatrix}\begin{array}{l}\text{60\% of homes have cable}\\ \text{25\% of homes have satelite}\\ \text{15\% of homes have no television}\end{array}$$

Initial State Vector

Probability Vector

Here $\mathbf{x}_0$ denotes the **initial state vector**. Each entry in $\mathbf{x}_0$ can be thought of as representing the probability that a household falls into one of the television-watching groups. In general, any vector that has nonnegative entries that add up to 1 is called a **probability vector**. Thus a stochastic matrix consists of columns that are probability vectors.

After one year has passed, the distribution of households changes. The new distribution is given by

▶ *A* is called a **transition matrix** because it is used to make the transition from one state vector to the next.

x1=Ax0=[0.800.050.050.100.900.100.100.050.85] [0.600.250.15]= [0.500.300.20]

That is, after one year, 50% have cable, 30% have satellite, and 20% have no television. Two and three years later we have, respectively,

x2=Ax1=[0.4250.3400.235]andx3=Ax2=[0.36880.37200.2592]

State Vector

Each vector in a sequence generated in this manner is called a **state vector** and is found by multiplying the stochastic matrix by the previous state vector. This process can be continued indefinitely—additional vectors are given below. (Some intermediate vectors are not shown to save space.)

▶ One can also compute state vectors using the formula $\mathbf{x}_1 = A^i x_o$ (as in Example 8, Section 3.2). However, calculating matrix powers A^2, A^3, … is more computationally intensive than using the recursive definition $\mathbf{x}_{i+1} = A\mathbf{x}_i$.

x4=[0.32660.39760.2758],x5=[0.29490.41810.2770],x6=[0.27120.43450.2943],x7=[0.25340.44760.2990],x8=[0.24000.45810.3019]x10=[0.22250.47320.3043],x12=[0.21270.48280.3045],x14=[0.20710.48900.3039],x16=[0.20400.49600.3030],x20=[0.20130.49170.3019]x24=[0.20040.49880.3008],x28=[0.20010.49950.3004],x32=

[0.20000.49980.3002],x36=[0.20000.49990.3001],x40=
[0.20000.50000.3000]

Markov Chain

In general, a sequence of vectors $\mathbf{x}_0$, $\mathbf{x}_1$, ⋯ generated in this way is called a **Markov Chain**.

Examining the state vectors $\mathbf{x}_0$, $\mathbf{x}_1$, ⋯, $\mathbf{x}_{40}$ above, we see that our sequence appears to be converging toward the vector

- Informally, by *converging* we mean that the entries in the sequence of vectors are getting closer and closer to fixed values.

x=[0.20.50.3] (1)

Since $\mathbf{x}_{i+1} = A\mathbf{x}_i$, if for large i we have $\mathbf{x}_{i+} \approx \mathbf{x}_i$, then this implies that $A\mathbf{x}_i \approx \mathbf{x}_i$. Hence the vector $\mathbf{x}$ that we are looking for should satisfy $A\mathbf{x} = \mathbf{x}$. Let's try this out for our A and $\mathbf{x}$ in 1.

Ax=[0.800.050.050.100.900.100.100.050.85] [0.20.50.2]=
[0.20.50.3]=x

Steady-State Vector

The vector $\mathbf{x}$ is called a **steady-state vector** for A.

Example 2

The computer support group at a large corporation maintains thousands of machines using a well-known operating system. As the computers age, some tend to become less reliable. Based on past records, if a given computer crashes this week, then there is a 92% chance that it will crash again next week. On the other hand, if a computer did not crash this week, then there is a 60% chance that it will not crash next week (see Figure 2). Suppose that 70% of computers crashed this week. What percentage will crash two weeks from now? Is there a steady-state vector? If so, what is it?

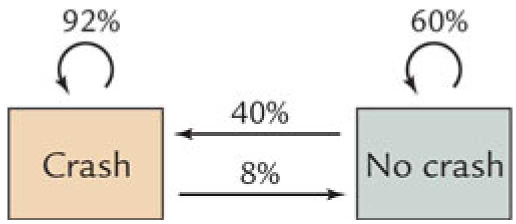


Figure 2 Transition percentages between crashing and noncrashing computers.

Solution The information given is summarized in the table and corresponding transition matrix *A* below.

		This Week	
		Crash	No Crash
Next Week	Crash	92%	40%
	No Crash	8%	60%

$$\Rightarrow A=\begin{bmatrix}0.92 & 0.40\\ 0.08 & 0.60\end{bmatrix}$$

This week 70% of computers crashed and 30% did not, giving an initial state vector of

$$x_0=\begin{bmatrix}0.70\\ 0.30\end{bmatrix}$$

The next two vectors in the sequence are

$$x_1=Ax_0=\begin{bmatrix}0.92 & 0.40\\ 0.08 & 0.60\end{bmatrix}\begin{bmatrix}0.70\\ 0.30\end{bmatrix}=\begin{bmatrix}0.764\\ 0.236\end{bmatrix}$$

and

$$x_2=Ax_1=\begin{bmatrix}0.92 & 0.40\\ 0.08 & 0.60\end{bmatrix}\begin{bmatrix}0.764\\ 0.236\end{bmatrix}=\begin{bmatrix}0.7973\\ 0.2027\end{bmatrix}$$

From $\mathbf{x}_2$ we see that two weeks from now we can expect 79.73% of computers to crash. Calculating more state vectors in the sequence gives us x3=[0.81460.1854],x4=[0.82360.1764],x5=[0.82830.1717],x6=[0.83070.1693]x7=[0.83200.1680],x8=[0.83260.1674],x9=[0.83300.1670],x10=[0.83310.1669]x11=[0.83320.1668],x12=[0.83330.1667],x13=[0.83330.1667],x14=[0.83330.1667]

This suggests a steady-state vector x=[5616]. Let’s test it out:

Ax=[0.920.400.080.60] [5616]=[5616]

This confirms our observation.

In practice, it may take many terms in the sequence for a steady-state vector to emerge, so computing lots of state vectors is usually not a practical way to find a steady-state vector. Fortunately, there is a direct algebraic method that we can use.

Finding Steady-State Vectors

We know that a steady-state vector **x** satisfies $A\mathbf{x} = \mathbf{x}$. Since $\mathbf{x} = I\mathbf{x}$, we have

Ax=x⇒Ax=Ix⇒Ax-Ix=0⇒(A-I)x=0

Thus a steady-state vector for A will satisfy the homogeneous system with coefficient matrix $A \times I$. For the matrix A in Example 2, we have

A=I=[-0.280.400.08-0.40]

The homogeneous system $(A - I)\mathbf{x} = \mathbf{0}$ has augmented matrix

[-0.080.4000.08-0.400]~[-.0.080.400000]

This leaves the single equation $-0.08\,x_1 + 0.40x_2 = 0$. Setting $x_2 = s$ and back-substituting gives the general solution

x=s[51]

Since **x** is to be a probability vector, the entries need to add to 1. Setting s=15+1=16 gives

x=16[52]=[5616]

which is the steady-state vector found in Example 2.

Example 3

Find a steady-state vector for the matrix *A* given in Example 1.

A=[0.800.050.050.100.900.100.100.050.85]

Solution The augmented matrix for the system ($A - I$)**x** = **0** and corresponding echelon form are

[-0.200.050.0500.10-0.100.1000.100.05-0.150]~[-0.200.050.0500-0.75-0.12500000]

Back substitution yields the general solution

x=s[0.20.50.3]

Setting *s* × 1 gives the steady-state vector and matches the vector in (1) found in Example 1.

Properties of Stochastic Matrices

The next theorem summarizes some properties of stochastic matrices. Proofs are left as exercises.

THEOREM 3.29 ▶

Let A be an $n \times n$ stochastic matrix and $\mathbf{x}_0$ a probability vector. Then

(a) If $\mathbf{x}_{i+1} = A\mathbf{x}_i$ for $i = 0, 1, 2, \dots$, then each of $\mathbf{x}_\mathrm{i}, \mathbf{x}_2, \dots$ in the Markov chain is a probability vector.

(b) If B is another $n \times n$ stochastic matrix, then the product AB is also an $n \times n$ stochastic matrix.

(c) For each of $i = 2, 3, 4, \dots$, A^i is an $n \times n$ stochastic matrix.

In the examples considered thus far, we could always find a steady-state vector for a given initial state vector. However, not all stochastic matrices have a steady-state vector for each initial state vector. For instance, if

$$A=\begin{bmatrix}0&1\\1&0\end{bmatrix} \text{ and } \mathbf{x}_0=\begin{bmatrix}1\\0\end{bmatrix}$$

then it is easy to verify that

$$\mathbf{x}_1=\begin{bmatrix}0\\1\end{bmatrix},\mathbf{x}_2=\begin{bmatrix}1\\0\end{bmatrix},\mathbf{x}_3=\begin{bmatrix}0\\1\end{bmatrix},\mathbf{x}_4=\begin{bmatrix}1\\0\end{bmatrix},\cdots$$

so that we will not reach a steady-state vector for this choice of $\mathbf{x}_0$. On the other hand, if we start with $\mathbf{x}_0=\begin{bmatrix}0.5\\0.5\end{bmatrix}$, then we have
Thus this choice for $\mathbf{x}_0$ yields a steady-state vector. In fact, for the matrix A, this choice for $\mathbf{x}_0$ is the *only* initial state vector that leads to a steady-state vector. (See Exercise 42.)

A stochastic matrix may have many different steady-state vectors, depending on the initial state vector. For instance, suppose

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

Then it can be verified (see Exercise 52) that

$$\text{Intial State: } \mathbf{x}_0=\begin{bmatrix}0.5\\0.25\\0.25\end{bmatrix}\Rightarrow\text{steady-state vector: } \mathbf{x}=\begin{bmatrix}0.625\\0.375\\0\end{bmatrix}$$

and

Intial State: x0=[0.20.60.2]⇒steady-state vector: x=[0.30.70]

In general for this case, solving the system $(A - I)\mathbf{x} = 0$ yields the general solution

x=s1[100]+s2[010]=[s1s20]

Thus being a stochastic matrix is not enough to ensure that there will be a unique steady-state vector. It turns out that we need one additional condition.

DEFINITION 3.30 ▶

Regular Matrix

Let A be a stochastic matrix. Then A is **regular** if for some integer $k \geq 1$ the matrix A^k has all strictly positive entries.

It can be verified that both A=[0110] and A=[101/3011/3001/3] are not regular. On the other hand, even though

B=[1/21/21/21/2 01/2 01/2 0] and B2=[1/21/21/21/41/21/41/4 01/4]

both have zero entries, B is regular because B^3 has all positive entries,

B3=[1/21/21/23/81/43/81/81/41/8]

THEOREM 3.31 ▶

Let A be a regular stochastic matrix. Then

(a) For any initial state vector $\mathbf{x}_0$, the Markov chain $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, ⋯ converges to a unique steady-state vector $\mathbf{x}$.

(b) The sequence $A, A^2, A^3, \ldots$ converges to the matrix $[\mathbf{x}\ \mathbf{x} \cdots \mathbf{x}]$, where **x** is the unique steady-state vector given in part (a).

Proof The proof of part (a) is beyond the scope of this book, but it can be found in texts on Markov chains. To prove part (b), we begin by noting that

$$A^n = A^{n-1}A = A^{n-1}[\mathbf{a}_1\ \mathbf{a}_2 \cdots \mathbf{a}_n] = [A^{n-1}\mathbf{a}_1\ A^{n-1}\mathbf{a}_2 \cdots A^{n-1}\mathbf{a}_n]$$

From part (a) we know that as n grows, $A^{n-1}\mathbf{a}_j$ converges to **x** for each of $j = 1, 2, \ldots, n$. Thus it follows that

$$A^n \to [\mathbf{x}\ \mathbf{x} \cdots \mathbf{x}]$$

completing the proof. ■ ■

Theorem 3.31 shows that a regular stochastic matrix will have a unique steady-state vector that is independent of the initial state vector, which explains why we had no trouble solving our earlier examples.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine if A is stochastic.

(a) $A=\begin{bmatrix} 0.35 & 0.75 \\ 0.65 & 0.25 \end{bmatrix}$

(b) A=[16121213012121212]

2. If possible, fill in the missing values to make A doubly stochastic.

(a) $A=\begin{bmatrix} a & 0.7 \\ 0.3 & b \end{bmatrix}$

(b) A=[0.40.2abd0.30.30.4c]

3. Find the state vector $\mathbf{x}_2$ for A=[0.40.70.60.3],x0=[0.50.5]

4. Find all steady-state vectors for A.

(a) A=[0.250.50.750.5]

(b) A=[0.20.50.50.40.500.400.5]

5. Determine if each statement is true or false, and justify your answer.

(a) A stochastic matrix must have positive entries.

(b) If A is a stochastic matrix, then so is A^2.

(c) Every stochastic matrix has at least one steady-state vector.

(d) If A is stochastic, then AA^T is doubly stochastic.

EXERCISES

Exercises 1–4: Determine if A is a stochastic matrix.

1. A=[0.20.60.80.4]

2. A=[1.50.15-0.50.85]

3. A=[15102501225012]

4. A=[31438121238123141412]

Exercises 5–8: Fill in the missing values to make A a stochastic matrix.

5. A=[a0.450.65b]

6. A=[a0.70.20.35b0.40.20.25c]

7. A=[21337ca3715313b710]

8. A=[a0.50.20.050.450.150.4d0.1bc0.2500.20.300.15]

Exercises 9–12: If possible, fill in the missing values to make A a doubly stochastic matrix.

9. A=[a0.30.3b]

10. A=[0.40.6ab]

11. A=[0.2a0.3b0.1cd0.40.2]

12. A=[a0.5b0.2c0.20.50.1d]

Exercises 13–16: Find the state vector $\mathbf{x}_3$ for the given stochastic matrix and initial state vector.

13. A=[0.20.60.80.4],x0=[0.20.8]

14. A=[0.50.30.50.7],x0=[0.30.7]

15. A=[13252535],x0=[1212]

16. A=[14373447],x0=[1323]

Exercises 17–20: Find all steady-state vectors for the stochastic matrix *A*.

17. A=[0.80.50.20.5]

18. A=[0.30.60.70.4]

19. A=[0.40.50.30.20.30.40.40.20.3]

20. A=[0.3000.2100.501]

Exercises 21–24: Determine if the stochastic matrix *A* is regular.

21. A=[10.400.6]

22. A=[0.300.71]

23. A=[0.70.20.10.30.80.4000.5]

24. A=[00.20.50.900.50.10.80]

FIND AN EXAMPLE Exercises 25–30: Find an example that meets the given specifications.

25. A 4 × 4 stochastic matrix.

26. A 4 × 4 doubly stochastic matrix.

27. A 2 × 2 stochastic matrix *A* that has [2/31/3] for a steady-state vector.

28. A 3 × 3 stochastic matrix *A* that has [0.50.250.25] for a steady-state vector.

29. A 3 × 3 stochastic matrix A and initial state vector $\mathbf{x}_0$ such that the Markov chain $A\mathbf{x}_0$, $A^2\mathbf{x}_0$, … does not converge to a steady-state vector.

30. A 3 × 3 stochastic matrix A that has exactly one initial state vector $\mathbf{x}_0$ that will generate a Markov chain with a steady-state vector.

TRUE OR FALSE Exercises 31–36: Determine if the statement is true or false, and justify your answer.

31.

(a) If A is a stochastic matrix, then so is A^T.

(b) If A is stochastic, then A^{-1} exists and is stochastic.

32.

(a) If A is an $n \times n$ stochastic matrix and

x=[1/n⋮1/n]thenAx =[1⋮1]

(b) A stochastic matrix and its inverse have the same steady-state vectors.

33.

(a) If A and B are stochastic $n \times n$ matrices, then AB^T is also stochastic.

(b) A regular matrix must have positive entries.

34.

(a) If A is a symmetric stochastic matrix, then A is doubly stochastic.

(b) The stochastic matrix A and the matrix $B = A^2$ have the same steady-state vectors.

35.

(a) All Markov chains converge to a steady-state vector.

(b) If A is a 3 × 3 stochastic matrix, then $A - I_3$ is not stochastic.

36.

(a) Every 2 × 2 stochastic matrix has at least one steady-state vector.

(b) If A and B are stochastic matrices, then so is 12(A+B).

37. Prove Theorem 3.29(a): Show that each state vector is a probability vector.

38. Prove Theorem 3.29(b): Show that the product of two stochastic matrices is a stochastic matrix. (HINT: Appeal to

Theorem 3.29(a).)

39. Prove Theorem 3.29(c): Show that if A is a stochastic matrix, then so is A^2, A^3, … . (HINT: Use induction and appeal to Theorem 3.29(b).)

40. Suppose that A is a regular stochastic matrix. Show that A^2 is also a regular stochastic matrix.

41. Let A=[abcd] be a doubly stochastic matrix. Prove that $a = d$ and $b = c$.

42. If A=[0110], prove that x0=[0.50.5] is the only initial state vector that will lead to a steady-state vector.

43. Let A be a regular stochastic matrix, and suppose that k is the smallest integer such that A^k has all strictly positive entries. Show that each of A^{k+1}, A^{k+2}, … will have strictly positive entries.

44. Let A be an upper or lower triangular stochastic matrix. Show that A is not regular.

45. Suppose that A=[α0(1-α)1], where 0, > α > 1.

(a) Explain why A is a stochastic matrix.

(b) Find a formula for A^k, and use it to show that A is not regular.

(c) As $k \to \infty$, what matrix does A^k converge to?

(d) Find the one steady-state vector for A. (Note that this shows that the converse of Theorem 3.31 does not hold: A stochastic matrix that is not regular can still have a unique steady-state vector.)

46. In an office complex of 1000 employees, on any given day some are at work and the rest are absent. It is known that if an employee is at work today, there is an 85% chance that she will be at work tomorrow, and if the employee is absent today, there is a 60% chance that she will be absent tomorrow. Suppose that today there are 760 employees at work.

(a) Find the transition matrix for this scenario.

(b) Predict the number that will be at work five days from now.

(c) Find the steady-state vector.

47. The star quarterback of a university football team has decided to return for one more season. He tells one person, who in turn tells someone else, and so on, with each person talking to someone who has not heard the news. At each step in this chain, if the message heard is “yes’’ (he is returning), then there is a 10% chance it will be changed to “no,’’ and if the message heard is “no,’’ then there is a 15% chance that it will be changed to “yes.’’

(a) Find the transition matrix for this scenario.

(b) Determine the probability that the sixth person in the chain hears the wrong news.

(c) Find the steady-state vector.

48. It has been claimed that the best predictor of today’s weather is yesterday’s weather. Suppose that in the town of Springfield, if it rained yesterday, then there is a 60% chance of rain today, and if it did not rain yesterday, then there is an 85% chance of no rain today.

(a) Find the transition matrix describing the rain probabilities.

(b) If it rained Tuesday, what is the probability of rain Thursday?

(c) If it did not rain Friday, what is the probability of rain Monday?

(d) If the probability of rain today is 30%, what is the probability of rain tomorrow?

(e) Find the steady-state vector.

49. Consumers in Shelbyville have a choice of one of two fast food restaurants, Krusty’s and McDonald’s. Both have trouble keeping customers. Of those who last went to Krusty’s, 65% will go to McDonald’s next time, and of those who last went to McDonald’s, 80% will go to Krusty’s next time.

(a) Find the transition matrix describing this situation.

(b) A customer goes out for fast-food every Sunday, and just went to Krusty’s.

(i) What is the probability that two Sundays from now she will go to McDonald’s?

(ii) What is the probability that three Sundays from now she will go to McDonald’s?

(c) Suppose a consumer has just moved to Shelbyville, and there is a 40% chance that he will go to Krusty’s for his first fast food outing. What is the probability that his third fast-food experience will be at Krusty’s?

(d) Find the steady-state vector.

50. An assembly line turns out two types of pastries, Chocolate Zots and Rainbow Wahoos. The pastries come out one at a time; 40% of the time, a Wahoo follows a Zot, and 25% of the time, a Zot follows a Wahoo.

(a) If a Zot has just emerged from the line, what is the probability that a Wahoo will come two pastries later?

(b) If a Zot has just emerged from the line, what is the probability that a Zot will come three pastries later?

(c) What is the long-term probability that a randomly emerging pastry will be a Wahoo?

51. A medium-size town has three public library branches, designated A, B, and C. Patrons checking out books can return them to any of the three branches, where the books stay until checked out again. History shows that books borrowed from each branch are returned to a given location based on the following probabilities:

		Borrowed		
		A	B	C
	A	0.4	0.1	0.2
Returned	B	0.3	0.7	0.7
	C	0.3	0.2	0.1

(a) If a book is borrowed from A, what is the probability that it ends up at C after two more circulations?

(b) If a book is borrowed from B, what is the probability that it ends up at B after three more circulations?

(c) What is the steady-state vector?

52. Let A=[101/3011/3001/3].
Numerically verify that each initial state vector $\mathbf{x}_0$ has the given steady-state vector **x**.

(a) x0=[0.50.250.25]⇒x =[0.6250.3750]

(b) x0=[0.20.60.2]⇒x =[0.30.70]

Exercises 53–54: Determine to six decimal places the steady-state vector corresponding to the given initial state vector. Also find the smallest integer k such that $\mathbf{x}_k = \mathbf{x}_{k+1}$ to 6 decimal places for all entries. (NOTE: Even if it is less computationally efficient, it may be easier to compute state vectors using powers of A instead of the recursive formula.)

53. A=[0.20.30.10.40.30.50.60.20.10.10.20.20.40.10.10.2],x0=[0.250.250.250.25]

54. A=[110.20.50.200.300.500.40.50.300.10],x0=[0.10.20.30.4]

Exercises 55–56: Use computational experimentation to find two initial state vectors that lead to different steady-state vectors. (NOTE: Even if it is less computationally efficient, it may be easier to compute state vectors using powers of A instead of the recursive formula.)

55. A=[0.5000.501000010.50.5000]

56. A=[100000.500.50010.500.500]

SUPPLEMENTARY EXERCISES

Exercises 1–4: Let $T(\mathbf{x}) = A\mathbf{x}$. Find $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$.

1. A=[-213-3],u1[3-1],u2=[25]

2. A=[30-2-143],u1[1-4],u2=[32]

3. A=[510-126],u1[1-23],u2=[30-1]

4. A=[-10432022-5],u1[0-2-1],u2=[242]

Exercises 5–6: Suppose that T is a linear transformation, with

T(u1)=[3-4],T(u2)=[71]

5. Find $T(\mathbf{u}_1 - \mathbf{u}_2)$

6. Find $T(2\mathbf{u}_1 - 3\mathbf{u}_2)$

Exercises 7–8: Suppose that T is a linear transformation, with

T(u1)=[2-1],T(u1)=[-63],T(u3)=[50]

7. Find $T(\mathbf{u}_1 + \mathbf{u}_2 - \mathbf{u}_3)$

8. Find $T(-\mathbf{u}_1 + 2\mathbf{u}_2 - \mathbf{u}_3)$

Exercises 9–12: Suppose that $T(\mathbf{x}) = A\mathbf{x}$ for the given A. Determine if T is one-to-one and if T is onto.

9. A=[-113-3]

10. A=[13-1262]

11. A=[213-214]

12. A=[30-21-24025]

Exercises 13–16: Perform the indicated computations when possible using the given matrices.

A=[4-215],B=[3-406],C=[-23126-4]

D=[3-2-5134],E[42103-32-54]

13. $B - A$; BC; DE

14. $5I_2 + A$; CD; EC

15. DA; E^2; A^3

16. $3D^T + C$; $AB + DC$; $B^2 + A^2$

Exercises 17–18: Let T_1 and T_2 be the linear transformations given below.

T1([x1x2])=[4x1-3x2-x1+5x2]T2([x1x2])=[6x1+2x2x1-5x2]

17. Find the matrix A such that $T_1(T_2(\mathbf{x})) = A\mathbf{x}$

18. Find the matrix A such that $T_1(T_1(\mathbf{x})) = A\mathbf{x}$

Exercises 19–20: Let and be the linear transformations given below.

T1([x1x2x3])=[x1-x2+x32x1+3x2-2x33x1+x3]T2([x1x2x3])=[-x1+2x2-x3 4x1-x23x1+x2-x3]

19. Find the matrix A such that $T_2(T_1(\mathbf{x})) = A\mathbf{x}$

20. Find the matrix A such that $T_2(T_2(\mathbf{x})) = A\mathbf{x}$

Exercises 21–24: Assume that A is a matrix with three rows. Find the elementary matrix E such that EA is the matrix resulting from A after the given row operation is performed.

21. $-5R_1 \to R_1$

22. R3↔R1

23. $7R_1 + R_3 \to R_3$

24. $-5R_2 + R_3 \to R_3$

Exercises 25–28: Assume that *A* is a matrix with four rows. Find the matrix *B* such that *BA* is the matrix resulting from *A* after the given row operations are performed.

25. -3R1+R2→R2, 5R4→R4

26. 2R2+R4→R4, R1↔R3

27. -R1→R1,R1↔R4,3R1+R3→R3

28. 5R1+R2→R2,-R1+R3→R3,-R2+R4→R4

Exercises 29–30: Use the Quick Formula to find the inverse of *A*.

29. A=[2315]

30. A=[-4068]

Exercises 31–36: If possible, find the inverse of *A*. If not possible, explain why.

31. A=[1-1-321-4020]

32. A=[32-2-1-231-24]

33. A=[15-34-26]

34. A=[37-1356]

35. A=[10-2001320-2110011]

36. A=[-11210201-112-1-2441]

Exercises 37–40: Find an LU factorization for *A*.

37. A=[15-38]

38. A=[3-4127]

39. A=[-2312-134-21]

40. A=[3-119200-56]

Exercises 41–44: Find an LDU factorization for *A*.

41. A=[23-87]

42. A=[-1372]

43. A=[-12021132-2]

44. A=[201-8516-14]

Exercises 45–48: Find all steady-state vectors for the stochastic matrix *A*.

45. A=[0.60.50.40.5]

46. A=[0.20.60.80.4]

47. A=[0.30.50.30.30.20.40.40.30.3]

48. A=[0.3100000.701]

Exercises 49–52: Determine if the stochastic matrix *A* is regular.

49. A=[0.400.61]

50. A=[1001]

51. A=[0.50.40.20.50.60.3000.5]

52. A=[00.30.50.8000.20.70.5]

CHAPTER 4

Subspaces

WIN-Intiative/Riser/Getty Images

Shown in the photo are solar panels in Nepal, near the Lobuche glacier and the Pyramid research station. These solar panels absorb light energy from the sun through photovoltaic cells to create electricity or heating. The use of solar power has become more popular, especially as the price has decreased. Current innovations in solar panels include continuing to make them more efficient so that they can take advantage of the full spectrum of the sun's light.

Subspaces are a special type of subset of Euclidean space $\mathbf{R}^n$. They arise naturally in connection with spanning sets, linear transformations, and systems of linear equations. Section 4.1 introduces subspaces and provides a general procedure for determining if a subset is a subspace. Section 4.2 introduces the concept of basis and a way to measure the size of a subspace. Section 4.3 connects the concept of subspaces to rows and columns of matrices. Section 4.4 describes how to change from one basis to another.

4.1 Introduction to Subspaces

In Section 2.2 we considered S = span{$\mathbf{u}_1$, $\mathbf{u}_2$}, where

u1=[211] and u2=[123] (1)

Recall that span{$\mathbf{u}_1$, $\mathbf{u}_2$}, the set of all linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$, forms a plane in $\mathbf{R}^3$ (Figure 1). This subset of $\mathbf{R}^3$ is an example of a *subspace*. In many ways, a subspace of $\mathbf{R}^n$ resembles $\mathbf{R}^m$ for some $m \leq n$.

DEFINITION 4.1 ▶

Subspace

A subset S of $\mathbf{R}^n$ is a **subspace** if S satisfies the following three conditions:

(a) S contains $\mathbf{0}$, the zero vector.

(b) If $\mathbf{u}$ and $\mathbf{v}$ are in S, then $\mathbf{u} + \mathbf{v}$ is also in S.

(c) If r is a real number and $\mathbf{u}$ is in S, then $r\mathbf{u}$ is also in S.

Closed Under Addition, Closed Under Scalar Multiplication

A subset of $\mathbf{R}^n$ that satisfies condition (b) above is said to be **closed under addition**, and if it satisfies condition (c), then it is **closed under scalar multiplication**. Closure under addition and scalar multiplication ensures that arithmetic performed on vectors in a subspace produce other vectors in the subspace.

▶ Geometrically, condition (a) says that the graph of a subspace must pass through the origin.

Is *S* a Subspace?

To determine if a given subset *S* is a subspace, an easy place to start is with condition (a) of Definition 4.1, which states that every subspace must contain **0**. A moment's thought reveals that this is equivalent to the statement

If 0 is not in a subset S, then S is not a subspace. (2)

Example 1

Let *S* consist of all solutions x= [x1x2] to the linear system

$$-3x1+2x2=17x1-5x2=-1$$

Is *S* a subspace of $\mathbf{R}^2$?

Solution We know that *S* is a subset of $\mathbf{R}^2$. However, note that x = [00] is *not* a solution to the given system. Thus **0** is not in *S*, and so *S* cannot be a subspace of $\mathbf{R}^2$.

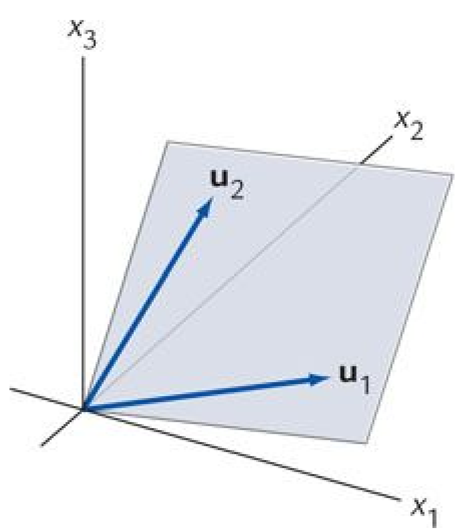


Figure 1 Span of {$\mathbf{u}_1$, $\mathbf{u}_2$}

Note that the converse of (2) is not true—just because **0** is in *S* does not guarantee that *S* is a subspace, because conditions (b) and (c) must also be satisfied.

Example 2

Let *S* be the subset of $\mathbf{R}^2$ consisting of the *x*-axis and *y*-axis (Figure 2). Show that *S* is not a subspace of $\mathbf{R}^2$.

Solution From Figure 2 we can see that **0** is in *S*, so (a) of Definition 4.1 is satisfied. Moreover, since the multiple of a point on a coordinate axis is on the same coordinate axis, (c) is also satisfied. However, (b) is not satisfied: [10] and [01] are both in *S*, but

$$[10]+[01]=[11]$$

is not in *S*, so *S* is not closed under addition.

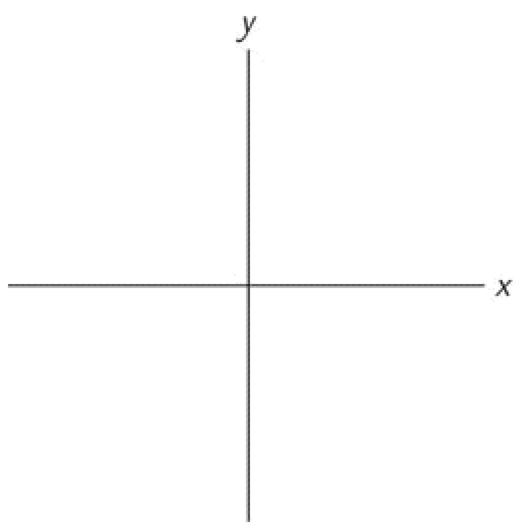


Figure 2 The coordinate axes are not a subspace of $\mathbf{R}^2$.

Example 3

Let S be the set of vectors [ab], where a and b are integers. Show that S is not a subspace.

Solution Since 0 is an integer, 0=[00] is in S, so (a) of Definition 4.1 is satisfied. The sum of two integers is again an integer, so S is closed under addition and (b) is satisfied. However, since [10] is in S but

$$2[10]=[20]$$

is not in S, (c) of Definition 4.1 is not satisfied. Therefore S is not a subspace.

This section opened with the statement that the span of two vectors forms a subspace of $\mathbf{R}^3$. This claim generalizes to the span of any finite set of vectors in $\mathbf{R}^n$ and provides a useful way to determine if a set of vectors is a subspace.

THEOREM 4.2

Let $S = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be in $\mathbf{R}^n$. Then S is a subspace of $\mathbf{R}^n$.

Theorem 4.2 also holds for the span of an infinite set of vectors. However, we do not require this case and including it would introduce additional technical issues, so we leave it out.

Proof To show that a subset is a subspace, we need to verify that the three conditions given in the definition are satisfied.

(a) Since $\mathbf{0} = 0\mathbf{u}_1 + \ldots + 0\mathbf{u}_m$, it follows that $\mathbf{0}$ is in $\text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} = S$.

(b) Suppose that **v** and **w** are in S. Then there exist scalars r_1, r_2, … r_m and s_1, s_2, … s_m such that

v= r1u1+r2u2+⋯rmumw= s1u1+s2u2+⋯smum

Thus

v + w=(r1u1+r2u2+rmum)+(s1u1+s2u2+smum) =(r1+s1)u1+(r2+s2)u2+⋯ (rm+sm)um

so $\mathbf{v} + \mathbf{w}$ is in span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\} = S$.

(c) If t is a real number, then taking **v** as in part (b), we have

tv=t(r1u1+r2u2+…rmum) =tr1u1+tr2u2+…trmum

so that $t\mathbf{v}$ is in span$\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} = S$.

Since parts (a)–(c) of the definition hold, S is a subspace.

Subspace Spanned, Subspace Generated

If S = span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$, then it is common to say that S is the **subspace spanned** (or **subspace generated**) by $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$.

DETERMINING IF *S* IS A SUBSPACE To determine if a subset S is a subspace, apply the following steps.

Step 1. Check if **0** is in S. If not, then S is not a subspace.
Step 2. If you can show that S is generated by a set of vectors, then S is a subspace by Theorem 4.2.
Step 3. Try to verify that conditions (b) and (c) of the definition are met. If so, then S is a subspace. If you cannot show that they hold, then you are likely to uncover a counterexample showing that they do not hold, which demonstrates that S is not a subspace.

Let's try this out on some examples.

Example 4

Determine if $S = \{\mathbf{0}\}$ and $S = \mathbf{R}^n$ are subspaces of $\mathbf{R}^n$.

Solution Since $\mathbf{0}$ is in both $S = \{\mathbf{0}\}$ and $S = \mathbf{R}^n$, Step 1 is no help, so we move to Step 2. Since $\{\mathbf{0}\} = \text{span}\{\mathbf{0}\}$, by Theorem 4.2 the set $S = \{\mathbf{0}\}$ is a subspace. We also have $\mathbf{R}^n = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{3}$$

Thus $\mathbf{R}^n$ is a subspace of itself. (These are sometimes called the **trivial subspaces** of $\mathbf{R}^n$.)

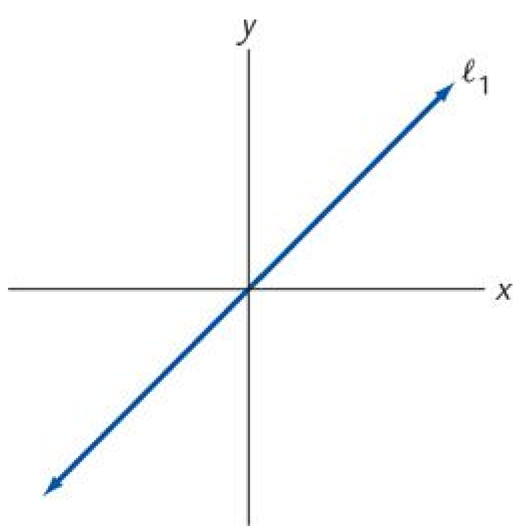


Figure 3 ℓ_1 is a subspace.

Example 5

Let ℓ_1 denote a line through the origin in $\mathbf{R}^2$ (Figure 3), and let ℓ_2 denote a line that does not pass through the origin in $\mathbf{R}^2$ (Figure 4). Do the points on ℓ_1 form a subspace? Do the points on ℓ_2 form a subspace?

Solution Since ℓ_1 passes through the origin, $\mathbf{0}$ is on ℓ_1, so Step 1 is not helpful. Moving to Step 2, suppose that we pick any nonzero vector $\mathbf{u}$ on ℓ_1. Then all points on ℓ_1 have the form $r\mathbf{u}$ for some scalar r. Thus $\ell_1 = \text{span}\{\mathbf{u}\}$, so ℓ_1 is a subspace.

Focusing now on the line ℓ_2, we note that this line does not contain $\mathbf{0}$, so ℓ_2 is not a subspace.

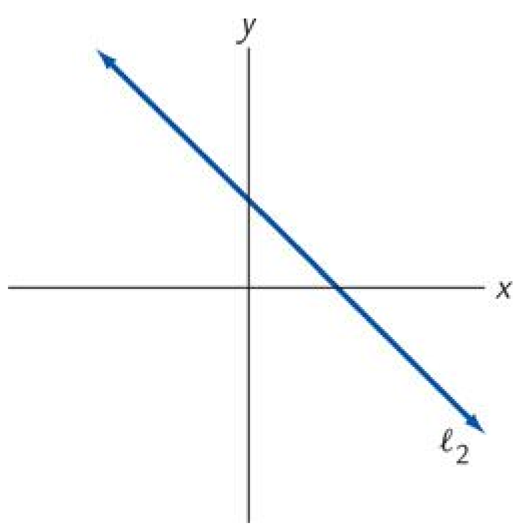


Figure 4 ℓ_2 does not pass through the origin, so it is not a subspace.

Example 6

Let A be the subset of $\mathbf{R}^3$ consisting of vectors of the form [2a−b3ba+5b] where a and b are real numbers. Show that S is a subspace.

Solution The set of vectors making up S can be expressed

$$[2a-b3ba+5b]=[2a-b0a+3ba+5b]=a[201]+b[-135]$$

Therefore

$$S=\mathrm{span}\{[201],[-135]\}$$

so S is a subspace of $\mathbf{R}^3$ by Theorem 4.2.

Example 7

Let S be the subset of $\mathbf{R}^3$ consisting of all vectors of the form

$$v=[v1v2v3]$$

such that $v_1 + v_2 + v_3 = 0$. Is S a subspace of $\mathbf{R}^3$?

Solution Starting with Step 1, we see that setting $v_1 = v_2 = v_3 = 0$ implies $\mathbf{0}$ is in S, so we cannot conclude anything. A spanning set for S is not immediately apparent (although we could find one), so let's skip to Step 3 and determine if conditions (b) and (c) of the definition are satisfied.
(b) Let u=[u1u2u3] and v=[v1v2v3] be in S. Then

$$u+v=[u1+v1u2+v2u3+v3]$$

and since

$$(u1+v1)+(u2+v2)+(u3+v3)=(u1+u2+u3)+(v1+v2+v3) =0+0=0$$

it follows that $\mathbf{u} + \mathbf{v}$ is in S.
(c) With v=[v1v2v3] in S as above, for any scalar r we have rv=[rv1rv2rv3]. Since

$$rv1+rv2+rv3=r(v1+v2+v3)=0$$

$r\mathbf{v}$ is also in S.

Since all conditions of the definition are satisfied, we conclude that S is a subspace of $\mathbf{R}^3$.

It is not hard to extend the result in Example 7 to $\mathbf{R}^n$. Let S be the set of all vectors of the form

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

such that $v_1 + \cdots + v_n = 0$. Then S is a subspace of $\mathbf{R}^n$ (see Exercise 15).

Homogeneous Systems and Null Spaces

The set of solutions to a homogeneous linear system forms a subspace. For instance, let A be the 3×4 matrix

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

Using our usual solution method, we find that all solutions to the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ have the form

$$\mathbf{x} = s_1 \begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + s_2 \begin{bmatrix} 1 \\ -3 \\ 0 \\ 1 \end{bmatrix}$$

where s_1 and s_2 can be any real numbers. Thus the set of solutions to $A\mathbf{x} = \mathbf{0}$ is equal to

$$\text{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

and so the set of solutions is a subspace of $\mathbf{R}^4$. This result generalizes to the set of solutions of any homogeneous linear system.

THEOREM 4.3 ▶

If A is an $n \times m$ matrix, then the set of solutions to the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ forms a subspace of $\mathbf{R}^m$.

Proof We verify the three conditions from Definition 4.1 to show that the set forms a subspace.

(a) Since $\mathbf{x} = \mathbf{0}$ is a solution to $A\mathbf{x} = \mathbf{0}$, the zero vector $\mathbf{0}$ is in the set of solutions.

(b) Suppose that $\mathbf{u}$ and $\mathbf{v}$ are both solutions to $A\mathbf{x} = \mathbf{0}$. Then

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

so that $\mathbf{u} + \mathbf{v}$ is in the set of solutions.

(c) Let $\mathbf{u}$ be a solution to $A\mathbf{x} = \mathbf{0}$, and let r be a scalar. Then

$$A(r\mathbf{u}) = r(A\mathbf{u}) = r\mathbf{0} = \mathbf{0}$$

and so $r\mathbf{u}$ is also in the set of solutions.

Since all three conditions of the definition are met, the set of solutions to $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbf{R}^n$. ■ ■

A subspace given by the set of solutions to a homogeneous linear system goes by a special name.

DEFINITION 4.4 ▶

Null Space

If A is an $n \times m$ matrix, then the set of solutions to $A\mathbf{x} = \mathbf{0}$ is called the **null space** of A and is denoted by null(A).

From Theorem 4.3 it follows that a null space is a subspace.

Example 8

Ethane burns in oxygen to produce carbon dioxide and steam. The chemical reaction is described using the notation

x1C2H6+x2O2→x3CO2+x4H2O

▶ Ethane is a gas similar to propane. Its primary use in the chemical industry is to make polyethylene, a common form of plastic.

where the subscripts on the elements indicate the number of atoms in each molecule. Describe the subspace of values that will balance this equation.

Solution To balance the equation, we need to find values for x_1, x_2, x_3, and x_4 so that the number of atoms for each element is the same on both sides. Doing so yields the linear system

2x1q −x3 =0(carbon atoms)6x1 −2x4=0(hydrogen atoms) 2x2−2x 3−x =0(oxygen atorms)

▶ Balancing chemical equations is also discussed in detail in Section 1.3.

Applying our usual methods, we find that the general solution to this system is

x1=2sx2=7sx3=4sx4=6sorx=s[2746]

where s can be any real number. Put another way, the set of solutions is equal to

span{[2746]}

and therefore the set is a subspace of $\mathbf{R}^4$.

Example 9

Let S be the set of vectors of the form [abc], where $a + b = c$. Show that S is a subspace of $\mathbf{R}^3$.

Solution The condition $a + b = c$ is equivalent to $a + b - c = 0$. Hence the set of vectors in S is the same as the set of solutions to the homogeneous linear equation

$$x1+x2-x3=0$$

This can be expressed as $A\mathbf{x} = \mathbf{0}$ where

$$A=[11-1]$$

Thus S is a subspace by Theorem 4.3.

Kernel and Range of a Linear Transformation

Kernel

There are two sets associated with every linear transformation T that are subspaces. Recall that the range of T is the set of all vectors $\mathbf{y}$ such that $T(\mathbf{x}) = \mathbf{y}$ for some $\mathbf{x}$ and is denoted by range(T). The **kernel** of T is the set of vectors such that $T(\mathbf{x}) = \mathbf{0}$, and is denoted by ker($T$) (see Figure 5). Theorem 4.5 shows that the range and kernel are subspaces.

THEOREM 4.5 ▶

Let $T : \mathbf{R}^m \to \mathbf{R}^n$ be a linear transformation. Then the kernel of T is a subspace of the domain $\mathbf{R}^m$ and the range of T is a subspace of the codomain $\mathbf{R}^n$.

Proof Because $T : \mathbf{R}^m \to \mathbf{R}^n$ is a linear transformation, it follows (Theorem 3.8, Section 3.1) that there exists an $n \times m$ matrix $A = [\mathbf{a}_1 \ldots \mathbf{a}_m]$ such that $T(\mathbf{x}) = A\mathbf{x}$. Thus $T(\mathbf{x}) = \mathbf{0}$ if and only if $A\mathbf{x} = \mathbf{0}$. This implies that

$$\ker(T)=\operatorname{null}(A)$$

and therefore by Theorem 4.3 the kernel of T is a subspace of the domain $\mathbf{R}^m$.

Now consider the range of T. By Theorem 3.3(b), we have

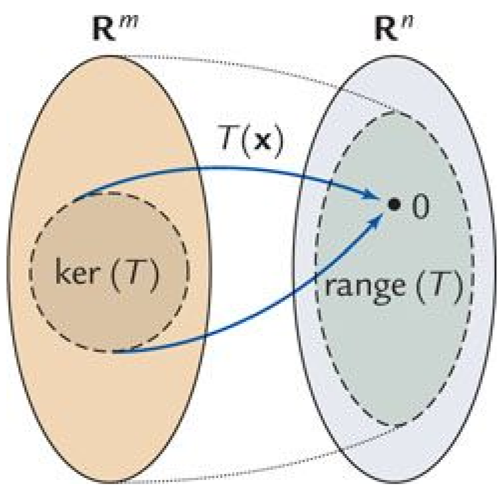


Figure 5 The kernel and range of T.

$$\operatorname{range}(T)=\operatorname{span}\{\mathbf{a}_1,\ldots,\mathbf{a}_m\}$$

Since range(T) is equal to the span of a set of vectors, by Theorem 4.2 the range of T is a subspace of the codomain $\mathbf{R}^n$. ■■

Example 10

Suppose that $T : \mathbf{R}^2 \to \mathbf{R}^3$ is defined by

$$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=\begin{bmatrix}x_1-2x_2\\-3x_1+6x_2\\2x_1-4x_2\end{bmatrix}$$

▶ Here T and A are closely related, making ker(T) and null(A) seem redundant. Later we will see a more general linear transformation T that is less closely associated with a matrix A.

Find ker(T) and range(T).

Solution We have $T(\mathbf{x}) = A\mathbf{x}$ for

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

As noted above, ker(T) = null(A). To find the null space of A, we solve the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. We have

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

which is equivalent to the single equation $x_1 - 2x_2 = 0$. If $x_2 = s_1$, then $x_1 = 2s_1$ so the general solution to $A\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x}=s_1\begin{bmatrix} 2 \\ 1 \end{bmatrix} \Rightarrow \ker(T)=\operatorname{span}\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$$

▶ Remember that the kernel is a subspace of the *domain*, whereas the range is a subspace of the *codomain*.

Because the range of T is equal to the span of the columns of A, we have

$$\operatorname{range}(T)=\operatorname{span}\{\mathbf{a}_1,\mathbf{a}_2\}=\operatorname{span}\left\{\begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix},\begin{bmatrix} -2 \\ 6 \\ -4 \end{bmatrix}\right\}=\operatorname{span}\left\{\begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}\right\}$$

because $\mathbf{a}_2 = -2\mathbf{a}_1$.

In Theorem 3.5 in Section 3.1 we showed that a linear transformation T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution. The next theorem formulates this result in terms of ker(T).

THEOREM 4.6 ▶

Let $T : \mathbf{R}^m \to \mathbf{R}^n$ be a linear transformation. Then T is one-to-one if and only if $\ker(T) = \{\mathbf{0}\}$.

The proof is covered in Exercise 64. As a quick application, in Example 10 we saw that $\ker(T) \neq \{\mathbf{0}\}$, so we can conclude from Theorem 4.6 that T is not one-to-one.

The Unifying Theorem, Version 4

Theorem 4.6 gives us another condition to add to the Unifying Theorem.

THEOREM 4.7 ▶

(THE UNIFYING THEOREM, VERSION 4) Let $\mathcal{S}$={a1,… ,an} be a set of n vectors in $\mathbf{R}^n$, let $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$, and let $T : \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent:

(a) $\mathcal{S}$ spans $\mathbf{R}^n$.

(b) $\mathcal{S}$ is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

(d) T is onto.

(e) T is one-to-one.

(f) A is invertible.

(g) $\ker(T) = \{\mathbf{0}\}$.

▶ This updates the Unifying Theorem, Version 3, from Section 3.3.

▶ Practice problems can also be used as additional examples.

Proof From The Unifying Theorem, Version 3, we know that (a) through (f) are equivalent. From Theorem 4.6 we know that T is one-to-one if and only if ker(T) = {**0**}, so (e) and (g) are equivalent. Thus (a)–(g) are all equivalent.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine if the given subsets are subspaces. If so, give a Proof If not, explain why not. Assume a and b are real numbers.

(a) The subset of $\mathbf{R}^3$ of vectors of the form [a23b].

(b) The subset of $\mathbf{R}^3$ of vectors of the form [0a−b−2a].

2. Find the null space for A.

(a) A=[13−230−4]

(b) A=[214−11513]

3. Suppose $T(\mathbf{x}) = A\mathbf{x}$. Determine if **b** is in ker(T) and if **c** is in range(T).

(a) A=[1339],b=[3−1],c=[25]

(b) A=[1−236−24],b=[3−2],c=[21−1]

4. Determine if each statement is true or false, and justify your answer.

(a) For a linear transformation T, ker(T) is a subset of range(T).

(b) Let $T : \mathbf{R}^3 \to \mathbf{R}^7$ be a linear transformation. Then ker(T) is a subspace of $\mathbf{R}^7$.

(c) Every subspace contains infinitely many nonzero vectors.

(d) If S is a subspace of $\mathbf{R}^n$, then so is S^c, the set of vectors *not* in S.

EXERCISES

Exercises 1–16: Determine if the described set is a subspace. If so, give a proof. If not, explain why not. Unless stated otherwise, a, b, and c are real numbers.

1. The subset of $\mathbf{R}^3$ consisting of vectors of the form [a0b].

2. The subset of $\mathbf{R}^3$ consisting of vectors of the form [aa0].

3. The subset of $\mathbf{R}^2$ consisting of vectors of the form [ab], where $a + b = 1$.

4. The subset of $\mathbf{R}^3$ consisting of vectors of the form [abc], where $a = b = c$.

5. The subset of $\mathbf{R}^4$ consisting of vectors of the form [a10b].

6. The subset of $\mathbf{R}^4$ consisting of vectors of the form [aa+b2a−b3b].

7. The subset of $\mathbf{R}^2$ consisting of vectors of the form [ab], where a and b are rational numbers.

8. The subset of $\mathbf{R}^3$ consisting of vectors of the form [abc], where $c = b - a$.

9. The subset of $\mathbf{R}^3$ consisting of vectors of the form [abc], where $abc = 0$.

10. The subset of $\mathbf{R}^2$ consisting of vectors of the form [ab], where $a^2 + b^2 \leq 1$.

11. The subset of $\mathbf{R}^3$ consisting of vectors of the form [abc], where $a \geq 0$, $b \geq 0$, and $c \geq 0$.

12. The subset of $\mathbf{R}^3$ consisting of vectors of the form [abc], where at most one of a, b, and c is nonzero.

13. The subset of $\mathbf{R}^2$ consisting of vectors of the form [ab], where $a \leq b$.

14. The subset of $\mathbf{R}^2$ consisting of vectors of the form [ab], where $|a| = |b|$.

15. The subset of $\mathbf{R}^n$ consisting of vectors of the form

v=[v1⋮vn]

such that $v_1 + \cdots + v^n = 0$.

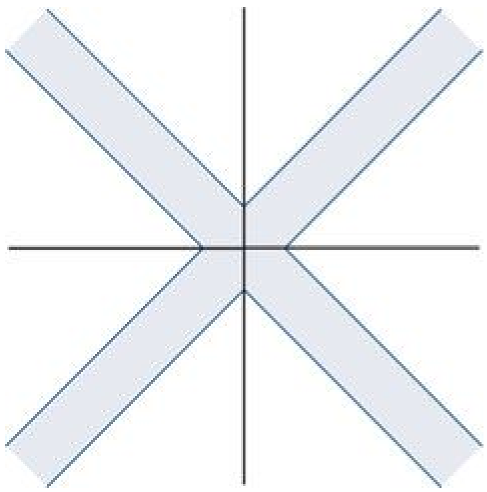
16. The subset of $\mathbf{R}^n$ (n even) consisting of vectors of the form

v=[v1⁝vn]

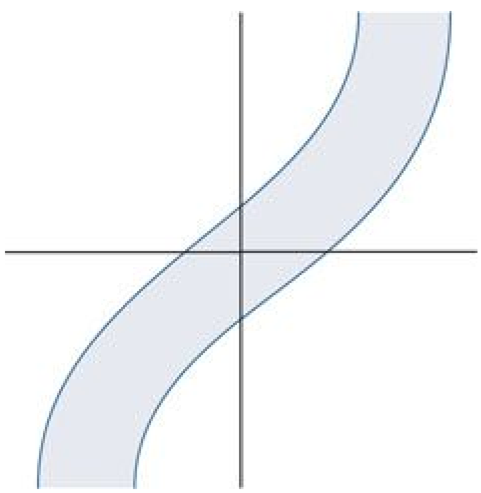
such that $v_1 - v_2 + v_3 - v_4 + v_5 - \cdots - v_n = 0$.

Exercises 17–20: The shaded region is not a subspace of $\mathbf{R}^2$. Explain why.

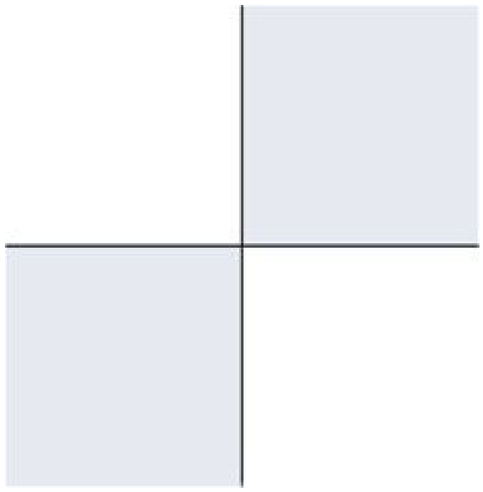
17.



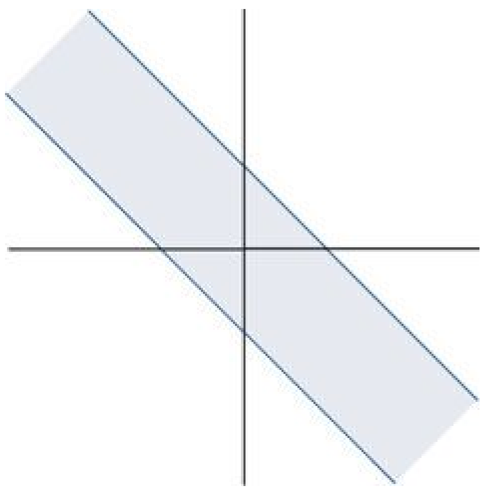
18.



19.



20.



Exercises 21–32: Find the null space for *A*.

21. A=[1−301]

22. A=[3564]

23. A=[10−5012]

24. A=[12−2014]

25. A=[1−22−25−7]

26. A=[30−4−162]

27. A=[13−2132]

28. A=[2−10−3151−5]

29. A=[1−11013003]

30. A=[120−3−4−12−23]

31. A=[11−21011−10002]

32. A=[1001021000101011]

Exercises 33–36: Let $T(\mathbf{x}) = A\mathbf{x}$ for the matrix A. Determine if the vector **b** is in the kernel of T and if the vector **c** is in the range of T.

33. A=[1−2−3−1],b=[21],c=[4−7]

34. A=[2−3014−2],b=[6411],c=[413]

35. A=[4−21327],b=[−52],c=[13]

36. A=[123456789],b=[1−21],c=[258]

FIND AN EXAMPLE Exercises 37–44: Find an example that meets the given specifications.

37. An infinite subset of $\mathbf{R}^2$ that is not a subspace of $\mathbf{R}^2$.

38. Two subspaces S_1 and S_2 of $\mathbf{R}^3$ such that $S_1 \cup S_2$ is not a subspace of $\mathbf{R}^3$.

39. Two nonsubspace subsets S_1 and S_2 of $\mathbf{R}^3$ such that $S_1 \cup S_2$ is a subspace of $\mathbf{R}^3$.

40. Two nonsubspace subsets S_1 and S_2 of $\mathbf{R}^3$ such that $S_1 \cap S_2$ is a subspace of $\mathbf{R}^3$.

41. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ such that range(T) = span {[11]}.

42. A linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^3$ such that range(T) = span {[1−12]}.

43. A linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ such that range(T) = $\mathbf{R}^3$.

44. A linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ such that range(T) = span {[314],[12−2]}.

TRUE OR FALSE Exercises 45–52: Determine if the statement is true or false, and justify your answer.

45.

(a) If A is an $n \times n$ matrix and $\mathbf{b} \neq \mathbf{0}$ is in $\mathbf{R}^n$, then the solutions to $A\mathbf{x} = \mathbf{b}$ do not form a subspace.

(b) If A is a 5×3 matrix, then null(A) forms a subspace of $\mathbf{R}^5$.

46.

(a) If A is a 4×7 matrix, then null(A) forms a subspace of $\mathbf{R}^7$.

(b) Let $T : \mathbf{R}^6 \to \mathbf{R}^3$ be a linear transformation. Then ker(T) is a subspace of $\mathbf{R}^6$.

47.

(a) Let $T : \mathbf{R}^5 \to \mathbf{R}^8$ be a linear transformation. Then ker(T) is a subspace of $\mathbf{R}^8$.

(b) Let $T : \mathbf{R}^2 \to \mathbf{R}^7$ be a linear transformation. Then range(T) is a subspace of $\mathbf{R}^2$.

48.

(a) Let $T : \mathbf{R}^3 \to \mathbf{R}^9$ be a linear transformation. Then range(T) is a subspace of $\mathbf{R}^9$.

(b) The union of two subspaces of $\mathbf{R}^n$ forms another subspace of $\mathbf{R}^n$.

49.

(a) The intersection of two subspaces of $\mathbf{R}^n$ forms another subspace of $\mathbf{R}^n$.

(b) Let S_1 and S_2 be subspaces of $\mathbf{R}^n$, and define S to be the set of all vectors of the form $\mathbf{s}_1 + \mathbf{s}_2$, where $\mathbf{s}_1$ is in S_1 and $\mathbf{s}_2$ is in S_2. Thus S is a subspace of $\mathbf{R}^n$.

50.

(a) Let S_1 and S_2 be subspaces of $\mathbf{R}^n$, and define S to be the set of all vectors of the form $\mathbf{s}_1 - \mathbf{s}_2$, where $\mathbf{s}_1$ is in S_1 and $\mathbf{s}_2$ is in S_2. Thus S is a subspace of $\mathbf{R}^n$.

(b) The set of integers forms a subspace of $\mathbf{R}$.

51.

(a) A subspace $S \neq \{\mathbf{0}\}$ can have a finite number of vectors.

(b) If $\mathbf{u}$ and $\mathbf{v}$ are in a subspace S, then every point on the line connecting $\mathbf{u}$ and $\mathbf{v}$ is also in S.

52.

(a) If S_1 and S_2 are subsets of $\mathbf{R}^n$ but *not* subspaces, then the union of S_1 and S_2 cannot be a subspace of $\mathbf{R}^n$.

(b) If S_1 and S_2 are subsets of $\mathbf{R}^n$ but *not* subspaces, then the intersection of S_1 and S_2 cannot be a subspace of $\mathbf{R}^n$.

53. Show that every subspace of $\mathbf{R}$ is either $\{\mathbf{0}\}$ or $\mathbf{R}$.

54. Suppose that S is a subspace of $\mathbf{R}^n$ and c is a scalar. Let cS denote the set of vectors $c\mathbf{s}$ where $\mathbf{s}$ is in S. Prove that cS is also a subspace of $\mathbf{R}^n$.

55. Prove that if $\mathbf{b} \neq \mathbf{0}$, then the set of solutions to $A\mathbf{x} = \mathbf{b}$ is not a subspace.

56. Describe the geometric form of all subspaces of $\mathbf{R}^2$.

57. Describe the geometric form of all subspaces of $\mathbf{R}^3$.

58. Some texts use just conditions (b) and (c) in Definition 4.1, along with S nonempty, as the definition of a subspace. Explain why this is equivalent to our definition.

59. Let A be an $n \times m$ matrix, and suppose that $\mathbf{y} \neq \mathbf{0}$ is in $\mathbf{R}^n$. Show that the set of all vectors $\mathbf{x}$ in $\mathbf{R}^m$ such that $A\mathbf{x} = \mathbf{y}$ is *not* a subspace of $\mathbf{R}^m$.

60. Let $A = [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3\ \mathbf{a}_4]$, and suppose that $\mathbf{x} = (2, -5, 4, 1)$ is in null(A). Write $\mathbf{a}_4$ as a linear combination of the other three vectors.

61. Let A be a matrix and $T(\mathbf{x}) = A\mathbf{x}$ a linear transformation. Show that $\ker(T) = \{\mathbf{0}\}$ if and only if the columns of A are linearly independent.

62. If T is a linear transformation, show that $\mathbf{0}$ is always in $\ker(T)$.

63. Prove that if $\mathbf{u}$ and $\mathbf{v}$ are in a subspace S, then so is $\mathbf{u} - \mathbf{v}$.

64. Prove Theorem 4.6: If T is a linear transformation, then T is one-to-one if and only if $\ker(T) = \{\mathbf{0}\}$.

Exercises 65–68: Use Example 8 as a guide to find the subspace of values that balances the given chemical equation.

65. Glucose ferments to form ethyl alcohol and carbon dioxide.

$$x1C6H12O6 \rightarrow x2C2H5OH+x3CO2$$

66. Methane burns in oxygen to form carbon dioxide and steam.

$$x1CH4+x2O2 \rightarrow x3CO2+x4H2O$$

67. An antacid (calcium hydroxide) neutralizes stomach acid (hydrochloric acid) to form calcium chloride and water.

$$x_1Ca(OH)_2+x_2HCl \rightarrow x_3CaCl_2+x_4H_2O$$

68. Ethyl alcohol reacts with oxygen to form vinegar and water.

$$x_1C_2H_5OH+x_2O_2 \rightarrow x_3HC_2H_3O_2+x_4H_2O$$

Exercises 69–72: Find the null space for the given matrix.

69. A=[17−2140301−2361−104]

70. A=[−10045262124032−5−102]

71. A=[3124502−12222−10310204]

72. A=[205−16244−1510411]

4.2 Basis and Dimension

In this section we combine the concepts of linearly independent sets and spanning sets to learn more about subspaces. Let $S = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a nonzero subspace of $\mathbf{R}^n$. Then every element $\mathbf{s}$ of S can be expressed as a linear combination

$$\mathbf{s} = r_1\mathbf{u}_1 + r_2\mathbf{u}_2 + \cdots + r_m\mathbf{u}_m$$

If $\mathbf{u}_1, \ldots, \mathbf{u}_m$, is a linearly dependent set, then by Theorem 2.15 we know that one of the vectors in the set—say, $\mathbf{u}_1$—is in the span of the remaining vectors. Thus it follows that every element of S can be written as a linear combination of $\mathbf{u}_2, \ldots, \mathbf{u}_m$ so that

$$S = \text{span}\{\mathbf{u}_2, \ldots, \mathbf{u}_m\}$$

If after eliminating $\mathbf{u}_1$ the remaining set of vectors is still linearly dependent, then we can repeat this process to eliminate another dependent vector. We can do this over and over, and since we started with a finite number of vectors the process must eventually lead us to a set that both spans S and is linearly independent. Such a set is particularly important and goes by a special name.

DEFINITION 4.8 ▶

Basis

A set $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a **basis** for a subspace S if

(a) $\mathcal{B}$ spans S.

(b) $\mathcal{B}$ is linearly independent.

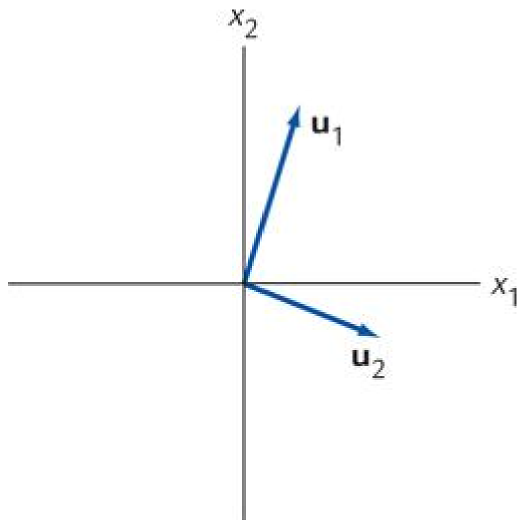


Figure 1 Any two nonzero vectors that do not lie on the same line form a basis for $\mathbf{R}^2$.

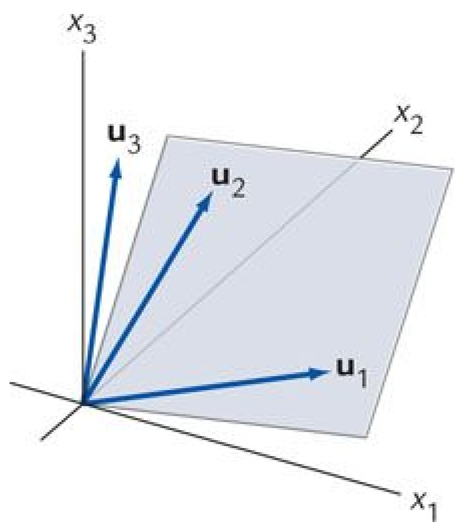


Figure 2 Any three nonzero vectors that do not lie on the same plane form a basis for $\mathbf{R}^3$.

Zero subspace

Figure 1 and Figure 2 show basis vectors for $\mathbf{R}^2$ and $\mathbf{R}^3$, respectively. One subspace does not have a basis: $S = \{\mathbf{0}\} = \text{span}\{\mathbf{0}\}$, the **zero subspace**. The set $\{\mathbf{0}\}$ is not linearly independent, so is not a basis. The zero subspace is the only subspace of $\mathbf{R}^n$ that does not have a basis.

Each basis has the following important property.

THEOREM 4.9 ▶

Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a basis for a subspace S. For every vector $\mathbf{s}$ in S there exists a unique set of scalars $s_1, \ldots, s_m$ such that

$$\mathbf{s} = s_1\mathbf{u}_1 + \ldots + s_m\mathbf{u}_m$$

Proof Because $\mathcal{B}$ is a basis for S, the vectors in $\mathcal{B}$ span S, so that every vector can be expressed as a linear combination of vectors in $\mathcal{B}$ in *at least* one way. To show that there can only be one way to express $\mathbf{s}$, let's suppose that there are two, say,

$$\mathbf{s} = r_1\mathbf{u}_1 + \cdots + r_m\mathbf{u}_m \quad \text{and} \quad \mathbf{s} = t_1\mathbf{u}_1 + \cdots + t_m\mathbf{u}_m$$

Then $r_1\mathbf{u}_1 + \cdots + r_m\mathbf{u}_m = t_1\mathbf{u}_1 + \cdots + t_m\mathbf{u}_m$, so that after reorganizing we have

$$(r_1 - t_1)\mathbf{u}_1 + \cdots + (r_m - t_m)\mathbf{u}_m = \mathbf{0}$$

Since $\mathcal{B}$ is a basis, it is also a linearly independent set, and therefore it must be that $r_1 - t_1 = 0, \ldots, r_m - t_m = 0$. Hence $r_1 = t_1, \ldots, r_m = t_m$, so that there is just one way to express $\mathbf{s}$ as a linear combination of the vectors in $\mathcal{B}$. ■ ■

This is important enough to repeat: Theorem 4.9 tells us that every vector in a subspace S can be expressed in *exactly* one way as a linear combination of vectors in a basis $\mathcal{B}$.

Finding a Basis

Frequently, a subspace S is described as the span of a set of vectors; that is, $S = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$. Example 1 demonstrates a way to find a basis in this situation. Before getting to the example, we pause to give a theorem that we will need soon. The proof is left as an exercise.

▶ Recall that two matrices *A* and *B* are equivalent if *A* can be transformed into *B* through a sequence of elementary row operations.

THEOREM 4.10 ▶

Let *A* and *B* be equivalent matrices. Then the subspace spanned by the rows of *A* is the same as the subspace spanned by the rows of *B*.

The next example shows one way to find a basis from a spanning set.

Example 1

Let *S* be the subspace of $\mathbf{R}^4$ spanned by the vectors

u1=[−1231],u2=[−6752],u3=[4−310]

Find a basis for *S*.

Solution Start by using the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ to form the *rows* of a matrix.

▶ Here $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ are the **rows** of *A*.

A=[u1u2u3]=[−1231−67524−310]

Next, use row operations to transform *A* into the equivalent matrix *B* that is in echelon form.

A=[−1231−67524−310]~[−1231051340000]=B

By Theorem 4.10, we know that the subspace spanned by the rows of B is the same as the subspace spanned by the rows of A, so the rows of B span S. Moreover, since B is in echelon form, the nonzero rows are linearly independent (see Exercise 62, Section 2.3). Thus the set

$$\{[-1231],[05134]\}$$

forms a basis for S.

Summarizing this method: To find a basis for $\mathcal{S}=\text{span}\{u1,\ldots,um\}$,

1. Use the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ to form the rows of a matrix A.
2. Transform A to echelon form B.
3. The nonzero rows of B give a basis for S.

If S is a nonzero subspace, then there are infinitely many different sets that can form a basis for S. Before proceeding, we pause to state the following useful result that will be used to show a second method for finding a basis for a subspace S. (The proof is left as an exercise.)

THEOREM 4.11 ▶

Suppose that $U = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_m]$ and $V = [\mathbf{v}_1 \ \ldots \ \mathbf{v}_m]$ are two equivalent matrices. Then any linear dependence that exists among the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ also exists among the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

For example, Theorem 4.11 tells us that

$$\text{if}\, 3v1-2v4+v6=5v2 \,\text{then}\, 3u1-2u4+u6=5u2$$

Theorem 4.11 gives us another way to find a basis from a spanning set.

Example 2

Let S be the subspace of $\mathbf{R}^4$ spanned by the vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ given in Example 1. Find a basis for S.

Solution This time we start by using the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ to form the *columns* of a matrix

▶ Here $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ are the **columns** of A.

A=[u1 u2 u3]=[−1−6427−3351120]

Using row operations to transform A to echelon form, we have

B=[v1 v2 v3]=[16−401−1000000]

The nice thing about the matrix B is that it is not hard to find the dependence relationship among the columns. For instance, we can readily verify that

2v1−v2=v3

Now we apply Theorem 4.11. Since $2\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{v}_3$, then we also have $2\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{u}_3$. Therefore

2[−1231]−[7−652]=[−3410]

For B we have

span{v1,v2,v3}=span{v1,v2}

and $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. Hence it follows that for A,

S=span{u1,u2,u3}=span{u1,u2}

and that $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent. Thus the set

{[−1231],[−6752]}

forms a basis for S.

Summarizing this method: To find a basis for $S = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$,

1. Use the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ to form the columns of a matrix A.
2. Transform A to echelon form B. The pivot columns of B will be linearly independent, and the other columns will be linearly dependent on the pivot columns.
3. The columns of A corresponding to the pivot columns of B form a basis for S.

The method in Example 1 will usually produce a subspace basis that is relatively "simple" in that the basis vectors will contain some zeros. The method in Example 2 produces a basis from a subset of the original spanning vectors, which is sometimes desirable. In general, each method will produce a different basis, showing that a basis need not be unique.

Dimension

Example 1 and Example 2 show that a subspace can have more than one basis. However, note that each basis has two vectors. Although a given nonzero subspace will have more than one basis, the next theorem shows that it has a fixed number of basis vectors.

THEOREM 4.12 ▸

If S is a subspace of $\mathbf{R}^n$, then every basis of S has the same number of vectors.

A proof of this theorem is given at the end of the section.

Since every basis for a subspace S has the same number of vectors, the following definition makes sense.

DEFINITION 4.13 ▶

Dimension

Let S be a subspace of $\mathbf{R}^n$. Then the **dimension** of S is the number of vectors in any basis of S.

Standard Basis

The zero subspace $S = \{\mathbf{0}\}$ has no basis and is defined to have dimension 0. At the other extreme, $\mathbf{R}^n$ is a subspace of itself, and in Example 4, Section 4.1, we showed that $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ spans $\mathbf{R}^n$. It is also clear that these vectors are linearly independent, so that $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ forms a basis—called the **standard basis**—of $\mathbf{R}^n$ (see Figure 3). Thus the dimension of $\mathbf{R}^n$ is n. It can be shown that $\mathbf{R}^n$ is the only subspace of $\mathbf{R}^n$ of dimension (see Exercise 53).

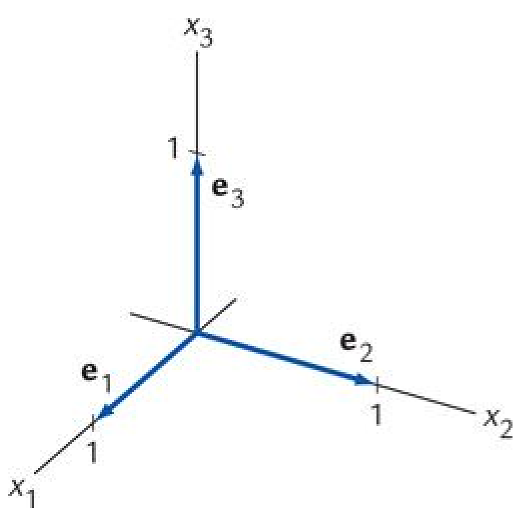


Figure 3 The standard basis for $\mathbf{R}^3$.

Example 3

Suppose that S is the subspace of $\mathbf{R}^5$ given by

$$S = \text{span}\left\{ \begin{bmatrix} -1 \\ 2 \\ 5 \\ -1 \\ -4 \end{bmatrix}, \begin{bmatrix} 3 \\ -6 \\ -15 \\ 3 \\ 12 \end{bmatrix}, \begin{bmatrix} -3 \\ 8 \\ 19 \\ -5 \\ -18 \end{bmatrix}, \begin{bmatrix} 5 \\ -3 \\ -11 \\ -2 \\ -1 \end{bmatrix} \right\}$$

Find the dimension of S.

Solution Since our set has four vectors, we know that the dimension of S will be 4 or less. To find the dimension, we need to find a basis for S. It makes no difference how we do this, so let's use the solution method given in Example 2. Our vectors form the columns of the matrix on the left, with an echelon form given on the right.

$$\begin{bmatrix} -1 & 3 & -3 & 5 \\ 2 & -6 & 8 & -3 \\ 5 & -15 & 19 & -11 \\ -1 & 3 & -5 & -2 \\ -4 & 12 & -18 & 1 \end{bmatrix} \sim \begin{bmatrix} -1 & 3 & -3 & 5 \\ 0 & 0 & 2 & 7 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since the first and third columns of the echelon matrix are the pivot columns, we conclude that the first and third vectors from the original set

$$\left\{ \begin{bmatrix} -1 \\ 2 \\ 5 \\ -1 \\ -4 \end{bmatrix}, \begin{bmatrix} -3 \\ 8 \\ 19 \\ -5 \\ -18 \end{bmatrix} \right\}$$

form a basis for S. Hence the dimension of S is 2.

In many instances it is handy to be able to modify a given set of vectors to serve as a basis. The following theorem gives two cases when this is possible.

THEOREM 4.14

Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set of vectors in a subspace $S \neq \{\mathbf{0}\}$ of $\mathbf{R}^n$.

(a) If $\mathcal{U}$ is linearly independent, then either $\mathcal{U}$ is a basis for S or additional vectors can be added to $\mathcal{U}$ to form a basis for S.

(b) If $\mathcal{U}$ spans S, then either $\mathcal{U}$ is a basis for S or vectors can be removed from $\mathcal{U}$ to form a basis for S.

Proof Taking part (a) first, if $\mathcal{U}$ also spans S, then we are done. If not, then select a vector s_1 from S that is not in the span of $\mathcal{U}$ and form a new set

$$\mathcal{U}1=\{u1,\dots ,um,s1\}$$

Then $\mathcal{U}_1$ must also be linearly independent, for if not then $\mathbf{s}_1$ would be in the span of $\mathcal{U}$. If $\mathcal{U}_1$ spans S, then we are done. If not, select a vector $\mathbf{s}_2$ that is not in the span of $\mathcal{U}_1$ and let

$$\mathcal{U}2=\{u1,\dots ,um,s1,s2\}$$

▶ The process cannot go on indefinitely. Since all of the vectors are in $\boldsymbol{R}^n$, no set can have more than n linearly independent vectors (see Theorem 2.14, Section 2.3).

As before, $\mathcal{U}2$ must be linearly independent. If $\mathcal{U}2$ spans S, then we are done. If not, repeat this procedure again and again, until we finally have a linearly independent set that also spans S, giving a basis.

For part (b), we start with a spanning set. All we need to do is employ the solution method from Example 2, which will give a subset of $\mathcal{U}$ that forms a basis for S. (Or we can use the method described at the beginning of the section, removing one vector at a time until reaching a basis.) ■ ■

Example 4

Expand the set $\mathcal{U}$={[11−2],[32−4]} to a basis for $\mathbf{R}^3$.

Solution Since $\mathbf{R}^3$ has dimension 3, we know that $\mathcal{U}$ does not have enough vectors to be a basis. We can see that the two vectors in $\mathcal{U}$ are linearly independent, so by Theorem 4.14(a) we can expand $\mathcal{U}$ to a basis of $\mathbf{R}^3$. We know that the standard basis {$\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$} forms a basis for $\mathbf{R}^3$, so that

R3=span{[11−2],[32−4],[100],[010],[001]}

Now we form the matrix

A=[1310012010−2−4001]

and then apply the solution method from Example 2, which will give us a basis for $\mathbf{R}^3$. Since we placed the vectors that we want to include in the left columns, we are assured that they will end up among the basis vectors. Employing our usual row operations, we find an echelon form equivalent to *A* is

A=[1310012010−2−4001]~[13100011−1000021]=B

Since the pivots are in the 1st, 2nd, and 4th columns of *B*, referring back to *A* we see that the vectors

{[11−2],[32−4],[010]}

must be linearly independent and span $\mathbf{R}^3$, so the set forms a basis for $\mathbf{R}^3$.

Example 5

The vector $\mathbf{x}_1$ is in the null space of *A*,

x1=[−1−52−4],A=[−33−6−6260−80−84−14−3−7−19210−2−14]

Find a basis for the null space that includes x_1.

Solution In Example 4, we were able to exploit the fact that we knew a basis for $\mathbf{R}^3$. Here we do not know a basis for the null space, so we start by determining the vector form of the general solution to $A\mathbf{x} = \mathbf{0}$ and use the vectors to form the initial basis. We skip the details and just report the news that

$$\text{null(A)=span\{[−1302],[−3120]\}} \tag{1}$$

From this point we follow the procedure in Example 4, by forming the matrix with our given vector $\mathbf{x}_1$ and the two basis vectors in (1), and then finding an echelon form.

$$\text{[−1−1−3−531202−420]∼[−1−1−3012000000]}$$

Since the pivots are in the first two columns, it follows that

$$\text{\{[−1−52−4],[−1302]\}} \tag{2}$$

forms a basis for the null space of A that contains $\mathbf{x}_1$.

Note that (2) is not the only basis containing $\mathbf{x}_1$. For instance, if we reverse the order of the two vectors in (1) and follow the same procedure, we end up with the basis

$$\text{\{[−1−52−4],[−3120]\}}$$

Nullity

The **nullity** of a matrix A is the dimension of the null space of A and is denoted by nullity(A). Thus in Example 5 we have nullity(A) = 2.

If we happen to know the dimension of a subspace S, then the following theorem makes it easier to determine if a given set forms a basis.

THEOREM 4.15 ▶

Let $\mathcal{U}=\{u1,\dots,um\}$ be a set of m vectors in a subspace S of dimension m. If $\mathcal{U}$ is either linearly independent or spans S, then $\mathcal{U}$ is a basis for S.

Proof First, suppose that $\mathcal{U}$ is linearly independent. If $\mathcal{U}$ does not span S, then by Theorem 4.14 we can add additional vectors to $\mathcal{U}$ to form a basis for S. But this gives a basis with more than m vectors, contradicting the assumption that the dimension of S equals m. Hence $\mathcal{U}$ also must span S and so is a basis.

A similar argument can be used to show that if $\mathcal{U}$ spans S then $\mathcal{U}$ is a basis. The details are left as an exercise. ■ ■

Example 6

Suppose that S is a subspace of $\mathbf{R}^3$ of dimension 2 containing the vectors in the set

$$\mathcal{U}=\{[-120],[371]\}$$

Show that $\mathcal{U}$ is a basis for S.

Solution Since S has dimension 2 and $\mathcal{U}$ has two vectors, by Theorem 4.15 all we need to do to show that $\mathcal{U}$ is a basis for S is verify that $\mathcal{U}$ is linearly independent or spans S. We do not know enough about S to show that $\mathcal{U}$ spans S, but since the two vectors are not multiples of each other, $\mathcal{U}$ is a linearly independent set. Hence we can conclude that $\mathcal{U}$ is a basis for S.

Theorem 4.16 and Theorem 4.17 present more properties of the dimension of a subspace that are useful in certain situations. The proofs are left as exercises.

THEOREM 4.16 ▶

Suppose that S_1 and S_2 are both subspaces of $\mathbf{R}^n$ and that S_1 is a subset of S_2. Then $\dim(S_1) \leq \dim(S_2)$, and $\dim(S_1) = \dim(S_2)$ only if $S_1 = S_2$.

THEOREM 4.17 ▶

Let $\mathcal{U}=\{u1,\ldots,um\}$ be a set of vectors in a subspace S of dimension k.

(a) If $m < k$, then $\mathcal{U}$ does not span S.

(b) If $m > k$, then $\mathcal{U}$ is not linearly independent.

The Unifying Theorem, Version 5

The results of this section give us another condition for the Unifying Theorem.

THEOREM 4.18 ▶

(THE UNIFYING THEOREM, VERSION 5) Let $\mathcal{S}=\{a1,\ldots,an\}$ be a set of n vectors in $\mathbf{R}^n$, let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$, and let $T : \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent:

(a) $\mathcal{S}$ spans $\mathbf{R}^n$.

(b) $\mathcal{S}$ is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

(d) T is onto.

(e) T is one-to-one.

(f) A is invertible.

(g) $\ker(T) = \{\mathbf{0}\}$.

(h) $\mathcal{S}$ is a basis for $\mathbf{R}^n$.

▶ This updates the Unifying Theorem, Version 4, given in Section 4.1.

Proof From the Unifying Theorem, Version 4, we know that (a) through (g) are equivalent. By Definition 4.8, (a) and (b) are equivalent to (h), completing the proof. ■ ■

Example 7

Let $x_1, x_2, \ldots, x_n$, be real numbers. The *Vandermonde matrix*, which arises in signal processing and coding theory, is given by

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{bmatrix}$$

Show that if $x_1, x_2, \ldots, x_n$ are distinct, then the columns of V form a basis for $\mathbf{R}^n$.

Solution By the Unifying Theorem, Version 5, we can show that the columns of V form a basis for $\mathbf{R}^n$ by showing that the columns are linearly independent. Given real numbers $a_0, a_1, \ldots, a_{n-1}$ we have

$$a_0\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + a_1\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \ldots + a_{n-1}\begin{bmatrix} x_1^{n-1} \\ x_2^{n-1} \\ \vdots \\ x_n^{n-1} \end{bmatrix} = \begin{bmatrix} a_0 + a_1x_1 + \cdots + a_{n-1}x_1^{n-1} \\ a_0 + a_1x_2 + \cdots + a_{n-1}x_2^{n-1} \\ \vdots \\ a_0 + a_1x_n + \cdots + a_{n-1}x_n^{n-1} \end{bmatrix} \tag{3}$$

If the polynomial $f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1}$, then the right side of (3) is equal to

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}$$

▶ One consequence of the *Fundamental Theorem of Algebra* (proved by Gauss) is that a polynomial of degree m can have at most m distinct roots.

This is the zero vector only if each of $x_1, x_2, \ldots, x_n$ are roots of the polynomial f. But since the roots are distinct and f has degree at most $n - 1$, the only way this can happen is if $f(x) = 0$, the zero polynomial. Hence $a_0 = \cdots = a_{n-1} = 0$, and so the columns of V are linearly independent. Therefore the columns of V form a basis for $\mathbf{R}^n$.

Proof of Theorem 4.12

We state the theorem again:

THEOREM 4.12 ▶

If S is a subspace of $\mathbf{R}^n$, then every basis of S has the same number of vectors.

Proof Suppose that we have a subspace S with two bases of different sizes. The argument that follows can be generalized (this is left as an exercise), but to simplify notation we assume that S has bases

$$\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2\} \quad \text{and} \quad \mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

Since $\mathcal{U}$ spans S, it follows that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_2$ can each be expressed as linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$,

$$\begin{aligned} \mathbf{v}_1 &= c_{11}\mathbf{u}_1 + c_{12}\mathbf{u}_2 \\ \mathbf{v}_2 &= c_{21}\mathbf{u}_1 + c_{22}\mathbf{u}_2 \\ \mathbf{v}_3 &= c_{31}\mathbf{u}_1 + c_{32}\mathbf{u}_2 \end{aligned} \tag{4}$$

Now consider the equation

$$a_1v_1+a_2v_2+a_3v_3=0 \tag{5}$$

Substituting into (5) from (4) for $\mathbf{v}_1$ $\mathbf{v}_2$, and $\mathbf{v}_3$ gives

$$0=a_1(c_{11}u_1+c_{12}u_2)+a_2(c_{21}u_1+c_{22}u_2)+a_3(c_{31}u_1+c_{32}u_2) = (a_1c_{11}+a_2c_{21}+a_3c_{31})u_1+(a_1c_{12}+a_2c_{22}+a_3c_{32})u_2$$

Since $\mathcal{U}$ is linearly independent, we must have

$$a_1c_{11}+a_2c_{21}+a_3c_{31}=0 \quad a_1c_{12}+a_2c_{22}+a_3c_{32}=0$$

Now view a_1, a_2 and a_3 as variables in this homogeneous system. Since there are more variables than equations, the system must have infinitely many solutions. But this means that there are nontrivial solutions to (5), which implies that $\mathcal{V}$ is linearly dependent, a contradiction. (Remember that $\mathcal{V}$ is a basis.) Hence our assumption that there can be bases of two different sizes is incorrect, so all bases for a subspace must have the same number of vectors.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Use the solution method from Example 1 to find a basis for the given subspace and give the dimension.
   (a) $S=\text{span}\{[2-7],[6-14]\}$
   (b) $S=\text{span}\{[1-13],[210],[-31-4]\}$
2. Use the solution method from Example 2 to find a basis for the given subspace and give the dimension.
   (a) $S=\text{span}\{[2-7],[6-14]\}$
   (b) $S=\text{span}\{[1-13],[210],[-31-4]\}$

3. Find a basis for the given subspace by deleting linearly dependent vectors, and give the dimension. Very little computation should be required.

(a) S=span{[2−6],[−13]}

(b) S=span{[1−23],[−24−6],[3−68]}

4. Expand the given set to a basis for $\mathbf{R}^3$.

(a) {[12−1]}

(b) {[−1−22],[120]}

5. Find a basis for the null space *A*, and find the nullity of *A*.

(a) A=[13−230−4]

(b) A=[214−11513]

6. Determine if each statement is true or false, and justify your answer.

(a) If *S* is a subspace of $\mathbf{R}^n$, then dim(*S*) is between 1 and *n*.

(b) Not every subspace of $\mathbf{R}^n$ has a dimension.

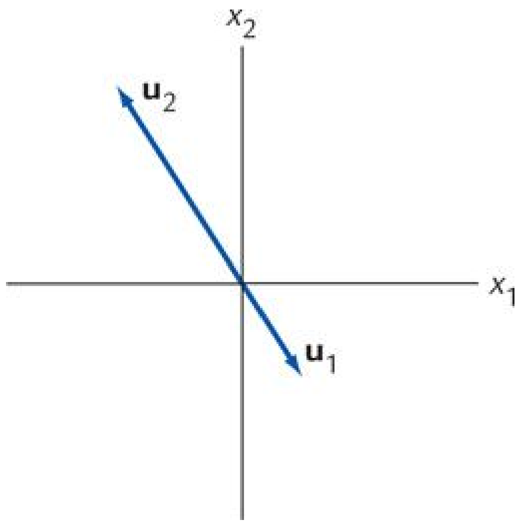
(c) A subspace of dimension two in $\mathbf{R}^3$ is a plane.

(d) If *A* is an $n \times m$ matrix, then $0 \leq \text{nullity}(A) \leq n$.

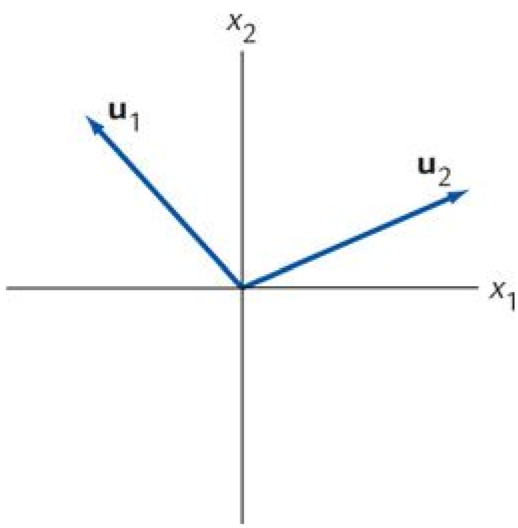
EXERCISES

Exercises 1–4: Determine if the vectors shown form a basis for $\mathbf{R}^2$. Justify your answer.

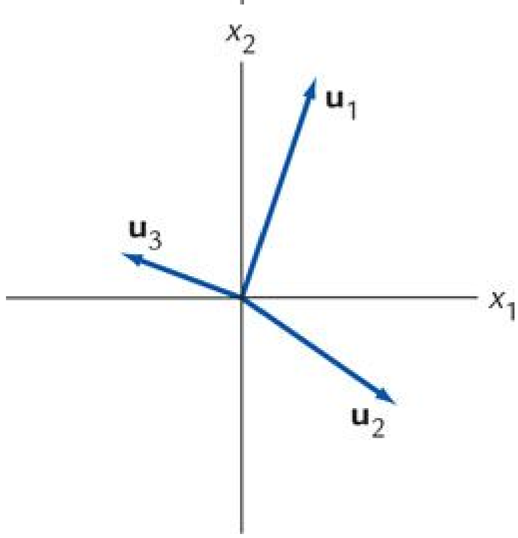
1.


x_2
$\mathbf{u}_2$
x_1
$\mathbf{u}_1$

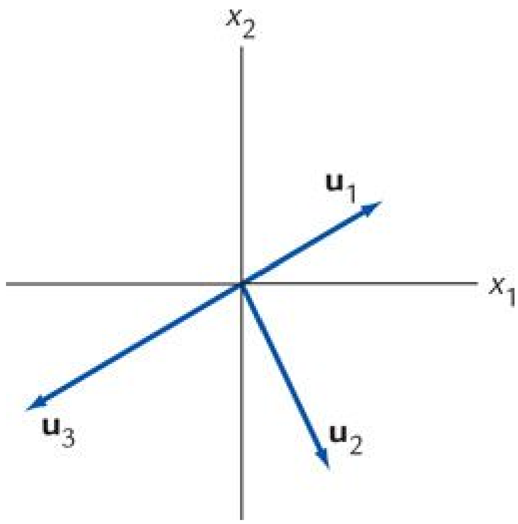
2.


x_2
$\mathbf{u}_1$
$\mathbf{u}_2$
x_1

3.


x_2
$\mathbf{u}_1$
$\mathbf{u}_3$
x_1
$\mathbf{u}_2$

4.



Exercises 5–10: Use the solution method from Example 1 to find a basis for the given subspace and give the dimension.

5. S=span{[1−4],[−520]}

6. S=span{[35],[9−2]}

7. S=span{[13−2],[241],[−11−8]}

8. S=span{[2−13],[4−12],[21−5]}

9. S=span{[1−23−2],[02−51],[2−21−3]}

10. S=span{[10−11],[2102],[0120],[31−13]}

Exercises 11–16: Use the solution method from Example 2 to find a basis for the given subspace and give the dimension.

11. S=span{[13],[4−12]}

12. S=span{[2−6],[−515]}

13. S=span{[124],[01−3],[3−2−1]}

14. S=span{[123],[375],[−1−31]}

15. S=span{[1−102],[2−597],[01−3−1]}

16. S=span{[1031],[42134],[2163],[−11−2−2]}

Exercises 17–22: Find a basis for the given subspace by deleting linearly dependent vectors, and give the dimension. Very little computation should be required.

17. S=span{[2−6],[−39]}

18. S=span{[12−3],[−186]}

19. S=span{[111],[222],[333]}

20. S=span{[1−11],[−55−5],[432]}

21. S=span{[000],[300],[210],[123]}

22. S=span{[1234],[5555],[4221],[6789]}

Exercises 23–24: Expand the given set to form a basis for $\mathbf{R}^2$.

23. {[1−3]}

24. {[04]}

Exercises 25–28: Expand the given set to form a basis for $\mathbf{R}^3$.

25. {[−121]}

26. {[105]}

27. {[13−2],[2−10]}

28. {[213],[326]}

Exercises 29–32: Find a basis for the null space of the given matrix and give nullity(A).

29. A=[−2−513]

30. A=[210111]

31. A=[1121001−3]

32. A=[10−21−10102000014]

FIND AN EXAMPLE Exercises 33–42: Find an example that meets the given specifications.

33. A set of four vectors in $\mathbf{R}^2$ such that, when two are removed, the remaining two are a basis for $\mathbf{R}^2$.

34. A set of three vectors in $\mathbf{R}^4$ such that, when one is removed and then two more are added, the new set is a basis for $\mathbf{R}^4$.

35. A subspace S of $\mathbf{R}^n$ with $\dim(S) = m$, where $0 < m < n$.

36. Two subspaces S_1 and S_2 of $\mathbf{R}^5$ such that $S_1 \subset S_2$ and $\dim(S_1) + 2 = \dim(S_2)$.

37. Two two-dimensional subspaces S_1 and S_2 of $\mathbf{R}^4$ such that $S_1 \cap S_2 = \{\mathbf{0}\}$.

38. Two three-dimensional subspaces S_1 and S_2 of $\mathbf{R}^5$ such that $\dim(S_1 \cap S_2) = 1$.

39. Two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in $\mathbf{R}^3$ that produce the same set of vectors when the methods of Example 1 and Example 2 are applied.

40. A subspace S of dimension two that includes the vector $(1, 3, -2)$.

41. A subspace S of dimension three that includes the vector $(-1, 0, 2, 1)$.

42. Three vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ in $\mathbf{R}^3$ that produce the same set of vectors when the methods of Example 1 and Example 2 are applied.

TRUE OR FALSE Exercises 43–50: Determine if the statement is true or false, and justify your answer.

43.

(a) If $\mathbf{s}_1$ and $\mathbf{s}_2$ are in the same basis for a subspace S, then so is $\mathbf{s}_1 + \mathbf{s}_2$.

(b) If S_1 is a proper subspace of another subspace S_2 and $\dim(S_2) = 1$, then S_1 $\{\mathbf{0}\}$.

44.

(a) If S_1 and S_2 are subspaces of $\mathbf{R}^n$ of the same dimension, then $S_1 = S_2$.

(b) If $S = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, then $\dim(S) = 3$.

45.

(a) If a set of vectors $\mathcal{U}$ spans a subspace S, then vectors can be added to $\mathcal{U}$ to create a basis for S.

(b) If a set of vectors $\mathcal{U}$ is linearly independent in a subspace S, then vectors can be added to $\mathcal{U}$ to create a basis for S.

46.

(a) If a set of vectors $\mathcal{U}$ spans a subspace S, then vectors can be removed from $\mathcal{U}$ to create a basis for S.

(b) If a set of vectors $\mathcal{U}$ is linearly independent in a subspace S, then vectors can be removed from $\mathcal{U}$ to create a basis for S.

47.

(a) Three nonzero vectors that lie in a plane in $\mathbf{R}^3$ might form a basis for $\mathbf{R}^3$.

(b) If S_1 is a subspace of dimension 3 in $\mathbf{R}^4$, then there cannot exist a subspace S_2 of $\mathbf{R}^4$ such that $S_1 \subset S_2 \subset \mathbf{R}^4$ but $S_1 \neq S_2$ and $S_2 \neq \mathbf{R}^4$.

48.

(a) The set $\{\mathbf{0}\}$ forms a basis for the zero subspace.

(b) $\mathbf{R}^n$ has exactly one subspace of dimension m for each of $m = 0, 1, 2, \ldots, n$.

49.

(a) Let $m > n$. Then any set $\mathcal{U}$={u1,u2,… ,um} in $\mathbf{R}^n$ can form a basis for $\mathbf{R}^n$ if the correct $m - n$ vectors are removed from $\mathcal{U}$.

(b) Let $m < n$. Then any set $\mathcal{U}$={u1,u2,… ,um} in $\mathbf{R}^n$ can form a basis for $\mathbf{R}^n$ if the correct $n - m$ vectors are added to $\mathcal{U}$.

50.

(a) If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a basis for $\mathbf{R}^3$, then span$\{\mathbf{u}_1, \mathbf{u}_2\}$ is a plane.

(b) The nullity of a matrix A is the same as the dimension of the subspace spanned by the columns of A.

51. Suppose that S_1 and S_2 are nonzero subspaces, with S_1 contained inside S_2. Suppose that $\dim(S_2) = 3$.

(a) What are the possible dimensions of S_1?

(b) If $S_1 \neq S_2$, then what are the possible dimensions of S_1?

52. Suppose that S_1 and S_2 are nonzero subspaces, with S_1 contained inside S_2. Suppose that $\dim(S_2) = 4$.

(a) What are the possible dimensions of S_1?

(b) If $S_1 \neq S_2$, then what are the possible dimensions of S_1?

53. Show that the only subspace of $\mathbf{R}^n$ that has dimension n is $\mathbf{R}^n$.

54. Explain why $\mathbf{R}^n$ ($n > 1$) has infinitely many subspaces of dimension 1.

55. Prove the converse of Theorem 4.9: If every vector $\mathbf{s}$ of a subspace S can be written uniquely as a linear combination of the vectors $\mathbf{s}_1, \ldots, \mathbf{s}_m$ (all in S), then the vectors form a basis for S.

56. Complete the proof of Theorem 4.15: Let $\mathcal{U}$={u1,… ,um} be a set of m vectors in a subspace S of dimension m. Show that if $\mathcal{U}$ spans S, then $\mathcal{U}$ is a basis for S.

57. Prove Theorem 4.16: Suppose that S_1 and S_2 are both subspaces of $\mathbf{R}^n$, with S_1 a subset of S_2. Then $\dim(S_1) \leq \dim(S_2)$, and $\dim(S_1) = \dim(S_2)$ only if $S_1 = S_2$.

58. Prove Theorem 4.17: Let $\mathcal{U}$={u1,… ,um} be a set of vectors in a subspace S of dimension k.

(a) If $m < k$, show that $\mathcal{U}$ does not span S.

(b) If $m > k$, show that $\mathcal{U}$ is not linearly independent.

59. Suppose that a matrix A is in echelon form. Prove that the nonzero rows of A are linearly independent.

60. If the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a basis for $\mathbf{R}^3$ an

A=[u1u2u3],

what is nullity(A)?

61. Suppose that S_1 and S_2 are subspaces of $\mathbf{R}^n$, with $\dim(S_1) = m_1$ and $\dim(S_2) = m_2$. If S_1 and S_2 have only $\mathbf{0}$ in common, then what is the maximum value of $m_1 + m_2$?

62. Prove Theorem 4.10: Let A and B be equivalent matrices. Then the subspace spanned by the rows of A is the same as the subspace spanned by the rows of B.

63. Prove Theorem 4.11: Suppose that $U = [\mathbf{u}_1 \dots \mathbf{u}_m]$ and $V = [\mathbf{v}_1 \dots \mathbf{v}_m]$ are two equivalent matrices. Then any linear dependence that exists among the vectors $\mathbf{u}_1, \dots, \mathbf{u}_m$ also exists among the vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$.

64. Give a general proof of Theorem 4.12: If S is a subspace of $\mathbf{R}^n$, then every basis of S has the same number of vectors.

C Exercises 65–66: Determine if the given set of vectors is a basis of $\mathbf{R}^3$. If not, then determine the dimension of the subspace spanned by the vectors.

65. {[2−15],[−34−2],[−5104]}

66. {[42−7],[−15−3],[37−9]}

Exercises 67–68: Determine if the given set of vectors is a basis of $\mathbf{R}^4$. If not, then determine the dimension of the subspace spanned by the vectors.

67. {[301−2],[2−450],[−2704],[−25−54]}

68. {[60−52],[5−113],[−341−5],[7−268]}

Exercises 69–70: Determine if the given set of vectors is a basis of $\mathbf{R}^5$. If not, then determine the dimension of the subspace spanned by the vectors.

69. {[11−111],[−1012−1],[21−212],[−2121−2],[12−101]}

70. {[12345],[23451],[34512],[45123],[51234]}

4.3 Row and Column Spaces

In Example 7 of Section 4.2, it was shown that if $x_1, \ldots, x_n$ are distinct real numbers, then the columns of the Vandermonde matrix

V=[1x1x12⋯x1n−11x2x22⋯x2n−1⋮⋮⋮⋱⋮1xnxn2⋯xnn−1]

form a basis for $\mathbf{R}^n$. But suppose that the x_i's are not distinct. Can we tell if the columns are linearly independent or linearly dependent? One result that we shall develop will make this question easy to answer.

In this section we round out our knowledge of subspaces of $\mathbf{R}^n$. As we have seen, subspaces arise naturally in the context of a matrix. For instance, suppose that

A=[1−275−2−1−9−7113−8−4]

Row Vectors

The **row vectors** of A come from viewing the rows of A as vectors. For the above matrix A, the set of row vectors is

{[1−275],[−2−1−9−7],[213−8−4]}

Column Vectors

Similarly, the **column vectors** of A come from viewing the columns of A as vectors, so in this case we have

{[1−21],[−2−113],[7−9−8],[5−7−4]}

Taking the span of the row or column vectors yields the subspaces defined below.

DEFINITION 4.19 ▶

Row Space

Column Space

Let A be an $n \times m$ matrix.

(a) The **row space** of A is the subspace of $\mathbf{R}^m$ spanned by the row vectors of A and is denoted by row(A).

(b) The **column space** of A is the subspace of $\mathbf{R}^n$ spanned by the column vectors of A and is denoted by col(A).

In Section 4.2 we proved Theorem 4.10 and Theorem 4.11, which concern the rows and columns of matrices and can be used to find a basis for a subspace. Theorem 4.20 is a reformulation of those theorems, stated in terms of row and column spaces.

THEOREM 4.20 ▶

Let A be a matrix and B an echelon form of A.

(a) The nonzero rows of B form a basis for row(A).

(b) The columns of A corresponding to the pivot columns of B form a basis for col(A).

Example 1

Find a basis and the dimension for the row space and the column space of A.

A=[1−275−2−1−9−7113−8−4]

Solution To use Theorem 4.20, we start by finding an echelon form of A, which is given by

$$A=\begin{bmatrix} 1 & -2 & 7 & 5 \\ -2 & -1 & -9 & -7 \\ 1 & 13 & -8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 7 & 5 \\ 0 & -5 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B$$

By Theorem 4.20(a), we know that a basis for the row space of A is given by the nonzero rows of B,

$$\left\{ \begin{bmatrix} 1 & -2 & 7 & 5 \end{bmatrix}, \begin{bmatrix} 0 & -5 & 5 & 3 \end{bmatrix} \right\}$$

By Theorem 4.20(b), we know that a basis for the column space of A is given by the columns of A corresponding to the pivot columns of B, which in this case are the first and second columns. Thus a basis for col(A) is

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 13 \end{bmatrix} \right\}$$

Since both row(A) and col(A) have two basis vectors, the dimension of both subspaces is 2.

In Example 1, the row space and the column space of A have the same dimension. This is not a coincidence.

THEOREM 4.21 ▸

For any matrix A, the dimension of the row space equals the dimension of the column space.

Proof Given a matrix A, use the usual row operations to find an equivalent echelon form matrix B. From Theorem 4.20(a), we know that the dimension of the row space of A is equal to the number of nonzero rows of B. Next note that each nonzero row of B has exactly one pivot, and that different rows have pivots in different columns. Thus the number of pivot columns equals the number of nonzero rows. But by Theorem 4.20(b), the number of pivot columns of B

equals the number of vectors in a basis for the column space of A. Thus the dimension of the column space is equal to the number of nonzero rows of B, and so the dimensions of the row space and column space are the same. ■ ■

Now let's return to the question about the Vandermonde matrix from the start of the section.

Example 2

Suppose that two or more of x_1, … , x_n are the same. Are the columns of

V=[1x1x12···x1n−11x2x11···x1n−1⋮⋮⋮⋱⋮1xnxn2···xnn−1]

linearly independent or linearly dependent?

Solution If two or more of $x_1,\ldots x_n$ are the same, then two or more of the rows of V are the same. Hence the rows of V are linearly dependent, so by the Unifying Theorem the rows of V do not span $\mathbf{R}^n$. Therefore the dimension of row(V) is less than n, and thus by Theorem 4.21 the dimension of col(V) is less than n. Finally, again by the Unifying Theorem applied to the columns of V, we conclude that the columns are linearly dependent.

Because the dimensions of the row and column spaces for a given matrix A are the same, the following definition makes sense.

DEFINITION 4.22 ▶

Rank of a Matrix

The **rank** of a matrix *A* is the dimension of the row (or column) space of *A*, and is denoted by rank(*A*).

Example 3

Find the rank and the nullity for the matrix

$$A=[1-230-12-47-333-683-8]$$

Solution Applying the standard row operation procedure to *A* yields the echelon form

$$A=[1-230-12-47-333-683-8]\sim[1-230-1001-3500000]=B$$

▶ Recall that the nullity is the dimension of the null space.

Since *B* has two nonzero rows, the rank of *A* is 2. To find the nullity of *A*, we need to determine the dimension of the subspace of solutions to $A\mathbf{x} = \mathbf{0}$. Adding a column of zeros to *A* and *B* gives the augmented matrix for $A\mathbf{x} = \mathbf{0}$ and the corresponding echelon form. (Why?)

$$[1-230-102-47-3303-683-80]\sim[1-230-10001-350000000]$$

The matrix on the right corresponds to the system

$$x1-2x3+3x3\ -x5=0\ x3-3x4+5x5=0 \qquad (1)$$

For this system, x_2, x_4, and x_5 are free variables, so we assign the parameters $x_2 = s_1$, $x_4 = s_2$, and $x_5 = s_3$. Back substitution gives us

$$x3=3x4-5x5=3s2-5s3x1=2x2-3x3+x5=2s1-3(3s2-5s3)+s3=2s1-9s2+16s3$$

▶ Once we know that the system $Ax = \mathbf{0}$ has three free variables, we can conclude that nullity(*A*) = 3. For the sake of completeness, we continue to the vector form of the solution.

In vector form, the general solution is

x=s1[21000]+s2[−90310]+s3[160−501]

The three vectors in the general solution form a basis for the null space, which shows that nullity(*A*) = 3.

Let's look at another example and see if a pattern emerges.

Example 4

Determine the rank and nullity for the matrix *A* given in Example 5 of Section 4.2.

Solution In Example 5, Section 4.2, we showed that a basis for the null space of *A* is given by

{[−1302],[−3120]}

so that nullity(*A*) = 2. Since we did not show an echelon form for *A* earlier, we report it at right.

A=
[−33−6−6260−80−8412−3−7−19210−2−14]∼[1−12202−1−3000000
000000]

From the echelon form, we see that the rank of *A* is 2.

Let's review what we have seen:

- Example 3: rank(*A*) = 2, nullity(*A*) = 3, total number of columns is 5.
- Example 4: rank(*A*) = 2, nullity(*A*) = 2, total number of columns is 4.

In both cases, rank(A) + nullity(A) equals the number of columns of A. This is not a coincidence.

THEOREM 4.23 ▶

(RANK—NULLITY THEOREM) Let A be an $n \times m$ matrix. Then

$$\text{rank}(A)+\text{nullity}(A)=m.$$

Proof Transform A to echelon form B.

- The rank of A is equal to the number of nonzero rows of B. Each nonzero row has a pivot, and each pivot appears in a different column. Hence the number of pivot columns equals rank(A).
- Every nonpivot column corresponds to a free variable in the system $A\mathbf{x} = \mathbf{0}$. Each free variable becomes a parameter, and each parameter is multiplied times a basis vector of null(A). (This is shown in detail in Example 3.) Therefore the number of nonpivot columns equals nullity(A).

Since the number of pivot columns plus the number of nonpivot columns must equal the total number of columns m, we have

$$\text{rank}(A)+\text{nullity}(A)=m$$

■■

Example 5

Suppose that A is a 5×13 matrix and that $T(\mathbf{x}) = A\mathbf{x}$. If the dimension of the kernel of T is 9, what is the dimension of the range of T?

Solution Since Theorem 4.23 is expressed in terms of the properties of a matrix A, we first convert the given information into equivalent statements about A. We are told that the dimension of $\ker(T)$ equals 9. Since $\ker(T) = \text{null}(A)$, then $\text{nullity}(A) = 9$. By Theorem 4.23, $m - \text{nullity}(A) = \text{rank}(A)$, so $\text{rank}(A) = 4$ because A has 13 columns. Recall that $\text{range}(T)$ is equal to the span of the columns of A (Theorem 3.3), which is the same as $\text{col}(A)$. Therefore the dimension of $\text{range}(T)$ is 4.

Example 6

Find a linear transformation T that has kernel equal to $\text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where

$$x1=[10-21],x2=[0132]$$

Solution Since T is a linear transformation, we know that there exists a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$. Since the kernel of T equals the null space of A, another way to state our problem is that we need a matrix A such that $\text{null}(A) = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$.

To get us started, since $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent (why?), they form a basis for $\text{null}(A)$, and so $\text{nullity}(A) = 2$. Moreover, A must have four columns because $\mathbf{x}_1$ and $\mathbf{x}_2$ are in $\mathbf{R}^4$. Thus $\text{rank}(A) = 4 - 2 = 2$ by the Rank–Nullity Theorem. This tells us that A must have at least two rows, so let's assume that A has the form and see if we can solve the problem.

$$A=[abcdefgh]$$

In order for $\mathbf{x}_1$ and $\mathbf{x}_2$ to be in $\text{null}(A)$, we must have $A\mathbf{x}_1 = \mathbf{0}$ and $A\mathbf{x}_2 = \mathbf{0}$. Computing the first entry of $A\mathbf{x}_1$ and $A\mathbf{x}_2$ and setting each equal to zero produces the linear system

$$a -2c+d=0 \quad b+3c+2d=0$$

The system is in echelon form, and after back substituting we find that the general solution is given by

$$[abcd]=s1[2-310]+s2[-1-201] \qquad (2)$$

There are many choices for s_1 and s_2, but let's make it easy on ourselves by setting $s_1 = 1$ and $s_2 = 0$, so that $a = 2$, $b = -3$, $c = 1$, and $d = 0$. This gives us half of A,

$$A=[2-310efgh]$$

In order to find e, f, g, and h, we could repeat the same analysis. However, we will just get the same answers, with e, f, g, and h replacing a, b, c, and d. So we can set $s_1 = 0$ and $s_2 = 1$ and use the second vector in (2) as the second row of A,

$$A=[2-310-1-201]$$

Since the two rows of A are linearly independent, we know that $\text{rank}(A) = 2$. This ensures that $\text{nullity}(A) = 2$, so that $\text{null}(A) = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$.

We wrap up this subsection with a theorem that relates row and column spaces to other topics that we previously encountered. The proofs of both parts are left as exercises.

THEOREM 4.24 ▶

Let A be an $n \times m$ matrix and $\mathbf{b}$ a vector in $\mathbf{R}^n$.

(a) The system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is in the column space of A.

(b) The system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if $\mathbf{b}$ is in the column space of A and the columns of A are linearly independent.

The Unifying Theorem, Version 6

We can add three more conditions to the Unifying Theorem based on our work in this section.

THEOREM 4.25 ▶

(THE UNIFYING THEOREM, VERSION 6) Let $\mathcal{S}$={a1,… ,an} be a set of n vectors in $\mathbf{R}^n$, let $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$, and let $T : \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent:

(a) $\mathcal{S}$ spans $\mathbf{R}^n$.

(b) $\mathcal{S}$ is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

(d) T is onto.

(e) T is one-to-one.

(f) A is invertible.

(g) $\ker(T) = \{\mathbf{0}\}$.

(h) $\mathcal{S}$ is a basis for $\mathbf{R}^n$.

(i) $\mathrm{col}(A) = \mathbf{R}^n$.

(j) $\mathrm{row}(A) = \mathbf{R}^n$.

(k) $\mathrm{rank}(A) = n$.

▶ This updates the Unifying Theorem, Version 5, given in Section 4.2.

▶ Practice problems can also be used as additional examples.

Proof From the Unifying Theorem, Version 5, we know that (a) through (h) are equivalent. Theorem 4.21 and Definition 4.22 imply that (i), (j), and (k) are equivalent, and by definition (a) and (i) are equivalent. Hence the 11 conditions are all one big equivalent family. ■

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find bases for the column space of A, the row space of A, and the null space of A. Verify that the Rank–Nullity Theorem holds.
   (a) A=[14−6−2−812]
   (b) A=[2103−1421032−1]
2. Suppose that A is a 7×12 matrix, and that the reduced echelon form of A has two zero rows.
   (a) What is the dimension of row(A)?
   (b) What is the dimension of col(A)?
   (c) What is nullity(A)?
   (d) What is rank(A)?
3. Suppose that A is a matrix with row(A) a subspace of $\mathbf{R}^8$ and rank(A) = 5. What is the nullity of A?
4. Determine if each statement is true or false, and justify your answer.
   (a) If A is not a square matrix, then rank(A) ≠ nullity(A).
   (b) If A is $n \times m$, then rank(A) = m.
   (c) If A is $n \times m$ with $n > m$, then nullity(A) > 0.
   (d) If A is $n \times m$ with $n < m$, then rank(A) ≤ n.

EXERCISES

Exercises 1–4: Find bases for the column space of A, the row space of A, and the null space of A. Verify that the Rank–Nullity Theorem holds. (To make your job easier, an equivalent echelon form is given for each matrix.)

1. A=[1−32−250−38−2]∼[10−1001−4000]
2. A=[10−4−3−211350151]∼[10−4−3015−10000]
3. A=[120−1−2−3−1414−24222−4]∼[102001−1000010000]
4. A=[1−25241−4021−20311]∼[100010001000000]

Exercises 5–8: Find bases for the column space of *A*, the row space of *A*, and the null space of *A*. Verify that the Rank–Nullity Theorem holds.

5. A=[1−222−23−1−20]
6. A=[12−1121−141−415]
7. A=[1320311711140]
8. A=[14−11311−14152328−22]

Exercises 9–12: Find all values of *x* so that rank(*A*) = 2.

9. A=[1−4−2x]
10. A=[234−1x0]
11. A=[−121311143x]
12. A=[−210701x910−31]
13. Suppose that *A* is a 6 × 8 matrix. If the dimension of the row space of *A* is 5, what is the dimension of the column space of *A*?
14. Suppose that *A* is a 9 × 7 matrix. If the dimension of col(*A*) is 5, what is the dimension of row(*A*)?
15. Suppose that *A* is a 4 × 7 matrix that has an echelon form with one zero row. Find the dimensions of the row space of *A*, the column space of *A*, and the null space of *A*.
16. Suppose that *A* is a 6 × 11 matrix that has an echelon form with two zero rows. Find the dimensions of the row space of *A*, the column space of *A*, and the null space of *A*.
17. A 8 × 5 matrix *A* has a null space of dimension 3. What is the rank of *A*?

18. A 5 × 13 matrix *A* has a null space of dimension 10. What is the rank of *A*?

19. A 7 × 11 matrix *A* has rank 4. What is the dimension of the null space of *A*?

20. A 14 × 9 matrix *A* has rank 7. What is the dimension of the null space of *A*?

21. Suppose that *A* is a 6 × 11 matrix and that $T(\mathbf{x}) = A\mathbf{x}$. If nullity(*A*) = 7, what is the dimension of the range of *T*?

22. Suppose that *A* is a 17 × 12 matrix and that $T(\mathbf{x}) = A\mathbf{x}$. If rank(*A*) = 8, what is the dimension of the kernel of *T*?

23. Suppose that *A* is a 13 × 5 matrix and that $T(\mathbf{x}) = A\mathbf{x}$. If *T* is one-to-one, then what is the dimension of the null space of *A*?

24. Suppose that *A* is a 5 × 13 matrix and that $T(\mathbf{x}) = A\mathbf{x}$. If *T* is onto, then what is the dimension of the null space of *A*?

25. Suppose that *A* is a 5 × 13 matrix. What is the maximum possible value for the rank of *A*, and what is the minimum possible value for the nullity of *A*?

26. Suppose that *A* is a 12 × 7 matrix. What is the minimum possible value for the rank of *A*, and what is the maximum possible value for the nullity of *A*?

Exercises 27–32: Suppose that *A* is a 9 × 5 matrix and that *B* is an equivalent matrix in echelon form.

27. If *B* has three nonzero rows, what is rank(*A*)?

28. If *B* has two pivot columns, what is rank(*A*)?

29. If *B* has three nonzero rows, what is nullity(*A*)?

30. If *B* has one pivot column, what is nullity(*A*)?

31. If rank(*A*) = 3, how many nonzero rows does *B* have?

32. If rank(*A*) = 1, how many pivot columns does *B* have?

33. Suppose that *A* is an $n \times m$ matrix, that col(*A*) is a subspace of $\mathbf{R}^7$, and that row(*A*) is a subspace of $\mathbf{R}^5$. What are the dimensions of *A*?

34. Suppose that A is an $n \times m$ matrix, with rank(A) = 4, nullity(A) = 3, and col(A) a subspace of $\mathbf{R}^5$. What are the dimensions of A?

FIND AN EXAMPLE Exercises 35–46: Find an example that meets the given specifications.

35. A 2×3 matrix A with nullity(A) = 1.

36. A 4×3 matrix A with nullity(A) = 0.

37. A 9×4 matrix A with rank(A) = 3.

38. A 5×7 matrix A with rank(A) = 4.

39. A matrix A with rank(A) = 3 and nullity(A) = 1.

40. A matrix A with rank(A) = 2 and nullity(A) = 2.

41. A 2×2 matrix A such that row(A) = col(A).

42. A 3×3 matrix A such that row(A) = col(A).

43. A 3×3 matrix whose null space is a plane.

44. A 3×3 matrix whose null space is a line.

45. A 3×3 matrix whose column space is a plane.

46. A 3×3 matrix whose row space is a line.

TRUE OR FALSE Exercises 47–52: Determine if the statement is true or false, and justify your answer.

47.

(a) If A is a matrix, then the dimension of the row space of A is equal to the dimension of the column space of A.

(b) If A is a square matrix, then row(A) = col(A).

48.

(a) The rank of a matrix A cannot exceed the number of rows of A.

(b) If A and B are equivalent matrices, then row(A) = row(B).

49.

(a) If A and B are equivalent matrices, then col(A) = col(B).

(b) If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system, then $\mathbf{b}$ is in row(A).

50.

(a) If $\mathbf{x}_0$ is a solution to $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x}_0$ is in row(A).

(b) If A is a 4×13 matrix, then the nullity of A could be equal to 5.

51.

(a) Suppose that A is a 9×5 matrix and that $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation. Then T can be onto.

(b) Suppose that A is a 9×5 matrix and that $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation. Then T can be one-to-one.

52.

(a) Suppose that A is a 4×13 matrix and that $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation. Then T can be onto.

(b) Suppose that A is a 4×13 matrix and that $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation. Then T can be one-to-one.

53. Prove that if A is an $n \times m$ matrix, then $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.

54. Prove that if A is an $n \times m$ matrix and $c \neq 0$ is a scalar, then $\operatorname{rank}(A) = \operatorname{rank}(cA)$.

55. Prove that if A is an $n \times m$ matrix and $\operatorname{rank}(A) < m$, then $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

56. Prove that if A is an $n \times m$ matrix and $\operatorname{rank}(A) < n$, then the reduced row echelon form of A has a row of zeros.

57. Suppose that A is an $n \times m$ matrix with $n \neq m$. Prove that either $\operatorname{nullity}(A) > 0$ or $\operatorname{nullity}(A^T) > 0$ (or both).

58. Prove Theorem 4.24: Let A be an $n \times m$ matrix and $\mathbf{b}$ a vector in $\mathbf{R}^n$.

(a) Show that the system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is in the column space of A.

(b) Show that the system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if $\mathbf{b}$ is in the column space of A and the columns of A are linearly independent.

Exercises 59–62: Determine the rank and nullity of the given matrix.

59. A=[1324−115−3−3−428−17−5]

60. A=[2−101521−4−1−4−16−8−5−29]

61. A=[482351919571335113]

62. A=[43215−1322136710316457−2−215]

4.4 Change of Basis

We have seen that there are numerous different bases for $\mathbf{R}^n$ (or a subspace of $\mathbf{R}^n$). In this section we develop a general procedure for changing from one basis to another.

▶ This section is optional and can be omitted without loss of continuity.

To develop a systematic procedure for changing from one basis to another, let's start with the standard basis for $\mathbf{R}^2$

$$\mathcal{S}=\{e1,e2\}=\{[10],[01]\}$$

Then, for example, if x=[3−2], we have

$$x=3e1-2e2$$

We can view the entries in [3−2] as the coefficients needed to write **x** as a linear combination of $\{\mathbf{e}_1, \mathbf{e}_2\}$.

Now suppose that

$$\mathcal{B}=\{u1,u2\}=\{[27],[14]\}$$

Then $\mathcal{B}$ is another basis for $\mathbf{R}^2$. It is not difficult to verify that if x= [3−2] then $\mathbf{x} = 14\mathbf{u}_1 - 25\mathbf{u}_2$. The compact notation that we use to express this relationship is

$$[\ x]\mathcal{B}=[14\text{-}25]$$

▶ Although we continue to use set notation for bases, in this section the order of the vectors in the basis also matters.

More generally, suppose that $\mathcal{B}$={u1,… ,un} forms a basis for $\mathbf{R}^n$. If $\mathbf{y} = y_1\mathbf{u}_1 + \dots + y_n\mathbf{u}_n$, then we write

$$[\ y]\mathcal{B}=[y1 \vdots yn]$$

Coordinate Vector

for the **coordinate vector of y with respect to** $\mathcal{B}$. As above, the coordinate vector contains the coefficients required to express **y** as a linear combination of the vectors in basis $\mathcal{B}$.

Now define the $n \times n$ matrix $U = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_n]$. Multiplying as usual, we have

$$U\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = y_1 u_1 + \cdots + y_n u_n$$

Hence multiplying by U transforms the coordinate vector with respect to $\mathcal{B}$ to the standard basis. Put symbolically, we have $\mathbf{y} = U[\mathbf{y}]_{\mathcal{B}}$.

Example 1

Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ -1 \end{bmatrix} \right\} \text{and} \ [x]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$$

Then $\mathcal{B}$ forms a basis for $\mathbf{R}^3$. Find **x** with respect to the standard basis $\mathcal{S}$.

Solution Start by setting $U = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 0 & 5 \\ -2 & 1 & -1 \end{bmatrix}$. Then we have

$$x = U[x]_{\mathcal{B}} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 0 & 5 \\ -2 & 1 & -1 \end{bmatrix}\begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ -1 \\ 6 \end{bmatrix}$$

Thus $x \begin{bmatrix} 8 \\ -1 \\ 6 \end{bmatrix}$ with respect to the standard basis.

When working with coordinate vectors, there is potential for confusion about which basis is in use.

NOTATION CONVENTION

If no subscript is given on a vector, then it is expressed with respect to the standard basis. For any other basis, a subscript will be included.

Change of Basis Matrix

The matrix U in Example 1, called a **change of basis matrix**, allows us to switch from a basis $\mathcal{B}$ to the standard basis $\mathcal{S}$. Example 2 shows how to go the other direction, from the standard basis $\mathcal{S}$ to another basis $\mathcal{B}$.

Example 2

Let

$$\mathbf{x}=\begin{bmatrix}3\\-2\end{bmatrix} \text{ and } \mathcal{B}=\{\mathbf{u}_1,\mathbf{u}_2\}=\left\{\begin{bmatrix}2\\7\end{bmatrix},\begin{bmatrix}1\\4\end{bmatrix}\right\}$$

as before, and set $U = [\mathbf{u}_1\ \mathbf{u}_2]$. Find the change of basis matrix from $\mathcal{S}$ to $\mathcal{B}$.

Solution To write $\mathbf{x}$ in terms of $\mathcal{B}$ we need to find x_1 and x_2 such that

$$x_1\mathbf{u}_1+x_2\mathbf{u}_2=\begin{bmatrix}3\\-2\end{bmatrix} \Rightarrow U\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}3\\-2\end{bmatrix}$$

Since the columns of U are linearly independent, U is invertible and hence the solution is

$$\begin{bmatrix}x_1\\x_2\end{bmatrix}=U^{-1}\begin{bmatrix}3\\-2\end{bmatrix}$$

▶ Recall the Quick Formula,
$\begin{bmatrix}a & b\\c & d\end{bmatrix}^{-1}=\frac{1}{ad-bc}\begin{bmatrix}d & -b\\-c & a\end{bmatrix}$

This shows that the change of basis matrix is U^{-1}. To test this, we compute (using the Quick Formula from Section 3.3)

$$U^{-1}=\begin{bmatrix}4 & -1\\-7 & 2\end{bmatrix}$$

Then

$$U^{-1}\begin{bmatrix}3\\-2\end{bmatrix}=\begin{bmatrix}4 & -1\\-7 & 2\end{bmatrix}\begin{bmatrix}3\\-2\end{bmatrix}=\begin{bmatrix}14\\-25\end{bmatrix}$$

which tells us that

$$[\ x]_{\mathcal{B}}=[14-25]$$

as we saw previously.

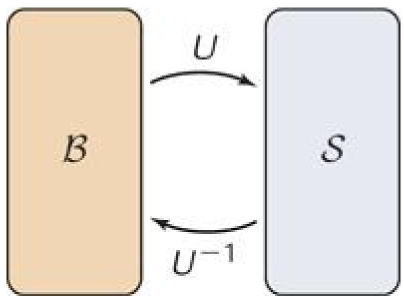


Figure 1 Change of basis between $\mathcal{S}$ and $\mathcal{B}$.

Example 2 illustrates a general fact: If U is the change of basis matrix from $\mathcal{B}$ to $\mathcal{S}$, then U^{-1} is the change of basis matrix from $\mathcal{S}$ to $\mathcal{B}$. This approach generalizes to R^n and is summarized in Theorem 4.26. The full proof is left as an exercise; a graphical depiction is given in Figure 1.

THEOREM 4.26 ▸

Let $\mathbf{x}$ be expressed with respect to the standard basis, and let $\mathcal{B}=\{u1,\dots,un\}$ be any basis for $\mathbf{R}^n$. If $U = [\mathbf{u}_1 \ \dots \ \mathbf{u}_n]$, then

(a) $x=U[\ x]_{\mathcal{B}}$

(b) $[\ x]_{\mathcal{B}}=U-1x$

Example 3

Let

$$x=[344] \text{and} \mathcal{B}=\{[101],[1-30],[212]\}$$

Find $[\mathbf{x}]_{\mathcal{B}}$, the coordinate vector of **x** with respect to the basis $\mathcal{B}$.

Solution We start by letting U be the matrix with columns given by the vectors in $\mathcal{B}$,

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

Then by Theorem 4.26, we have

$$[\mathbf{x}]_{\mathcal{B}}=U^{-1}\mathbf{x}=\begin{bmatrix}-6 & -2 & 7\\ 1 & 0 & -1\\ 3 & 1 & -3\end{bmatrix}\begin{bmatrix}3\\ 4\\ 4\end{bmatrix}=\begin{bmatrix}2\\ -1\\ 1\end{bmatrix}$$

Two Nonstandard Bases

Now suppose that we have nonstandard bases $\mathcal{B}_1=\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ and $\mathcal{B}_2=\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ for $\mathbf{R}^n$. How do we get from $[\mathbf{x}]_{\mathcal{B}_1}$ to $[\mathbf{x}]_{\mathcal{B}_2}$—that is, from **x** expressed with respect to $\mathcal{B}_1$ to **x** expressed with respect to $\mathcal{B}_2$?

The simple solution uses two steps. We apply Theorem 4.26 twice, first to go from $[\mathbf{x}]_{\mathcal{B}_1}$ to $[\mathbf{x}]_{\mathcal{S}}$, and then to go from $[\mathbf{x}]_{\mathcal{S}}$ to $[\mathbf{x}]_{\mathcal{B}_2}$. Matrix multiplication is used to combine the steps.

THEOREM 4.27 ▸

Let $\mathcal{B}_1=\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ and $\mathcal{B}_2=\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ be bases for $\mathbf{R}^n$. If $U = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_n]$ and $V = [\mathbf{v}_1 \ \ldots \ \mathbf{v}_n]$, then

(a) $[\mathbf{x}]_{\mathcal{B}_2}=V^{-1}U[\mathbf{x}]_{\mathcal{B}_1}$

(b) $[\mathbf{x}]_{\mathcal{B}_1}=U^{-1}V[\mathbf{x}]_{\mathcal{B}_2}$

Proof If $U = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_n]$ and $V = [\mathbf{v}_1 \ \ldots \ \mathbf{v}_n]$, then by Theorem 4.26 we know then

$$[\mathbf{x}]_{\mathcal{S}}=U[\mathbf{x}]_{\mathcal{B}_1}\quad\text{and}\quad[\mathbf{x}]_{\mathcal{B}_2}=V^{-1}[\mathbf{x}]_{\mathcal{S}}$$

Combining these gives

$$[\mathbf{x}]_{\mathcal{B}_2} = V^{-1}[\mathbf{x}]_{\mathcal{S}} = V^{-1}(U[\mathbf{x}]_{\mathcal{B}_1}) = V^{-1}U[\mathbf{x}]_{\mathcal{B}_1}$$

Thus the change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_2$ is $V^{-1}U$. The change of basis matrix from $\mathcal{B}_2$ to $\mathcal{B}_1$ is the inverse,

$$(V^{-1}U)^{-1} = U^{-1}V$$

A graphical depiction is given in Figure 2. ■ ■

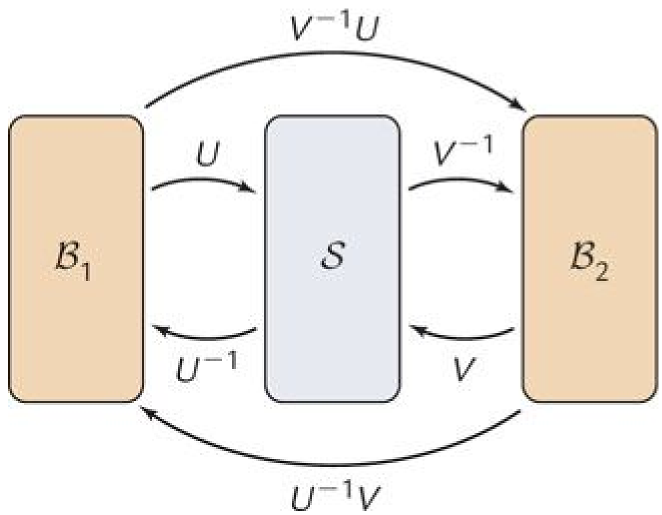


Figure 2 Change of basis between $\mathcal{B}_1$ and $\mathcal{B}_2$.

Example 4

Suppose that

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 7 \end{bmatrix} \right\} \text{ and } \mathcal{B}_2 = \left\{ \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$$

Find $[\mathbf{x}]_{\mathcal{B}_2}$ if $[\mathbf{x}]_{\mathcal{B}_1} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$.

Solution We start by setting $U = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$ and $V = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix}$. Then by Theorem 4.27, the change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_2$ is

$$V^{-1}U = \begin{bmatrix} -3 & 2 \\ 5 & -3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ -4 & -11 \end{bmatrix}$$

Hence it follows that

$$[\,x]\mathcal{B}2=V-1U[\,x]\mathcal{B}1=[38-4-11][-14]=[29-40]$$

We can check our results by converting $[\,x]\mathcal{B}1$ and $[\,x]\mathcal{B}2$ to the standard basis. We have

$$U[\,x]\mathcal{B}1=[1237][-14]=[725]\text{and}V[\,x]\mathcal{B}2=[3253][29-40]=[725]$$

which confirms that $[\,x]\mathcal{B}1=[\,x]\mathcal{B}2$.

Example 5

Suppose that

$$\mathcal{B}1=\{[113],[142],[216]\}\text{and}\mathcal{B}2=\{[101],[1-30],[212]\}$$

Find $[\,x]\mathcal{B}1$ if $[\,x]\mathcal{B}2=[321]$.

Solution We start by setting

$$U=[112141326]\text{and}V=[1120-31102]$$

By Theorem 4.27 we have

$$[\,x]\mathcal{B}1=U-1V[\,x]\mathcal{B}2 =[-222730-110-1-3][1120-31102][321]=[-1291660]$$

Change of Basis in Subspaces

Suppose that a subspace S of $\mathbf{R}^n$ has two bases $\mathcal{B}1=\{u1,\ldots ,uk\}$ and $\mathcal{B}2=\{v1,\ldots ,vk\}$. There exists a change of basis matrix, but it cannot be found using Theorem 4.27, because the matrices U and V are not square and so are not invertible.

THEOREM 4.28 ▶

Let S be a subspace of $\mathbf{R}^n$ with bases $\mathcal{B}_1=\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}$ and $\mathcal{B}_2=\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$. If

$$C=\begin{bmatrix}[\mathbf{u}_1]_{\mathcal{B}_2} & \cdots & [\mathbf{u}_k]_{\mathcal{B}_2}\end{bmatrix}$$

then $[\mathbf{x}]_{\mathcal{B}_2}=C[\mathbf{x}]_{\mathcal{B}_1}$.

Proof For a vector $\mathbf{x}$ in S, there exist scalars $x_1, \ldots, x_k$ such that

$$\mathbf{x}=x_1\mathbf{u}_1+\cdots+x_k\mathbf{u}_k \Rightarrow [\mathbf{x}]_{\mathcal{B}_1}=\begin{bmatrix}x_1\\ \vdots\\ x_k\end{bmatrix}$$

Therefore

$$C[\mathbf{x}]_{\mathcal{B}_1}=\begin{bmatrix}[\mathbf{u}_1]_{\mathcal{B}_2} & \cdots & [\mathbf{u}_k]_{\mathcal{B}_2}\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_k\end{bmatrix} =x_1[\mathbf{u}_1]_{\mathcal{B}_2}+\cdots+x_k[\mathbf{u}_k]_{\mathcal{B}_2} =[x_1\mathbf{u}_1+\cdots+x_k\mathbf{u}_k]_{\mathcal{B}_2} =[\mathbf{x}]_{\mathcal{B}_2} \qquad \text{(See Exercise 46)}$$

■ ■

When $S = \mathbf{R}^n$, the matrix C in Theorem 4.28 is equal to the matrix $V^{-1}U$ given in Theorem 4.27 (see Exercise 47).

Example 6

Let

$$\mathcal{B}_1=\left\{\begin{bmatrix}1\\-5\\8\end{bmatrix},\begin{bmatrix}3\\-8\\3\end{bmatrix}\right\} \text{ and } \mathcal{B}_2=\left\{\begin{bmatrix}1\\-3\\2\end{bmatrix},\begin{bmatrix}-1\\2\\1\end{bmatrix}\right\}$$

be two bases of a subspace S of $\mathbf{R}^3$. Find the change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_2$ and find $[\mathbf{x}]_{\mathcal{B}_2}$ if $[\mathbf{x}]_{\mathcal{B}_1}=\begin{bmatrix}3\\-1\end{bmatrix}$.

Solution To apply Theorem 4.28, we need to express each vector $\mathcal{B}_1$ in terms of the vectors $\mathcal{B}_2$. The system

[1−58]=c11[1−32]+c21[−121]

has solution c_{11} = 3 and c_{21} = 2, so that [1−58]=[32]ℬ2. Similarly, the system

[3−83]=c12[1−32]+c22[−121]

has solution c_{12} = 2 and c_{22} = −1, so that [3−83]=[2−1]ℬ2. Thus

C=[322−1]

By Theorem 4.28,

[x]ℬ2=C[x]ℬ1=[322−1][3−1]=[77]

▶ Practice problems can also be used as additional examples.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Convert the coordinate vector [x]ℬ from the given basis ℬ to the standard basis.
   (a) ℬ={[3−7],[26]},[x]ℬ=[−14]
   (b) ℬ={[126],[10−5],[−234]},[x]ℬ=[30−1]
2. Find the change of basis matrix from the standard basis 𝒮 to ℬ, and then convert **x** to the coordinate vector with respect to ℬ.
   (a) ℬ={[12],[25]},x=[1−3]
   (b) ℬ={[11−1],[2−10],[100]},x=[21−1]
3. Find the change of basis matrix from ℬ1 to ℬ2.
   (a) ℬ1={[12],[35]}, ℬ2={[41],[31]}
   (b) ℬ1={[10−1],[210],[10−2]}, ℬ2={[120],[111],[−220]}

4. Determine if each statement is true or false, and justify your answer.

(a) If A is the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}2$, then A must be nonsingular.

(b) If A is the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}2$, then A^{-1} is the change of basis matrix from $\mathcal{B}2$ to $\mathcal{B}1$

(c) If $\mathcal{B}1 = \mathcal{B}2$, then the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}2$ is I.

(d) If $\mathcal{B}1$ and $\mathcal{B}2$ are bases of a three-dimensional subspace S of $\mathbf{R}^5$, then the change of basis matrix A from $\mathcal{B}1$ to $\mathcal{B}2$ has dimensions 3 × 3.

EXERCISES

Exercises 1–6: Convert the coordinate vector $[\ x]\mathcal{B}$ from the given basis $\mathcal{B}$ to the standard basis.

1. $\mathcal{B}=\{[3-2],[25]\}$, $[\ x]\mathcal{B}=[1-1]$

2. $\mathcal{B}=\{[-53],[-21]\}$, $[\ x]\mathcal{B}=[-23]$

3. $\mathcal{B}=\{[43],[21]\}$, $[\ x]\mathcal{B}=[2-4]$

4. $\mathcal{B}=\{[-61],[5-3]\}$, $[\ x]\mathcal{B}=[-32]$

5. $\mathcal{B}=\{[1-2-1],[-120],[2-13]\}$, $[\ x]\mathcal{B}=[121]$

6. $\mathcal{B}=\{[031],[123],[0-12]\}$, $[\ x]\mathcal{B}=[11-2]$

Exercises 7–12: Find the change of basis matrix from the standard basis $\mathcal{S}$ to $\mathcal{B}$, and then convert $\mathbf{x}$ to the coordinate vector with respect to $\mathcal{B}$.

7. $\mathcal{B}=\{[12],[13]\}$, $x=[3-1]$

8. $\mathcal{B}=\{[54],[11]\}$, $x=[12]$

9. $\mathcal{B}=\{[-21],[5-3]\}$, $x=[1-1]$

10. $\mathcal{B}=\{[75],[43]\}$, $x=[4-3]$

11. $\mathcal{B}=\{[100],[-121],[1-10]\}$, $x=[12-1]$

12. $\mathcal{B}=\{[22-1],[-1-21],[1-11]\}$, $x=[-212]$

Exercises 13–18: Find the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}2$.

13. $\mathcal{B}1$={[21],[11]},$\mathcal{B}2$={[57],[34]}
14. $\mathcal{B}1$={[21],[32]},$\mathcal{B}2$={[31],[72]}
15. $\mathcal{B}1$={[−104],[233],[1−1−2]},$\mathcal{B}2$={[13−1],[01−1],[370]}
16. $\mathcal{B}1$={[230],[1−12],[−415]},$\mathcal{B}2$={[132],[−2−5−4],[123]}
17. $\mathcal{B}1$={[3−55],[−14−3]},$\mathcal{B}2$={[2−12],[13−1]}
18. $\mathcal{B}1$={[5−63],[−143]},$\mathcal{B}2$={[2−13],[−350]}

Exercises 19–24: Find the change of basis matrix from $\mathcal{B}2$ to $\mathcal{B}1$.

19. $\mathcal{B}1$={[21],[11]},$\mathcal{B}2$={[57],[34]}
20. $\mathcal{B}1$={[21],[32]},$\mathcal{B}2$={[31],[72]}
21. $\mathcal{B}1$={[−11−1],[102],[−250]},$\mathcal{B}2$={[−213],[201],[41−1]}
22. $\mathcal{B}1$={[−1−31],[14−2],[−2−3−2]},$\mathcal{B}2$={[141],[420],[31−2]}
23. $\mathcal{B}1$={[−142],[231]},$\mathcal{B}2$={[173],[−453]}
24. $\mathcal{B}1$={[−311],[2−54]},$\mathcal{B}2$={[−56−3],[1−99]}
25. For $\mathcal{B}1$ and $\mathcal{B}2$ in Exercise 13, find [x]$\mathcal{B}2$ if [x]$\mathcal{B}1$=[2−1].
26. For $\mathcal{B}1$ and $\mathcal{B}2$ in Exercise 16, find [x]$\mathcal{B}2$ if [x]$\mathcal{B}1$=[13−2].
27. For $\mathcal{B}1$ and $\mathcal{B}2$ in Exercise 17, find [x]$\mathcal{B}2$ if [x]$\mathcal{B}1$=[25].
28. For $\mathcal{B}1$ and $\mathcal{B}2$ in Exercise 19, find [x]$\mathcal{B}1$ if [x]$\mathcal{B}2$=[12].
29. For $\mathcal{B}1$ and $\mathcal{B}2$ in Exercise 21, find [x]$\mathcal{B}1$ if [x]$\mathcal{B}2$=[−113].
30. For $\mathcal{B}1$ and $\mathcal{B}2$ in Exercise 24, find [x]$\mathcal{B}1$ if [x]$\mathcal{B}2$=[−23].
31. Suppose that $\mathcal{B}1$={u1,u2} and $\mathcal{B}2$={u2,u1} are bases of R2. Find [x]$\mathcal{B}2$ if [x]$\mathcal{B}1$=[ab].
32. Suppose that $\mathcal{B}1$={u1,u2,u3} and $\mathcal{B}2$={u2,u3,u1} are bases of $\mathbf{R}^3$. Find [x]$\mathcal{B}1$ if [x]$\mathcal{B}2$=[abc].

FIND AN EXAMPLE Exercises 33–38: Find an example that meets the given specifications.

33. A basis $\mathcal{B}$ of $\mathbf{R}^2$ such that [13]$\mathcal{B}$=[−21].
34. A basis $\mathcal{B}$ of $\mathbf{R}^3$ such that [31−2]$\mathcal{B}$=[125].
35. Bases $\mathcal{B}1$ and $\mathcal{B}2$ of $\mathbf{R}^2$ such that [2−2]$\mathcal{B}1$=[41]$\mathcal{B}2$.

36. Bases $\mathcal{B}1$ and $\mathcal{B}2$ of $\mathbf{R}^3$ such that [30−1]$\mathcal{B}1$=[2−41]$\mathcal{B}2$.

37. Bases $\mathcal{B}1$ and $\mathcal{B}2$ of $\mathbf{R}^2$ with change of basis matrix

C=[1327] from $\mathcal{B}1$ to$\mathcal{B}2$.

38. Bases $\mathcal{B}1$ and $\mathcal{B}2$ of $\mathbf{R}^3$ with change of basis matrix

C=[1−12−230141] from $\mathcal{B}2$ to $\mathcal{B}1$.

TRUE OR FALSE Exercises 39–42: Determine if the statement is true or false, and justify your answer.

39.

(a) If U is a change of basis matrix between bases $\mathcal{B}1$ and $\mathcal{B}2$ of $\mathbf{R}^n$, then U must be an $n \times n$ matrix.

(b) A change of basis matrix from one basis of $\mathbf{R}^n$ to another basis of $\mathbf{R}^n$ is unique.

40.

(a) If U is a change of basis matrix between bases $\mathcal{B}1$ and $\mathcal{B}2$ of $\mathbf{R}^n$, then U must be invertible.

(b) Any change of basis matrix must have linearly independent columns.

41.

(a) If $\mathcal{B}1$={u1,u2} and $\mathcal{B}2$={2u1,5u2} are bases of the same subspace, then the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}2$ is [2005].

(b) Every change of basis matrix must be square.

42.

(a) Let $\mathcal{B}1$={u1,u2} and $\mathcal{B}2$={v1,v2} be two bases of $\mathbf{R}^2$, and suppose that $\mathbf{u}_1 = a\mathbf{v}_1 + b\mathbf{v}_2$ and $\mathbf{u}_2 = c\mathbf{v}_1 + d\mathbf{v}_2$. Then the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}2$ is [abcd].

(b) If C_1 is the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}2$ and C_2 is the change of basis matrix from $\mathcal{B}2$ to $\mathcal{B}3$, then C_1C_2 is the change of basis matrix from $\mathcal{B}1$ to $\mathcal{B}3$.

43. Let $\mathcal{B}$ be a basis. Prove that $[\mathbf{u} + \mathbf{v}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}}$ for vectors $\mathbf{u}$ and $\mathbf{v}$.

44. Let $\mathcal{B}$ be a basis. Prove that $[c\mathbf{u}]_{\mathcal{B}} = c[\mathbf{u}]_{\mathcal{B}}$, where $\mathbf{u}$ is a vector and c a scalar.

45. Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathbf{x}) = [\mathbf{x}]_{\mathcal{B}}$ for a basis $\mathcal{B}$. Prove that T is a linear transformation.

46. Let $\mathcal{B}$ be a basis, $c_1, \dots, c_k$ be scalars, and $\mathbf{u}_1, \dots, \mathbf{u}_k$ be vectors. Prove that

c1[u1]$\mathcal{B}$+⋯+ck[uk]$\mathcal{B}$=[c1u1+⋯ckuk]$\mathcal{B}$

47. Prove that if $S = \mathbf{R}^n$, then the matrix C in Theorem 4.28 is equal to the matrix $V^{-1}U$ given in Theorem 4.27.

48. Let $\mathcal{B}$1={u1,… ,un} and $\mathcal{B}$2={v1,… ,vn} be bases for $\mathbf{R}^n$, and set $U = [\mathbf{u}_1 \dots \mathbf{u}_n]$ and $V = [\mathbf{v}_1 \dots \mathbf{v}_n]$. Show that the change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_2$ can be found by extracting the right half of the row-reduced echelon form of $[V\ U]$.

Exercises 49–54: Find the change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_1$

49. $\mathcal{B}$1={[359],[4−27],[211−6]}$\mathcal{B}$2={[33−7],[258],[−4−61]}

50. $\mathcal{B}$1={[642],[50−5],[−332]}$\mathcal{B}$2={[3−2−1],[439],[−815]}

51. $\mathcal{B}$1={[101−1],[2−2−53],[500−2],[37−9−2]}$\mathcal{B}$2={[34−21],[7−2−10],[5−323],[403−1]}

52. $\mathcal{B}$1={[7−2−23],[−3011],[04−47],[5431]}$\mathcal{B}$2={[357−8],[−3−4−5−6],[2014],[1742]}

53. $\mathcal{B}$1={[−1042−5],[−550−4],[−852−6]}$\mathcal{B}$2={[210−1],[−302−2],[6−320]}

54. $\mathcal{B}$1={[7−7−45],[−534−5],[−12117−9]}$\mathcal{B}$2={[3−2−12],[4−1−13],[−543−4]}

SUPPLEMENTARY EXERCISES

Exercises 1–4: Determine if the described set is a subspace. If so, prove that it is a subspace. If not, explain why not. Assume that *a*, *b*, and *c* are real numbers.

1. The set of vectors of the form [0a2c]

2. The set of vectors of the form [a+bc+1]

3. The set of vectors of the form [a+ba+cb+c]

4. The set of vectors of the form [ab], where $a - b = 0$.

Exercises 5–8: Find the null space for *A*.

5. A=[26−13]

6. A=[1−24−103]

7. A=[−1543−21]

8. A=[326−1124−23]

Exercises 9–12: Let $T(\mathbf{x}) = A\mathbf{x}$. Determine if **b** is in ker(*T*) and if **c** is in range(*T*).

9. A=[−153−15],b=[2−10],c=[−315]

10. A=[−131321],b=[1−21],c=[25]

11. A=[1−2−12−24],b=[63],c=[−132]

12. A=[10−3−12512−1],b=[14−2],c=[−153]

Exercises 13–18: Find a basis for *S* and determine its dimension.

13. S=span{[−124],[312]}

14. S=span{[32−5],[−6−410]}

15. S=span{[21−4],[15−2],[4−7−8]}

16. S=span{[−1201],[2−32−1],[1041]}

17. *S* is the set of vectors of the form [a−b0], where *a* and *b* are real numbers.

18. *S* is the set of vectors of the form [a+4ba−b2a+3b], where *a* and *b* are real numbers.

Exercises 19–22: Find a basis for the null space of *A* (if a basis exists) and determine nullity(*A*).

19. A=[−48−132−4]

20. A=[142−1310146]

21. A=[−131259−23−1]

22. A=[12−2102422331]

Exercises 23–26: Find bases (if they exist) for the row, column, and null spaces of *A*. Verify that the Rank-Nullity Theorem is correct.

23. A=[−152131]

24. A=[21−2−15337−1]

25. A=[1−321−23]

26. A=[02−11−22230−231]

Exercises 27–30: Suppose $T(\mathbf{x}) = A\mathbf{x}$. Find bases (if they exist) for ker(*T*), range(*T*), and the row, column, and null spaces of *A*.

27. A=[−113211335]

28. A=[11−2−2−2455−10]

29. A=[14−12−53]

30. A=[02−21−23−131−231]

Exercises 31–32: Find **x** with respect to the standard basis corresponding to the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ and basis $\mathcal{B}$.

31. $\mathcal{B}$={[2−1],[35]}, [x]$\mathcal{B}$=[1−3]

32. $\mathcal{B}$={[210],[113],[−102]}, [x]$\mathcal{B}$=[23−1]

Exercises 33–34: Find the coordinate vector $\mathbf{x}_B$ corresponding to **x** and basis $\mathcal{B}$.

33. $\mathcal{B}$={[12],[25]}, x=[3−2]

34. $\mathcal{B}$={[110],[213],[312]}, x=[2−11]

TRUE OR FALSE Exercises 35–40: Determine if the statement is true or false, and justify your answer.

35.

(a) If $T(\mathbf{x}) = A\mathbf{x}$ then null(A) = ker(T).

(b) If $T(\mathbf{x}) = A\mathbf{x}$ then null(A) is a subspace of range(T).

36.

(a) If $T(\mathbf{x}) = A\mathbf{x}$ and A has a row of zeros, then ker(T) = $\{\mathbf{0}\}$.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ and A has a column of zeros, then ker(T) = $\{\mathbf{0}\}$.

37.

(a) If A is a 3×8 matrix, then null(A) is a subspace of $\mathbf{R}^8$.

(b) If A is a 7×4 matrix, then nullity(A) ≤ 4.

38.

(a) If S_1 and S_2 are nonzero subspaces of $\mathbf{R}^5$, then $\dim(S_1 \cap S_2) = \min\{\dim(S_1), \dim(S_2)\}$.

(b) If A is a square matrix, then row(A) = col(A).

39.

(a) If S is a nontrivial subspace, then S = col(A) for some matrix A.

(b) Any set of linearly independent vectors in $\mathbf{R}^n$ can be extended to a basis for $\mathbf{R}^n$.

40.

(a) Every subspace S of $\mathbf{R}^n$ has a basis with at most vectors.

(b) If A is a 5×9 matrix with rank(A) = 3, then nullity(A) = 2.

CHAPTER 5

Determinants

Peter Zelei Images/Moment/Getty Images

Similar in principle to the modern wind turbines of today, in the past windmills were used to harness the natural power of the wind. the windmills shown here, located in Kinderdijk in the Netherlands, are a popular tourist site and designated as a UNeSCO World heritage Site. these 19 windmills, dating back to the eighteenth century, were built to pump water. Windmills were also used to mill, or grind, grain.

The *determinant* is a function that takes a matrix as input and produces a real number as output. Determinants have a rich history and a variety of useful interpretations. In Section 5.1 we define the determinant and find formulas for the determinant for certain special types of matrices. The properties of the determinant are further developed in Section 5.2, including how row operations and matrix arithmetic influence the determinant. In Section 5.3, we see how to use the determinant to find the solution to a linear system and a matrix inverse, and how determinants give us information about the behavior of linear transformations.

5.1 The Determinant Function

The determinant of a square matrix A combines the entries of A to produce a single real number. There are several different interpretations and characterizations of the determinant. The development that we give here is guided by an important property of the determinant: that a square matrix is invertible exactly when it has a nonzero determinant.

Let's start with the easiest case. Let $A = [a_{11}]$ be a 1×1 matrix. Then A is invertible exactly when $a_{11} \neq 0$. This brings us to our first definition.

DEFINITION 5.1 ▶

Determinant of a 1 × 1 Matrix

Let $A = [a_{11}$ be a 1×1 matrix. Then the **determinant** of A is given by

$$\det(A)=a_{11}$$

Thus a 1×1 matrix A is invertible if and only if $\det(A) \neq 0$.

Next, suppose that

$$A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{1}$$

In Section 3.3 we developed the "Quick Formula" for the inverse of A, which says that

$$A^{-1}=\frac{1}{a_{11}a_{22}-a_{12}a_{21}}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{2}$$

provided that $a_{11}a_{22} - a_{12}a_{21} \neq 0$. If $a_{11}a_{22} - a_{12}a_{21} = 0$, then (2) is undefined and A has no inverse. Hence $a_{11}a_{22} - a_{12}a_{21}$ is nonzero exactly when A is invertible, so we take this as the definition of the determinant for a 2×2 matrix A.

DEFINITION 5.2 ▶

Determinant of a 2 × 2 Matrix

Let A be the 2×2 matrix in (1). Then the **determinant** of A is given by

$$\det(A) = a_{11}a_{22} - a_{12}a_{21} \tag{3}$$

Example 1

Let $A = \begin{bmatrix} 3 & 5 \\ -1 & 4 \end{bmatrix}$. Find $\det(A)$ and determine if A is invertible.

Solution Applying Definition 5.2, we have

$$\det(A) = (3)(4) - (5)(-1) = 17$$

Since $\det(A) \neq 0$, we can conclude that A is invertible.

The definition of the determinant is more complicated for larger matrices. Let's consider the 3×3 case. Suppose

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{4}$$

▶ The row operations are (in order performed):
$a_{11}R_2 \to R_2 \quad a_{11}R_3 \to R_3 \quad -a_{21}R_1 + R_2 \to R_2 \quad -a_{31}R_1 + R_3 \to R_3$

In order for A to be invertible, at least one of a_{11}, a_{21}, or a_{31} must be nonzero. For the moment, assume that $a_{11} \neq 0$. Applying the row operations listed in the margin, we have

$$A\sim[a11a12a130(a11a22-a12a21)(a11a23-a13a21)0(a11a32-a12a31)(a11a33-a13a31)]$$

Since $a_{11} \neq 0$, the matrix on the right is invertible exactly when

$$[(a11a22-a12a21)(a11a23-a13a21)(a11a32-a12a31)(a11a33-a13a31)]$$

is invertible. The determinant of this 2 × 2 matrix is

$$(a11a22-a12a21)(a11a33-a13a31)-(a11a23-a13a21)(a11a32-a12a31)=a11[a11a22a33+a12a23a31+a13a21a32-a11a23a32-a12a21a33-a13a22a31]$$

Since $a_{11} \neq 0$ is invertible if and only if the expression in brackets is nonzero. Other than a sign change, the bracketed expression is the same if we start with a_{21} or a_{31}. Furthermore, this expression is zero if $a_{11} = a_{21} = a_{31} = 0$. Hence the term in brackets is nonzero when A is invertible, so we use it for the determinant.

DEFINITION 5.3 ▶

Determinant of a 3 × 3 Matrix

Let A be the 3 × 3 matrix in (4). Then the **determinant** of A is given by

$$\det(A)=a11a22a33+a12a23a31+a13a21a32-a11a23a32-a12a21a33-a13a22a31 \tag{5}$$

Example 2

Find det(A) for

$$A=[-31255-842-5]$$

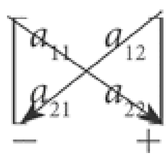
Solution From (5) we have

det(A)=(-3)(5)(-5)+(1)(-8)(4)+(2)(5)(2) -(-3)(-8)(2)-(1)(5)(-5)-(2)(5)(4) =75-32+20-48+25-40=0

Note that this implies that *A* is not invertible.

The Shortcut Method

For 2 × 2 and 3 × 3 matrices there exist nice visual aids for computing determinants that we refer to as *the Shortcut Method*. For the 2 × 2 case, start by drawing diagonal arrows through the terms of the matrix, labeled with + and − as shown below. Multiply the terms of each arrow, and then add or subtract as indicated by the + or −.



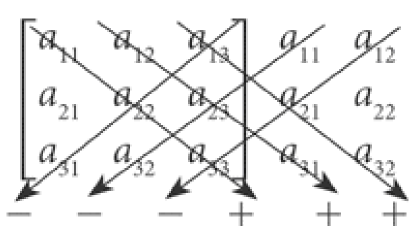
$$\Rightarrow \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

▶ We replace[ ] with | | around a matrix to indicate the determinant. For example,

|4237|=(4)(7)-(2)(3)=22

Note that this matches the formula given in (3).

For a 3 × 3 matrix, we write down the matrix, copy the left two columns to the right, and then draw six diagonal arrows with labels as shown.



▶ **Warning!** The Shortcut Method does not work for 4 × 4 or larger matrices.

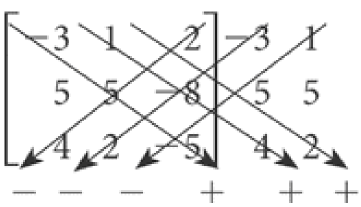
As in the 2 × 2 case, for each arrow we multiply terms and then add or subtract based on the labels. This yields a11a22a33+a12a23a31+a13a21a32-a13a22a31-a11a23a32-a12a21a33

which matches the formula given in (5).

Example 3

Find det(A) from Example 2 using the Shortcut Method.

Solution Adding on the two extra columns and drawing the diagonals, we have



$$\Rightarrow \quad \det(A) = (75 - 32 + 20) - (40 + 48 - 25) = 0$$

which matches what we found earlier.

Our formulas for determinants may appear unconnected, but in fact the 3 × 3 determinant is related to the 2 × 2 determinant. To see how, we start by reorganizing (5) and factoring out common terms,

det(A)=a11(a22a33-a23a32)-a12(a21a33-a23a31)+a13(a21a32-a22a31) (6)

Each expression in parentheses can be viewed as the determinant of a 2 × 2 matrix. For instance,

a22a33-a23a32=|a22a23a32a33|

Combining this observation with (6) gives us

det(A)=a11|a22a23a32a33|-a12|a21a23a31a33|+a13|a21a22a31a32| (7)

Each of these 2 × 2 matrices can be found within A by crossing out the row and column containing a_{11}, a_{12}, and a_{13}, respectively, then forming 2 × 2 matrices from the entries that remain.

$$a_{11} \Rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad a_{12} \Rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad a_{13} \Rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Minor

In general, if A is an $n \times n$ matrix, then M_{ij} denotes the $(n - 1) \times (n - 1)$ matrix that we get from A after deleting the row and column containing a_{ij}. The determinant $\det(M_{ij})$ is called the **minor** of a_{ij}.

Example 4

Suppose

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

Find M_{23}, M_{42}, and the associated minors.

Solution The term a_{23} is located in row 2 and column 3 of A. To find M_{23}, we cross our row 2 and column 3,

$$\begin{bmatrix} 4 & -1 & 1 & 0 \\ 1 & 7 & 3 & 5 \\ 0 & -3 & -2 & 1 \\ 2 & 4 & 8 & -1 \end{bmatrix} \Rightarrow \quad M_{23} = \begin{bmatrix} 4 & -1 & 0 \\ 0 & -3 & 1 \\ 2 & 4 & -1 \end{bmatrix}$$

Similarly, M_{42} is found by deleting row 4 and column 2,

$$\begin{bmatrix} 4 & -1 & 1 & 0 \\ 1 & 7 & 3 & 5 \\ 0 & -3 & -2 & 1 \\ 2 & 4 & 8 & -1 \end{bmatrix} \Rightarrow M_{42} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 5 \\ 0 & -2 & -1 \end{bmatrix}$$

The minors are

det(M23)=(4)|-314-1|-(-1)|012-1|+(0)|0-324| =(4)(3-4)-(-1)(0-2)+(0)(0-(-6))=-6

and

det(M42)=(4)|35-21|-(-1)|1501|+(0)|130-2| =(4)(3-(-10))-(1)(1-0)+(0)(-2-0)=51

Referring to (7), we see that our formula for the determinant of a 3 × 3 matrix can be expressed in terms of minors as

det(A)=a11det(M11)-a12det(M12)+a13det(M13)

Cofactor

This formula can be further simplified with the introduction of C_{ij}, the **cofactor** of a_{ij}, given by

Cij=(-1)i+jdet(Mij)

▶ The matrix below shows the sign of $(-1)^{i+j}$ by location:
[+-+- ⋯ -+-+ ⋯ +-+- ⋯ -+-+ ⋯ ⋮⋮⋮⋮⋱]

Thus we have

det(A)=a11C11+a12C12+a13C13 (8)

The formula (8) provides us with a model for the general definition of the determinant.

DEFINITION 5.4 ▶

Determinant

Let A be the $n \times n$ matrix

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \tag{9}$$

Then the **determinant** of A is

$$\det(A)=a_{11}C_{11}+a_{12}C_{12}+\cdots+a_{1n}C_{1n} \tag{10}$$

where $C_{11}, \ldots, C_{1n}$ are the cofactors of $a_{11}, \ldots, a_{1n}$, respectively. When $n = 1$, $A = [a_{11}]$ and $\det(A) = a_{11}$.

Definition 5.4 is an example of a *recursive* definition, because the determinant of an $n \times n$ matrix is defined in terms of the determinants of $(n - 1) \times (n - 1)$ matrices.

Example 5

Use Definition 5.4 and cofactors to find $\det(A)$ for

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

Solution The first step is to find the cofactors C_{11}, C_{12}, and C_{13}:

$$\begin{aligned} C_{11}&=(-1)^{1+1}\begin{vmatrix} 4 & 0 \\ 1 & 2 \end{vmatrix}=(1)((4)(2)-(0)(1))=8 \\ C_{12}&=(-1)^{1+2}\begin{vmatrix} 1 & 0 \\ 3 & 2 \end{vmatrix}=(-1)((1)(2)-(0)(3))=-2 \\ C_{13}&=(-1)^{1+3}\begin{vmatrix} 1 & 4 \\ 3 & 1 \end{vmatrix}=(1)((1)(1)-(4)(3))=-11 \end{aligned}$$

By Definition 5.4 we have

$$\det(A)=2(8)+(-1)(-2)+3(-11)=-15$$

▶ The determinants for the cofactors were calculated using the formula (3).

We can use induction and Definition 5.4 to prove that $\det(I_n) = 1$.

THEOREM 5.5 ▶

For $n \geq 1$, we have $\det(I_n) = 1$.

Proof We shall carry out this proof by induction on, the number of rows. First suppose that $n = 1$. Then

$$I_1 = [\ 1] \Rightarrow \det(I_1) = 1$$

so that the theorem is true in this case. Next suppose that $n \geq 2$ and that the theorem is true for I_{n-1}. (This is the *induction hypothesis*.) Since the top row of has a single 1 followed by zeros, by Definition 5.4 we have

▶ This chapter contains theorems and exercises using proof by induction. If you are unfamiliar with this method of proof or are just a bit rusty, consult the appendix "Reading and Writing Proofs," available on the book website.

$$\det(I_n) = (1)C_{11} + (0)C_{12} + \cdots + (0)C_{1n} = C_{11} = (-1)^2 \det(M_{11}) = \det(M_{11})$$

Since M_{11} is the matrix we get from deleting the first row and column of I_n, we have

$$M_{11} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = I_{n-1}$$

By the induction hypothesis we know $\det(I_{n-1}) = 1$. Therefore

$$\det(I_n) = \det(M_{11}) = \det(I_{n-1}) = 1$$

The two parts of the induction proof are verified, so the proof is complete. ■■

A remarkable fact about the general definition of the determinant given in Definition 5.4 is that it has the same property as the determinant of 2 × 2 and 3 × 3 matrices, namely, that for any $n \times n$ matrix A, $\det(A)$ is nonzero exactly when A is invertible.

THEOREM 5.6 ▶

Let A be an $n \times n$ matrix. Then A is invertible if and only if $\det(A) \neq 0$.

We have already seen that Theorem 5.6 is true when $n = 1$, 2, and 3. The proof of Theorem 5.6 for larger matrices is given in Section 5.2.

Theorem 5.6 allows us to add a condition involving the determinant to the Unifying Theorem.

THEOREM 5.7 ▶

(THE UNIFYING THEOREM, VERSION 7) Let $\mathcal{S} = \{a1, \cdots, an\}$ be a set of n vectors in $\mathbf{R}^n$, let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$, and let $T: \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathrm{x}) = A\mathbf{x}$. Then the following are equivalent:

(a) $\mathcal{S}$ spans $\mathbf{R}^n$.

(b) $\mathcal{S}$ is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

(d) T is onto.

(e) T is one-to-one.

(f) A is invertible.

(g) $\ker(T) = \{\mathbf{0}\}$.

(h) $\mathcal{S}$ is a basis for $\mathbf{R}^n$.

(i) col(A) = $\mathbf{R}^n$.

(j) row(A) = $\mathbf{R}^n$.

(k) rank(A) = n.

(l) det(A) ≠ 0.

Proof From the Unifying Theorem, Version 6 in Section 4.3, we know that (a) through (k) are equivalent. Theorem 5.6 tells us that (f) and (l) are equivalent, and so we conclude that all 12 conditions are equivalent. ■■

▶ This updates the Unifying Theorem, Version 6, given in Section 4.3.

Example 6

Show that the set

A={[3-15],[-207],[431]}

forms a basis for $\mathbf{R}^3$.

Solution Let

A=[3-24-103571]

By the Unifying Theorem, Version 7, A is a basis for $\mathbf{R}^3$ if and only if det(A) ≠ 0. Applying the Shortcut Method, we find that

$$\begin{array}{cccccc} \left[\begin{matrix} 3 & -2 & 4 \\ -1 & 0 & 3 \\ 5 & 7 & 1 \end{matrix}\right. & & & & & \\ \end{array}\begin{matrix} 3 & -2 \\ -1 & 0 \\ 5 & 7 \end{matrix} \quad \Rightarrow \quad \det(A) = (0 - 30 - 28) - (0 + 63 + 2) = -123$$

$$- \quad - \quad - \quad + \quad + \quad +$$

Since det(A) ≠ 0, A is a basis for $\mathbf{R}^3$.

Cofactor Expansion

In our definition of the determinant, we use cofactors for the entries along the top row of the matrix. The next theorem allows us to generalize to entries in other rows or columns. The proof is omitted.

THEOREM 5.8

Cofactor Expansions

Let A be the $n \times n$ matrix in (9). Then

(a) $\det(A)=a_{i1}C_{i1}+a_{i2}C_{i2}+\cdots+a_{in}C_{in}$ (Expand across row i)

(b) $\det(A)=a_{1j}C_{1j}+a_{2j}C_{2j}+\cdots+a_{nj}C_{nj}$ (Expand down column j)

where C_{ij} denotes the cofactor of a_{ij}. Thse formulas are referred to collectively as the **cofactor expansions**.

Theorem 5.8 tells us that we can compute the determinant by taking cofactors along *any* row or column of the matrix. This theorem is handy in cases where a matrix has a row or column containing several zeros, because we can save ourselves some work.

The 3 × 3 determinants in Example 7 were found using the Shortcut Method.

Example 7

Find det(A) for

A=[-214-110-125-121003-1]

Solution The cofactor expansion down the 2nd column is

$$\det(A)=a_{12}C_{12}+a_{22}C_{22}+a_{32}C_{32}+a_{42}C_{42}$$

Since $a_{22} = 0$ and $a_{42} = 0$, there is no need to calculate C_{22} and C_{42}. The cofactors C_{12} and C_{32} are given by

$$C12=(-1)1+2|1-1252103-1|=(-1)(20)=-20C32=(-1)3+2|-24-11-1203-1|=(-1)(11)=-11$$

Pulling everything together, we have

$$det(A)=1·C12+(-1)·C32=1(-20)+(-1)(-11)=-9$$

Other than the amount of work involved, it makes no difference which row or column we choose. Expanding along the 4th row of A, we have

$$det(A)=a41C41+a42C42+a43C43+a44C44$$

Since $a_{41} = a_{42} = 0$, we need only compute C_{43} and C_{44}, which are

$$C43=(-1)4+3|-21-11025-11|=(-1)(6)=-6C44=(-1)4+4|-21410-15-12|=(1)(-9)=-9$$

Therefore

$$det(A)=3·C43+(-1)·C44=3(-6)+(-1)(-9)=-9$$

as before.

Matrices that have certain special forms or characteristics have determinants that are easy to compute.

Example 8

Find det(A) for

$$A=[-127-580341-900-2411000-4500001]$$

Solution Note that A is upper triangular. Let’s take advantage of all the zeros in the first column. Since the first entry in that column is the only one that is nonzero, the cofactor expansion down the first column is

$$\det(A)=(-1)C_{11}=(-1)(-1)^2\det(M_{11})=(-1)\begin{vmatrix}3 & 4 & 1 & -9\\ 0 & -2 & 4 & 11\\ 0 & 0 & -4 & 5\\ 0 & 0 & 0 & 1\end{vmatrix}$$

For the remaining determinant $\det(M_{11})$, we again use cofactor expansion down the first column, giving us

$$\det(A)=(-1)(3)C_{11}=(-1)(3)\begin{vmatrix}-2 & 4 & 11\\ 0 & -4 & 5\\ 0 & 0 & 1\end{vmatrix}$$

The remaining 3 × 3 determinant can be computed using the Shortcut Method, which produces only one nonzero term, the product of the diagonal $(-2)(-4)(1)$. Therefore

$$\det(A)=(-1)(3)(-2)(-4)(1)=-24$$

In Example 8, A is a square triangular matrix and $\det(A)$ is equal to the product of the diagonal terms. This suggests the next theorem.

THEOREM 5.9 ▸

If A is a triangular $n \times n$ matrix, then $\det(A)$ is the product of the terms along the diagonal.

The proof of Theorem 5.9 is left as Exercise 76.

Recall that if we interchange the rows and columns of a matrix A, we get A^T, the transpose of A. Interestingly, taking the transpose has no effect on the determinant.

THEOREM 5.10 ▸

Let A be a square matrix. Then $\det(A^T) = \det(A)$.

Proof We use induction. First note that if $A = [a_{11}]$ is a 1×1 matrix, then $A = A^T$ so that $\det(A) = \det(A^T)$.

▶ For

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

the transpose is

$$A^T=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}$$

Now for the induction hypothesis: Suppose that $n \geq 2$ and that the theorem holds for all $(n-1) \times (n-1)$ matrices. If A is an $n \times n$ matrix, then the cofactor expansion along the top row of A gives us

$$\det(A)=a_{11}\det(M_{11})-a_{12}\det(M_{12})+\cdots+(-1)^{n+1}a_{1n}\det(M_{1n})$$

Since $M_{11}, \ldots, M_{1n}$ are all $(n-1) \times (n-1)$ matrices, by the induction hypothesis we have $\det(M_{11})=\det(M_{11}^T), \ldots, \det(M_{1n})=\det(M_{1n}^T)$. Hence

$$\det(A)=a_{11}\det(M_{11}^T)-a_{12}\det(M_{12}^T)+\cdots+(-1)^{n+1}a_{1n}\det(M_{1n}^T) \quad (11)$$

Next note that the first *column* of A^T has entries $a_{11}, \ldots, a_{1n}$. Thus the right side of (11) also gives the cofactor expansion down the first column of A^T, and so it follows that $\det(A^T) = \det(A)$. ■■

Example 9

Verify that $\det(A) = \det(A^T)$ for

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

Solution Applying the Shortcut Method twice, we find that

$$\det(A)=\begin{vmatrix} 2 & 3 & -1 \\ 4 & 2 & 0 \\ 5 & -2 & -4 \end{vmatrix}=(-16+0+8)-(-10+0-48)=50$$

$$\det(A^T)=\begin{vmatrix} 2 & 4 & 5 \\ 3 & 2 & -2 \\ -1 & 0 & -4 \end{vmatrix}=(-16+8+0)-(-10+0-48)=50$$

Hence $\det(A) = \det(A^T)$.

THEOREM 5.11 ▶

Let A be a square matrix.

(a) If A has a row or column of zeros, then $\det(A) = 0$.

(b) If A has two identical rows or columns, then $\det(A) = 0$.

Proof The proofs of both parts of this theorem are left as Exercise 77 and Exercise 78, respectively. ■ ■

Example 10

Show that $\det(A) = 0$ and $\det(B) = 0$, where

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

Solution Since the third column of A consists of zeros, by Theorem 5.11(a) we have $\det(A) = 0$. Since rows 1 and 3 of B are identical, by Theorem 5.11(b) we have $\det(B) = 0$.

Multiplying matrices and computing determinants are both processes requiring numerous arithmetic operations. However, the relationship between the determinant of the product of two matrices and the product of the individual determinants is remarkably simple, as shown in Theorem 5.12.

THEOREM 5.12 ▶

If A and B are both $n \times n$ matrices, then

$$\det(AB)=\det(A)\det(B)$$

A proof of Theorem 5.12, and further discussion of the determinant of products of matrices, is given in Section 5.2.

Example 11

Verify that det(AB) = det(A) det(B) for the matrices

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

Solution Starting with A and B, we have

$$\det(A)=(2)(1)-(-4)(-1)=-2 \text{ and } \det(B)=(3)(1)-(-1)(2)=5$$

Hence det(A) det(B) = −10. Computing AB, we have

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

Therefore det(AB) = (−2)(2) − (−6)(−1) = −10 = det(A) det(B).

COMPUTATIONAL COMMENTS

Except when a matrix has mostly zero entries, computing determinants using cofactor expansion is slow for even a modest-sized matrix. Working recursively eventually generates a lot of 3 × 3 determinants that all require evaluation. For an $n \times n$ matrix, the number of multiplications needed is about $n!$. Thus, for example, a 20 × 20 matrix will require about 20! = 2,432,902,008,176,640,000 multiplications, far more than is remotely practical. In the next section, we see how to use row operations to speed things up.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find C_{23} and C_{31} for the given matrix.

(a) A=[13-230-45-12]

(b) A=[20-12411231032213]

2. Determine if $T(\mathbf{x}) = A\mathbf{x}$ is an invertible linear transformation for the given matrix.

(a) A=[31022-2413]

(b) A=[52103120-13424012]

3. Find the value(s) of a such that A is invertible.

(a) A=[3a13]

(b) A=[a120-a31-22]

4. Verify that $\det(A) = \det(A^T)$.

(a) A=[2-516]

(b) A=[12-5-3142-12]

5. Determine all values λ of such that $\det(A - \lambda I_2) = 0$.

(a) A=[231-4]

(b) A=[1411]

6. Determine if each statement is true or false.

(a) If A has a row of zeroes, then A is not invertible.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ and $\det(A) = 0$, then T is one-to-one.

(c) If A is 4×4, then $\det(3A) = 3\det(A)$.

(d) If $A \sim B$, then $\det(A) = \det(B)$.

EXERCISES

Exercises 1–6: Find M_{23} and M_{31} for the given matrix A.

1. A=[70-4362515]

2. A=[-531622440]

3. A=[61-15023071114312]

4. A=[0-24051-1002-103616]

5. A=[4321061205322445100322410]

6. A=[10101379458100225341612 33]

Exercises 7–10: Find C_{13} C_{22} for the given matrix *A*.

7. A=[2130-14401]
8. A=[6102-1-3341]
9. A=[612430111]
10. A=[0-1-1321402]

Exercises 11–18: Find the determinant for the given matrix *A* in two ways, by using cofactor expansion (a) along the row of your choosing, and (b) along the column of your choosing. Use the determinant to decide if *T*(**x**) = *A***x** is invertible.

11. A=[12-30405-10]
12. A=[-200450306]
13. A=[-11-12032014010-13-1]
14. A=[2130120142010120]
15. A=[-110023320-105314-1]
16. A=[025430-1011-21-2030]
17. A=[4210103-1120-10010120300101]
18. A=[1-1010300-11-110-200012020-201]

Exercises 19–26: When possible, use the Shortcut Method to compute det(*A*). If the Shortcut Method is not applicable, explain why.

19. A=[46-12]
20. A=[51-3-2]
21. A=[12-1310]
22. A=[6-11022]
23. A=[31-1204161]
24. A=[223-14131-2]
25. A=[612131000221123-1]
26. A=[21-123110512143-32]

Exercises 27–34: Find all values of *a* such that the given matrix is not invertible. (HINT: Think determinants, not row operations.)

27. A=[236a]
28. A=[12aa3]
29. A=[aa3-1]
30. A=[a-32a]
31. A=[1-130a-2243]
32. A=[-12a01130-1]
33. A=[1a-2-101a3-4]
34. A=[04aa130a1]

Exercises 35–40: Find det(A). No cofactor expansions are required, but you should explain your answer.

35. A=[20005-10072101337114]
36. A=[34570-25-900160005]
37. A=[610402031006610-7]
38. A=[2001000041020123]
39. A=[21623-24121623524]
40. A=[14314324320160-12-11211201-1]

Exercises 41–44: Verify that det(A) = det(A^T).

41. A=[3-241]
42. A=[6123]
43. A=[0712314-1-1]
44. A=[-1212-10104]

Exercises 45–48: Determine all possible real values of λ such that det($A - \lambda I_2$) = 0 for the matrix A.

45. A=[2453]
46. A=[0352]
47. A=[10-51]
48. A=[3-62-1]

Exercises 49–52: Find all possible real values λ such that $\det(A - \lambda I_3) = 0$.

49. A=[100530-47-2]

50. A=[-5-2-3006004]

51. A=[0201022-10]

52. A=[0121112-10]

Exercises 53–56: For each matrix A, first compute $\det(A)$. Then interchange two rows of your choosing and compute the determinant of the resulting matrix. Form a conjecture about the effect of row interchanges on determinants.

53.

(a) A=[35-24]

(b) A=[12-130201-1]

54.

(a) A=[1021]

(b) A=[2211-12100]

55.

(a) A=[310123021]

(b) A=[4-10021111]

56.

(a) A=[0-11-121403]

(b) A=[32101100-2]

Exercises 57–60: For each matrix A, first compute $\det(A)$. Then multiply a row of your choosing by 3 and compute the determinant of the resulting matrix. Form a conjecture about the effect on determinants of multiplying a row times a scalar.

57.

(a) A=[35-24]

(b) A=[12-130201-1]

58.

(a) A=[1021]

(b) A=[2211-12100]

59.

(a) A=[310123021]

(b) A=[4-10021111]

60.

(a) A=[0-11-121403]

(b) A=[32101100-2]

FIND AN EXAMPLE Exercises 61–68: Find an example that meets the given specifications.

61. A 2 × 2 matrix A with det(A) = 12.

62. A 3 × 3 matrix A with det(A) = 21.

63. A 2 × 2 matrix A with nonzero entries and det(A) = −3.

64. A 3 × 3 matrix A with nonzero entries and det(A) = 5.

65. A 3 × 3 matrix A with

M11=[046-3],M23=[5-126]

66. A 4 × 4 matrix A with

M14=[314450-173],M33=[1-21310-176]

67. A 3 × 3 matrix A with cofactors $C_{12} = -8$ and $C_{31} = -5$.

68. A 3 × 3 matrix A with cofactors $C_{22} = -2$, $C_{11} = 6$, and $C_{32} = -2$.

TRUE OR FALSE Exercises 69–72: Determine if the statement is true or false, and justify your answer.

69.

(a) Every matrix A has a determinant.

(b) If A is an $n \times n$ matrix, then each cofactor of A is an $(n - 1) \times (n - 1)$ matrix.

70.

(a) If A is an $n \times n$ matrix with all positive entries, then det(A) > 0.

(b) If A is an $n \times n$ matrix such that = $C_{i1} = \ldots = C_{in} = 0$ for some i, then det(A) = 0.

71.

(a) If A is an upper triangular $n \times n$ matrix, then det(A) ≠ 0.

(b) If A is a diagonal matrix, then M_{ij} is also diagonal for all i and j.

72.

(a) If the cofactors of an $n \times n$ matrix A are all nonzero, then $\det(A) \neq 0$.

(b) If A and B are 2×2 matrices, then $\det(A - B) = \det(A) - \det(B)$.

73. Let (x_1, y_1) and (x_2, y_2) be two distinct points in the plane. Prove that

|xy1x1y11x2y21|=0

gives an equation for the line passing through (x_1, y_1) and (x_2, y_2).

74. Find a general formula for the determinant of

A=[0 ⋯ 00a1n0 ⋯ 0a2(n-1)a2n0 ⋯ a3(n-2)a3(n-1)a3n⋮ ⋯ ⋮⋮⋮an1 ⋯ an(n-2)an(n-1)ann]

75. Let

A=[a11a12 ⋯ a1na21a22 ⋯ a2n⋮⋮⋱⋮an1an2 ⋯ ann]

Let $C_{j1}, \ldots, C_{jn}$ be the cofactors of A along row j. For $i \neq j$ prove that

ai1Cj1+ai2Cj2+…+ainCjn=0

76. Use induction to complete the proof of Theorem 5.9: If A is an $n \times n$ lower triangular matrix, then $\det(A)$ is the product of the terms along the diagonal of A.

77. Prove Theorem 5.11(a): Let A be a square matrix. If A has a row or column of zeros, then $\det(A) = 0$.

78. Prove Theorem 5.11(b): Let A be a square matrix. If A has two identical rows or columns, then $\det(A) = 0$. (HINT: Use induction.)

Exercises 79–82: Find $\det(A)$.

79. A=[3-40521-710-32258-2-1]

80. A=[0379-14-1029-4323-3-2]

81. A=[3500201-2-3-27-2-10041114-5-1053]

82. A=[3212378913-1-23-2-13696242194]

5.2 Properties of the Determinant

At the end of Section 5.1, we noted that computing determinants using cofactor expansion is too slow for use with even modest-sized matrices. In this section we show how to use row operations to make computing determinants more efficient. We will also develop additional properties of the determinant.

Instead of using cofactor expansion to compute the determinant, it is typically faster to first convert the matrix to echelon form using row operations and then multiply the terms on the diagonal. Example 1 examines the influence of row operations on the determinant.

Example 1

Suppose that A=[2-14-63-3150]. Compare det(*A*) with det(*B*), where *B* is the matrix
we get from *A* after performing the given row operation.

(a) Interchange Row 1 and Row 3(R1↔R3).

(b) Multiply Row 2 by 13(13R2→R2).

(c) Add −2 times Row 1 to Row 3 ($-2R_1 + R_3 \to R_3$).

Solution We use the Shortcut Method to compute the determinants in this example, starting with

det(A)=(0+3-120)-(12+0-30)=-99

For part (a), we interchange Row 1 with Row 3 and then compute the determinant.

B=[150-63-32-14]⇒det(B)=(12-30+0)-(0-120+3)=99

Hence det(*A*) = −det(*B*), so interchanging two rows changed the sign of the determinant. For part (b), multiplying the second row of

A by 13 and then computing the determinant gives us

B=[2-14-21-1150]⇒det(B)=(0+1-40)-(4+0-10)=-33

Thus det(A) = 3 det(B). For part (c), we add −2 times Row 1 to Row 3 and then compute the determinant.

B=[2-14-63-3-37-8]⇒det(B)=(-48-9-168)-(-36-48-42)=-99

This time det(A) = det(B), so adding a multiple of one row to another did not change the determinant.

▶ We already noted that it takes about n! multiplications to compute the determinant of an $n \times n$ matrix using cofactor expansion. By contrast, using row operations requires roughly n^3 multiplications. The difference is modest for small matrices but highly significant for larger matrices. For instance, $20^3 = 8 \times 10^3$ while $20! \approx 2.43 \times 10^{18}$.

Theorem 5.13 summarizes the influence of row operations on determinants.

THEOREM 5.13 ▶

Let A be a square matrix.
(a) Suppose that B is produced by interchanging two rows of A. Then det(A) = −det(B).
(b) Suppose that B is produced by multiplying a row of A by c. Then det(A)=1c. det(B).
(c) Suppose that B is produced by adding a multiple of one row of A to another. Then det(A) = det(B).

▶ Theorem 5.13 is also true if the rows are replaced with columns.

A proof of Theorem 5.13 is given at the end of the section.

▶ This step could be combined with the one that follows. The steps are separated to make it easier to track the effect of the row operations on the determinant.

Example 2

Use row operations together with Theorem 5.13 to find det(A) for

A=[-214-110-125-121003-1]

Solution In Example 7 of Section 5.1, we used cofactor expansion to show that det(A) = −9. Here we use row operations to transform A to triangular form while applying Theorem 5.13 to track the effect of the row operations on det(A).

To reduce A to triangular form, we start with the first column. It is handy to have a 1 in the pivot position, so we start by interchanging the first two rows.

A=[-214-110-125-121003-1]R1↔R2~[10-12-214-15-121003-1]=A1

By Theorem 5.13(a) we have det(A) = −det(A_1). Next, we introduce zeros down the first and second columns (the row operations are combined for brevity) with

A1=[10-12-214-15-121003-1]2R1+R2→R2-5R1+R3→R3R2+R3→R3~[10-120123009-6003-1]=A2

By Theorem 5.13(c), none of these row operations changes the determinant, so that det(A_1) = det(A_2) and hence det(A) = −det(A_1) = −det(A_2).

Next we multiply the third row of A_2 by -13 to introduce a −3 in the pivot position of the third column.

A2=[10-120123009-6003-1] -13R3→R3~ [10-12012300-32003-1]=A3

By Theorem 5.13(b) we have det(A_2) = −3 det(A_3), so that

$$\det(A)=-\det(A2)=-(-3)\det(A3)=3\det(A3)$$

The last step is to introduce a zero at the bottom of the third column of A_3.

$$A3=[10-12012300-32003-1]\ R3+R4\rightarrow R4\sim [10-12012300-320001]=A4$$

Since this row operation has no effect on the determinant, we have $\det(A_3) = \det(A_4)$ and so $\det(A) = 3\det(A_3) = 3\det(A_4)$. Since A_4 is a triangular matrix, by Theorem 5.9

$$\det(A4)=(1)(1)(-3)(1)=-3$$

and hence $\det(A) = 3\det(A_4) = -9$, matching the answer we obtained using cofactor expansion.

Example 3

Use row operations and Theorem 5.13 to find $\det(A)$ for

$$A=[1-23-13-6111-24-942-481]$$

Solution We proceed just as in Example 2. Starting with the first column, we introduce zeros with the row operations

$$[1-23-13-6111-24-942-481]-3R1+R2\rightarrow R22R1+R3\rightarrow R3-2R1+R4\rightarrow R4\ \sim[1-23-1002400-320023]$$

By Theorem 5.13(c), none of these row operations changes the determinant, so that

$$\det(A)=[1-23-1002400-320023]$$

We can continue row operations to triangular form, but the zero in the a_{22} position will remain. Hence the product of the diagonal terms of any eventual triangular matrix will be zero. Since det(A) is a multiple of this product, it must be that det(A) = 0.

Example 4

Suppose

A=[abcdefghi]andB=[3g3h3i(2g+d)(2h+e)(2i+f)abc]

Find det(B) if det(A) = 5.

Solution We have

[3g3h3i(2g+d)(2h+e)(2i+f)abc]13R1→R1~[ghi(2g+d)(2h+e)(2i+f)abc] -2R1+R2→R2~[ghjdefabc] R1↔R3~[abcdefghi]=A

The row operation 13R1→R1 multiplies det(B) by 3 and the row operation R1↔R3 multiplies det(B) by −1. Therefore

det(B)=(3)(-1)det(A)=-15.

Theorem 5.13 gives us the tools to prove Theorem 5.6 from Section 5.1. The theorem is stated again below.

THEOREM 5.6 ▸

Let A be an $n \times n$ matrix. Then A is invertible if and only if det(A) $\neq$ 0.

Proof We can transform any square matrix A to an upper triangular matrix B using two types of row operations:

- Row interchanges, which only change the sign of the determinant.
- Adding a multiple of one row to another, which does not change the determinant.

Therefore $\det(A) = \pm\det(B)$. Since B is triangular, $\det(B)$ is equal to the product of the diagonal terms of B. Thus $\det(A) \neq 0$ exactly when all the diagonal terms of B are nonzero.

When the diagonal terms of B are all nonzero, all of the pivots are also nonzero. Hence transforming B to reduced echelon form will yield the identity matrix I_n, which implies A is invertible. On the other hand, if there is a zero among the diagonal terms of B, then $\det(A) = 0$. In this case, the zero pivot means that the reduced echelon form of A cannot be equal to I_n, so that A is not invertible. Hence A is invertible if and only if $\det(A) \neq 0$.

Example 5

Suppose that $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation, with

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

Determine if T is invertible.

Solution The linear transformation T is invertible if and only if A is invertible, so by Theorem 5.6 T is invertible exact when $\det(A) \neq 0$. Applying row operations to A, we have

$$A=\begin{bmatrix} 1 & 5 & 3 & -4 \\ -1 & 2 & -3 & 4 \\ 0 & 4 & 2 & -3 \\ 2 & 4 & 2 & -2 \end{bmatrix} \begin{matrix} R_1+R_2 \to R_2 \\ -2R_1+R_4 \to R_4 \\ \sim \end{matrix} \begin{bmatrix} 1 & 5 & 3 & -4 \\ 0 & 7 & 0 & 0 \\ 0 & 4 & 2 & -3 \\ 0 & -6 & -4 & 6 \end{bmatrix}=B$$

These row operations do not change the determinant, so $\det(A) = \det(B)$. With all of the zeros in B, it makes applying cofactor expansion attractive.

$$\det(B)=(1)\begin{vmatrix} 7 & 0 & 0 \\ 4 & 2 & -3 \\ -6 & -4 & 6 \end{vmatrix}=(1)(7)\begin{vmatrix} 2 & -3 \\ -4 & 6 \end{vmatrix}=(1)(7)(0)=0$$

Therefore det(A) = 0 so we conclude that A is not invertible. Hence T is not invertible.

Determinants of Products

We now return to Theorem 5.12 from Section 5.1, which says that the determinant of the product of two matrices is equal to the product of the individual determinants. The theorem is stated again below.

THEOREM 5.12

If A and B are both $n \times n$ matrices, then det(AB) = det(A) det(B).

We will get to the proof of Theorem 5.12 shortly. First, note that an interesting consequence of Theorem 5.12 is that while generally $AB \neq BA$, it is true that det(AB) = det(BA), because

$$\det(AB)=\det(A)\det(B)=\det(B)\det(A)=\det(BA)$$

Here we used the fact that det(A) det(B) = det(B) det(A) because multiplication of real numbers is commutative. Let's look at a specific example illustrating Theorem 5.12.

Example 6

Show that $AB \neq BA$ but det(AB) = det(A) det(B) = det(BA) for the matrices

$$A=[1325] \text{ and } B=[4-5-32]$$

Then show that det($A + B$) $\neq$ det(A) + det(B).

Solution Starting with A and B, we have

$$\det(A)=(1)(5)-(3)(2)=-1 \quad \text{and} \quad \det(B)=(4)(2)-(-5)(-3)=-7$$

Hence $\det(A)\det(B) = 7$. We also have

$$AB=\begin{bmatrix}1&3\\2&5\end{bmatrix}\begin{bmatrix}4&-5\\-3&2\end{bmatrix}=\begin{bmatrix}-5&1\\-7&0\end{bmatrix} \Rightarrow \det(AB)=0-(-7)=7$$

and

$$BA=\begin{bmatrix}4&-5\\-3&2\end{bmatrix}\begin{bmatrix}1&3\\2&5\end{bmatrix}=\begin{bmatrix}-6&-13\\1&1\end{bmatrix} \Rightarrow \det(BA)=-6-(-13)=7$$

Therefore $\det(AB) = \det(A)\det(B) = \det(BA)$. Turning to $A + B$, we have

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

Thus $\det(A + B) = (5)(7) - (-2)(-1) = 33$ and $\det(A) + \det(B) = -8$, so $\det(A + B) \neq \det(A) + \det(B)$.

▶ **Warning!** In general $\det(A + B) \neq \det(A) + \det(B)$

To prove Theorem 5.12, we start with a special case involving elementary matrices. Recall that elementary matrices, introduced in Section 3.2, are square matrices E such that the product EA performs an elementary row operation on A.

THEOREM 5.14 ▶

If E and B are both $n \times n$ matrices and E is an elementary matrix, then $\det(EB) = \det(E)\det(B)$.

Proof Suppose that E is an elementary matrix corresponding to interchanging two rows. Then E is equal to I_n after the same two rows have been interchanged. Since $\det(I_n) = 1$, it follows from

Theorem 5.13(a) that det(E) = −1. As the product EB is the same as B with two rows interchanged, we have (again by Theorem 5.13(a)) det(EB) = −det(B). Therefore

$$\det(EB)=-\det(B)=\det(E)\det(B)$$

completing the proof for this type of elementary matrix. The proofs for the other two types of elementary matrices are similar and are left as exercises. ■ ■

We now use Theorem 5.14 to prove Theorem 5.12.

Proof of Theorem 5.12 First, if A is singular, then so is AB (see Exercise 68, Section 3.3), so that det(A) det(B) = 0 and det(AB) = 0, proving the theorem in this case. Now suppose that A is nonsingular and hence has an inverse. Then there exists a sequence of row operations that will transform A into I_n. Let E_1, E_2, … E_k denote the corresponding elementary matrices, with E_1 inducing the first row operation, the E_2 second, and so on. Then E_k … E_2 E_1 A = I_n, so that

$$A=(E_k \cdots E_2E_1)^{-1}=E_1^{-1}E_2^{-1} \cdots E_k^{-1}$$

It is not hard to verify that the inverse of an elementary matrix is another elementary matrix (see Exercise 64), so that A is the product of elementary matrices. By repeatedly applying Theorem 5.14, we have

$$\det(AB)=\det(E_1^{-1} \cdots E_2^{-1} \cdots E_k^{-1}B) =\det(E_1^{-1})\det(E_2^{-1} \cdots E_k^{-1}B) \; \vdots =\det(E_1^{-1})\det(E_2^{-1}) \cdots \det(E_k^{-1})\det(B) =\det(E_1^{-1}E_2^{-1} \cdots E_k^{-1})\det(B)=\det(A)\det(B)$$

so that det(AB) = det(A) det(B). ■ ■

Theorem 5.15 is an immediate consequence of Theorem 5.12.

THEOREM 5.15 ▶

Let A be an $n \times n$ invertible matrix. Then

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

Proof Since A is invertible, A^{-1} exists and $AA^{-1} = I_n$. Therefore

$$1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$$

with the last equality holding by Theorem 5.12. Since $\det(A)\det(A^{-1}) = 1$, we have

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

■■

Determinants of Partitioned Matrices

Suppose that we have the 5 × 5 partitioned matrix $P = \begin{bmatrix} A & 0_{23} \\ 0_{32} & D \end{bmatrix}$ with blocks

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } D = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}$$

▶ Partitioned matrices were introduced in Section 3.2. Recall that 0_{nm} denotes an $n \times m$ matrix with all entries equal to zero.

Our goal is to find a formula for $\det(P)$ in terms of $\det(A)$ and $\det(D)$. *Thinking about 2 × 2 matrices, we have*

$$\begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} = ad$$

which suggests the possibility that $\det(P) = \det(A)\det(D)$. To see if this is true, we employ cofactor expansion across the top row of P, which produces

det(P)=a11|a220000d11d12d130d21d22d230d31d32d33|-
a12|a210000d11d12d130d21d22d230d31d32d33|

Applying cofactor expansion across the top rows of both determinants, we have

det(P)=a11a22|d11d12d13d21d22d23d31d32d33|-
a12a21|d11d12d13d21d22d23d31d32d33| =(a11a22-
a12a21)det(D)=det(A)det(D)

Hence it is true that det(P) = det(A) det(D). Looking back over our computations, we see that we would have arrived at the same formula even if

P=[A023CD]

for any 3 × 2 matrix C, because the entries of C would not contribute to the cofactor expansions. This is again consistent with

|a0cd|=ad

A similar argument using cofactor expansions down columns can be used to show that

|AB032D|=det(A)det(D)

for any 2 × 3 matrix B (see Exercise 66). These observations generalize to partitioned square matrices of higher dimension.

THEOREM 5.16 ▶

Let P be a partitioned n × n matrix of the form

P=[AB0D]orP=[A0CD]

where A and D are square block submatrices. Then det(P) = det(A) det(D).

A proof can be formulated using induction on the dimension of P. This is left as Exercise 67.

It is tempting to speculate that

|ABCD|=det(A)det(D)-det(B)det(C)

when A, B, C, and D are square submatrices, but this turns out not to be true in general. See Exercises 33–34 for counterexamples.

Proof of Theorem 5.13

Proof We take each part of the theorem in turn.

(a) Suppose that the matrix B results from A by interchanging two adjacent rows R_i and R_{i+1}. Using cofactor expansion along R_i of A gives

det(A)=ai1(-1)i+1det(Mij)+ ⋯ +ain(-1)i+ndet(Min)

Now suppose that we compute det(B) by using cofactor expansion along of R_{i+1} B. Since we interchanged R_i and R_{i+1} to get B, the entries of R_{i+1} of B are the same as those of R_i of A, as are the matrices M_{ij} corresponding to these entries. Hence

det(B)=ai1(-1)i+2det(Mi1)+⋯+ain(-1)i+1+ndet(Min) =(-1)
{ai1(-1)i+1det(Mi1)+ ⋯ +ain(-1)i+ndet(Min)}=-det(A)

Keeping this in mind, let's consider the general case. Suppose that B results from A by the operation Ri↔Rj (interchanging rows R_i and R_j), where for convenience we assume that $i < j$. is operation can be accomplished by two sequences of interchanges of adjacent rows. Start with the sequence Ri↔Ri+1, Ri↔Ri+2, … , Rj-1↔Rj. When these $j - i$ interchanges are complete, the elements of rows $i + 1$ through j are shifted up one row, and the elements in row i are moved to row j (see Figure 1).

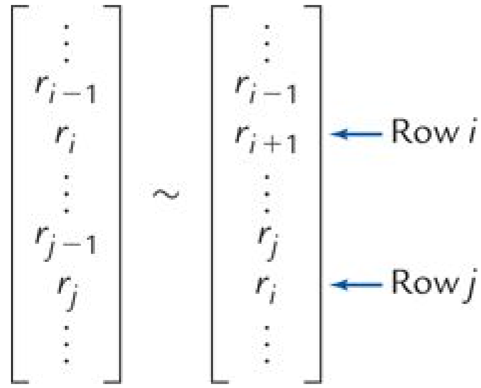


Figure 1 The result of the first $j - i$ adjacent row interchanges.

We shift the elements originally in row j up to row i with the sequence of $j - i - 1$ interchanges of adjacent rows $R_{j-2} \leftrightarrow R_{j-1}$, $R_{j-3} \leftrightarrow R_{j-2}, \ldots$ $R_i \leftrightarrow R_{i+1}$ (see Figure 2). At this point we have $R_i \leftrightarrow R_j$, and all the other rows are back where they started.

Returning to the relationship between det(A) and det(B), by our earlier observation each interchange of adjacent rows multiplied the determinant by −1. Since there are a total of $2(j - i) - 1$ such interchanges, we have

$$\det(B) = (-1)^{2(j-i)-1} \det(A) = -\det(A)$$

as stated in part (a) of the theorem.

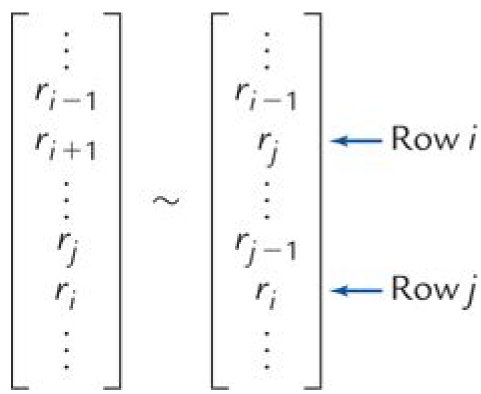


Figure 2 The result of the second $j - i - 1$ adjacent row interchanges.

(b) Suppose that B is produced by multiplying row i of A by a scalar c. Using cofactor expansion along row i of B, we find that

$$\det(B) = c a_{i1} C_{i1} + \cdots + c a_{in} C_{in} = c(a_{i1} C_{i1} + \cdots + a_{in} C_{in}) = c \det(A)$$

so that det(A)=1cdet(B).

(c) Suppose that B results from applying to A the row operation $cR_i + R_j \to R_j$. Using cofactor expansion along row j of B yields

det(B)=(cai1+aj1)Cj1+ ⋯ +(cain+ajn)Cjn =
(aj1Cj1+ ⋯ +ajnCjn)+c(ai1Cj1+ ⋯ +ainCjn)

The term in the left parentheses is the cofactor expansion along row j of A and so is equal to det(A). The term in the right parentheses is equal to zero (see Exercise 75 in Section 5.1). Hence

det(B)=det(A)+c(0)=det(A)

completing the proof of part (c) and the theorem. ■ ■

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Compute det(A).

(a) A=[32-10-2100-5].

(b) A=[000-2001-102-172137].

2. Suppose A and B are 3 × 3 with $\det(A) = 3$ and $\det(B) = -1$. Compute the following.

(a) $\det(A^4)$

(b) $\det(-2A^2)$

(c) $\det(AB^2)$

(d) $\det(3A^2B)$

3. The given row operations, when performed on a matrix A, result in the given matrix B. Find the determinant of A, and decide if A is invertible.

(a) R1↔R2-R1+R2→R2⇒B=[270-3]

(b) R2↔R1-3R1+R3→R34R2+R3→R3⇒B=[2-140-23005]

4. Determine if each statement is true or false, and justify your answer.
   (a) An elementary row operation can change det(A) from nonzero to zero.
   (b) There are no 2 × 2 matrices A such that det(A)= det(A^2).
   (c) If A and B are 3 × 3 matrices, then det(A − B) = det(A) − det(B).
   (d) If A is invertible and B is not, then AB is invertible.

EXERCISES

Exercises 1–6: Compute the determinant of A by using row operations to reduce to echelon form, as illustrated in Example 2.

1. A=[28-1-3]
2. A=[9-132]
3. A=[1-1-3-226-3-310]
4. A=[-42-21013-11]
5. A=[0100100000010010]
6. A=[-32-167-32-11000-2100-1]

Exercises 7–14: The given row operations, when performed on a matrix A, result in the given matrix B. Find the determinant of A and decide if A is invertible.

7. R1↔R22R1+R2→R2⇒B=[1-50-4]
8. -3R1→R1-R1+R2→R2⇒B=[1703]
9. R3↔R1-2R1+R2→R25R2+R3→R3⇒B=[1-490-32007]
10. (1/2)R1→R15R1+R2→R2R2↔R3-4R2+R3→R3⇒B=[20-70-11006]
11. -7R2→R2-3R1+R3→R35R2+R3→R3⇒B=[-62-100100-4]
12. R1→R32R1+R3→R3R2↔R3-3R3→R3⇒B=[1-2-004300-2]
13. R1→R4-5R1+R3→R3-2R1+R4→R4R2→R3⇒B=[1052012000210004]

14. R2↔R3R2+R3→R3-7R2+R4→R4R3↔R4-5R4→R4⇒B=[1200031100160005]

Exercises 15–18: Suppose that

det(A)=|abcdefghi|=3

and find the determinant of the given matrix.

15. [defabcghi]

16. [defghiabc]

17. [abc-2d-2e-2fa+gb+hc+i]

18. [abc5g5h5idef]

Exercises 19–22: Verify that det(A) det(B) = det(AB) and that det(A + B) ≠ det(A) + det(B).

19. A=[231-4],B=[0-137]

20. A=[-1524],B=[7321]

21. A=[20-1110011],B=[221204-131]

22. A=[00-2203111],B=[1-201-1314-1]

23. Suppose that A is a square matrix with det(A) = 3. Find each of the following:

(a) det(A^2)

(b) det(A^4)

(c) det(A^2 A^T)

(d) det(A^{-1})

24. Suppose that B is a square matrix with det(B) = −2. Find each of the following:

(a) det(B^3)

(b) det(B^5)

(c) det(BB^T)

(d) det($(B^{-1})^3$)

25. Suppose that A and B are $n \times n$ matrices with det(A) = 3 and det(B) = −2. Find each of the following:

(a) det(A^2B^3)

(b) $\det(AB^{-1})$

(c) $\det(B^3 A^T)$

(d) $\det(A^2 B^3 B^T)$

26. Suppose that A and B are $n \times n$ matrices with $\det(A) = 4$ and $\det(B) = -3$. Find each of the following:

(a) $\det(B^3 A^2)$

(b) $\det(A^{-2} B^4)$

(c) $\det(A^2 B^T)$

(d) $\det(B^{-2} A^3 A^T)$

Exercises 27–32: Partition the matrix A in order to compute $\det(A)$.

27. A=[1-3002500005-10033]

28. A=[-2020013100020400000-3200012]

29. A=[4211315900320010]

30. A=[1267923825001210021300321]

31. A=[9300310039228131]

32. A=[23000-1500067132-4103105210]

Exercises 33–34: For the given matrices A, B, C, and D verify that |ABCD|≠det(A)det(D)-det(B)det(C).

33. A=[1204],B=[-1051],C=[0-2-13],D=[2140]

34. A=[-1024],B=[01-32],C=[320-3],D=[1-151]

Exercises 35–40: Determine if a unique solution exists for the given linear system.

35. 6x1-5x2=12-2x1+7x2=0

36. 10x1-5x2=5-4x1+2x2=-3

37. 3x1+2x2+7x3=0-3x3=-3-x2-4x3=13

38. -2x1+5x2-10x3=4x1-2x2+3x3=-17x1-17x2+34x3=-16

39. x1+x2-2x3=-33x1-2x2+2x3=96x1-7x2-x3=4

40. x1-3x2+2x3=4-2x1+7x2-2x3=-74x1-13x2+7x3=12

FIND AN EXAMPLE Exercises 41–46: Find an example that meets the given specifications.

41. A nonzero 2×2 matrix A such that $3\det(A) = \det(3A)$.

42. A nonzero 3×3 matrix A such that $-2\det(A) = \det(-2A)$.

43. Find 2×2 matrices A and B, both nonzero, such that $\det(A + B) = \det(A) + \det(B)$. (NOTE: This identity is not generally true, but there are examples where it holds.)

44. Find 3×3 matrices A and B, both nonzero, such that $\det(A + B) = \det(A) + \det(B)$. (NOTE: This identity is not generally true, but there are examples where it holds.)

45. Find a 3×3 matrix A such that $\det(A) = 1$ and all entries of A are nonzero. (HINT: Start with an upper triangular matrix that has the specified determinant, then use row operations to obtain A.)

46. Find a 4×4 matrix A such that $\det(A) = 1$ and all entries of A are nonzero. (HINT: Start with an upper triangular matrix that has the specified determinant, then use row operations to obtain A.)

TRUE OR FALSE Exercises 47–52: Determine if the statement is true or false, and justify your answer.

47.

(a) Interchanging the rows of a matrix has no effect on its determinant.

(b) If $\det(A) \neq 0$, then the columns of A are linearly independent.

48.

(a) If E is an elementary matrix, then $\det(E) = 1$.

(b) If A and B are $n \times n$ matrices, then $\det(A + B) = \det(A) + \det(B)$.

49.

(a) If A is a 3×3 matrix and $\det(A) = 0$, then $\text{rank}(A) = 0$.

(b) If A is a 4×4 matrix and $\det(A) = 4$, then $\text{nullity}(A) = 0$.

50.

(a) Suppose A, B, and S are $n \times n$ matrices, and that S is invertible. If $B = S^{-1}AS$, then $\det(A) = \det(B)$.

(b) If A is an $n \times n$ matrix with all entries equal to 1, then $\det(A) = n$.

51.

(a) Suppose that A is a 4×4 matrix and that B is the matrix obtained by multiplying the third column of A by 2. Then $\det(B) = 2\det(A)$.

(b) If A is an invertible matrix, then at least one of the submatrices M_{ij} of A is also invertible.

52.

(a) For any matrix A, $\det(A^T A)$ is defined.

(b) For any matrix A, $\det(A^T A) = (\det(A))^2$.

Exercises 53–62: Assume that A is an $n \times n$ matrix.

53. Prove that if A has two identical rows, then $\det(A) = 0$.

54. Prove that if A has two identical columns, then $\det(A) = 0$. (HINT: Apply Theorem 5.10 and Exercise 53.)

55. Prove that $\det(A^T A) \geq 0$.

56. Suppose that $\det(A) = 2$. Prove that A^{-1} cannot have all integer entries.

57. Prove that $\det(-A) = (-1)^n \det(A)$.

58. Prove that $\det(cA) = c^n \det(A)$.

59. Suppose that A is *idempotent*, which means $A = A^2$. What are the possible values of $\det(A)$?

60. Suppose that A is *skew symmetric*, which means $A = -A^T$. Show that if n is odd, then $\det(A) = 0$.

61. Prove that if n is odd, then $A^2 \neq -I_n$. (HINT: Compare determinants of A^2 and $-I_n$.)

62. Show that if the entries of each row of A add to zero, then $\det(A) = 0$. (HINT: Think linear independence and The Unifying Theorem.)

63. Prove that a square matrix A has an echelon form B such that $\det(A) = \pm\det(B)$.

64. Suppose that E is an elementary matrix. Show that E^{-1} is also an elementary matrix.

65. This exercise completes the proof of Theorem 5.14. Let B be an $n \times n$ matrix and E be an $n \times n$ elementary matrix.

(a) Suppose that E corresponds to multiplying a row by a scalar c. Show that $\det(EB) = \det(E)\det(B) = c\det(B)$.

(b) Suppose that E corresponds to adding a multiple of one row to another. Show that $\det(EB) = \det(E)\det(B) = \det(B)$.

66. Prove that

$$\begin{vmatrix} A & B \\ 0_{32} & D \end{vmatrix} = \det(A)\det(D)$$

where A is a 2×2 matrix, B is a 2×3 matrix, D is a 3×3 matrix, and 0_{32} is a 3×2 matrix with all entries equal to zero.

67. Prove Theorem 5.16: Let M be a partitioned $n \times n$ matrix of either of the forms

$$M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \quad \text{or} \quad M = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix}$$

where A and D are square block submatrices. Then $\det(M) = \det(A)\det(D)$. (HINT: Show the formula holds for the first form of M using induction on the number of rows of M, and then take the transpose to show that the formula holds for the second form.)

68. The *Vandermonde matrix* is given by

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

(a) For the Vandermonde matrix with $n = 3$, show that

$$\begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix} = (x_3 - x_2)(x_3 - x_1)(x_2 - x_1)$$

(b) For any $n > 1$, prove that

$$\det(V) = \prod_{1 \le i \le j \le n} (x_j - x_i)$$

Exercises 69–70: Verify *Sylvester's determinant theorem*, which states that

$$\det(I_m + AB) = \det(I_n + BA)$$

for any $m \times n$ matrix A and $n \times m$ matrix B.

69. A=[3-264052-915-1-4]B=[0-3261442-8305]

70. A=[623-1470-245-82490]B=[24-6052-377028-935]

5.3 Applications of the Determinant

In this section we consider a few applications of the determinant, beginning with a method for using determinants to find the solution to the linear systems $A\mathbf{x} = \mathbf{b}$ when A is an invertible square matrix. From the Unifying Theorem, Version 7, we know that in this case there will be a unique solution.

▶ This section is optional and can be omitted without loss of continuity.

Before stating the theorem, we need to introduce some notation. If $A = [\mathbf{a}_1 \quad \mathbf{a}_2 \ \ldots \ \mathbf{a}_n]$ is an $n \times n$ matrix and $\mathbf{b}$ is in $\mathbf{R}^n$, then let A_i denote the matrix A after replacing $\mathbf{a}_i$ with $\mathbf{b}$. That is,

Ai=[a1 ⋯ ai-1bai+1 ⋯ an]

For instance, if

A=[4-120375-2-52040611]andb=[892-3]

then

A1=[8-120975-22204-3611]andA3=[4-180379-2-522406-31]

Matrices of this type are used in the next theorem.

THEOREM 5.17 ▶

(CRAMER'S RULE) Let A be an invertible $n \times n$ matrix. Then the components of the unique solution to $A\mathbf{x} = \mathbf{b}$ are given by

$$x_i = \frac{\det(A_i)}{\det(A)} \quad \text{for } i = 1, 2, \ldots, n$$

The proof of Theorem 5.17 is given at the end of this section. For now, let's look at an example.

Example 1

Use Cramer's Rule to find the solution to the system

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

Solution The system is equivalent to $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 5 \\ -2 \\ -1 \end{bmatrix}$$

We have

$$A_1 = \begin{bmatrix} 5 & 1 & 0 \\ -2 & 2 & 1 \\ -1 & -1 & 2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 3 & 5 & 0 \\ -1 & -2 & 1 \\ 0 & -1 & 2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 3 & 1 & 5 \\ -1 & 2 & -2 \\ 0 & -1 & -1 \end{bmatrix}$$

Computing determinants gives us $\det(A) = 17$, $\det(A_1) = 28$, $\det(A_2) = 1$, and $\det(A_3) = -8$. Therefore, by Cramer's Rule, the solution to $A\mathbf{x} = \mathbf{b}$ is

$$x_1 = \frac{\det(A_1)}{\det(A)} = \frac{28}{17}, \quad x_2 = \frac{\det(A_2)}{\det(A)} = \frac{1}{17}, \quad x_3 = \frac{\det(A_3)}{\det(A)} = -\frac{8}{17}$$

▶ The determinants in Example 1 were computed using the Shortcut Method.

Cramer's Rule is easy to implement when the coefficient matrix A is 2×2 or 3×3. Unfortunately, for larger systems all the required determinants generally make this method computationally impractical. (See the "Computational Comment" at the end of Section 5.1.)

Inverses from Determinants

In Chapter 3 we showed how to adapt our row operation algorithm for finding the solutions to a linear system to determine the inverse of a square matrix. Here we do something similar, using Cramer's Rule to develop a formula for finding inverses. We start with a statement of the formula and then explain why it works.

Cofactor Matrix

For an $n \times n$ matrix A, the **cofactor matrix** is given by

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

Adjoint Matrix

where the cofactors C_{ij} are as defined in Section 5.1. Now we define the **adjoint** of A, denoted adj(A), by

$$\text{adj}(A)=C^T=\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

THEOREM 5.18

If A is an invertible matrix, then

$$A^{-1}=\frac{1}{\det(A)}\text{adj}(A) \tag{1}$$

Proof We prove this theorem by showing that $A\left(\frac{1}{\det(A)}\text{adj}(A)\right)=I_n$. Start by forming the product

$$A(\text{adj}(A))=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

The entry in row i and column j of $A(\text{adj}(A))$ is

$$a_{i1}C_{j1}+a_{i2}C_{j2}+\cdots+a_{in}C_{in} \tag{2}$$

When $i = j$, (2) is the cofactor expansion across row i of A and is equal to det(A) by Theorem 5.8, Section 5.1. If $i \neq j$, then (2) is equal to zero (see Exercise 75 in Section 5.1). Hence we have

$$A(\operatorname{adj}(A)) = \det(A) I_n \Rightarrow A\left(\frac{1}{\det(A)}\operatorname{adj}(A)\right) = I_n$$

and so $A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A)$.

Example 2

Use Theorem 5.18 to find the inverse of the matrix

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

Solution From Example 1 we know that det(A) = 17. Since A is 3×3, A has nine cofactors. Four of them are

$$C_{11} = (-1)^2 \begin{vmatrix} 2 & 1 \\ -1 & 2 \end{vmatrix} = 5 \qquad C_{21} = (-1)^3 \begin{vmatrix} 1 & 0 \\ -1 & 2 \end{vmatrix} = -2 \qquad C_{31} = (-1)^4 \begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix} = 1 \qquad C_{12} = (-1)^3 \begin{vmatrix} -1 & 1 \\ 0 & 2 \end{vmatrix} = 2$$

The remaining five are computed similarly, yielding $C_{22} = 6$, $C_{32} = -3$, $C_{13} = 1$, $C_{23} = 3$, and $C_{33} = 7$. Filling out the adjoint of A gives the inverse

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A) = \frac{1}{17}\begin{bmatrix} 5 & -2 & 1 \\ 2 & 6 & -3 \\ 1 & 3 & 7 \end{bmatrix} = \begin{bmatrix} \frac{5}{17} & -\frac{2}{17} & \frac{1}{17} \\ \frac{2}{17} & \frac{6}{17} & -\frac{3}{17} \\ \frac{1}{17} & \frac{3}{17} & \frac{7}{17} \end{bmatrix}$$

In Section 3.3, we encountered the “Quick Formula” for computing the inverse of a 2×2 matrix. This is revisited in the next example.

Example 3

Use Theorem 5.18 to find the inverse of the matrix

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

Solution We have det(A) = $ad - bc$, which we know must be nonzero for an inverse to exist. The cofactors of A are

$$C_{11}=d, \quad C_{21}=-b, \quad C_{12}=-c, \quad C_{22}=a$$

Therefore

$$A^{-1}=\frac{1}{\det(A)}\operatorname{adj}(A)=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

which matches the Quick Formula.

Area and Determinants

There is a striking relationship between area, determinants, and linear transformations. Although we focus on $\mathbf{R}^2$ here, the results developed are also true in higher dimensions.

Let $\mathcal{S}$ denote the unit square in the first quadrant of $\mathbf{R}^2$, and suppose that T: $\mathbf{R}^2 \to \mathbf{R}^2$ is a linear transformation. Let $A=[\mathbf{a}_1\ \mathbf{a}_2]$ be the 2 × 2 matrix such that $T(\mathbf{x}) = A\mathbf{x}$. If $\mathcal{P}=T(\mathcal{S})$ denotes the image of $\mathcal{S}$ under T and A is invertible, then $\mathcal{P}$ is a parallelogram in $\mathbf{R}^2$ (see Exercise 65 in Section 3.1). An example is shown in Figure 1.

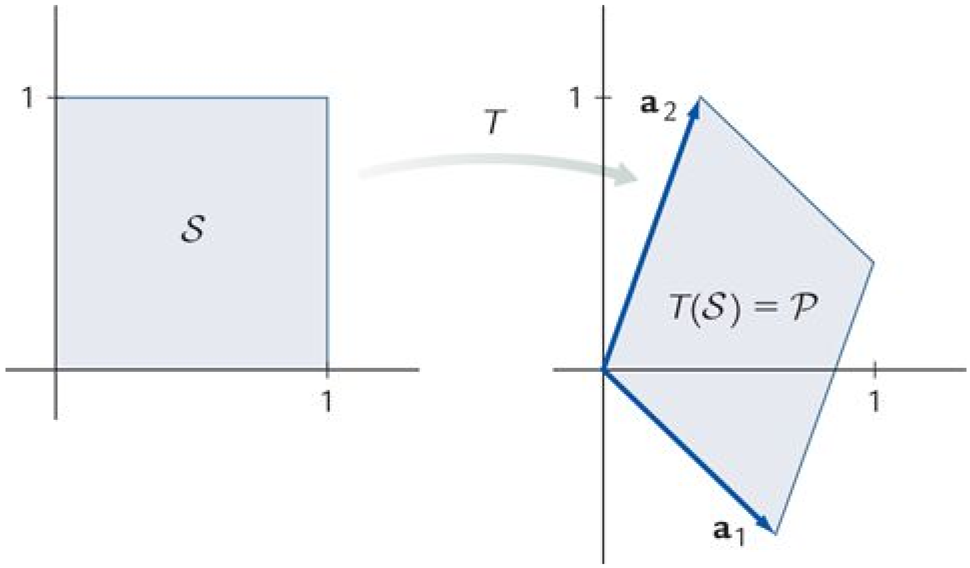


Figure 1 The square $\mathcal{S}$ and $\mathcal{P}$=T($\mathcal{S}$), where $T(\mathbf{x}) = A\mathbf{x}$ and $A = [\mathbf{a}_1\ \mathbf{a}_2]$.

Theorem 5.19 shows how the area of $\mathcal{P}$ is related to the determinant of A.

THEOREM 5.19 ▶

Let $\mathcal{S}$ be the unit square in the first quadrant of $\mathbf{R}^2$, and let T: $\mathbf{R}^2 \rightarrow \mathbf{R}^2$ with $T(\mathbf{x}) = A\mathbf{x}$. If $\mathcal{P}$=T($\mathcal{S}$) is the image of $\mathcal{S}$ under T, then

$$\text{area}(\mathcal{P})=|\det(A)| \tag{3}$$

where area ($\mathcal{P}$) denotes the area of $\mathcal{P}$.

Proof Let A=[a1a2]. First suppose that the columns of A are linearly dependent. Then T($\mathcal{S}$) is a line segment (see Exercise 65, Section 3.1) so that area($\mathcal{P}$)=0. We also have $\det(A) = 0$ by the Unifying Theorem, Version 7, and so (3) is true in this case.

Next suppose that the columns of A are linearly independent, so that $\mathcal{P}$ is a parallelogram. Rotate $\mathcal{P}$ about the origin through the angle θ so that the rotated image of $\mathbf{a}_1$ ends up on the x-axis and the rotated parallelogram $\mathcal{P}^*$ is above the x-axis (see Figure 2).

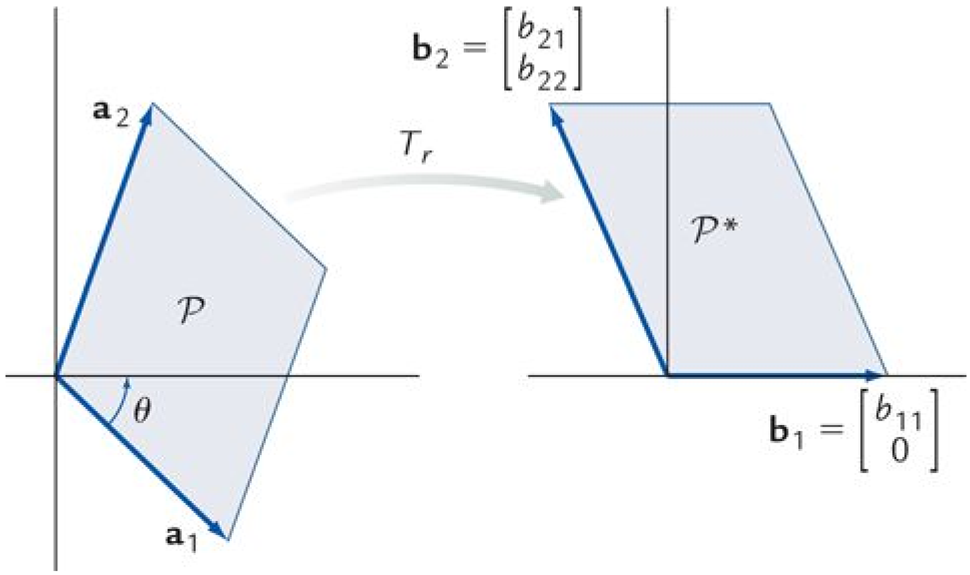


Figure 2 Rotating $\mathcal{P}$ about the origin by an angle θ yields a new region $\mathcal{P}^*$ of equal area.

Rotation preserves area, so that

$$\text{area}(\mathcal{P}) = \text{area}(\mathcal{P}^*) \tag{4}$$

The rotation of $\mathcal{P}$ to $\mathcal{P}^*$ is achieved with the linear transformation $T_r(\mathbf{x}) = C\mathbf{x}$, where (see Section 3.1)

$$C = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

If $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix}$, where $\mathbf{b}_1$ and $\mathbf{b}_2$ are as in Figure 2, then formulas from geometry tell us that

▶ It is possible that $b_{11} < 0$. Absolute values are included so that (5) is true in general.

$$\text{area}(\mathcal{P}^*) = |b_{11} b_{22}| = |\det(B)| \tag{5}$$

If $T_1(\mathbf{x}) = B\mathbf{x}$, then

$$T_1(e_1)=[b_1\ b_2][1\ 0]=b_1=T_r(a_1)=T_r(T(e_1))\quad T_1(e_2)=[b_1\ b_2][0\ 1]=b_2=T_r(a_2)=T_r(T(e_2))$$

Since $\{\mathbf{e}_1, \mathbf{e}_2\}$ is a basis for $\mathbf{R}^2$, it must be that $T_1(\mathbf{x})$ and $T_r(T(\mathbf{x}))$ are the same linear transformation. Since $T_1(\mathbf{x}) = B\mathbf{x}$ and $T_r(T(\mathbf{x})) = CA\mathbf{x}$, we have $B = CA$. Therefore

$$\det(B)=\det(CA)=\det(C)\det(A)=\det(A) \tag{6}$$

Combining (4), (5), and (6), we conclude that $\text{area}(\mathcal{P})=|\det(A)|$. ■ ■

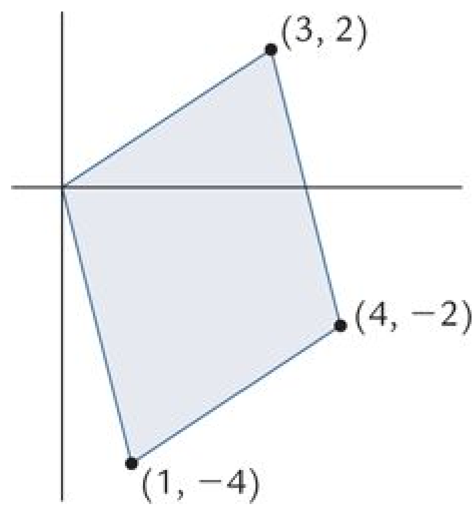


Figure 3 A parallelogram in $\mathbf{R}^2$.

Example 4

Use Theorem 5.19 to find the area of the parallelogram in Figure 3.

Solution Let $A=\begin{bmatrix}3 & 1\\ 2 & -4\end{bmatrix}$. Then $T(\mathbf{x}) = A\mathbf{x}$ will map $\mathcal{S}$, the unit square in the first quadrant, onto $\mathcal{P}$. Hence, by Theorem 5.19,

$$\text{area}(\mathcal{P})=|\det(A)|=|-14|=14$$

We now generalize Theorem 5.19 to arbitrary regions of finite area.

THEOREM 5.20 ▸

Let $\mathcal{D}$ be a region of finite area in $\mathbf{R}^2$, and suppose that T: $\mathbf{R}^2 \to \mathbf{R}^2$ with $T(\mathbf{x}) = A\mathbf{x}$. If $T(\mathcal{D})$ denotes the image of $\mathcal{D}$ under T, then

$$\text{area}\,(T(\mathcal{D}))=|\det(A)|\cdot\text{area}(\mathcal{D}) \qquad (7)$$

Proof We give a complete proof for rectangular regions $\mathcal{R}$, and after that we sketch the method of proof for general regions $\mathcal{D}$.

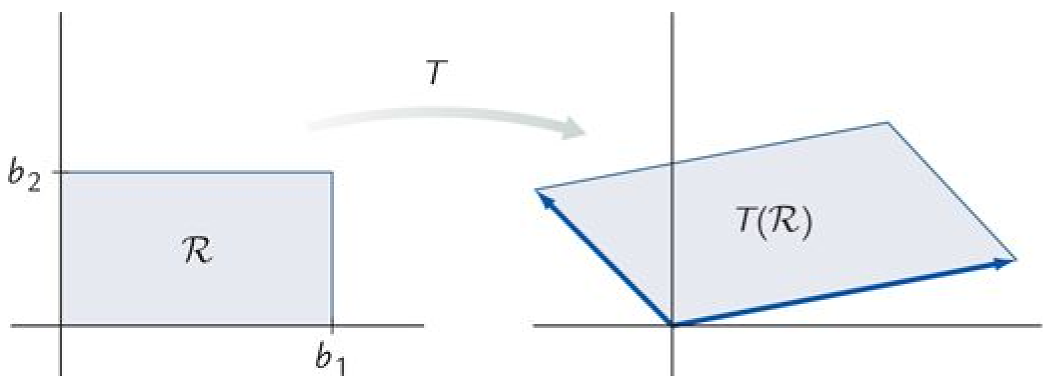


Figure 4 A rectangular region $\mathcal{R}$ and its image $T(\mathcal{R})$.

Our proof strategy is to find the linear transformation T_1 such that T1(S)=T(R) ($\mathcal{S}$ denotes the unit square) and then apply Theorem 5.19 (Figure 4). If for $T_B(\mathbf{x}) = B^{\mathbf{x}}$ for

B=[b100b2]

then TB($\mathcal{S}$)=$\mathcal{R}$ (see Figure 5 and Exercise 47).

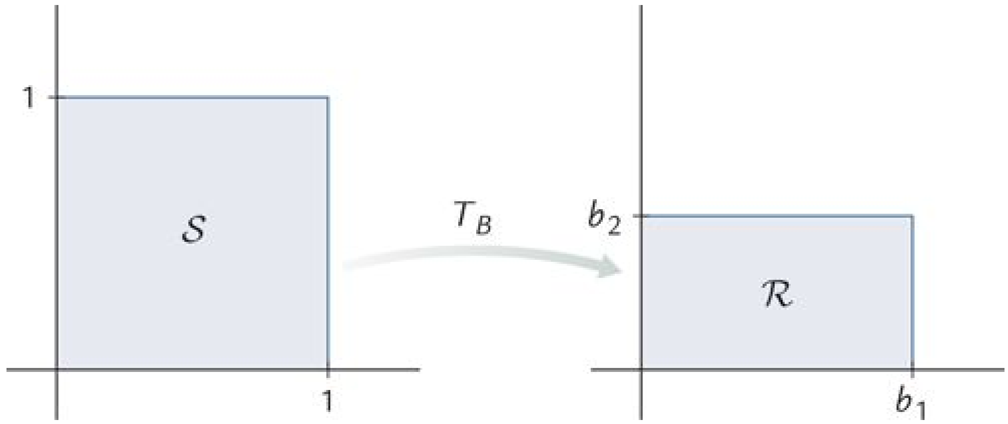


Figure 5 The image of S under T_B is the rectangular region $\mathcal{R}$.

Since T($\mathcal{R}$) is the image $\mathcal{R}$ under T, then T($\mathcal{R}$) is also the image of $\mathcal{S}$ under the composition

$$T1(x)=T(TB(x))$$

That is, T1($\mathcal{S}$) = T($\mathcal{R}$). Since $T(\mathbf{x}) = A\mathbf{x}$, we have $T_1(\mathbf{x}) = T(T_B(\mathbf{x})) = AB\mathbf{x}$. Hence, by Theorem 5.19, we have

$$\text{area}(T(\mathcal{R}))=|\det(AB)| =|\det(A)\cdot\det(B)| =|b1b2\cdot\det(A)| =|\det(A)| \cdot\text{area}(\mathcal{R})$$

▶ The area of R is $b_1 b_2$.

Thus (7) is true in this case.

An arbitrary rectangular region $\mathcal{R}$* with sides parallel to the coordinate axes (see Figure 6) has the form v+$\mathcal{R}$, where **v** is a fixed vector and $\mathcal{R}$ is a rectangular region of the type in Figure 4.

Since each vector **r*** in $\mathcal{R}$* has the form **v** + **r** for some **r** in R, we have

$$T(\mathbf{r}^*) = T(\mathbf{v} + \mathbf{r}) = T(\mathbf{v}) + T(\mathbf{r})$$

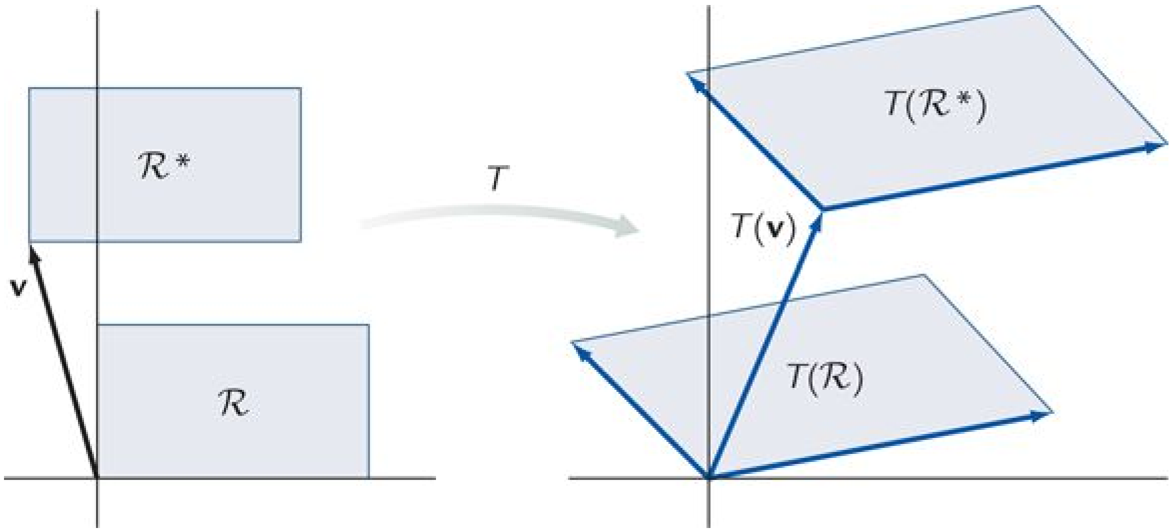


Figure 6 A rectangular region $\mathcal{R}^*$ and its image $T(\mathcal{R}^*)$.

Hence $T(\mathcal{R}^*)$ is a translation of $T(\mathcal{R})$, with $T(\mathcal{R}^*)=T(v)+T(\mathcal{R})$. As translation does not change area, we have $\text{area}(\mathcal{R}^*)=\text{area}(\mathcal{R})$ and area $\text{area}(T(\mathcal{R}^*))=\text{area}(T(\mathcal{R}))$. Therefore

$$\begin{aligned} \text{area}(T(\mathcal{R}^*))=\text{area}(T(\mathcal{R})) &=|\det(A)|\cdot\text{area}(\mathcal{R}) \\ \text{(by the previous case)} &=|\det(A)|\cdot\text{area}(\mathcal{R}^*) \end{aligned}$$

so (7) also holds for translated rectangles. ■ ■

With the formal proof for rectangles complete, now suppose that $\mathcal{D}$ is a general region such as the one shown in Figure 7. Then (7) also holds for this case. The proof uses techniques found in calculus —we give the basic elements of the argument here.

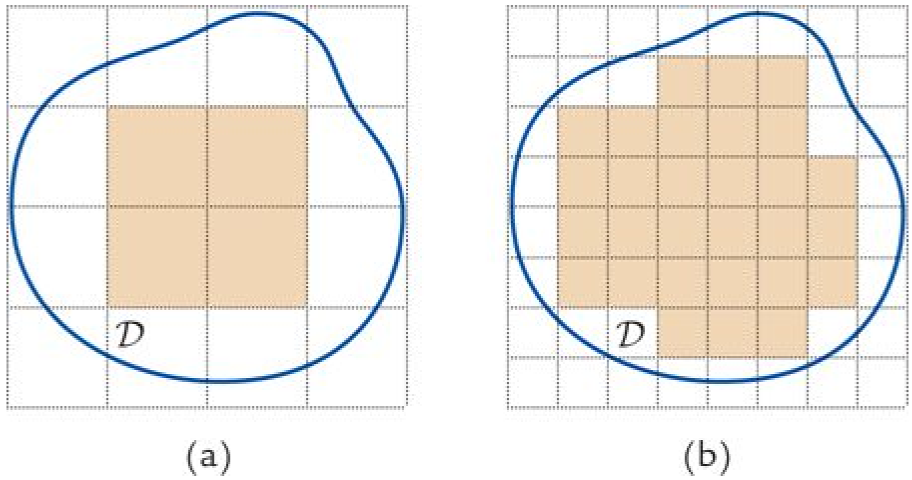


Figure 7 (a) A grid of squares over $\mathcal{D}$. (b) In general, a finer grid of squares over $\mathcal{D}$ gives a more accurate area estimate.

We can approximate the area of $\mathcal{D}$ by superimposing a grid of squares and then counting the number of squares that are inside $\mathcal{D}$ (see Figure 7). Multiplying this count by the area of a single square—which we can easily calculate—gives the area approximation. We can make the approximation as accurate as we like by making the grid of squares suffciently fine.

When we apply a linear transformation T to $\mathcal{D}$, each grid square is mapped to a corresponding parallelogram (see Figure 8).

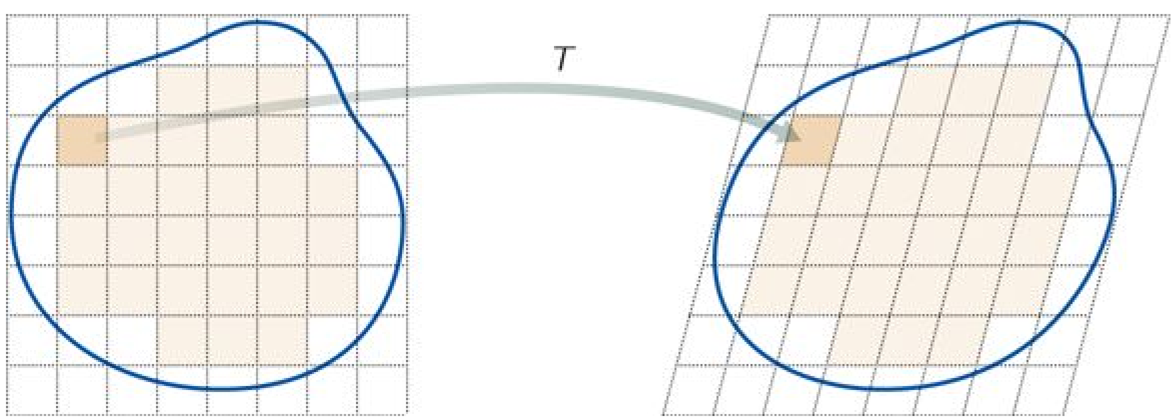


Figure 8 A grid of squares over $\mathcal{D}$ and the corresponding parallelograms over T($\mathcal{D}$).

There is a one-to-one correspondence between grid squares within $\mathcal{D}$ and parallelograms within T($\mathcal{D}$). Furthermore, from (7) we have

$$(\text{area of parallelogram}) = |\det(A)| \cdot (\text{area of grid square})$$

The grid squares provide an approximation to the area of $\mathcal{D}$ and the parallelograms provide an approximation to the area of $T(\mathcal{D})$. Because there are the same number of each, we have

$$\text{area}(T(\mathcal{D})) \approx |\det(A)| \cdot \text{area}(\mathcal{D})$$

In general, the approximation gets better as the grid of squares becomes finer, so that as the size of the grid squares shrinks, in the limit we get (7).

Theorem 5.20 has a higher-dimensional analog, stated below without proof. See Exercises 33–36 for a brief discussion of this theorem in $\mathbf{R}^3$.

THEOREM 5.21 ▸

Let $\mathcal{D}$ be a region of finite volume in $\mathbf{R}^n$, and suppose that $T: \mathbf{R}^n \to \mathbf{R}^n$ with $T(\mathbf{x}) = A\mathbf{x}$. If $T(\mathcal{D})$ denotes the image of $\mathcal{D}$ under T, then

$$\text{volume}(T(\mathcal{D})) = |\det(A)| \cdot \text{volume}(\mathcal{D}) \tag{8}$$

Example 5

Use Theorem 5.20 and the known area of the unit circle $\mathcal{U}$ to find the area of the ellipse $\mathcal{E}$ in Figure 9.

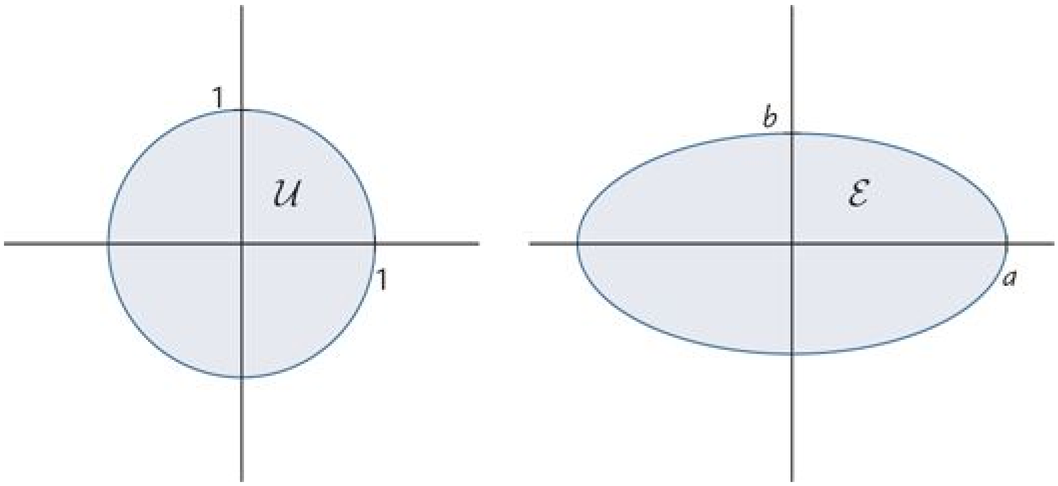


Figure 9 The region inside the unit circle $\mathcal{U}$ is related by a linear transformation to the region inside the ellipse $\mathcal{E}$.

Solution The regions inside the unit circle $\mathcal{U}$ and the ellipse $\mathcal{E}$ are given by the set of all (x_1, x_2) such that

$$x_1^2+x_2^2\le 1 \qquad \left(\frac{x_1}{a}\right)^2+\left(\frac{x_2}{b}\right)^2\le 1 \qquad \text{Unit circle } \mathcal{U} \quad \text{Ellipse } \mathcal{E}$$

The linear transformation

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} ax_1 \\ bx_2 \end{bmatrix} = \underbrace{\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

gives a one-to-one correspondence from $\mathcal{U}$ to $\mathcal{E}$ (see Exercise 48). Since $T(\mathbf{x}) = A\mathbf{x}$ and the area of $\mathcal{U}$ is π, by Theorem 5.20 we have

$$\text{area}(\mathcal{E})=|\det(A)|\cdot\text{area}(\mathcal{U})=ab\pi$$

Proof of Theorem 5.17 (Cramer's Rule)

Proof Since A is invertible, the system $A\mathbf{x} = \mathbf{b}$ has a unique solution. If $I_{n,i}$ denotes the $n \times n$ identity matrix with the ith column replaced by (see margin), then we have

$$I_{n,i} = \begin{array}{c} \begin{bmatrix} 1 & 0 & \cdots & x_1 & \cdots & 0 \\ 0 & 1 & \cdots & x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & x_i & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & x_n & \cdots & 1 \end{bmatrix} \\ \text{Column } i \end{array}$$

$$AI_{n,i}=A[e_1 \cdots x \cdots e_n] =[Ae_1 \cdots Ax \cdots Ae_n] =[a_1 \cdots b \cdots a_n]=A_i$$

where we replace $A\mathbf{x}$ with $\mathbf{b}$. Since $AI_{n,i} = A_i$, it follows that $\det(AI_{n,i}) = \det(A_i)$. Applying Theorem 5.12, we have $\det(AI_{n,i}) = \det(A) \det(I_{n,i})$, so that $\det(I_{n,i}) = \det(A)/\det(A_i)$. On the other hand, cofactor expansion along the ith row of $I_{n,i}$ yields

$$\det(I_{n,i})=x_i\det(I_{n-1})=x_i$$

Therefore

Practice problems can also be used as additional examples.

$$x_i=\det(A_i)\det(A)$$

as stated in the theorem. ■ ■

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Use Cramer's Rule to find the solution to the given system.

(a) $5x_1-3x=7 -x_1+2x_2=-4$

(b) 2x1+x2-2x3=1x1-3x2+5x3=1-x1-4x2+2x3=-6

2. Find adj(A) and use it to find A^{-1}.

(a) A=[133-4]

(b) A=[214150211]

3. Sketch the parallelogram with the given vertices, then determine its area using determinants.

(a) (0, 0), (3, 1), (5, 4), (2, 3)

(b) (0, 0), (2, 3), (3, 7), (1, 4)

4. Determine if each statement is true or false, and justify your answer.

(a) Cramer's Rule works if and only if a given system has the same number of equations and variables.

(b) There is no 2 × 2 matrix A such that adj(A) = A.

(c) If A is 2 × 2, then adj(3A) = 3adj(A).

(d) If $A \sim B$, then adj(A) = adj(B).

EXERCISES

Exercises 1–6: Determine if Cramer's Rule can be applied to find the solution for the linear system, and if so, then find the solution.

1. 6x1-5x2=12-2x1+7x2=0

2. 10x1-5x2=5-4x1+2x2=-3

3. 3x1+2x2+7x3=0   -3x3=-3 -x2-4x3=13

4. -2x1+5x2-10x3=4x1-2x2+3x3=-17x1-17x2+34x3=-16

5. x1+x2-2x3=-33x1-2x2+2x3=96x1-7x2-x3=4

6. x1-3x2+2x3=4-2x1+7x2-2x3=-74x1-13x2+7x3=12

Exercises 7–12: Find the value of x_2 in the unique solution of the linear system.

7. -2x1+3x2=3-3x1-7x2=-1

8. x1-4x2=11-3x1+x2=-2

9. 3x1 +2x3=1 3x2+2x3=32x1+3x2+x3=-4

10. x1-x2 =03x1-x2-3x3=2x1-3x2-2x3=-3

11. 3x2 -3x4=12x1-x2+3x3-3x4=-2-2x1+3x2+2x3+2x4=02x1 +2x3+x4=-1

12. 3x1-3x2 -3x4=5-x1 +2x3+x4=0x1 +3x3 =3 -2x2+3x3+3x4=1

Exercises 13–18: For the matrix *A*, find adj(*A*) and then use it to compute A^{-1}.

13. A=[2537]

14. A=[1716]

15. A=[010001100]

16. A=[201002110]

17. A=[121012001]

18. A=[300120111]

Exercises 19–22: Sketch the parallelogram with the given vertices, then determine its area using determinants.

19. (0, 0), (2, 3), (5, 1), (7, 4)

20. (0, 0), (2, 7), (4, −5), (6, 2)

21. (2, 3), (−1, 4), (5, 7), (2, 8)

22. (3, −1), (0, −2), (5, −6), (2, −7)

Exercises 23–28: Find the area of T($\mathcal{D}$) for $T(\mathbf{x}) = A\mathbf{x}$.

23. $\mathcal{D}$ is the rectangle with vertices (2, 2), (7, 2), (7, 5), (2, 5), and A=[3-152].

24. $\mathcal{D}$ is the rectangle with vertices (−3, 4), (5, 4), (5, 7), (−3, 7), and A=[-2734].

25. $\mathcal{D}$ is the parallelogram with vertices (0, 0), (5, 1), (2, 4), (7, 5), and A=[1425].

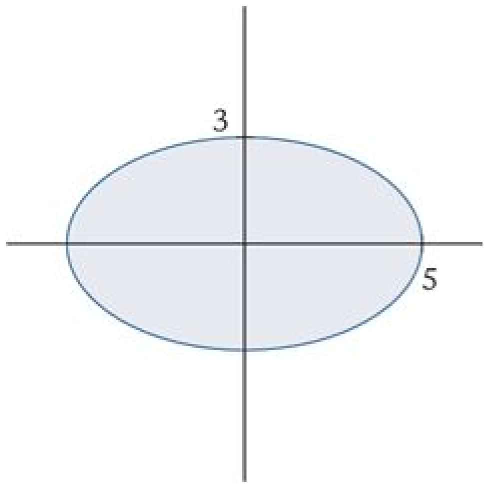
26. $\mathcal{D}$ is the parallelogram with vertices (0, 0), (−2, 3), (3, 5), (1, 8), and A=[5291].

27. $\mathcal{D}$ is the parallelogram with vertices (1, 2), (6, 4), (2, 6), (7, 8), and A=[1425].

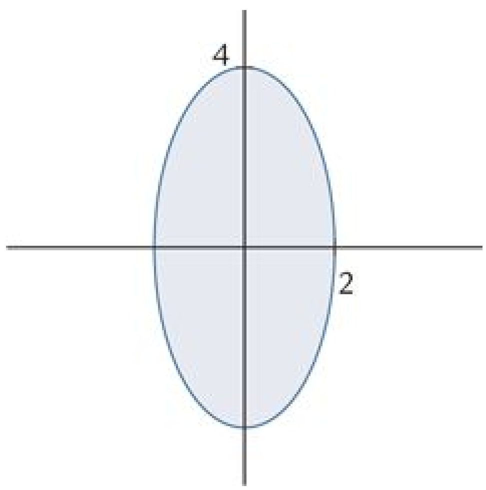
28. $\mathcal{D}$ is the parallelogram with vertices (−2, 1), (−4, 4), (1, 6), (−1, 9), and A=[5291].

Exercises 29–32: Find a linear transformation T that gives a one-to-one correspondence between the unit circle and the given ellipse.

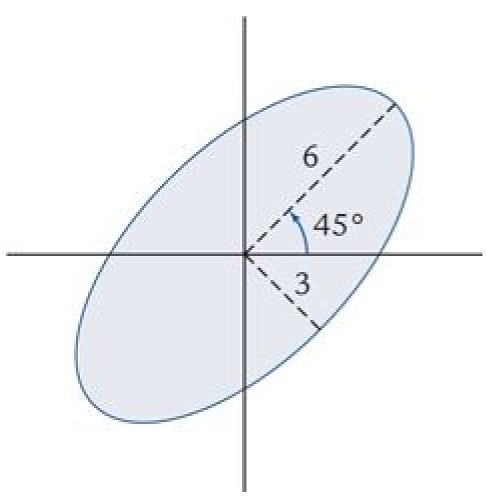
29.



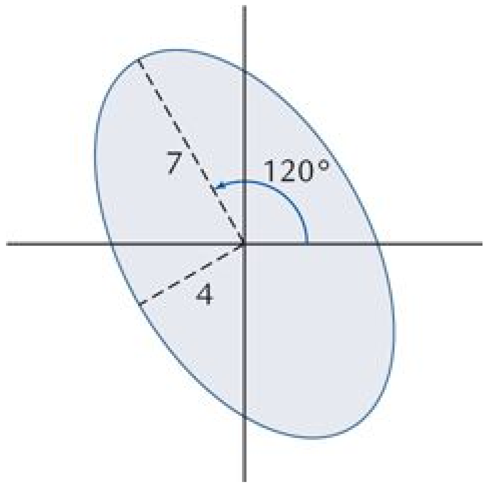
30.



31.



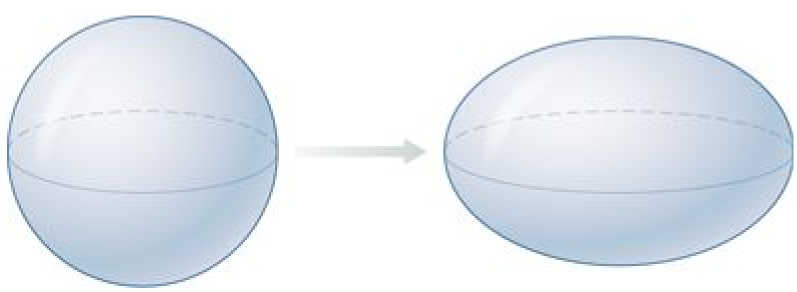
32.



Exercises 33–36: In three dimensions, Theorem 5.21 states that if $\mathcal{D}$ is a region of finite volume in $\mathbf{R}^3$, and T: $\mathbf{R}^3 \to \mathbf{R}^3$ with $T(\mathbf{x}) = A\mathbf{x}$, then

$$\text{volume }(T(\mathcal{D}))=|\det(A)|\cdot\text{volume}(\mathcal{D}) \tag{9}$$

The image of the unit sphere (below left) under a linear transformation is an *ellipsoid* (example below right).

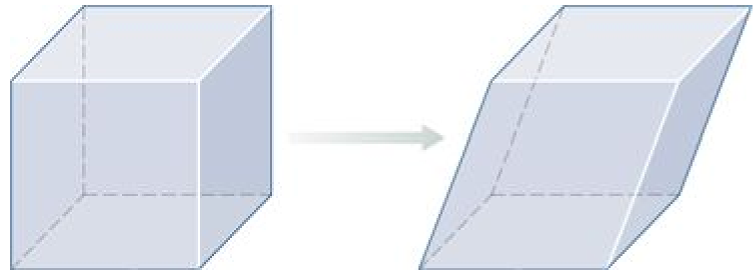


In Exercises 33–34, use (9) to determine the volume of the described ellipsoid.

33. An ellipsoid centered at the origin with axis intercepts $x = \pm 4$, $y = \pm 3$, and $z = \pm 5$.

34. An ellipsoid centered at the origin with axis intercepts $x = \pm 2$, $y = \pm 6$, and $z = \pm 4$.

The image of the unit cube (below left) under a linear transformation is a *parallelepiped* (example below right).



In Exercises 35–36, use (9) to determine the volume of the described parallelepiped.

35. The parallelepiped with sides described by the vectors

[352],[613],[204]

36. The parallelepiped with sides described by the vectors

[-135],[042],[611]

FIND AN EXAMPLE Exercises 37–42: Find an example that meets the given specifications.

37. A linear system with two equations and two unknowns that is consistent but cannot be solved with Cramer's Rule.

38. A linear system with three equations and three unknowns that is consistent but cannot be solved with Cramer's Rule.

39. A parallelogram that has vertices with integer coordinates and area 5.

40. A parallelogram with area 7 that has vertices with integer coordinates that are not on the coordinate axes.

41. A 2 × 2 matrix *A* such that adj(A)=[5-3-21].

42. A 2 × 2 matrix *A* such that adj(A)=[2357].

TRUE OR FALSE Exercises 43–46: Determine if the statement is true or false, and justify your answer.

43.

(a) Cramer's Rule can be used to find the solution to any system that has the same number of equations as unknowns.

(b) If *A* is a square matrix with integer entries, then so is adj(*A*).

44.

(a) If A is a 3 × 3 matrix, then adj(2A) = 2adj(A).

(b) If A is a square matrix that has all positive entries, then so does adj(A).

45.

(a) If A is a 2 × 2 matrix, S is the unit square, and $T(\mathbf{x}) = A\mathbf{x}$, then T($\mathcal{S}$) is a parallelogram of nonzero area.

(b) If A is an $n \times n$ matrix with det(A) = 1, then $A-^{1}$ = adj(A).

46.

(a) Suppose that A is an invertible $n \times n$ matrix with integer entries. If det(A) = 1, then A^{-1} also has integer entries.

(b) If A is a square matrix, then $(\text{adj}(A))^T = \text{adj}(A^T)$.

47. Prove that the linear transformation $T(\mathbf{x}) = B\mathbf{x}$ with

B=[b100b2]

gives a one-to-one correspondence between the interior of the unit square $\mathcal{S}$ and the interior of the rectangle $\mathcal{R}$ shown in Figure 5.

48. Prove that the linear transformation $T(\mathbf{x}) = A\mathbf{x}$ with

A=⌊a00b⌋

gives a one-to-one correspondence between the interior of the unit circle $\mathcal{U}$ and the interior of the ellipse $\mathcal{E}$ shown in Figure 9.

49. Suppose that A is a 3 × 3 matrix with det(A) = −2. Show that A · adj(A) is a 3 × 3 diagonal matrix, and determine the diagonal terms.

50. Prove that if A is an $n \times n$ matrix with linearly independent columns, then so is adj(A).

51. Show that if A is an $n \times n$ symmetric matrix, then adj(A) is also symmetric.

52. Prove that if A is an $n \times n$ diagonal matrix, then so is adj(A).

53. Suppose that A is an $n \times n$ matrix and c is a scalar. Prove that adj(cA) = c^{n-1} adj(A).

54. Prove that if A is an invertible $n \times n$ matrix, then det(adj(A)) = $(\det(A))^{n-1}$.

55. Suppose that A is an invertible $n \times n$ matrix. Show that $(\text{adj}(A))^{-1} = \text{adj}(A^{-1})$.

56. Suppose that A is an invertible $n \times n$ matrix and that both A and A^{-1} have integer entries. Show that $\det(A) = \pm 1$.

57. Prove that if A is a diagonal matrix, then so is adj(A).

58. In this problem, we show that

area($\mathcal{T}$)=12|det(A)|

where area($\mathcal{T}$) is the area of the triangle $\mathcal{T}$ with vertices (x_1, y_1), (x_2, y_2), and (x_3, y_3) (Figure 10) and

A=[x1y11x2y21x3y31]

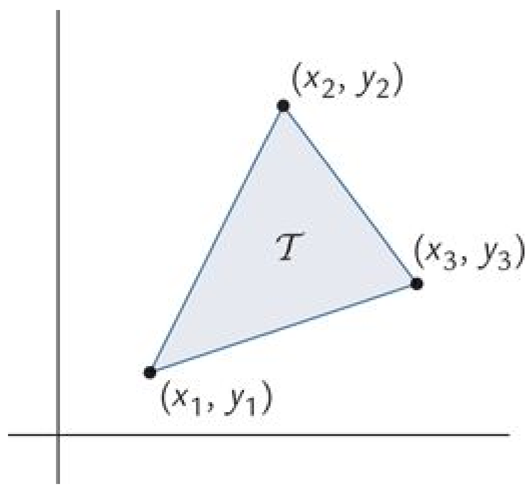


Figure 10 The triangle $\mathcal{T}$.

(a) Explain why area($\mathcal{T}$)=area($\mathcal{T}^*$), where $\mathcal{T}^*$ is the triangle with vertices $(0, 0)$, $(x_2 - x_1, y_2 - y_1)$, and $(x_3 - x_1, y_3 - y_1)$ (Figure 11).

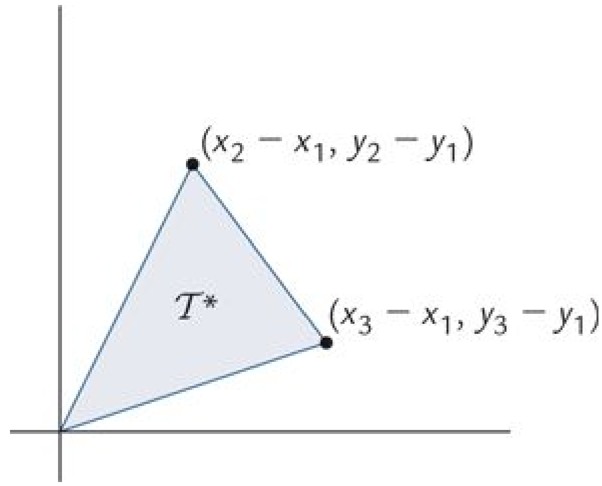


Figure 11 The triangle $\mathcal{T}^*$.

(b) Show that ($\mathcal{T}^*$)=12|det(B)|, where

B=[(x2-x1)(y2-y1)(x3-x1)(y3-y1)]

HINT: Apply Theorem 5.19 to compute the area of the parallelogram in Figure 12, and use the fact that det(B) = det(B^T).

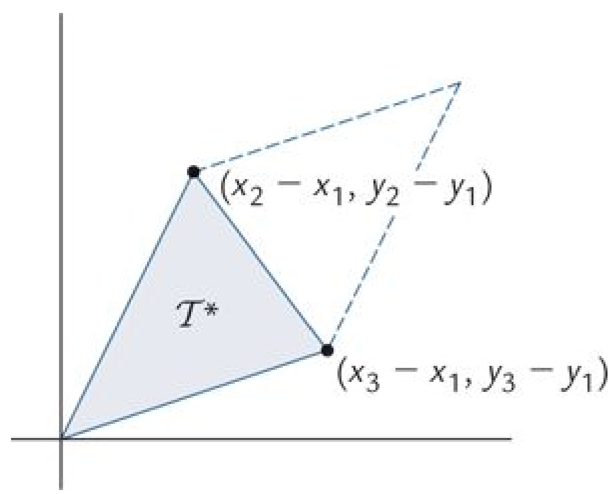


Figure 12 $\mathcal{T}^*$ and parallelogram.

(c) Show that det(B) = det(C), where

C=[x1y11(x2-x1)(y2-y1)0(x3-x1)(y3-y1)0]

(d) Use row operations to show that det(C) = det(A), and from this conclude that area($\mathcal{T}$)=12|det(A)|.

C Exercises 59–62: Use Cramer's Rule to find the solution to the system.

59. -x1+7x2+5x3=136x1-2x2+x3=93x1+11x2-9x3=4

60. 8x1-5x2 =6-2x1-4x2+8x3=-135x1+7x2-11x3=17

61. 3x1+5x2-x3-4x4=-5-2x1-4x2-3x3+7x4=0x1+2x2+4x3-2x4=24x1+x2-5x3-x4=4

62. -5x1+3x2+2x3+x4=2x1-7x2-5x3+7x4=-34x1+x2+x3+2x4=0-4x1-11x2 -5x4=-9

Exercises 63–66: For the given matrix A, find adj(A) and then use it to compute A^{-1}.

63. A=[4-25830-179]

64. A=[037-362511-1]

65. A=[425-1-230657211301-5]

66. A=[8-211-5353044-43192]

SUPPLEMENTARY EXERCISES

Exercises 1–4: Find M_{21} and M_{13} for A.

1. A=[36-814-2430]

2. A=[6624-2-3415]

3. A=[15-2-2-124374-726118]

4. A=[00040023025-33617]

Exercises 5–8: Find det(A) using cofactor expansion.

5. A=[3-814]

6. A=[103-2126-1-2]

7. A=[132-1056-22]

8. A=[403-1-105312613412]

Exercises 9–12: Find all values of such that det(A) = 0.

9. A=|2a14|

10. A=|-5a2aa|

11. A=[1-1a-1041-21]

12. A=[a-1300-21a2]

Exercises 13–14: Find all values of λ such that $\det(A - \lambda I_2) = 0$.

13. A=[-3124]

14. A=[12-45]

Exercises 15–16: Find all values of λ such that $\det(A - \lambda I_3) = 0$.

15. A=[11-1201-110]

16. A=[-2212-12-314]

Exercises 17–20: Convert *A* to echelon form, then find det(*A*) by tracking the row operations and computing the determinant of the echelon form.

17. A=[1523]

18. A=[3-2-12]

19. A=[231-12-2213]

20. A=[41032-113-2]

Exercises 21–24: Suppose that *A* is converted to echelon form *B* using the row operations given. Find det(*A*).

21. R1↔R2-3R1+R2→R2⇒B=[2-305]

22. -5R1↔R1-2R1+R2→R2⇒B=[140-2]

23. R3↔R2-R1+R3→R34R2+R3→R3⇒B=[3-670-15005]

24. (1/4)R2→R2-4R1+R3→R3R2↔R1-R2+R3→R3⇒B=[-10-4025007]

Exercises 25–28: Determine if Cramer’s Rule can be applied to find the solution for the given linear system, and if so, then find the solution.

25. 5x1-2x2=43x1+x2=-3

26. 6x1-4x2=6-3x1+2x2=-5

27. 3x1+x2+2x3=1-2x3=-2x2+x3=5

28. x1+2x2-4x3=13x1-x2+2x3=-22x1-4x2+3x3=0

Exercises 29–32: Find adj(*A*) and from it A^{-1}.

29. A=[1225]

30. A=[2-147]

31. A=[1-12-24-1220]

32. A=[4-431-2-125-4]

Exercises 33–38: Determine if the statement is true or false, and justify your answer.

33.

(a) If A is 3×4, then it must be that $\det(A) = 0$.

(b) If A is 5×5, then $\det(-2A) = -2\det(A)$.

34.

(a) It is not possible to compute $\operatorname{adj}(A)$ if A is not invertible.

(b) If $T(\mathbf{x}) = A\mathbf{x}$ and is not one-to-one, then $\det(A) = 0$.

35.

(a) If $\det(A) = 3$ and $\det(B) = 2$, then $\det(A + B) = 5$.

(b) If $\det(A) = 7$, then A^{-1} exists and $\det(A^{-1}) = 7$.

36.

(a) If $\det(A) \neq 0$, then $\operatorname{nullity}(A) = 0$.

(b) If $\det(A) = 0$, then $\operatorname{rank}(A) = 0$.

37.

(a) There are linear systems where Cramer's Rule can be used to find only some, but not all, of the variable values.

(b) For a square matrix A with $\det(A) \neq 0$, the cofactor matrix is equal to $\operatorname{adj}(A)$.

38.

(a) $\det(A) = \det(A^T)$.

(b) If A is a 5×5 matrix with $\operatorname{rank}(A) = 3$, then $\det(A) = 2$.

CHAPTER 6

Eigenvalues and Eigenvectors

Chris Sattlberger/Blend Images/Getty Images

Unlike the solar panels we saw earlier, the solar power plant pictured here uses concentrated solar power, with a power tower located in the center and surrounded by mirrors called heliostats. The heliostats reflect solar energy in a concentrated beam to a receiver at the top of the tower where the sun's heat is collected and converted into electrical power through the use of a heat engine and power generator.

The focus of this chapter is *eigenvalues* and *eigenvectors*, which are characteristics of square matrices and linear transformations. Eigenvalues and eigenvectors arise in a wide range of fields, including finance, quantum mechanics, image processing, and mechanical engineering.

In Section 6.1 we define eigenvalues and eigenvectors and develop an algebraic method for finding them. In Section 6.2 we show how to use eigenvalues and eigenvectors to diagonalize a matrix. Section 6.3 covers eigenvalues and eigenvectors involving complex numbers, and Section 6.4 focuses on solving systems of differential equations by using eigenvalues and eigenvectors. Algebraic methods often are not practical for large matrices, so in Section 6.5 we describe numerical methods for finding eigenvalues and eigenvectors.

6.1 Eigenvalues and Eigenvectors

Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be a linear transformation, with $T(\mathbf{x}) = A\mathbf{x}$ for a 2×2 matrix A. For a given vector $\mathbf{u}$ in $\mathbf{R}^2$, we can think of the multiplication $A\mathbf{u}$ as changing the direction and length of $\mathbf{u}$. Figure 1 shows $\mathbf{u}$ and $A\mathbf{u}$ for several vectors $\mathbf{u}$ and A=[1120].

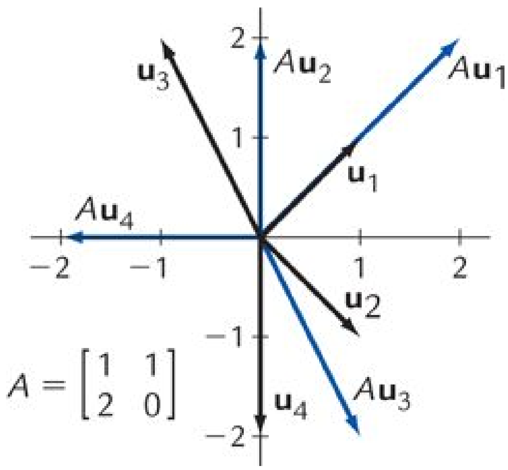


Figure 1 Plots of $\mathbf{u}$ and $A\mathbf{u}$ for the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, and $\mathbf{u}_4$.

▶ Note that $\mathbf{u}_1$ and $A\mathbf{u}_1$ are parallel, which means that they point in the same (or opposite) direction. This is also true of $\mathbf{u}_3$ and $A\mathbf{u}_3$.

Of particular importance in applications and in analyzing the behavior of a linear transformation are those vectors $\mathbf{u}$ such that $\mathbf{u}$ and $A\mathbf{u}$ are parallel. In Figure 1, $\mathbf{u}_1$ and $\mathbf{u}_3$ are such vectors. Algebraically, $\mathbf{u}$ and $A\mathbf{u}$ are parallel if there exists a scalar λ such that $A\mathbf{u} = \lambda\mathbf{u}$. Nonzero vectors that satisfy this equation are called *eigenvectors*.

DEFINITION 6.1 ▶

Eigenvector, Eigenvalue

Let A be an $n \times n$ matrix. Then a nonzero vector $\mathbf{u}$ is an **eigenvector** of A if there exists a scalar λ such that

$$Au=\lambda u \qquad (1)$$

The scalar λ is called an **eigenvalue** of A.

- An eigenvalue λ can be equal to zero, but an eigenvector **u** must be a nonzero vector.

When λ and **u** are related as in equation (1), we say that λ is the eigenvalue associated with **u** and that **u** is an eigenvector associated with λ.

Example 1

Let A=[3542]. Determine if each of

u1=[54],u2=[4−1],andu3=[−11]

is an eigenvector of A. For those that are, find the associated eigenvalue.

Solution Starting with $\mathbf{u}_1$, we have

Au1=[3542] [54]=[3528]=7[54]=7u1

Thus $A\mathbf{u}_1 = 7\mathbf{u}_1$, so $\mathbf{u}_1$ is an eigenvector of A with associated eigenvalue $\lambda = 7$. Calculating $A\mathbf{u}_2$, we have

Au2=[3542] [4−1]=[714]

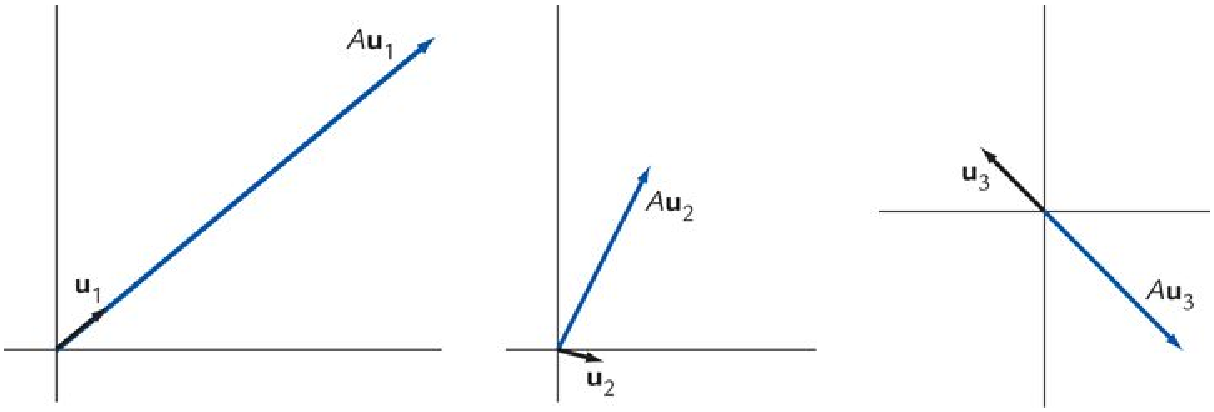


Figure 2 Graphs of vectors from Example 1. If **u** and A**u** are parallel vectors, then **u** is an eigenvector.

Since $A\mathbf{u}_2$ is not a multiple of $\mathbf{u}_2$, this tells us that $\mathbf{u}_2$ is not an eigenvector of A. Finally,

Au3=[3542] [−11]=[2−2]=−2[−11]=−2u3

Since $A\mathbf{u}_3 = -2\mathbf{u}_3$, it follows that $\mathbf{u}_3$ is an eigenvector of A with associated eigenvalue $\lambda = -2$. See Figure 2 for graphs of vectors.

Referring back to $\mathbf{u}_1$ in Example 1, suppose u4=3u1=3[54]=[1512]. Then we have

Au4=[3542] [1512]=[10584]=7[1512]=7u4

Therefore $\mathbf{u}_4 = 3\mathbf{u}_1$ is also an eigenvector of A associated with $\lambda = 7$. We could have used any nonzero scalar in place of 3 and achieved the same result, so any nonzero multiple of $\mathbf{u}_1$ is also an eigenvector of A associated with $\lambda = 7$. Theorem 6.2 generalizes this observation.

THEOREM 6.2 ▸

Let A be a square matrix, and suppose that **u** is an eigenvector of A associated with eigenvalue λ. Then for any scalar $c \neq 0$, $c\mathbf{u}$ is also an eigenvector of A associated with λ.

▶ We require $c \neq 0$ because eigenvectors must be nonzero.

Proof Let **u** be an eigenvector of A with associated eigenvalue λ. Then for any scalar $c \neq 0$, we have

$$A(c\mathbf{u}) = c(A\mathbf{u}) = c(\lambda\mathbf{u}) = \lambda(c\mathbf{u})$$

so that $c\mathbf{u}$ is also an eigenvector of A associated with eigenvalue λ. ■ ■

▶ We address the problem of finding the eigenvalues shortly.

Finding Eigenvectors

Here we consider the problem of finding the eigenvectors associated with a known eigenvalue. Let's start with an example.

Example 2

Take as known that $\lambda = 3$ and $\lambda = 2$ are eigenvalues for

A=[44−214−136−1]

Find the eigenvectors associated with each eigenvalue.

Solution Starting with $\lambda = 3$, we need to find all nonzero vectors **u** such that $A\mathbf{u} = 3\mathbf{u}$. Since $3\mathbf{u} = 3I_3\mathbf{u}$, our equation becomes

Au=3I3u⇒Au−3I3u=0⇒(A−3I3)u=0

Thus we need to find the solutions to the homogeneous system with coefficient matrix

A−3I3=[44−214−136−1]−[300030003]=[14−211−136−4]

Forming an augmented matrix and performing the indicated row operations, we have

$$[14-211-136-4|000]-R1+R2\rightarrow R2-3R1+R3\rightarrow R3-2R2+R3\rightarrow R3\sim[14-20-31000|000]$$

After back substitution, we find that the system $(A - 3I_3)\mathbf{u} = \mathbf{0}$ has general solution

$$u=s[213]$$

We can verify that **u** is an eigenvector associated with $\lambda = 3$ by computing

$$Au=[44-214-136-1](s[213])=s[44-214-136-1]\ [213]=s[639]=3u$$

The procedure is the same for $\lambda = 2$. This time we need to find the general solution to the homogeneous system $(A - 2I_3)\mathbf{u} = \mathbf{0}$, where

$$A-2I3=[44-214-136-1]-[200020002]=[24-212-136-3]$$

The augmented matrix and corresponding echelon form are

$$[24-212-136-3|000]R1\leftrightarrow R2-2R1+R2\rightarrow R2-3R1+R3\rightarrow R3\sim[12-1000000|000]$$

After back substitution, we find that this system has general solution

$$u=s1[-210]+s2[101] \tag{2}$$

As long as s_1 or s_2 (or both) is nonzero, then **u** will be an eigenvector associated with $\lambda = 2$. We can check by computing

$$Au=A(s1[-210]+s2[101])=s1A[-210]+s2A[101]\ =s1[-420]+s2[202]=2(s1[-210]+s2[101])=2u$$

Recall that a set of vectors is a subspace if it contains **0**, is closed under addition, and is closed under scalar multiplication.

If the zero vector is included, the eigenvectors in Example 2 associated with each eigenvalue form a subspace. This is always true of the set of eigenvectors associated with a given eigenvalue.

THEOREM 6.3 ▶

Let A be an $n \times n$ matrix with eigenvalue λ. Let S denote the set of all eigenvectors associated with λ, together with the zero vector $\mathbf{0}$. Then S is a subspace of $\mathbf{R}^n$.

▶ We can also think of S as the null space of the matrix $A - \lambda I_n$.

Proof We show S is a subspace by verifying the three required conditions of Definition 4.1. First, by definition $\mathbf{0}$ is in S. Second, Theorem 6.2 tells us that if $\mathbf{u}$ is an eigenvector associated with λ, then so is $c\mathbf{u}$ for $c \neq 0$. If $c = 0$, then $c\mathbf{u} = \mathbf{0}$ is in S, so S is closed under scalar multiplication. Third, if $\mathbf{u}_1$ and $\mathbf{u}_2$ are both eigenvectors associated with λ, then

$$A(\mathbf{u}_1+\mathbf{u}_2)=A\mathbf{u}_1+A\mathbf{u}_2=\lambda\mathbf{u}_1+\lambda\mathbf{u}_2=\lambda(\mathbf{u}_1+\mathbf{u}_2)$$

so that $\mathbf{u}_1 + \mathbf{u}_2$ is also an eigenvector associated with λ. Therefore S is closed under addition, and so S is a subspace. ■■

DEFINITION 6.4 ▶

Eigenspace

Let A be a square matrix with eigenvalue λ. The subspace of all eigenvectors associated with λ, together with the zero vector, is called the **eigenspace** of λ.

▶ Each distinct eigenvalue has its own associated eigenspace.

For instance, in Example 2 we see from (2) that the set

$$\{[-210],[101]\}$$

forms a basis for the eigenspace of $\lambda = 2$.

Finding Eigenvalues

Let's review what we have learned so far. If we know an eigenvalue λ for a given $n \times n$ matrix A, then we can find the associated eigenvectors by solving the linear system $A\mathbf{u} = \lambda\mathbf{u}$, or equivalently, the homogeneous system

$$(A-\lambda In)u=0 \tag{3}$$

If we know the eigenvalue, then this is a problem that we know how to solve. Finding the eigenvalues is a different problem that we have not yet considered. The next theorem shows how to use determinants to find eigenvalues.

THEOREM 6.5 ▶

Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof λ is an eigenvalue of A if and only if there exists a nontrivial solution to $A\mathbf{u} = \lambda\mathbf{u}$. This is equivalent to the existence of a nontrivial solution to (3), which by the Unifying Theorem, Version 7 is true if and only if $\det(A - \lambda I_n) = 0$. ■ ■

Example 3

Find the eigenvalues for $A=[336-4]$.

Solution Our strategy is to determine the values of λ that satisfy $\det(A - \lambda I_2) = 0$. We have

$$A-\lambda I2=[336-4]-[\lambda 00\lambda]=[(3-\lambda)36(4-\lambda)]$$

Next, we compute the determinant,

$$\det(A-\lambda I_2)=(3-\lambda)(-4-\lambda)-18=\lambda^2+\lambda-30$$

Setting $\det(A - \lambda I) = 0$, we have

$$\lambda^2+\lambda-30=0 \Rightarrow (\lambda-5)(\lambda+6)=0 \Rightarrow \lambda=5 \text{ or } \lambda=-6$$

Thus the eigenvalues for A are $\lambda = 5$ and $\lambda = -6$.

Characteristic Polynomial, Characteristic Equation

The polynomial that we get from $\det(A - \lambda I)$ is called the **characteristic polynomial** of A, and the equation $\det(A - \lambda I) = 0$ is called the **characteristic equation**. The eigenvalues for a matrix A are given by the roots of the characteristic equation. In this section we focus on real roots; complex roots are covered in Section 6.3.

Example 4

Find the eigenvalues and a basis for each eigenspace for the matrix

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

Solution We start by finding the eigenvalues of A by computing

$$\det(A-\lambda I_2)=\begin{vmatrix} (2-\lambda) & -1 \\ -1 & (2-\lambda) \end{vmatrix} =(2-\lambda)^2-1 =\lambda^2-4\lambda+3=(\lambda-1)(\lambda-3)$$

Therefore the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 3$. Next, we find the eigenvectors, starting with those associated with $\lambda_1 = 1$. We find the associated eigenvectors by solving the homogeneous linear system $(A - 1 \cdot I_2)\mathbf{u} = (A - I_2)\mathbf{u} = \mathbf{0}$. Since

$$A-I_2=\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

the augmented matrix and corresponding echelon form are

$$\left[\begin{array}{cc|c} 1 & -1 & 0 \\ -1 & 1 & 0 \end{array}\right] \xrightarrow{R_1+R_2\to R_2} \sim \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Back substitution gives us

General Solution: u1=s[11]⇒Basis for Eigenspace of λ1=1:{[11]}

The eigenvectors associated with $\lambda_2 = 3$ are found by solving the homogeneous linear system $(A - 3I_2)\mathbf{u} = \mathbf{0}$. We have

A−3I2=[−1−1−1−1]

so the augmented matrix and corresponding echelon form are

[−1−1−11|00]−R1+R2→R2~[−1−100|00]

Back substitution gives us

General Solution: u2=s[−11]⇒Basis for Eigenspace of λ2=3:{[−11]}

Example 5

Find the eigenvalues and a basis for each eigenspace of

A=[1−332−22200]

Solution We determine the eigenvalues of A by calculating the characteristic polynomial,

det(A−λI3)=|(1−λ)−332(−2−λ)220(0−λ)|=−λ3−λ2+2λ=−λ(λ+2)(λ−1)

Thus the eigenvalues are $\lambda_1 = -2$, $\lambda_2 = 0$, and $\lambda_3 = 1$. Taking them in order, the eigenvectors associated with $\lambda_1 = -2$ are found by solving the homogeneous system $(A + 2I_3)\mathbf{u} = \mathbf{0}$. The augmented matrix and echelon form are

[3−33202202|000]R1↔R2−32R1+R2→R2−R1+R3→R3~[2020−3000
0|000]

Back substitution gives us

General Solution: u1=s[10−1]⇒Basis for Eigenspace of λ1=−2: {[10−1]}

For $\lambda_2 = 0$, the homogeneous system is $(A - 0 \cdot I_2)\mathbf{u} = A\mathbf{u} = \mathbf{0}$. The augmented matrix and echelon form are

[1−332−22202|000]−2R1+R1→R2−R1+R3→R3−32R2+R3→R3~[1−3304−4000|000]

This time back substitution yields

General Solution: u2=s[011]⇒Basis for Eigenspace of λ2=0:{[011]}

The final eigenvalue is $\lambda_3 = 1$. The homogeneous system is $(A - I_3)\mathbf{u} = \mathbf{0}$, and the augmented matrix and echelon form are

[0−332−3220−1|000]R1↔R3−R1+R2→R2−R2+R3→R3~[20−10−33000|000]

With back substitution, we find

General Solution: u3=s[122]⇒Basis for Eigenspace of λ3=1:{[122]}

Example 6

Find the eigenvalues and a basis for each eigenspace of

A=[1−21−101−1−23]

Solution We start out by finding the eigenvalues for A by computing

det(A−λI3)=|(1−λ)−21−1−λ1−1−2(3−λ)|=−λ3+4λ2−4λ=−λ(λ−2)2

From the factored form we see that our matrix has two distinct eigenvalues, $\lambda_1 = 0$ and $\lambda_2 = 2$.

Now we find the eigenvectors. Starting with those associated with $\lambda_1 = 0$, we solve the homogeneous system $(A - 0 \cdot I_3)\mathbf{u} = A\mathbf{u} = \mathbf{0}$. The augmented matrix and echelon form are

[1−21−101−1−23|000]R+R2→R3R1+R3→R3−2R2+R3→R3~[1−210−22000|000]

Back substitution produces

General Solution: u1=s[111]⇒Basis for Eigenspace of λ1=0:{[111]}

For the eigenvalue $\lambda_2 = 2$ we form the augmented matrix for the system $(A - 2I_3)\mathbf{u} = \mathbf{0}$ and reduce to echelon form,

[−1−21−1−21−1−21|000]−R1+R2→R2−R1+R3→R3~[1−21000000|000]

This time back substitution produces

General Solution: u2=s1[101]+s2[−210]⇒Basis for Eigenspace of λ2=2:{[101],[−210]}

Multiplicities

In Example 6 the factored form of the characteristic polynomial is

−λ(λ−2)2=−(λ−0)1(λ−2)2

The *multiplicity* of an eigenvalue is equal to its factor's exponent. In this case, we say that $\lambda = 0$ has multiplicity 1 and $\lambda = 2$ has multiplicity 2.

Multiplicity

In general, for a polynomial $P(x)$, a root α of $P(x) = 0$ has **multiplicity** if $P(x) = (x - \alpha)^r Q(x)$ with $Q(\alpha) \neq 0$. When discussing eigenvalues, the phrase "λ has multiplicity r" means that λ is a root of the characteristic polynomial with multiplicity r.

Reviewing our previous examples, we see that the dimension of the eigenspaces matched the multiplicities of the associated eigenvalues. This happens most of the time, but not always.

▶ Put another way, the multiplicity is the number of times a root is repeated.

Example 7

Find the eigenvalues and a basis for each eigenspace of

$$A=[02-11-101-20]$$

Solution The characteristic polynomial of A is

$$\det(A-\lambda I3)=|-\lambda 2-11(-1-\lambda)01-2-\lambda|=-\lambda 3-\lambda 2+\lambda+1=-(\lambda-1)(\lambda+1)2$$

Thus A has two distinct eigenvalues, $\lambda_1 = -1$ (multiplicity 2) and $\lambda_2 = 1$ (multiplicity 1).

To find the eigenvectors associated with $\lambda_1 = -1$, we solve the homogeneous system $(A + I_3)\mathbf{u} = \mathbf{0}$. The augmented matrix and echelon form are

$$[12-11001-21|000]-R1+R2\rightarrow R2-R1+R3\rightarrow R3-2R2+R3\rightarrow R3\sim[12-1 0-22000|000]$$

Back substitution produces

$$\text{General Solution: } u1=s[012]\Rightarrow\text{Basis for Eigenspace of } \lambda 1=-1:\{[012]\}$$

Note that although the eigenvalue $\lambda_1 = -1$ has multiplicity 2, the eigenspace has dimension 1.

For $\lambda_2 = 1$, the augmented matrix for the system $(A - I_3)\mathbf{u} = \mathbf{0}$ and echelon form are

$$[-12-11-201-2-1|000]R1+R2\rightarrow R2R1+R3\rightarrow R3-2R2+R3\rightarrow R3\sim[-12 -100-1000|000]$$

Back substitution gives us

General Solution: $u_2 = s\begin{bmatrix}2\\1\\0\end{bmatrix} \Rightarrow$ Basis for Eigenspace of $\lambda_2 = 1$: $\left\{\begin{bmatrix}2\\1\\0\end{bmatrix}\right\}$

As demonstrated in Example 7, it is possible for the dimension of an eigenspace to be less than the multiplicity of the associated eigenvalue. However, the opposite cannot happen.

THEOREM 6.6 ▸

Let A be a square matrix with eigenvalue λ. Then the dimension of the associated eigenspace is less than or equal to the multiplicity of λ.

The proof is beyond the scope of this book and is omitted.

The methods developed in this section work well for small matrices, but they can be impractical for large complicated matrices. For instance, suppose

$$A = \begin{bmatrix} 245 & -254 & -252 & -46 & -224 \\ 161 & -168 & -174 & -32 & -148 \\ -39 & 40 & 45 & 7 & 38 \\ 27 & -28 & -32 & -6 & -26 \\ 110 & -113 & -110 & -21 & -101 \end{bmatrix} \tag{4}$$

Computing the determinant by hand to find the characteristic polynomial for A is not easy. Using computer software, we find that

$$\det(A - \lambda I_5) = -\lambda^5 + 15\lambda^4 - 3\lambda^3 - 287\lambda^2 - 192\lambda + 468$$

This polynomial is challenging to factor. In fact, there is no general algorithm for factoring polynomials of degree 5 or more. (Not even on a computer.) Thus we cannot find the eigenvalues and eigenvectors using our existing methods.

Numerous applications require eigenvalues and eigenvectors from large matrices. In principle, we could use numerical methods to find approximations to the eigenvalues, and use these to find the eigenvectors. However, for various reasons this does not work well in practice. Instead, there are algorithms that lead directly to approximations to the eigenvectors, bypassing the need to first find the eigenvalues. These are described in Section 6.5, where we analyze the matrix A in (4).

The Unifying Theorem, Version 8

Although $\mathbf{u} = \mathbf{0}$ is not allowed as an eigenvector, it is fine to have $\lambda = 0$ as an eigenvalue. From Theorem 6.5, we know that $\lambda = 0$ is an eigenvalue of an $n \times n$ matrix A if and only if $\det(A - 0I_n) = \det(A) = 0$. Put another way, $\lambda = 0$ is *not* an eigenvalue of A if and only if $\det(A) \neq 0$. This observation provides us with another condition for the Unifying Theorem.

THEOREM 6.7 ▶

(THE UNIFYING THEOREM, VERSION 8) Let $\mathcal{S}=$ {a1,… ,an} be a set of n vectors in $\mathbf{R}^n$, let $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$, and let $T : \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent:

(a) $\mathcal{S}$ spans $\mathbf{R}^n$.

(b) $\mathcal{S}$ is linearly independent.

(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$ in $\mathbf{R}^n$.

(d) T is onto.

(e) T is one-to-one.

(f) A is invertible.

(g) $\ker(T) = \{\mathbf{0}\}$.

(h) $\mathcal{S}$ is a basis for $\mathbf{R}^n$.

(i) $\text{col}(A) = \mathbf{R}^n$.

(j) $\text{row}(A) = \mathbf{R}^n$.

(k) $\text{rank}(A) = n$.

(l) $\det(A) \neq 0$.

(m) $\lambda = 0$ is not an eigenvalue of A.

Proof From the Unifying Theorem, Version 7, we know that (a) through (l) are equivalent. The comments above show that (l) and (m) are equivalent, and so it follows that all 13 conditions are equivalent.

Example 8

Show that $\lambda = 0$ is an eigenvalue for the matrix

$$A=[3−15210412]$$

Solution From the Unifying Theorem, Version 8, $\lambda = 0$ is an eigenvalue of A if and only if $\det(A) = 0$. By the Shortcut Method (Section 5.1), we have

$$\det(A)=(6+0+10)−(20−4+0)=0$$

so $\lambda = 0$ is an eigenvalue of A.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine which of $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are eigenvectors for A. For those that are, determine the associated eigenvalue.

$$A=[−2−42−212425],x1=[−201],x2=[−231],x3=[121]$$

2. Find a basis for the eigenspace of A with associated eigenvalue λ.

$$A=[2−4−1−1],\lambda=3$$

3. Find the eigenvalues and eigenvectors for $A=[3122]$.
4. Find the eigenvalues and eigenvectors for $A=[100200−211]$.

5. Find a 2 × 2 matrix A that has eigenvalues $\lambda_1 = -2$ and $\lambda_2 = 5$.

6. Determine if each statement is true or false.
   (a) Every square matrix A has **0** for an eigenvector.
   (b) If $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors associated with an eigenvalue λ, then so is $\mathbf{u}_1 + \mathbf{u}_2$.
   (c) If $n \times n$ matrix A has eigenvalue $\lambda = 0$, then the column space of A is equal to $\mathbf{R}^n$.
   (d) If $A \sim B$ then A and B have the same eigenvalues.

EXERCISES

Exercises 1–6: Determine which of $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are eigenvectors for the matrix A. For those that are, determine the associated eigenvalue.

1. A=[1322],x1=[−32],x2=[1−1],x3=[−2−2]
2. A=[−1203],x1=[02],x2=[13],x3=[12]
3. A=[2720−100−21],x1=[−311],x2=[−201],x3=[100]
4. A=[3−10−130−112],x1=[111],x2=[110],x3=[12−1]
5. A=[6−3100310−6600−33−23], x1=[1100],x2=[12−10], x3=[11−32]
6. A=[55188218−66−90−7−1−2−10], x1=[110−2],x2=[1100], x3=[12−2−1]

Exercises 7–10: Use the characteristic polynomial to determine if λ is an eigenvalue for the matrix A.

7. A=[27−16],λ=3
8. A=[1542],λ=6
9. A=[02020022−2],λ=−2
10. A=[63−1−4−11186−4],λ=1

Exercises 11–20: Find a basis for the eigenspace of A associated with the eigenvalue λ.

11. A=[1−315],λ=4
12. A=[−243−1],λ=−5

13. $A=[6-102-3],\lambda=2$
14. $A=[-11-12-89],\lambda=-3$
15. $A=[6-374154-39],\lambda=4$
16. $A=[-25-7-211-13-25-7],\lambda=-4$
17. $A=[5-122222-15],\lambda=6$
18. $A=[-12-7-1022-1022],\lambda=9$
19. $A=[11-3-3813-5-582-2-20-3330],\lambda=-4$
20. $A=[15-3-15821-9-1782-2-60-33154],\lambda=-8$

Exercises 21–30: Find the characteristic polynomial, the eigenvalues, and a basis for each eigenspace for the matrix *A*.

21. $A=[204-3]$
22. $A=[2611]$
23. $A=[1-22-3]$
24. $A=[-281-4]$
25. $A=[300120-45-1]$
26. $A=[101100000]$
27. $A=[2510-3-12144]$
28. $A=[0-3-1-1213-9-4]$
29. $A=[-10005-20003102011]$
30. $A=[0010010010000001]$

FIND AN EXAMPLE Exercises 31–36: Find an example that meets the given specifications.

31. A 2×2 matrix *A* with eigenvalues $\lambda = 1$ and $\lambda = 2$.
32. A 2×2 matrix *A* with eigenvalues $\lambda = -3$ and $\lambda = 0$.
33. A 3×3 matrix *A* with eigenvalues $\lambda = 1$, $\lambda = -2$, and $\lambda = 3$.
34. A 3×3 matrix *A* with eigenvalues $\lambda = -1$ (multiplicity 2) and $\lambda = 4$.
35. A 2×2 matrix that has no real eigenvalues.
36. A 4×4 matrix that has no real eigenvalues.

TRUE OR FALSE Exercises 37–42: Determine if the statement is true or false, and justify your answer.

37.

(a) An eigenvalue λ must be nonzero, but an eigenvector **u** can be equal to the zero vector.

(b) The dimension of an eigenspace is always less than or equal to the multiplicity of the associated eigenvalue.

38.

(a) If **u** is a eigenvector of A, then **u** is a multiple of $A\mathbf{u}$.

(b) If λ_1 and λ_2 are eigenvalues of a matrix, then so is $\lambda_1 + \lambda_2$.

39.

(a) If A is a diagonal matrix, then the eigenvalues of A lie along the diagonal.

(b) If 0 is an eigenvalue of an $n \times n$ matrix A, then the columns of A span $\mathbf{R}^n$.

40.

(a) If 0 is an eigenvalue of A, then nullity(A) > 0.

(b) Row operations do not change the eigenvalues of a matrix.

41.

(a) If 0 is the only eigenvalue of A, then A must be the zero matrix.

(b) If 0 is an eigenvalue of A, then the constant term of the characteristic polynomial of A is also 0.

42.

(a) If D is an $n \times n$ diagonal matrix, then each column of I_n is an eigenvector of D.

(b) If A and B are 3×3 matrices with the same eigenvalues (counting multiplicity), then $A = B$.

Exercises 43–46: The Cayley-Hamilton theorem states that a square matrix A satisfies its characteristic equation. For instance if A has characteristic polynomial $\lambda^2 - 3\lambda + 8$, then $A^2 - 3A + 8I = 0_{22}$. Confirm the Cayley-Hamilton theorem for A.

43. A=[1−45−3]

44. A=[5214]

45. A=[3−221−34−451]

46. A=[33−2−1201−5−4]

47. Suppose that A is a square matrix with characteristic polynomial $(\lambda - 3)^3(\lambda - 2)^2(\lambda + 1)$.

(a) What are the dimensions of A?

(b) What are the eigenvalues of A?

(c) Is A invertible?

(d) What is the largest possible dimension for an eigenspace of A?

48. Suppose that A is a square matrix with characteristic polynomial $-\lambda(\lambda - 1)^3(\lambda + 2)^3$.

(a) What are the dimensions of A?

(b) What are the eigenvalues of A?

(c) Is A invertible?

(d) What is the largest possible dimension for an eigenspace of A?

49. Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Prove that if 0 is not an eigenvalue of A, then T is onto.

50. Prove that if λ is an eigenvalue of A, then 4λ is an eigenvalue of $4A$.

51. Prove that if $\lambda = 1$ is an eigenvalue of an $n \times n$ matrix A, then $A - I_n$ is singular.

52. Prove that if $\mathbf{u}$ is an eigenvector of A, then $\mathbf{u}$ is also an eigenvector of A^2.

53. Prove that $\mathbf{u}$ cannot be an eigenvector associated with two distinct eigenvalues λ_1 and λ_2 of A.

54. Prove that if 5 is an eigenvalue of A, then 25 is an eigenvalue of A^2.

55. If $\mathbf{u} = \mathbf{0}$ was allowed to be an eigenvector, then which values of λ would be associated eigenvalues? (This is one reason why eigenvectors are defined to be nonzero.)

56. Suppose that A is a square matrix that is either upper or lower triangular. Show that the eigenvalues of A are the diagonal terms of A.

57. Let A be an invertible matrix. Prove that if λ is an eigenvalue of A with associated eigenvector $\mathbf{u}$, then λ^{-1} is an eigenvalue of A^{-1} with associated eigenvector $\mathbf{u}$.

58. Let A=[abcd]. Find a formula for the eigenvalues of A terms of a, b, c, and d. (HINT: The quadratic formula can be handy here.)

59. Suppose that A and B are both $n \times n$ matrices, and that $\mathbf{u}$ is an eigenvector for both A and B. Prove that $\mathbf{u}$ is an eigenvector for the product AB.

60. Suppose that A is an $n \times n$ matrix with eigenvalue λ and associated eigenvector **u**. Show that for each positive integer k, the matrix A^k has eigenvalue λ^k and associated eigenvector **u**.

61. Suppose that the entries of each row of a square matrix A add to zero. Prove that $\lambda = 0$ is an eigenvalue of A.

62. Suppose that A=[abcd], where a, b, c, and d satisfy $a + b = c + d$. Show that $\lambda_1 = a + b$ and $\lambda_2 = a - c$ are both eigenvalues of A.

63. Suppose that A is a square matrix. Prove that if λ is an eigenvalue of A, then λ is also an eigenvalue of A^T. (HINT: Recall that $\det(A) = \det(A^T)$.)

64. Suppose that A is an $n \times n$ matrix and c is a scalar. Prove that if λ is an eigenvalue of A with associated eigenvector **u**, then $\lambda - c$ is an eigenvalue of $A - cI_n$ with associated eigenvector **u**.

65. Suppose that the entries of each row of a square matrix A add to c for some scalar c. Prove that $\lambda = c$ is an eigenvalue of A.

66. Let A be an $n \times n$ matrix.

    (a) Prove that the characteristic polynomial of A has degree n.

    (b) What is the coefficient on λ^n in the characteristic polynomial?

    (c) Show that the constant term in the characteristic polynomial is equal to $\det(A)$.

    (d) Suppose that A has eigenvalues $\lambda_1, \ldots, \lambda_n$ that are all real numbers. Prove that $\det(A) = \lambda_1 \lambda_2 \ldots \lambda_n$.

In Exercises 67–70, find the eigenvalues and bases for the eigenspaces of A.

67. A=[00−2−111652041−20−21]

68. A=[−20−914184017−18−28179−10−11−17−91415]

69. A=[1001−3323−16−32−240−19−91401−55−100−13−3]

70. A=[5021−16143−3−60−3−332023−2−40−2−12]

6.2 Diagonalization

If D is a diagonal matrix, then it is relatively easy to analyze the behavior of the linear transformation $T(\mathbf{x}) = D\mathbf{x}$ because for $\mathbf{x}$ in $\mathbf{R}^n$,

$$D=\begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{bmatrix} \Rightarrow D\mathbf{x}=\begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}=\begin{bmatrix} d_{11}x_1 \\ d_{22}x_2 \\ \vdots \\ d_{nn}x_n \end{bmatrix}$$

In this section we develop a procedure for expressing a square matrix A as the product of three matrices. The process is called *diagonalizing* A, because the middle matrix in the product is diagonal.

DEFINITION 6.8 ▶

Diagonalizable Matrix

An $n \times n$ matrix A is **diagonalizable** if there exist $n \times n$ matrices D and P, with D diagonal and P invertible, such that

$$A=PDP^{-1}$$

Because D is a diagonal matrix, expressing $A = PDP^{-1}$ makes it easier to analyze the linear transformation $T(\mathbf{x}) = A\mathbf{x}$. Diagonalizing A also allows for more efficient computation of matrix powers A^2, A^3, … , which arise in modeling systems that evolve over time. Matrix powers are discussed at the end of the section.

Example 1

Let $A = \begin{bmatrix} -2 & 2 \\ -6 & 5 \end{bmatrix}$. Show that if

$$P = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

then $A = PDP^{-1}$ and hence A is diagonalizable.

Solution Applying the Quick Formula for the inverse of a 2 × 2 matrix (Section 3.3) gives us $P^{-1} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}$. Thus we have

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

Therefore A is diagonalizable.

Now let's develop a method for finding matrices P and D that diagonalize an $n \times n$ matrix A. Suppose that A has n linearly independent eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Define

$$P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Since $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n$ are eigenvectors of A, we have

$$A\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{for } i = 1, \ldots, n$$

Hence it follows that

$$AP = A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{u}_1 & A\mathbf{u}_2 & \cdots & A\mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix}$$

We also have

$$PD = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix}$$

Thus $AP = PD$. Since $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are linearly independent, P is an invertible matrix so that

$$A = A(PP^{-1}) = (AP)P^{-1} = (PD)P^{-1} = PDP^{-1}$$

and A is diagonalizable.

Example 2

Find matrices P and D to diagonalize $A=\begin{bmatrix}3 & 1\\ -2 & 0\end{bmatrix}$.

Solution We start by finding the eigenvalues and eigenvectors of A. Beginning with the eigenvalues, we have

$$\det(A-\lambda I_2)=\begin{vmatrix}3-\lambda & 1\\ -2 & -\lambda\end{vmatrix}=(3-\lambda)(-\lambda)-(-2)=\lambda^2-3\lambda+2=(\lambda-2)(\lambda-1)$$

Thus we have eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$. Starting with $\lambda_1 = 1$, the homogeneous system $(A - I_2)\mathbf{u} = \mathbf{0}$ has augmented matrix and echelon form

$$\left[\begin{array}{cc|c}2 & 1 & 0\\ -2 & -1 & 0\end{array}\right]\xrightarrow{R_1+R_2\to R_2}\sim\left[\begin{array}{cc|c}2 & 1 & 0\\ 0 & 0 & 0\end{array}\right]$$

Back substitution gives us the associated eigenvector

$$\mathbf{u}_1=\begin{bmatrix}1\\ -2\end{bmatrix}$$

For $\lambda_2 = 2$, the system $(A - 2I_2)\mathbf{u} = \mathbf{0}$ has augmented matrix and echelon form

$$\left[\begin{array}{cc|c}1 & 1 & 0\\ -2 & -2 & 0\end{array}\right]\xrightarrow{2R_1+R_2\to R_2}\sim\left[\begin{array}{cc|c}1 & 1 & 0\\ 0 & 0 & 0\end{array}\right]$$

This time back substitution gives us the eigenvector

$$\mathbf{u}_2=\begin{bmatrix}1\\ -1\end{bmatrix}$$

Now we define P and D, with

$$P=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}=\begin{bmatrix}1 & 1\\ -2 & -1\end{bmatrix}\quad\text{and}\quad D=\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$$

To check our work, we find P^{-1} using the Quick Formula, and then compute the product

PDP−1=[11−2−1] [1002] [−1−121]=[31−20]=A

We can use the diagonalization formula to construct a matrix A that has specified eigenvalues and eigenvectors.

Example 3

Find a 2 × 2 matrix A that has eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$ and corresponding eigenvectors u1=[53] and u2=[32].

Solution We start by defining the diagonalization matrices P and D and multiply to find A. Let

P=[u1u2]=[5332]andD=[λ100λ2]=[−1002]

We have P−1=[2−3−35] from the Quick Formula for 2 × 2 matrix inverses, so that

A=PDP−1=[5332] [−1002] [2−3−35]=[−2845−1829]

To verify that A has the required specifications, we compute

Au1=[−2845−1829] [53]=[−5−3]=(−1)[53]=λ1u1Au2=
[−2845−1829] [32]=[64]=(2)[32]=λ2u2

Hence $\{\lambda_1, \lambda_2\}$ are eigenvalues of A associated with eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2\}$.

Many square matrices are diagonalizable, but not all. The next theorem tells us exactly when a matrix is diagonalizable.

THEOREM 6.9 ▸

An $n \times n$ matrix A is diagonalizable if and only if A has eigenvectors that form a basis for $\mathbf{R}^n$.

Proof We have seen how to diagonalize A if A has eigenvectors that form a basis for $\mathbf{R}^n$, so half of the proof is done. For the other half, suppose that A is diagonalizable, with

$$A=PDP^{-1} \tag{1}$$

where $\mathbf{p}_1, \dots \mathbf{p}_n$ are the columns of P and $d_{11}, \dots, d_{nn}$ are the diagonal entries of D. Since P is invertible, the columns $\mathbf{p}_1, \dots \mathbf{p}_n$ of P are nonzero and linearly independent.
Multiplying by P on the right of both sides of (1), we have $AP = PD$. Since column i of AP is equal to $A\mathbf{p}_i$ and column i of PD is equal to $d_{ii}\mathbf{p}_i$, we have

$$A\mathbf{p}_i=d_{ii}\mathbf{p}_i$$

Therefore $\mathbf{p}_i$ is an eigenvector of A with associated eigenvalue d_{ii}. Since P is invertible, A has eigenvectors that form a basis for $\mathbf{R}^n$. ■ ■

The proof tells us both when A is diagonalizable and how A is diagonalized. The diagonal elements of D must be the eigenvalues and the columns of P must be associated eigenvectors. Of course, there are many possibilities for the associated eigenvectors, so the diagonalization is not unique. However, we now have a way to find a diagonalization of an $n \times n$ matrix A.

Summary: Diagonalizing an $n \times n$ matrix A

Find the eigenvalues and the associated linearly independent eigenvectors.

- If A has n linearly independent eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$, then A is diagonalizable, with $P=[\mathbf{u}_1 \cdots \mathbf{u}_n]$ and the diagonal entries of D given by the corresponding eigenvalues.
- If there are not n linearly independent eigenvectors, then A is not diagonalizable.

Note that the order of the eigenvalues in D does not matter, as long as it matches the order of the corresponding eigenvectors in P.

The next theorem tells us that eigenvectors associated with distinct eigenvalues must be linearly independent. This theorem comes in handy when trying to diagonalize a matrix.

THEOREM 6.10 ▸

If $\{\lambda_1, \ldots, \lambda_k\}$ are distinct eigenvalues of a matrix A, then any set of associated eigenvectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent.

Proof Suppose that the set of eigenvectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly dependent. Since all eigenvectors are nonzero, it follows from Theorem 2.15 and Exercise 64 of Section 2.3 that one of the eigenvectors can be written as a linear combination of a linearly independent subset of the remaining eigenvectors, with the coefficients nonzero and unique for the given subset. Thus, without loss of generality, let $c_2, \ldots c_j$ be nonzero scalars such that

$$\mathbf{u}_1 = c_2\mathbf{u}_1 + \cdots + c_j\mathbf{u}_j \qquad (2)$$

Then

$$\lambda_1\mathbf{u}_1 = A\mathbf{u}_1 = A(c_2\mathbf{u}_2 + \cdots + c_j\mathbf{u}_j) = c_2A\mathbf{u}_2 + \cdots + c_jA\mathbf{u}_j = c_2\lambda_2\mathbf{u}_2 + \cdots + c_j\lambda_j\mathbf{u}_j$$

If $\lambda_1 \neq 0$, then

$$u1=c2(\lambda2\lambda1)u2+\cdots+cj(\lambda j\lambda1)uj$$

which is a different linear combination equal to $\mathbf{u}_1$, contradicting the uniqueness of (2). If $\lambda_1 = 0$, then

$$c2\lambda2u2+\cdots+cj\lambda juj=0$$

Since $c_2, \ldots, c_j$ and $\lambda_2, \ldots, \lambda_j$ are nonzero, this contradicts the linear independence of $\{\mathbf{u}_2, \ldots, \mathbf{u}_j\}$. Hence either way we reach a contradiction, so it must be that the set of eigenvectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent. ■ ■

Example 4

If possible, diagonalize the matrix A=[111−2−2−100−1].

Solution We start by finding the eigenvalues by factoring the characteristic polynomial of A,

$$\det(A-\lambda I3)=-\lambda3-2\lambda2-\lambda=-\lambda(\lambda+1)2$$

Thus we have eigenvalues $\lambda_1 = 0$ and $\lambda_2 = -1$. Starting with $\lambda_1 = 0$, the augmented matrix for the system $(A - 0I_3)\mathbf{u} = \mathbf{0}$ and the corresponding echelon form are

$$[111-2-2-100-1|000]R1+R2\rightarrow R2R2+R3\rightarrow R3\sim[111001000|000]$$

Back substitution can be used to show that a basis for the eigenspace associated with $\lambda_1 = 0$ is {[1−10]}.

For $\lambda_2 = -1$, the augmented matrix and echelon form are

$$[211-2-1-1000|000]R1+R2\rightarrow R2\sim[211000000|000]$$

Back substituting shows that the eigenspace of $\lambda_2 = -1$ has dimension 2 and basis $\left\{\begin{bmatrix}1\\-2\\0\end{bmatrix}, \begin{bmatrix}1\\0\\-2\end{bmatrix}\right\}$. By Theorem 6.10 we know that eigenvectors associated with distinct eigenvalues are linearly independent. Hence the set

$$\left\{\begin{bmatrix}1\\-1\\0\end{bmatrix}, \begin{bmatrix}1\\-2\\0\end{bmatrix}, \begin{bmatrix}1\\0\\-2\end{bmatrix}\right\}$$

is linearly independent and thus forms a basis for $\mathbf{R}^3$. Since there are two linearly independent eigenvectors associated with $\lambda_2 = -1$, this eigenvalue appears twice in *D*. We have

$$D = \begin{bmatrix}0&0&0\\0&-1&0\\0&0&-1\end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix}1&1&1\\-1&-2&0\\0&0&-2\end{bmatrix}$$

Note that the eigenvalues along the diagonal of *D* are in columns corresponding to the columns of *P* containing the associated eigenvectors. We check that *D* and *P* are correct by computing

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

and

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

Since $AP = PD$ and P is invertible, we have $A = PDP^{-1}$.

The next theorem provides a set of conditions required for a matrix to be diagonalizable.

THEOREM 6.11

Suppose that an $n \times n$ matrix A has only real eigenvalues. Then A is diagonalizable if and only if the dimension of each eigenspace is equal to the multiplicity of the corresponding eigenvalue.

▶ See Section 6.1 for the definition of the multiplicity of an eigenvector.

Proof This theorem follows from things that we already know:

- An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. (This is from Theorem 6.9.)
- Each eigenspace has a dimension no greater than the multiplicity of its associated eigenvalue. (This is from Theorem 6.6 in Section 6.1.)
- If A is an $n \times n$ matrix, then the multiplicities of the eigenvalues sum to n. (This is because the degree of the characteristic polynomial is equal to n, and the multiplicities must add up to the degree.)
- Vectors from distinct eigenspaces are linearly independent. (This is from Theorem 6.10.)

Pulling these together, we see that A is diagonalizable when the dimension of each eigenspace is as large as possible. Otherwise there will not be enough linearly independent eigenvectors to form a basis. If the dimension of each eigenspace is as large as possible, then since vectors from distinct eigenspaces are linearly independent we know that there will be enough linearly independent eigenvectors to form a basis for $\mathbf{R}^n$. ■ ■

Example 5

If possible, diagonalize the matrix A=[365323−5−6−7].

Solution The characteristic polynomial for A is

det(A−λI3)=−(λ+2)2(λ−2)

giving us eigenvalues $\lambda_1 = -2$ and $\lambda_2 = 2$. For $\lambda_1 = -2$, the augmented matrix for the system $(A + 2I_3)\mathbf{u} = \mathbf{0}$ and the corresponding echelon form are

$$\left[\begin{array}{rrr|r} 5 & 6 & 5 & 0 \\ 3 & 4 & 3 & 0 \\ -5 & -6 & -5 & 0 \end{array}\right] \begin{array}{c} -\frac{3}{5}R_1+R_2 \to R_2 \\ R_1+R_3 \to R_3 \\ \sim \end{array} \left[\begin{array}{rrr|r} 5 & 6 & 5 & 0 \\ 0 & \frac{2}{5} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Back substitution shows that the eigenspace associated with $\lambda_1 = -2$ has basis $\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right\}$.

This eigenspace has dimension 1, which is less than the multiplicity of $\lambda_1 = -2$. By Theorem 6.11, we know immediately that A is not diagonalizable.

We usually will need to find the eigenvectors in order to determine if a matrix is diagonalizable, but there is a case where knowing only the eigenvalues is enough.

THEOREM 6.12

If A is an $n \times n$ matrix with n distinct real eigenvalues, then A is diagonalizable.

Proof Every eigenvalue has an eigenvector, ensuring that the associated eigenspace has dimension at least 1. On the other hand, if the eigenvalues are distinct, then each has multiplicity 1, so that each eigenspace must have dimension 1—there is no other option. Thus by Theorem 6.11 A is diagonalizable.

Example 6

If possible, diagonalize the matrix $A = \begin{bmatrix} 5 & 0 & 0 \\ -4 & 3 & 0 \\ 1 & -3 & -2 \end{bmatrix}$.

Solution Since A is lower triangular, the eigenvalues lie along the main diagonal (Exercise 56, Section 6.1), with $\lambda_1 = 5$, $\lambda_2 = 3$, and $\lambda_3 = -2$. These are distinct, so we know from Theorem 6.12 that A is diagonalizable. Bases for the associated eigenspaces are (computations are not shown)

$$\lambda_1 = 5 \Rightarrow \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}, \quad \lambda_2 = 3 \Rightarrow \left\{ \begin{bmatrix} 0 \\ -5 \\ 3 \end{bmatrix} \right\}, \quad \lambda_3 = -2 \Rightarrow \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Therefore $A = PDP^{-1}$, with

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

Matrix Powers

Suppose that A is diagonalizable, with $A = PDP^{-1}$. Then

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PD^2P^{-1} \\ A^3 &= A(A^2) = (PDP^{-1})(PD^2P^{-1}) = PD(P^{-1}P)D^2P^{-1} = PD^3P^{-1} \end{aligned}$$

and in general

▶ This formula for A^k can be formally proved using induction.

$$A^k = PD^kP^{-1}$$

Next, note that

$$D = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_{nn} \end{bmatrix} \Rightarrow D^k = \begin{bmatrix} d_{11}^k & 0 & \cdots & 0 \\ 0 & d_{22}^k & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_{nn}^k \end{bmatrix}$$

For example, if $D = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix}$, then

$$\begin{aligned} D^2 &= \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} (-3)^2 & 0 \\ 0 & 2^2 \end{bmatrix} \\ D^3 &= \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} (-3)^2 & 0 \\ 0 & 2^2 \end{bmatrix} = \begin{bmatrix} (-3)^3 & 0 \\ 0 & 2^3 \end{bmatrix} \end{aligned}$$

and so on. Hence we see that although directly calculating A^k can take many computations, calculating D^k is relatively easy.

▶ A is an example of a *probability matrix*, because the entries are nonnegative and each column adds to 1. Probability matrices are discussed in Section 3.5.

Example 7

Suppose that A=[13292379]. Find P and D so that $A = PDP^{-1}$, and then use this to give a formula for A^k.

Solution Leaving out the computational details, the eigenvalues and associated eigenvectors of A are

λ1=1⇒u1=[13],λ2=19⇒u2=[−11]

Therefore if

P=[1−131]andD=[10019]

then $A = PDP^{-1}$. To compute A^k, we use

Ak=PDkP−1=
[1−131] [1k00(19)k] [1414−3414] =14[1+3(19)k1−(19)k3−3(19)k3+(19)k]

Since (19)k→0 as $k \to \infty$, it follows that Ak→[14143434] as $k \to \infty$.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Compute A^4 if $A = PDP^{-1}$.
   (a) P=[1225],D=[100−1]
   (b) P=[1211−1002−1],D=[−100010002]
2. Find a matrix A that has the given eigenvalues and corresponding eigenvectors.
   (a) λ1=2⇒{[11]},λ2=−1⇒{[12]}

(b) $\lambda 1=-2\Rightarrow\{[-101]\}$,
$\lambda 2=2\Rightarrow\{[1-11]\},\lambda 3=1\Rightarrow\{[110]\}$

3. Diagonalize *A* if possible.
   (a) $A=[-3-2106]$
   (b) $A=[-2-82000-1-21]$
4. Find a 3 × 3 matrix that has all positive entries and three distinct eigenvalues.
5. Suppose that a 5 × 5 diagonalizable matrix has four distinct eigenvalues. What are the dimensions of the associated eigenspaces? Explain your answer.

EXERCISES

Exercises 1–4: Compute A^5 if $A = PDP^{-1}$.

1. $P=[4311],D=[200-1]$
2. $P=[2173],D=[100-3]$
3. $P=[1310-1200-1],D=[10002000-1]$
4. $P=[11-1101102],D=[300010001]$

Exercises 5–8: Find a matrix *A* that has the given eigenvalues and eigenvectors.

5. $\lambda 1=1\Rightarrow\{[23]\},\lambda 2=-1\Rightarrow\{[35]\}$
6. $\lambda 1=3\Rightarrow\{[47]\},\lambda 2=1\Rightarrow\{[12]\}$
7. $\lambda 1=-1\Rightarrow\{[110]\},\lambda 2=0\Rightarrow\{[121]\},\lambda 3=1\Rightarrow\{[-111]\}$
8. $\lambda 1=2\Rightarrow\{[131]\},\lambda 2=1\Rightarrow\{[21-1],[021]\}$

Exercises 9–18: Diagonalize *A* if possible.

9. $A=[1-201]$
10. $A=[-2200]$
11. $A=[7-84-5]$
12. $A=[7-102-2]$

13. $A=[1210-3-2242]$
14. $A=[4-1-2-6348-2-4]$
15. $A=[01-11011-12]$
16. $A=[353-5-7-3331]$
17. $A=[1001020000300004]$
18. $A=[100011001000-101-1]$

Exercises 19–22: Compute A^{1000} for the matrix A.

19. $A=[-34-23]$
20. $A=[5-42-1]$
21. $A=[7-84-5]$
22. $A=[7-102-2]$
23. Suppose that a 4 × 4 diagonalizable matrix has two distinct eigenvalues, one with an eigenspace of dimension 2. What is the dimension of the other eigenspace?
24. Suppose that a 7 × 7 diagonalizable matrix has three distinct eigenvalues, one with an eigenspace of dimension 1 and another with an eigenspace of dimension 2. What is the dimension of the third eigenspace?

FIND AN EXAMPLE Exercises 25–30: Find an example that meets the given specifications.

25. A 2 × 2 matrix that is diagonalizable but not invertible.
26. A 3 × 3 matrix that is diagonalizable but not invertible.
27. A 2 × 2 matrix that is invertible but is not diagonalizable.
28. A 3 × 3 matrix that is invertible but is not diagonalizable.
29. A 3 × 3 diagonalizable (but not diagonal) matrix that has three distinct eigenvalues.
30. A 3 × 3 diagonalizable (but not diagonal) matrix that has two distinct eigenvalues.

TRUE OR FALSE Exercises 31–34: Determine if the statement is true or false, and justify your answer.

31.

(a) Suppose a square matrix A has only real eigenvalues. If each eigenspace of A has dimension equal to the multiplicity of the associated eigenvalue, then A is diagonalizable.

(b) If an $n \times n$ matrix A has distinct eigenvectors, then A is diagonalizable.

32.

(a) If A is not invertible, then A is not diagonalizable.

(b) If A is diagonalizable, then so is A^k for $k = 2, 3, \ldots$.

33.

(a) If A is a diagonalizable $n \times n$ matrix, then rank(A) = n.

(b) If A and B are diagonalizable $n \times n$ matrices, then so is AB.

34.

(a) If A and B are diagonalizable $n \times n$ matrices, then so is $A + B$.

(b) If A is a diagonalizable $n \times n$ matrix, then there exist eigenvectors of A that form a basis for $\mathbf{R}^n$.

35. Suppose that $\lambda_1 \neq \lambda_2$ are eigenvalues of a 2×2 matrix A with associated eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$. Prove that det(U) ≠ 0 for U=[u1u2].

36. Prove that if A is diagonalizable and $c \neq 0$ is a scalar, then cA is also diagonalizable.

37. Prove that if A is diagonalizable, then there are infinitely many distinct matrices P and D such that $A = PDP^{-1}$.

38. Suppose that A is an $n \times n$ matrix with eigenvectors that form a basis for $\mathbf{R}^n$. Prove that there exists an invertible matrix Q such that QAQ^{-1} is a diagonal matrix.

39. Prove that if A is diagonalizable, then so is A^T.

40. Suppose that A is a matrix that can be diagonalized using matrices P and D as in Definition 6.8. Prove that det(A) = det(D).

41. Suppose that A and B are $n \times n$ matrices that can both be diagonalized using the same matrix P. Prove that $AB = BA$.

42. Suppose that A is a diagonalizable matrix with distinct nonzero eigenvalues. Prove that A^2 has positive eigenvalues.

Exercises 43–46: Diagonalize *A* if possible.

43. A=[30−2−1−125460−5−3−6042]

44. A=[0−1−1−11032−4−8−2−1−5−4220]

45. A=[2−1010−430−5−45−2−11−1−620−4−22−1010]

46. A=[3−532−35−73116−62−150001300002]

6.3 Complex Eigenvalues and Eigenvectors

Up until now we have only considered eigenvalues and eigenvectors made up of real numbers. However, there are characteristic polynomials that have roots that are not real numbers. For instance, the matrix

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

▶ This section is optional. However, complex numbers and eigenvalues are revisited in optional Sections 6.4, 11.4, and 11.5.

has characteristic polynomial

$$|A-\lambda I_2|=(3-\lambda)^2+4$$

This characteristic polynomial has no real roots, but it has two complex roots. Eigenvalues and eigenvectors involving complex numbers have useful applications, which we will get to later in this section and in the next. We start here with a brief review of the properties of complex numbers. Feel free to skip to the next subsection if you are already familiar with the basics of complex numbers.

Complex Numbers

We are not going to fully develop the complex numbers, but instead just focus on the aspects that are needed later. If a and b are real numbers, then a typical complex number has the form

$$z = a + ib$$

where i satisfies $i^2 = -1$, making i a square root of -1. Here a is called the **real part** of z, denoted by $\mathrm{Re}(z)$, and b is the **imaginary part**, denoted by $\mathrm{Im}(z)$. (Note that both the real and imaginary parts are real numbers.) The set of all complex numbers is denoted by **C**.

To add complex numbers $z_1 = a_1 + ib_1$ and $z_2 = a_2 + ib_2$, we just add the real parts and the imaginary parts separately,

$$z_1+z_2=(a_1+ib_1)+(a_2+ib_2)=(a_1+a_2)+i(b_1+b_2)$$

For example, if $z_1 = 3 + i$ and $z_2 = 2 + 3i$, then

$$z_1+z_2=(3+i)+(2+3i)=(3+2)+i(1+3)=5+4i$$

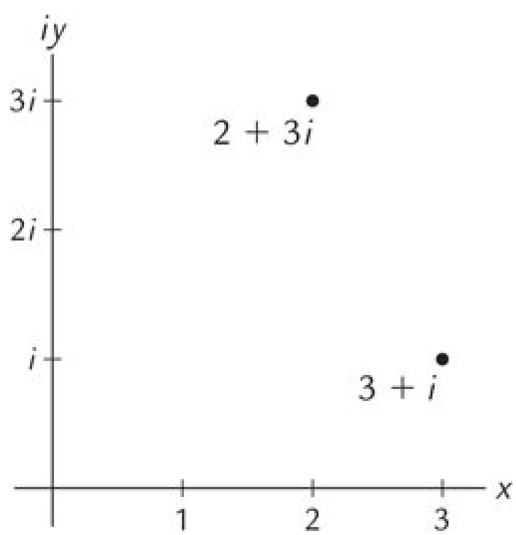


Figure 1 $z_1 = 3 + i$ and $z_2 = 2 + 3i$.

Adding complex numbers is similar to adding vectors in $\mathbf{R}^2$, with each component added separately.

We can also represent complex numbers geometrically just as we do vectors in $\mathbf{R}^2$, with $\mathrm{Re}(z)$ on the x-axis and $\mathrm{Im}(z)$ on the y-axis. For example, z_1 and z_2 from above are shown in Figure 1.

The product of complex numbers is found by multiplying term by term and then simplifying using the identity $i^2 = -1$,

$$z_1z_2=(a_1+ib_1)(a_2+ib_2)=a_1a_2+ia_1b_2+ia_2b_1+i^2b_1b_2=(a_1a_2-b_1b_2)+i(a_1b_2+a_2b_1)$$

For our two complex numbers $z_1 = 3 + i$ and $z_2 = 2 + 3i$, we have

$$z_1z_2=(3+i)(2+3i)=((3)(2)-(1)(3))+i((3)(3)+(2)(1))=3+11i$$

Modulus

An alternate way to represent a complex number is to use polar coordinates. Define the **modulus** of a complex number $z = a + ib$ by

$$|z|=a^2+b^2.$$

Argument

The modulus extends the absolute value to complex numbers and gives the distance from z to the origin. The **argument** of z, denoted by $\arg(z)$, is the angle θ (in radians) in the counter clockwise direction from the positive x-axis to the ray from the origin to z (see Figure 2). Note that the argument is not unique, because we can always add or subtract multiples of 2π.

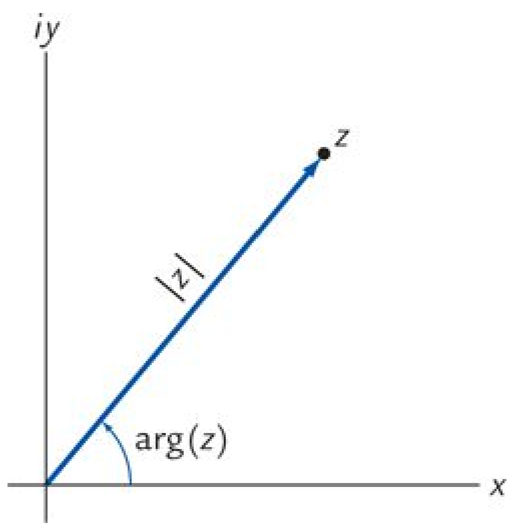


Figure 2 $|z|$ and $\arg(z)$.

If $r = |z|$ and $\theta = \arg(z)$, then we can express z in polar form as

$$z=r(\cos(\theta)+i\sin(\theta))$$

For instance, if r = 5 and $\theta = \pi/3$, then we can convert to rectangular coordinates by evaluating, with

$$z=5(\cos(\pi/3)+i\sin(\pi/3))=52+i532$$

Converting from rectangular to polar coordinates is depicted in Figure 3. For $z = a + ib$, we set r=|z|=a2+b2 and have

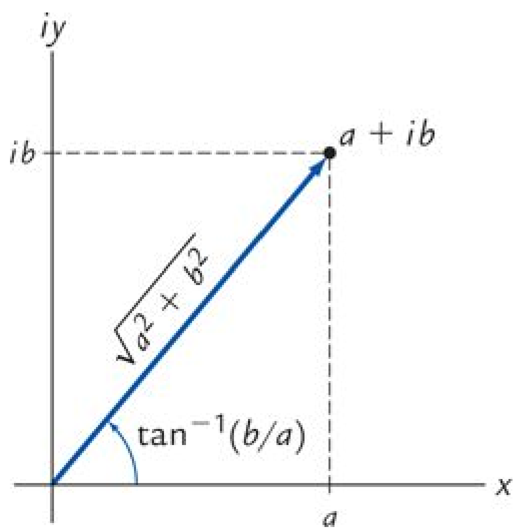


Figure 3 Converting from rectangular to polar coordinates.

$$\tan(\theta)=ba$$

For example, in the case of $z_1 = 3 + i$, we have

$$r=32+12=10 \text{and} \theta=\tan-1(13)\approx 0.3218 \text{ radians}$$

An interesting formula arises when multiplying complex numbers written in polar form. If

$$z1=r1(\cos(\theta 1)+i\sin(\theta 1)) \text{and} z2=r2(\cos(\theta 2)+i\sin(\theta 2))$$

then we have

$$z_1z_2=r_1(\cos(\theta_1)+i\sin(\theta_1))\cdot r_2(\cos(\theta_2)+i\sin(\theta_2)) =r_1r_2\{(\cos(\theta_1)\cos(\theta_2)-\sin(\theta_1)\sin(\theta_2)+i(\cos(\theta_1)\sin(\theta_2)+\cos(\theta_2)\sin(\theta_1))\} =r_1r_2(\cos(\theta_1+\theta_2)+i\sin(\theta_1+\theta_2))$$

with the last line following from trigonometric identities. This formula tells us that

$$|z_1z_2|=|z_1||z_2| \text{and} \arg(z_1z_2)=\arg(z_1)\arg(z_2)$$

Applying this repeatedly (by induction) to $z = r(\cos(\theta) + i \sin(\theta))$, we find that for each positive integer k,

$$z^k=r^k(\cos(k\theta)+i\sin(k\theta))$$

In the special case where $z = \cos(\theta) + i \sin(\theta)$ (that is, $r = 1$), this yields *DeMoivre's Formula*,

$$(\cos(\theta)+i\sin(\theta))^k=\cos(k\theta)+i\sin(k\theta)$$

For $z = a + ib$, the exponential function e^x extends to the complex numbers by the definition

$$e^{a+ib}-e^a(\cos(b)+\sin(b))$$

Note that if $z = a$ is real, then $e^z = e^a$ reduces to the usual exponential function. On the other hand, if $z = ib$ is purely imaginary, then

$$e^{ib}=\cos(b)+i\sin(b)$$

and $|e^{ib}| = 1$.

Complex Conjugate

The **complex conjugate** of $z = a + ib$ is given by $\bar{z}=a-ib$. Complex conjugation distributes across addition and multiplication, so that if z and w are complex numbers, then

$$\overline{z+w}=\bar{z}+\bar{w} \text{and} \overline{zw}=\bar{z}\cdot\bar{w}$$

One interesting consequence of these properties is that for polynomials with real coefficients, complex roots come in conjugate pairs. that is, if

$$f(z)=a_nz^n+\cdots+a_1z+a_0$$

has real coefficients and $f(z_0) = 0$, then $f(\overline{z}_0)=0$. To see why, note that $\overline{x}=x$ for any real number x. If $f(z_0) = 0$, then

$$0=\overline{0}=\overline{f(z_0)}=\overline{a_nz_0^n+\cdots+a_1z_0+a_0}=\overline{a}_n\overline{z_0^n}+\cdots+\overline{a_1z_0}+\overline{a_0}=a_n(\overline{z}_0)^n+\cdots+a_1\overline{z}_0+a_0=f(\overline{z}_0)$$

which tells us that $f(\overline{z}_0)=0$ as well. For example, if $f(z) = z^2 + 2z + 4$, then the solutions to $f(z) = 0$ can be found by applying the quadratic formula,

$$z=\frac{-2\pm\sqrt{2^2-4(1)(4)}}{2(1)}=\frac{-2\pm\sqrt{-12}}{2}=-1\pm i\sqrt{3}$$

Therefore we have two solutions, $z=-1+i\sqrt{3}$ and the conjugate $\overline{z}=-1-i\sqrt{3}$.

One benefit of expanding from the real numbers to the complex numbers is that we can completely factor polynomials. Specifically, any polynomial $f(z)$ of degree n with real or complex coefficients can be factored completely to

$$f(z)=c(z-z_1)(z-z_2)\cdots(z-z_n)$$

where c, z_1, … z_n are complex numbers.

Given an $n \times n$ matrix A, we know that the characteristic polynomial will have degree n. When working in the complex numbers, the characteristic polynomial has exactly n roots (counting multiplicities), so that A must have exactly n eigenvalues (again, counting multiplicities).

Complex Eigenvalues and Eigenvectors

Now that we have refreshed our knowledge of complex numbers, let's return to our opening problem.

Example 1

Find the eigenvalues and associated eigenvectors for the matrix

$$A=[3-413]$$

Solution We know that $\det(A - \lambda I_2) = (3 - \lambda)^2 + 4$. Setting this equal to 0 and solving for λ, we have

$$(3-\lambda)2=-4\Rightarrow 3-\lambda=\pm-4\pm 2i\Rightarrow \lambda=3\pm 2i$$

In general, it is algebraically messy to find complex eigenvectors by hand. However, it is manageable for 2×2 matrices. For the eigenvector $\lambda_1 = 3 - 2i$, we have

$$A-\lambda 1I2=A-(3-2i)I2=[2i-412i]$$

The augmented matrix of $(A - \lambda_1 I_2)\mathbf{u} = 0$ and the corresponding echelon form are

$$[2i-412i|00]12iR1+R2\rightarrow R2\sim[2i-400|00]$$

The echelon form is equivalent to the equation $2ix_1 - 4x_2 = 0$. A nontrivial solution is $x_1 = 2$ and $x_2 = i$, which gives us the eigenvector u1=[2i]. Similar calculations applied to $\lambda_2 = 3 + 2i$ can be used to produce the associated eigenvector u2=[2−i].

▶ For a vector **z**, the complex conjugate z‾ means that we take the complex conjugate for each entry of **z**. An analogous definition holds for A‾, the complex conjugate of the matrix A.

The eigenvalues in Example 1 are a conjugate pair, with λ2=λ‾1. The corresponding eigenvectors are similarly related, with u2=u‾1.

This is true for any square matrix with real entries.

THEOREM 6.13 ▶

Suppose that A is a real matrix with eigenvalue λ and associated eigenvector **u**. Then $\bar{\lambda}$ is also an eigenvalue of A, with associated eigenvector $\bar{\mathbf{u}}$.

Proof If A is a real matrix, then the characteristic polynomial has real coefficients. Previously we showed that complex roots of polynomials with real coefficients come in conjugate pairs. Thus, if λ is a complex eigenvalue of a matrix A, then so is $\bar{\lambda}$.

Next suppose that **u** is an eigenvector of A associated with λ. Since A has real entries, we have $A = \bar{A}$, so that

$$A\bar{\mathbf{u}} = \bar{A}\bar{\mathbf{u}} = \overline{A\mathbf{u}} = \overline{\lambda\mathbf{u}} = \bar{\lambda}\bar{\mathbf{u}}$$

Hence $\bar{\mathbf{u}}$ is an eigenvector of A associated with eigenvalue $\bar{\lambda}$. ■ ■

Example 2

Find the eigenvalues and associated eigenvectors for the matrix

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

Solution Starting with the characteristic polynomial, we have

$$\det(A - \lambda I_3) = -\lambda^3 + 5\lambda^2 - 17\lambda + 13 = -(\lambda - 1)(\lambda^2 - 4\lambda + 13)$$

Thus one eigenvalue is $\lambda_1 = 1$. Applying the quadratic formula to the quadratic term shows the other eigenvalues are $\lambda_2 = 2 + 3i$ and $\lambda_3 = \bar{\lambda}_2 = 2 - 3i$.

For the eigenvectors associated with $\lambda_1 = 1$, the augmented matrix and corresponding echelon form are

[−23−4−22−4112|000]−R1+R2→R212R1+R3→R352R2+R3→R3~
[−23−40−10000|000]

▶ The row operations are

−23+3iR1+R2→R2,13+3iR1+R3→R3,−1+3i4R2+R3→R3

Back substitution yields the eigenvector u1=[20−1]. For $\lambda_2 = 2 + 3i$, we have

[(−3−3i)3−4−2(1−3i)
−411(1−3i)|000]~[(−3−3i)3−40−2i−(8+4i)/3000|000]

After back substitution and scaling we find that u2=[−1+5i−2+4i 3].

There is no need for row operations to find $\mathbf{u}_3$. Since λ3=λ¯2, we know from Theorem 6.13 that u3=u¯2=[−1−5i−2−4i 3].

Rotation-Dilation Matrices

One application of complex eigenvalues and eigenvectors is in analyzing the behavior of a special class of 2×2 matrices. For illustrative purposes we distribute eight vectors $\mathbf{x}_1, \ldots, \mathbf{x}_8$ evenly around the unit circle as shown in Figure 4(a). Now define the matrix

A=[1−221]

Figure 4(b) shows the vectors $A\mathbf{x}_1, \ldots, A\mathbf{x}_8$.

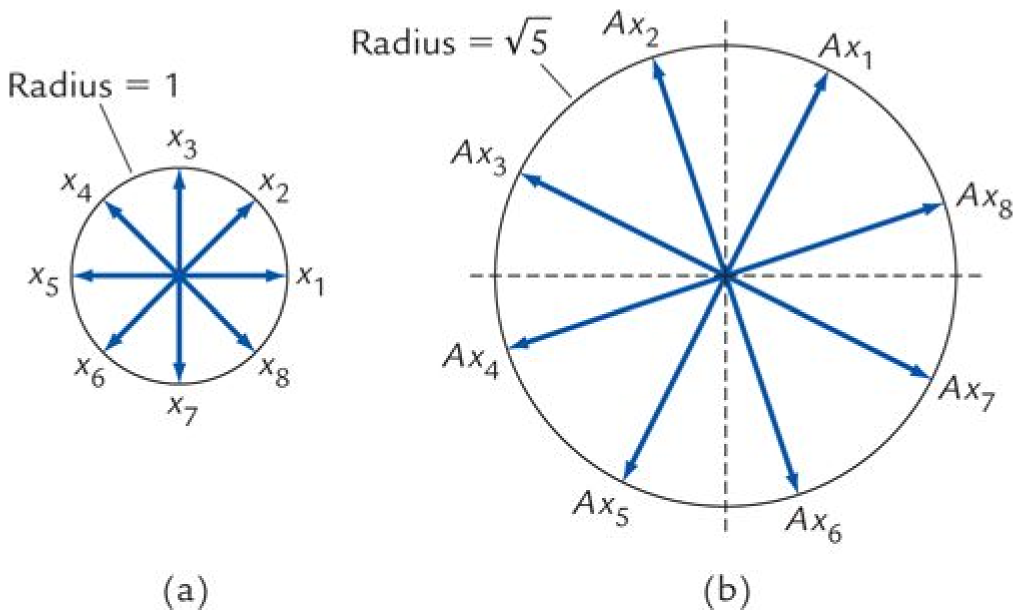


Figure 4 (a) The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_8$ are evenly spaced around the unit circle. (b) The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_8$ after multiplication by A, which rotates each by 63.43° and dilates each by 5. (NOTE: Figures are not drawn to scale.)

Comparing $\mathbf{x}_i$ with $A\mathbf{x}_i$, we see that multiplication by A causes each $\mathbf{x}_i$ to be rotated and dilated by the same amount. It turns out that the same thing will happen for any vector $\mathbf{x}$ when multiplied by A.

This rotation–dilation behavior happens with any real matrix of the form

$$A=\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

To see why, first note that if $x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, then

$$Ax=\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} ax_i-bx_2 \\ bx_1+ax_2 \end{bmatrix}$$

It can be shown that $\lambda = a + ib$ is an eigenvalue of A (see Exercise 43). Next note that if we let $x = x_1 + ix_2$, then

$$\lambda x=(a+ib)(x_1+ix_2)=(ax_1-bx_2)+i(bx_1+ax_2)$$

Thus the components of $A\mathbf{x}$ match the real and imaginary parts of λx. If we think of the real and imaginary parts as components of λx, then the products $A\mathbf{x}$ and λx can be viewed as equivalent. From the properties of the products of complex numbers, we know that

$$\arg(\lambda x)=\arg(\lambda)+\arg(x)(\text{Rotation of x in C by the angle }\arg(\lambda))\ |\lambda x|=|\lambda||x|=a2+b2|x|(\text{Dilation of x in C by the multiple}|\lambda|)$$

Therefore the eigenvalue tells us the amount of rotation and dilation induced by A. Returning to our matrix

$$A=[1-221]$$

we have $\lambda = 1 + 2i$, so that $A\mathbf{x}$ will produce

$$\text{Rotation by }\arg(\lambda)=\tan-1(2/1)\approx 63.43\text{ degreesDilation by }|\lambda|=12+22=5$$

Note that this is consistent with Figure 4.

Example 3

Determine the rotation and dilation that result from multiplying $\mathbf{x}$ in $\mathbf{R}^2$ by

$$A=[7-447]$$

Solution An eigenvalue of A is $\lambda = 7 + 4i$, so that we have

$$\text{Rotation by }\tan-1(4/7)\approx 0.5191\text{ radiansDilation by }72+42=65$$

The Hidden Rotation–Dilation Matrix

As it happens, *any* 2×2 real matrix with complex eigenvalues has a rotation–dilation hidden within it. Finding this rotation–dilation requires a procedure reminiscent of diagonalization. We start by illustrating using the matrix from the beginning of this section,

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

We have previously shown that $\mathbf{u}=\begin{bmatrix}2\\ i\end{bmatrix}$ is an eigenvector of *A*. Now form the matrix

$$P=[\operatorname{Re}(\mathbf{u}) \quad \operatorname{Im}(\mathbf{u})]=\begin{bmatrix}2 & 0\\ 0 & 1\end{bmatrix}$$

where Re(**u**) and Im(**u**) denote the vectors formed by taking the real and imaginary parts of each component of **u**. Then we compute

$$P^{-1}AP=\begin{bmatrix}\frac{1}{2} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}3 & -4\\ 1 & 3\end{bmatrix}\begin{bmatrix}2 & 0\\ 0 & 1\end{bmatrix}=\begin{bmatrix}3 & -2\\ 2 & 3\end{bmatrix}=B$$

Thus $A = PBP^{-1}$, where *B* is a rotation–dilation matrix. Note that the first row of *B* corresponds to the real and imaginary parts of the eigenvalue $\lambda = 3 - 2i$ of *A* associated with **u**. This example is generalized in the next theorem.

THEOREM 6.14 ▶

Let *A* be a nonzero real 2 × 2 matrix with complex eigenvalue $\lambda = a - ib$ and associated eigenvector **u**. If $P = [\operatorname{Re}(\mathbf{u})\ \operatorname{Im}(\mathbf{u})]$, then

$$A=PBP^{-1}$$

where $B=\begin{bmatrix}a & -b\\ b & a\end{bmatrix}$ is a rotation–dilation matrix.

We do not give a proof here, but most of the pieces required are covered in Exercise 45. The form PBP^{-1} suggests viewing the transformation $A\mathbf{x}$ as the composition of transformations. The first is a change to the basis {Re(**u**), Im(**u**)}, the second is a rotation–dilation, and the third is a change back to the standard basis. Note that the rotation–dilation is not applied to **x**, but to the coordinate vector of **x** relative to the basis {Re(**u**), Im(**u**)}.

Example 4

Find the hidden rotation–dilation matrix within

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

Solution The characteristic polynomial is $\det(A - \lambda I_2) = \lambda^2 - 4\lambda + 13$. A quick application of the quadratic formula reveals that one of the eigenvalues is $\lambda = 2 - 3i$. (What is the other?) The usual procedure yields an associated eigenvector $\mathbf{u}=\begin{bmatrix}1+3i\\ 2\end{bmatrix}=\begin{bmatrix}1\\ 2\end{bmatrix}+i\begin{bmatrix}3\\ 0\end{bmatrix}$. Applying Theorem 6.14 with $\lambda = 2 - 3i$, we have

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

Clearly B has the form of a rotation–dilation matrix. We can check our calculations with the computation

$$PBP^{-1}=\begin{bmatrix}1 & 3\\ 2 & 0\end{bmatrix}\begin{bmatrix}2 & -3\\ 3 & 2\end{bmatrix}\left(-\frac{1}{6}\begin{bmatrix}0 & -3\\ -2 & 1\end{bmatrix}\right)=\begin{bmatrix}1 & 5\\ -2 & 3\end{bmatrix}=A$$

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Let $z_1 = 2 + i$ and $z_2 = 5 - 3i$. Compute the following:

(a) $z_1 - z_2$

(b) $4z_2 + 3z_1$

(c) z_2/z_1

(d) $|z_1|$

(e) $\bar{z}_2$

(f) $\mathrm{Re}(z_1 z_2)$

2. Find the eigenvalues and a basis for each eigenspace for A.

(a) $A=\begin{bmatrix}2 & -3\\ 3 & 2\end{bmatrix}$

(b) A=[12−54]

3. Determine the rotation and dilation for *A*.

(a) A=[4−114]

(b) A=[25−52]

4. Determine the rotation and dilation within *A*.

(a) A=[1−131]

(b) A=[25−51]

EXERCISES

Exercises 1–2: Suppose that $z_1 = 5 - 2i$ and $z_2 = 3 + 4i$ are both complex numbers.

1. Compute each of the following:

(a) $z_1 + z_2$

(b) $2z_1 + 3z_2$

(c) $z_1 - z_2$

(d) $z_1 \, z_2$

2. Compute each of the following:

(a) $-4z_1$

(b) $3z_1 + 2z_2$

(c) $z_2 - z_1$

(d) $(-5z_1)(3z_2)$

Exercises 3–8: Find the eigenvalues and a basis for each eigenspace for *A*.

3. A=[31−51]

4. A=[1−123]

5. A=[1−213]

6. A=[13−31]

7. A=[42−12]

8. A=[52−5−1]

Exercises 9–14: Determine the rotation and dilation for *A*.

9. A=[2−112]

10. A=[3−223]

11. A=[1−111]

12. A=[3−443]

13. A=[43−34]

14. A=[52−25]

Exercises 15–20: Find the rotation–dilation matrix within *A*.

15. A=[31−51]

16. A=[1−123]

17. A=[1−213]

18. A=[13−31]

19. A=[42−12]

20. A=[52−5−1]

21. Suppose that some of the roots of a degree 5 polynomial with real coefficients are 2, $1 + 2i$, and $3 - i$. What are the other roots, and what are the multiplicities of all roots?

22. Suppose that some of the roots of a degree 7 polynomial with real coefficients are -3, $-2 + i$, $5 + i$, and i. What are the other roots, and what are the multiplicities of all roots?

FIND AN EXAMPLE Exercises 23–30: Find an example that meets the given specifications.

23. A complex number z such that $|z| = 3$ and $\mathrm{Re}(z) = 2\mathrm{Im}(z)$.

24. A complex number z such that $|z| = 5$ and $2\mathrm{Re}(z) = -\mathrm{Im}(z)$.

25. A rotation–dilation matrix *A* that rotates vectors by 90° and dilates vectors by 2.

26. A rotation–dilation matrix *A* that rotates vectors by −45° and dilates vectors by 12.

27. A 2×2 matrix *A* that is not itself a rotation–dilation matrix but does have the rotation–dilation matrix B=[1−221] hidden within it.

28. A 2 × 2 matrix *A* that is not itself a rotation–dilation matrix but does have the rotation–dilation matrix B=[3−113] hidden within it.

29. A 2 × 2 matrix *A* that has complex entries but real eigenvalues.

30. A 4 × 4 matrix with real entries and only complex eigenvalues.

TRUE OR FALSE Exercises 31–36: Determine if the statement is true or false, and justify your answer.

31.

(a) If z and w are complex numbers, then $|zw| = |z||w|$.

(b) A 2 × 2 matrix *A* with real eigenvalues has a rotation–dilation matrix hidden within it.

32.

(a) If z and w are complex numbers, then $|z + w| = |z||w|$.

(b) If $z \neq 0$ is a complex number, then $z\overline{z}$ is a real number.

33.

(a) (a) If *A* is a square matrix with real entries and complex eigenvectors, then *A* has complex eigenvalues.

(b) If z is a complex number, then $|z|=|\overline{z}|$.

34.

(a) If z and w are complex numbers, then $|z + w| = |z| + |w|$.

(b) The amount of dilation imparted by a rotation–dilation matrix *A* is equal to $|\lambda|$, where λ is an eigenvalue of *A*.

35.

(a) If **z** is a complex vector, then $z=\overline{\overline{z}}$.

(b) If *A* is a rotation–dilation matrix, then the angle between a vector $\mathbf{x}_0$ and $A\mathbf{x}_0$ is the same for all nonzero $\mathbf{x}_0$ in $\mathbf{R}^2$.

36.

(a) If A=[a−bba] is a rotation–dilation matrix that has eigenvalues λ_1 and λ_2, then $|\lambda_1| = |\lambda_2|$.

(b) If *A* is a 2 × 2 matrix with complex entries, then the eigenvalues of *A* cannot be complex conjugates.

37. If z and w are complex numbers, prove that

(a) $\overline{z+w}=\overline{z}+\overline{w}$

(b) $\overline{zw}=\overline{z}\cdot\overline{w}$

38. Suppose z is a complex number.

(a) Prove that $z\overline{z}=|z|^2$.

(b) If w is also complex, use (a) to show that $\frac{w}{z}=\frac{w\overline{z}}{|z|^2}$ for $z \neq 0$.

(c) Use part (b) to simplify $\frac{2+i}{4-3i}$ to the form $a + ib$, where a and b are real.

39. If c is a complex scalar and **v** is a vector with complex entries, prove that $\overline{c}\cdot\overline{\mathbf{v}}=\overline{c\mathbf{v}}$.

40. If A is a matrix with complex entries and **v** is a vector with complex entries, prove that $\overline{A}\,\overline{\mathbf{v}}=\overline{A\mathbf{v}}$.

41. Suppose that λ is complex and t is real. Prove that $e^{\overline{\lambda}t}=\overline{e^{\lambda t}}$.

42. If z is complex, prove that

(a) $\frac{1}{2}(z+\overline{z})=\mathrm{Re}(z)$

(b) $\frac{1}{2i}(z-\overline{z})=\mathrm{Im}(z)$

43. Prove that $\lambda = a + ib$ is an eigenvalue of the matrix

$$A=\begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

44. Prove that if n is odd and A is a real $n \times n$ matrix, then there exists a nonzero vector **u** such that $A\mathbf{u} = c\mathbf{u}$, where c is a real number.

45. In this exercise we prove that $AP = PC$ for the matrices in Theorem 6.14. (This combined with Exercise 46 proves the theorem.)

(a) Show that $A(\mathrm{Re}(\mathbf{u})) = \mathrm{Re}(A\mathbf{u})$ and $A(\mathrm{Im}(\mathbf{u})) = \mathrm{Im}(A\mathbf{u})$. (HINT: Recall that A is a real matrix.)

(b) Use (a) and the identity $\mathbf{u} = \mathrm{Re}(\mathbf{u}) + i\mathrm{Im}(\mathbf{u})$ to show that

$$A(\mathrm{Re}(\mathbf{u}))=a\mathrm{Re}(\mathbf{u})+b\mathrm{Im}(\mathbf{u}),\quad A(\mathrm{Im}(\mathbf{u}))=-b\mathrm{Re}(\mathbf{u})+a\mathrm{Im}(\mathbf{u})$$

(HINT: Recall that **u** is an eigenvector of A with eigenvalue $\lambda = a - ib$.)

(c) Apply (b) to show that the columns of AP are the same as the columns of PC, and conclude $AP = PC$.

46. In this exercise we prove that the columns of P in Theorem 6.14 are linearly independent and therefore P is invertible. The proof is by contradiction: Suppose that $\mathrm{Re}(\mathbf{u})$ and $\mathrm{Im}(\mathbf{u})$ are linearly dependent. Then there exists a real scalar c such that $\mathrm{Re}(\mathbf{u}) = c\mathrm{Im}(\mathbf{u})$.

(a) Prove that

$$\mathbf{u} = \mathrm{Re}(\mathbf{u}) + i\,\mathrm{Im}(\mathbf{u}) = (c + i)\,\mathrm{Im}(\mathbf{u})$$

and from this show $\lambda\mathbf{u} = \lambda(c + i)\mathrm{Im}(\mathbf{u})$.

(b) Show that $\mathrm{Re}(\lambda\mathbf{u}) = c\,\mathrm{Im}(\lambda\mathbf{u})$ by evaluating $A(\mathrm{Re}(\mathbf{u}))$ and $A(c\,\mathrm{Im}(\mathbf{u}))$ and setting the results equal to each other. (HINT: Use (a) from Exercise 45 and the fact that **u** is an eigenvector with eigenvalue λ.)

(c) Show that $\lambda\mathbf{u} = \mathrm{Re}(\lambda\mathbf{u}) + i\,\mathrm{Im}(\lambda\mathbf{u})$, and combine this with the result from (b) to prove that $\lambda\mathbf{u} = (c + i)\mathrm{Im}(\lambda\mathbf{u})$.

(d) Prove that $\lambda(c + i)\mathrm{Im}(\mathbf{u}) = (c + i)\mathrm{Im}(\lambda\mathbf{u})$. Show that $\lambda\,\mathrm{Im}(\mathbf{u}) = \mathrm{Im}(\lambda\mathbf{u})$, and explain why this implies λ is a real number.

(e) Explain why λ being a real number is a contradiction, and from this complete the proof.

Exercises 47–50: Find the complex eigenvalues and a basis for each associated eigenspace for the given matrix.

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

48. $A = \begin{bmatrix} 4 & -3 & 1 \\ 2 & 2 & 7 \\ 1 & -4 & 2 \end{bmatrix}$

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

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

6.4 Systems of Differential Equations

In a variety of applications, systems of equations arise involving one or more functions and the derivatives of those functions. One example can be found in a simplified model of the concentration of insulin and glucose in an individual. Insulin is a hormone that reduces glucose concentrations.

▶ This section is optional and can be omitted without loss of continuity.

Suppose that $y_1(t)$ and $y_2(t)$ give the deviation from normal of insulin and glucose concentrations, respectively. Then the rates of change y1′(t) and y2′(t) of insulin and glucose concentrations are related by

▶ For instance, $y_2(t) = 10$ indicates a glucose level of 10 mg/dl above normal.

$$y2'(t)=ay1(t)+by2(t)y2'(t)=cy1(t)+dy(t)$$

where *a*, *b*, *c*, and *d* are constants. This is an example of a system of linear differential equations. We will return to this example shortly, after developing a method for finding the solutions to such systems.

If $y = y(t)$, one of the simplest *differential equations* is

▶ Here we assume a basic familiarity with differential equations, and so only provide a brief background.

$$y'=ay \tag{1}$$

where a is a constant. Setting $y = ce^{at}$ for any constant$c$, we have $y' = ace^{at}$, so that our function y satisfies $y' = ay$. Our function is called a *solution* to the differential equation, and in fact, the only solutions to this differential equation have this form.

In this section we describe how to find the solutions to a system of linear first-order differential equations, which has the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3 + \cdots + a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + a_{23}y_3 + \cdots + a_{2n}y_n \\ y_3' &= a_{31}y_1 + a_{32}y_2 + a_{33}y_3 + \cdots + a_{3n}y_n \\ \vdots\ & \quad\ \vdots \qquad\ \vdots \qquad \vdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + a_{n3}y_3 + \cdots + a_{nn}y_n \end{aligned} \tag{2}$$

We assume that $y_1 = y_1(t), \ldots, y_n = y_n(t)$ are each differentiable functions. The system is linear because the functions are linearly related, and it is first-order because the derivatives that appear are first order only. If we denote

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{y}' = \begin{bmatrix} y_2' \\ \vdots \\ y_n' \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

then the system (2) can be expressed compactly as $\mathbf{y}' = A\mathbf{y}$. This matrix equation resembles the differential equation (1), which suggests that a solution to our system might have the form

$$\mathbf{y} = \begin{bmatrix} u_1 e^{\lambda t} \\ u_2 e^{\lambda t} \\ \vdots \\ u_n e^{\lambda t} \end{bmatrix} = e^{\lambda t}\mathbf{u}$$

where $\mathbf{u}$ is a constant vector. If $\mathbf{y}$ is so defined, then $\mathbf{y}' = \lambda e^{\lambda t}\mathbf{u} = e^{\lambda t}(\lambda\mathbf{u})$ and $A\mathbf{y} = A(e^{\lambda t}\mathbf{u}) = e^{\lambda t}(A\mathbf{u})$. Since $e^{\lambda t} \neq 0$, this tells us that is a solution to $\mathbf{y}' = A\mathbf{y}$ exactly when

$$A\mathbf{u} = \lambda\mathbf{u}$$

That is, $\mathbf{y} = e^{\lambda t}\mathbf{u}$ is a solution when λ is an eigenvalue of A with associated eigenvector $\mathbf{u}$.

In many cases, an $n \times n$ matrix A will have n linearly independent eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ with associated eigenvalues $\lambda_1, \ldots, \lambda_n$. If we form the linear combination

$$\mathbf{y} = c_1 e^{\lambda_1 t}\mathbf{u}_1 + \cdots + c_n e^{\lambda_n t}\mathbf{u}_n$$

where $c_1, \ldots, c_n$ are constants, then $\mathbf{y}' = c_1\lambda_1 e^{\lambda_1 t}\mathbf{u}_1 + \cdots + c_n\lambda_n e^{\lambda_n t}\mathbf{u}_n$ and

$$A\mathbf{y}=A(c_1e^{\lambda_1t}\mathbf{u}_1+\cdots+c_ne^{\lambda_nt}\mathbf{u}_n)=c_1e^{\lambda_1t}A\mathbf{u}_1+\cdots+c_ne^{\lambda_nt}A\mathbf{u}_n=c_1\lambda_1e^{\lambda_1t}\mathbf{u}_1+\cdots+c_n\lambda_ne^{\lambda_nt}\mathbf{u}_n=\mathbf{y}'$$

Thus $\mathbf{y}$ is a solution to $\mathbf{y}' = A\mathbf{y}$. It turns out that all solutions to this type of system of differential equations will have this form. The set of all solutions is called the **general solution** for the system. This is summarized in the next theorem.

THEOREM 6.15 ▶

Suppose that $\mathbf{y}' = A\mathbf{y}$ is a first-order linear system of differential equations. If A is an $n \times n$ diagonalizable matrix, then the general solution to the system is given by

$$\mathbf{y}=c_1e^{\lambda_1t}\mathbf{u}_1+\cdots+c_ne^{\lambda_nt}\mathbf{u}_n$$

where $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are n linearly independent eigenvectors with associated eigenvalues $\lambda_1, \ldots, \lambda_n$, and $c_1, \ldots, c_n$ are constants.

Note that if A is diagonalizable, then there must be n linearly independent eigenvectors. Also, the eigenvalues may be repeated to reflect multiplicities.

Example 1

The concentrations of insulin and glucose in an individual interact with each other and vary over time. A mathematical model for the concentrations of insulin and glucose in an individual is given by the system

$$\begin{aligned} y_1' &= -0.05y_1 + 0.225y_2 \\ y_2' &= -0.3y_1 - 0.65y_2 \end{aligned}$$

Find the general solution for this system.

Solution The coefficient matrix is $A=\begin{bmatrix}-0.05 & 0.225\\ -0.3 & -0.65\end{bmatrix}$. Applying our standard methods, we find that the eigenvalues and eigenvectors are

$$\lambda_1=-0.5 \Rightarrow \mathbf{u}_1=\begin{bmatrix}-1\\ 2\end{bmatrix}, \quad \lambda_2=-0.2 \Rightarrow \mathbf{u}_2=\begin{bmatrix}3\\ -2\end{bmatrix}$$

By Theorem 6.15, the general solution is

$$\mathbf{y}=c_1e^{-0.5t}\mathbf{u}_1+c_2e^{-0.2t}\mathbf{u}_2=c_1e^{-0.5t}\begin{bmatrix}-1\\ 2\end{bmatrix}+c_2e^{-0.2t}\begin{bmatrix}3\\ -2\end{bmatrix}$$

The functions giving the insulin and glucose concentrations are

$$\begin{aligned} y_1&=-c_1e^{-0.5t}+3c_2e^{-0.2t}\\ y_2&=2c_1e^{-0.5t}-2c_2e^{-0.2t}\end{aligned}$$

Arms Races

After the end of World War I, Lewis F. Richardson (who pioneered the use of mathematics in meteorology) proposed a model to describe the evolution of an arms race between two countries. Here we consider a simplified version of this model. Let $y_1 = y_1(t)$ and $y_2 = y_2(t)$ denote the quantity of arms held by two different nations. The derivatives y_1' and y_2' represent the rate of change in the size of each nation's arsenal. Each nation is concerned about security against the other and acquires arms in proportion to those held by its opponent. There is also a cost of acquiring arms, which tends to reduce the rate of additions to each nation's arsenal in proportion to arsenal size. These factors are incorporated into the system of differential equations

$$\begin{aligned} y_1'&=-dy_1+ey_2\\ y_2'&=fy_1-gy_2\end{aligned} \tag{3}$$

The constants d, e, f, and g are all positive and depend on the particular situation, with e and f dictated by the degree of fear that each country has of the other and d and g the level of aversion to additional spending on arms in each country.

Example 2

Suppose that we have two countries in an arms race modeled by the system of differential equations

$$\begin{aligned} y_1' &= -3y_1 + 2y_2 \\ y_2' &= 3y_1 - 2y_2 \end{aligned}$$

Find the general solution for this system. Then find the formula for $y_1(t)$ and $y_2(t)$ if $y_1(0) = 5$ and $y_2(0) = 15$.

Solution The coefficient matrix for this system is $A = \begin{bmatrix} -3 & 2 \\ 3 & -2 \end{bmatrix}$. Applying our usual methods for finding the eigenvalues and eigenvectors, we have

$$\lambda_1 = -5 \Rightarrow \mathbf{u}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \lambda_2 = 0 \Rightarrow \mathbf{u}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Therefore, by Theorem 6.15, the general solution for this system is

$$\mathbf{y} = c_1 e^{-5t}\mathbf{u}_1 + c_2 e^{0}\mathbf{u}_2 = c_1 e^{-5t}\begin{bmatrix} -1 \\ 1 \end{bmatrix} + c_2\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Extracting the two functions y_1 and y_2, we have

$$\begin{aligned} y_1 &= -c_1 e^{-5t} + 2c_2 \\ y_2 &= c_1 e^{-5t} + 3c_2 \end{aligned}$$

Evaluating these functions at $t = 0$ and using the equations $y_1(0) = 5$ and $y_2(0) = 15$ yield the system

$$\begin{aligned} -c_1 + 2c_2 &= 5 \\ c_1 + 3c_2 &= 15 \end{aligned}$$

This linear system has unique solution $c_1 = 3$ and $c_2 = 4$. Hence

$$\begin{aligned} y_1 &= -3e^{-5t} + 8 \\ y_2 &= 3e^{-5t} + 12 \end{aligned}$$

The next example shows what we do if the coefficient matrix for a first-order linear system has repeated real eigenvalues.

Example 3

Find the general solution for the system

y1′=2y1+ y2− y3y2′=2y1+3y2−2y3y3′=−3y1−3y1+4y3

Solution Here the coefficient matrix is given by

A=[21−123−2−3−34]

The characteristic polynomial for A is $-(\lambda - 7)(\lambda - 1)^2$, and bases for the eigenspaces are

λ1=7⇒Basis: {[1+−3]},λ2=1⇒Basis: {[1+−3],[1−10]}

Although $\lambda_2 = 1$ has multiplicity 2, since the eigenspace also has dimension 2, Theorem 6.15 still applies. (The case where an eigenspace has less than maximal dimension is more complicated and not included here.) We just repeat the term corresponding to $\lambda_2 = 1$ twice, once for each of the linearly independent eigenvectors. The general solution is

y=c1e7t[12−3]+c2et[101]+c3et[1−10]

Writing the individual functions, we have

y1= c1e7t+(c2+c3)ety2= 2c1e7t−c3ety3=−3c1e7t+c2et

Complex Eigenvalues

Suppose that our system of linear differential equations has a real coefficient matrix A with complex eigenvalues. We can express the general solution just as we did with real eigenvalues. However, when λ and associated eigenvector $\mathbf{u}$ are both complex, the product $e^{\lambda t}\mathbf{u}$

typically is as well. It is generally preferable to have solutions free of complex terms, and with a bit of extra thought we can.

To get us started, we recall a few properties of complex numbers. (All are given in Section 6.3.)

- $e^{a+ib} = e^{a}(\cos(b) + i\sin(b))$
- 12(z¯+z)=Re(z)
- eλ¯t=eλt¯
- 12i(z¯−z)=Im(z)

Now let A be a real matrix with complex eigenvalue $\lambda = a + ib$ and associated eigenvector **u**. Then λ¯=a−ib is also an eigenvalue and has associated eigenvector u¯. Instead of taking linear combinations of $e^{\lambda t}\mathbf{u}$ and eλ¯tu¯ for the general solution, we take linear combinations of

y1=12(eλ¯tu¯+eλtu)=Re(eλtu)y2=12i(eλ¯tu¯+eλtu)==Im(eλtu)

Note that both $\mathrm{Re}(e^{\lambda t}\mathbf{u})$ and $\mathrm{Im}(e^{\lambda t}\mathbf{u})$ are real-valued, so that the general solution will be made up of real-valued functions. To find $\mathrm{Re}(e^{\lambda t}\mathbf{u})$ and $\mathrm{Im}(e^{\lambda t}\mathbf{u})$, we compute

eλtu=eat+ibtu =eat(cos(bt)+isin(bt))(Re(u)+iIm(u)) =eat(cos(bt)Re(B) −sin(bt)Im(u))+ieat(sin(bt)Re(u)+cos(bt)Im(u))

Separating the real and imaginary parts, we have

y1=eat(cos(bt)Re(u)−sin(bt)Im(u))y2=eat(sin(bt)Re(u)+cos(bt)Im(u)) (4)

Example 4

Find the general solution for the system

y1′=6y1−5y1y2′=5y1−2y2

Solution Here the coefficient matrix is

$$A=[6-55-2]$$

The characteristic polynomial for this matrix is $\det(A-\lambda I_2)=\lambda^2-4\lambda+13$. Applying the quadratic formula and the usual matrix manipulations yields the eigenvalues and eigenvectors

$$\lambda1=2+3i\Rightarrow u1=[4+3i\ 5\ ],\lambda2=2-3i\Rightarrow u1=[4-3i\ 5\ ]$$

The two eigenvalues are a complex conjugate pair, so we use the formulas for $\mathbf{y}_1$ and $\mathbf{y}_2$ given in (4) to find

$$y1=e2t(\cos(3t)\mathrm{Re}(u1)-\sin(3t)\mathrm{Im}(u1))=e2t(\cos(3t)[45]-\sin(3t)[30])y2=e2t(\sin(3t)\mathrm{Re}(u1)+\cos(3t)\mathrm{Im}(u1))=e2t(\sin(3t)[45]+\cos(3t)[30])$$

Hence the general solution is

$$y=c1y1+c2y2=c1e2t(\cos(3t)[45]-\sin(3t)[30])+c2e2t(\sin(3t)[45]+\cos(3t)[30])$$

The individual functions are

$$y1=((4c1+3c2)\cos(3t)+(4c2-3c1)\sin(3t))e2ty2=(5c1\cos(3t)+5c2\sin(3t))e2t$$

The next example features a system with a combination of real and complex eigenvalues.

Example 5

Find the general solution for the system

$$y1'=-10y1+6y2-3y3y2'=-12y1+6y2-5y3y3'=8y1-4y2+3y3$$

Solution The coefficient matrix is

$$A=[-106-3-126-58-43]$$

and the characteristic polynomial is

$$\det(A-\lambda I3)=-\lambda3-\lambda2-4\lambda-4=-\lambda2(\lambda+1)-4(\lambda+1)=-(\lambda2+4)(\lambda+1)$$

giving us eigenvalues $\lambda = -1, \pm 2i$. The associated eigenvectors are

$$\lambda1=-1\Rightarrow u1=[-1-11],\lambda2=2i\Rightarrow u2=[9-3i12-2i\ -8],\lambda3=-2i\Rightarrow u3=[9+3i12+2i\ -8]$$

We treat the real and complex eigenvalues separately and form a linear combination of the components at the end. For $\lambda_1 = -1$, we have

$$y1=e-tu1=e-t[-1-11]$$

For the complex conjugate pair $\pm 2i$, we have

$$y2=e0(\cos(2t)\mathrm{Re}(u2)-\sin(2t)\mathrm{Im}(u2))=\cos(2t)[912-8]-\sin(2t)[-3-20]y3=e0(\sin(2t)\mathrm{Re}(u2)+\cos(2t)\mathrm{Im}(u2))=\sin(2t)[912-8]+\cos(2t)[-3-20]$$

The general solution is then

$$y=c1y1+c2y2+c3y3$$

for $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ above.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. The coefficient matrix for a system of linear differential equations of the form **y′**=A**y** has the given eigenvalues and eigenspace bases. Find the general solution for the system.
   (a) $\lambda1=4\Rightarrow\{[1-2]\},\lambda2=-4\Rightarrow\{[52]\}$
   (b) $\lambda1=-2\Rightarrow\{[1-21]\},\lambda2=1\Rightarrow\{[-101],[2-30]\}$
   (c) $\lambda1=i\Rightarrow\{[3+2i1+i]\},\lambda2=-i\Rightarrow\{[3-2i1-i]\}$
2. Find the general solution for the system.
   (a) $y1'=y1-2y2y2'=4y1+y2$
   (b) $y1'=y1+2y2-y3y2'=-3y1-y2+y3y3'=5y1+2y2-2y3$
3. Find the solution for the system that satisfies the conditions at t = 0.
   (a) $y1'=\quad y1+2y2, \qquad y1(0)=-2y2'=-y1-y2, \quad y2(0)=1$
   (b) $y1'=y1-3y2-y3,\ y1(0)=-2y2'=2y1+y2+2ys,\ y2(0)=1y3'=3y1-2y2+y3\ y3(0)=3$

EXERCISES

Exercises 1–10: The coefficient matrix for a system of linear differential equations of the form **y′**=A**y** has the given eigenvalues and eigenspace bases. Find the general solution for the system.

1. $\lambda1=-1\Rightarrow\{[11]\},\lambda2=2\Rightarrow\{[1-1]\}$
2. $\lambda1=1\Rightarrow\{[2-1]\},\lambda2=3\Rightarrow\{[31]\}$
3. $\lambda1=2\Rightarrow\{[431]\},\lambda2=-2\Rightarrow\{[120],[231]\}$
4. $\lambda1=3\Rightarrow\{[110]\},\lambda2=0\Rightarrow\{[151],[214]\}$
5. $\lambda1=2i\Rightarrow\{[1+i2-i]\},\lambda2=-2i\Rightarrow\{[1-i2+i]\}$
6. $\lambda1=3+i\Rightarrow\{[2ii]\},\lambda2=-2-i\Rightarrow\{[-2ii]\}$
7. $\lambda1=4\Rightarrow\{[315]\},\lambda2=1+i\Rightarrow\{[4+i\ -2i3+i]\},\lambda3=1-i\Rightarrow\{[4-i\ \ 2i3-i]\}$
8. $\lambda1=-1\Rightarrow\{[103]\},\lambda2=3i\Rightarrow\{[2-i1+i\ \ 7i]\},\lambda3=-3i\Rightarrow\{[2+i1-2i\ -7i]\}$
9. $\lambda1=1\Rightarrow\{[6250]\},\lambda2=1+i\Rightarrow\{[3+2i\ 6\ 2-3i\ -5i]\},\lambda3=4\Rightarrow\{[1232]\},\lambda4=1-3i\Rightarrow\{[3-2i\ 6\ 2+3i\ 5i\ ]\}$

10. $\lambda_1=-2\Rightarrow\{[3012]\},\lambda_2=4i\Rightarrow\{[0-5\ 4+3i\ 3i\ 1-i]\},\lambda_3=0\Rightarrow\{[4211]\},\lambda_4=-4\Rightarrow\{[\ -5\ 4-3i\ -3i\ 1+i]\}$

Exercises 11–18: Find the general solution for the system.

11. $y_1'=y_1+4y_2\quad y_2'=y_1+\ \ y_2$

12. $y_1'=4y_1+2y_2\quad y_2'=6y_1+3y_2$

13. $y_1'=7y_1-8y_2\quad y_2'=4y_1-5y_2$

14. $y_1'=7y_1-10y_2\quad y_2'=2y_1-\ \ 2y_2$

15. $y_1'=3y_1+y_2\quad y_2'=-2y_1+y_2$

16. $y_1'=y_1-y_2\quad y_2'=2y_1+3y_2$

17. $y_1'=7y_1+2y_2-8y_3\quad y_2'=-3y_1\ \ +3y_3\quad y_3'=6y_1+2y_2-7y_3$

18. $y_1'=\ \ y_2-y_3\quad y_2'=y_1\ \ +y_3\quad y_3'=y_1-y_2+2y_3$

Exercises 19–24: Find the solution for the system that satisfies the conditions at *t* = 0.

19. $y_1'=8y_1-10y_2,\ y_1(0)=4\quad y_2'=5y_1-7y_2,\ y_2(0)=1$

20. $y_1'=-4y_1+10y_2,\ y_1(0)=1\quad y_2'=-3y_1+7y_2,\ y_2(0)=1$

21. $y_1'=4y_1+3y_2,\ y_1(0)=2\quad y_2'=-3y_1+y_2,\ y_2(0)=-1$

22. $y_1'=2y_1+4y_2,\ y_1(0)=-1\quad y_2'=-2y_1-2y_2,\ y_2(0)=3$

23. $y_1'=2y_1-y_2-y_3,\ y_1(0)=-1\quad y_2'=6y_1+3y_2-2y_3,\ y_2(0)=0\quad y_3'=6y_1-2y_2-3y_3,\ y_3(0)=-4$

24. $y_1'=3y_1-2y_2,\ \ y_1(0)=1\quad y_2'=5y_1-3y_2,\ \ y_2(0)=-1\quad y_3'=y_1+\ y_2-y_3,\ \ y_3(0)=2$

Exercises 25–26: The given system of linear differential equations models the concentrations of insulin and glucose in an individual, as described earlier in this section. Find the general solution for the system.

25. $y_1'=-0.1y_1+0.2y_2\quad y_2'=-0.3y_1-0.6y_2$

26. $y_1'=-0.44y_1+0.12y_2\quad y_2'=-0.08y_1-0.16y_2$

27. For the system of differential equations given in Exercise 25, suppose that it is known that at time *t* = 0 the concentrations of insulin and glucose, respectively, are

$$y_1(0)=10, y_2(0)=20$$

Find a formula for $y_1(t)$ and $y_2(t)$.

28. For the system of differential equations given in Exercise 29, suppose that it is known that at time $t = 0$ the concentrations of insulin and glucose, respectively, are

$$y_1(0)=15, y_2(0)=50$$

Find a formula for $y_1(t)$ and $y_2(t)$.

Exercises 29–30: The given system of linear differential equations models a two-country arms race, as described in this section. Find the general solution for the system, and provide a brief interpretation of the results.

29. $y_1'=-3y_1+5y_2 \quad y_2'=-4y_1-4y_2$

30. $y_1'=y_1+y_2 \quad y_2'=4y_1+y_2$

31. For the system of differential equations given in Exercise 29, suppose that it is known that the initial quantity of arms in each country's arsenal is

$$y_1(0)=1, y_2(0)=2$$

Find a formula for $y_1(t)$ and $y_2(t)$.

32. For the system of differential equations given in Exercise 30, suppose that it is known that the initial quantity of arms in each country's arsenal is

$$y_1(0)=4, y_2(0)=1$$

Find a formula for $y_1(t)$ and $y_2(t)$.

FIND AN EXAMPLE Exercises 33–38: Find an example that meets the given specifications.

33. A system of two first-order linear differential equations that has general solution $y_1 = c_1e^{-3t}$ and $y_2 = c_2e^{2t}$.

34. A system of two first-order linear differential equations that has general solution $y_1 = c_1e^t$ and $y_2 = c_2e^{-2t}$.

35. A system of two first-order linear differential equations that has general solution $y_1 = 2c_1e^t - c_2e^{-2t}$ and $y_2 = -3c_1e^t + 2c_2e^{-2t}$. (HINT: Section 6.2 contains an example showing how to construct a matrix with specific eigenvalues and eigenvectors.)

36. A system of two first-order linear differential equations that has general solution $y_1 = 3c_1e^{-t} + 7c_2e^{4t}$ and $y_2 = c_1e^{-t} + c_2e^{4t}$. (HINT: Section 6.2 contains an example showing how to construct a matrix with specific eigenvalues and eigenvectors.)

37. A system of three first-order linear differential equations with general solution that is made up of linear combinations of e^t, e^{-2t}, and e^{5t}.

38. A system of two first-order linear differential equations with general solution that is made up of linear combinations of only trigonometric functions.

TRUE OR FALSE Exercises 39–40: Determine if the statement is true or false, and justify your answer.

39.

(a) The general solution to $y' = ky$ is $y = ce^{kt}$.

(b) Every system of linear differential equations $\mathbf{y}' = A\mathbf{y}$ can be solved using the methods presented in this section.

40.

(a) The solution to any system of linear differential equations always includes a real exponential function e^{ct} for some nonzero constant c.

(b) If $\mathbf{y}_a$ and $\mathbf{y}_b$ are solutions to the system of linear differential equations $\mathbf{y}' = A\mathbf{y}$, then so is a linear combination $c_a\mathbf{y}_a + c_b\mathbf{y}b$.

C Exercises 41–44: Find the general solution for the system.

41. y1′=2y1−3y2+6y3y2′=−3y1−5y2+7y3y3′=4y1+4y2+4y3

42. y1′=7y1+2y2+ y3y2′= y1+6y2− y3y3′=2y1+3y2+ 4y3

43. y1′=3_y1−y2 +5y4y2′=−2y1+3y2+7y3 y3′=4y1−3y2−y3−y4y4′ =5y1+2y3+y3+y4

44. $y_1' = -y_1 - 5y_2 + 4y_3 + y_4 \quad y_2' = y_1 + y_2 - 3y_3 - 2y_4 \quad y_3' = -3y_1 + y_2 - 6y_3 - -7y_4 \quad y_4' = 2y_1 - y_2 - y_3 - 5y_4$

Exercises 45–48: Find the solution for the system that satisfies the conditions at $t = 0$.

45. $y_1' = 3y_1 - y_2 + 4y_3,\ y_1(0) = -1 \quad y_2' = -2y_1 - 6y_2 + y_3,\ y_2(0) = -4 \quad y_3' = 4y_1 + 5y_2 + 5y_3,\ y_3(0) = 3$

46. $y_1' = -4y_1 + 7y_2 + 2y_3,\ y_1(0) = -1 \quad y_2' = 2y_1 + 4y_2 + 3y_3,\ y_2(0) = -5 \quad y_3' = 3y_1 - 2y_2 + 4y_3,\ y_3(0) = 2$

47. $y_1' = -2y_1 + 3y_2 - 4y_3 + 5y_4,\ y_1(0) = 7 \quad y_2' = -y_1 + 2y_2 - 3y_3 + 4y_1,\ y_2(0) = 2 \quad y_3' = 4y_1 - 3y_2 - 2y_3 + y_4,\ y_3(0) = -2 \quad y_4' = 5y_1 + 6y_2 + 7y_3 + 8y_4,\ y_4(0) = -5$

48. $y_1' = 7y_1 - 3y_2 + 5y_3 - y_4,\ y_1(0) = 2 \quad y_2' = 4y_1 + 2y_2 - 3y_3 + 6y_1,\ y_2(0) = -9 \quad y_3' = -y_1 + 4y_2 - 4y_4,\ y_3(0) = -4 \quad y_4' = 3y_2 - 4y_3 - 2y_4,\ y_4(0) = 3$

6.5 Approximation Methods

In Section 6.1 we saw how to use the characteristic polynomial to find the eigenvalues (and then the eigenvectors) for matrices. Such methods work fine for the small matrices considered there, but they are not practical for many larger matrices.

▶ This section is optional and can be omitted without loss of continuity.

Since large matrices turn up in all kinds of applications, we need another way to find eigenvalues and eigenvectors. Several related approaches for dealing with large matrices are described in this section. All have their basis in the Power Method, so we start with that.

The Power Method

The Power Method is an iterative algorithm that gets its name from how it is implemented. Given a square matrix A, we start with a fixed vector $\mathbf{x}_0$ and compute the sequence $A\mathbf{x}_0$, $A^2\mathbf{x}_0$, $A^3\mathbf{x}_0$, $\cdots$. Remarkably, in many cases the resulting sequence of vectors will
approach an eigenvector of A. We begin with an example and leave the discussion of when and why the Power Method works for later in this section.

Suppose that and let

A=[1322],x0=[10] (1)

and let

x1=Ax0,x2=Ax1=A2x0,x3=Ax2=A3x0,⋯

Table 1 gives $\mathbf{x}_0$ to $\mathbf{x}_7$ for A and $\mathbf{x}_0$ above.

Table 1 $\mathbf{x}_k = A^k\mathbf{x}_0$ **for** $k = 0, 1, \ldots, 7$

k	0	1	2	3	4	5	6	7
$\mathbf{x}_k$	[10]	[12]	[76]	[2526]	[103102]	[409410]	[16391638]	[65536554]

Often the components of $\mathbf{x}_k$ grow with k when forming this type of sequence. Since scalar multiples of eigenvectors are still eigenvectors, we control the size of each vector in the sequence by scaling.

DEFINITION 6.16 ▶

THE POWER METHOD: For each $k \geq 0$,

Scaling Factor

(a) Let s_k denote the largest component (in absolute value) of $A\mathbf{x}_k$. (We call s_k a **scaling factor**.)

(b) Set xk+1=1skAxk.

Repeat (a) and (b) to generate the sequence $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, ….

Scaling in this way ensures that the largest component (in absolute value) of each vector is either 1 or −1. For example, starting with $\mathbf{x}_0$ and A in (1), we compute $\mathbf{x}_1$ by

$$A\mathbf{x}_0=\begin{bmatrix}1&3\\2&2\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}\Rightarrow s_0=2\Rightarrow \mathbf{x}_1=\frac{1}{2}A\mathbf{x}_0=\begin{bmatrix}0.5\\1\end{bmatrix}$$

We then compute $\mathbf{x}_2$ by

$$A\mathbf{x}_1=\begin{bmatrix}1&3\\2&2\end{bmatrix}\begin{bmatrix}0.5\\1\end{bmatrix}=\begin{bmatrix}3.5\\2\end{bmatrix}\Rightarrow s_1=3.5\Rightarrow \mathbf{x}_2=\frac{1}{3.5}A\mathbf{x}_1=\begin{bmatrix}1\\0.8571\end{bmatrix}$$

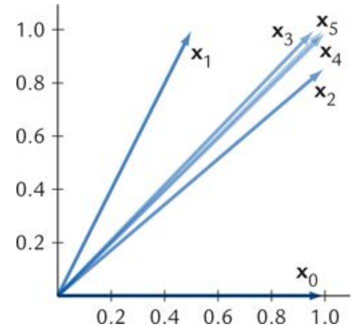


Figure 1 Vectors $\mathbf{x}_0, \ldots, \mathbf{x}_5$ from Table 2.

Table 2 shows $\mathbf{x}_0$ to $\mathbf{x}_7$, together with the scaling factors.

Table 2 **The Power Method Applied to A and $\mathbf{x}_0$ in (1)**

k	0	1	2	3	4	5	6	7
$\mathbf{x}_k$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0.5\\1\end{bmatrix}$	$\begin{bmatrix}1.000\\0.8571\end{bmatrix}$	$\begin{bmatrix}0.9615\\1.000\end{bmatrix}$	$\begin{bmatrix}1.000\\0.9903\end{bmatrix}$	$\begin{bmatrix}0.9976\\1.000\end{bmatrix}$	$\begin{bmatrix}1.000\\0.9994\end{bmatrix}$	$\begin{bmatrix}0.9998\\1.0\end{bmatrix}$
s_k	2	3.5	3.714	3.962	3.981	3.998	3.999	4.000

The entries in Table 2 and Figure 1 suggest that the sequence of vectors is getting closer and closer to $\mathbf{u}=\begin{bmatrix}1\\1\end{bmatrix}$. Computing $A\mathbf{u}$, we find that

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

Hence $\mathbf{u}=\begin{bmatrix}1\\1\end{bmatrix}$ is an eigenvector of A with associated eigenvalue of $\lambda = 4$.

In Table 2, not only does the sequence of vectors converge to an eigenvector, but the sequence of scaling factors converges to the associated eigenvalue. This latter observation makes sense, because our scaling rule $\mathbf{x}_{k+1}=\frac{1}{s_k}A\mathbf{x}_k$ can be expressed as

$$A\mathbf{x}_k=s_k\mathbf{x}_{k+1}$$

Since the vectors are getting closer to an eigenvector (and each other), it follows that s_k must be getting closer to an eigenvalue.

Typically, any nonzero $\mathbf{x}_0$ can be used as an initial vector for the Power Method. (However, see the note in the Computational Comments at the end of the section.) If an approximate value of an eigenvector is known, then using it for $\mathbf{x}_0$ can speed convergence.

Example 1

Find an eigenvector and associated eigenvalue for the matrix

$$A=\begin{bmatrix}245&-254&-252&-46&-224\\161&-168&-174&-32&-148\\-39&40&45&7&38\\27&-28&-32&-6&-26\\110&-113&-110&-21&-101\end{bmatrix}$$

▶ The matrix A is from (4) in Section 6.1. (3)

▶ To save space, in this section we sometimes write vectors horizontally instead of vertically.

Solution We apply the Power Method starting with $\mathbf{x}_0 = (1, 1, 1, 1, 1)$. Table 3 gives the vectors $\mathbf{x}_1$, … $\mathbf{x}_9$.

Table 3 **The Power Method Applied to A in Example 1**

k	$\mathbf{x}_k$	s_k
1	(1.0000, 0.6798, −0.1714, 0.1224, 0.4426)	10.74
2	(1.0000, 0.6693, −0.1717, 0.1124, 0.4431)	13.83
3	(1.0000, 0.6688, −0.1683, 0.1126, 0.4438)	12.94
4	(1.0000, 0.6674, −0.1676, 0.1116, 0.4442)	13.07
5	(1.0000, 0.6671, −0.1670, 0.1114, 0.4443)	13.01
6	(1.0000, 0.6668, −0.1668, 0.1112, 0.4444)	13.01
7	(1.0000, 0.6668, −0.1667, 0.1112, 0.4444)	13.00
8	(1.0000, 0.6667, −0.1667, 0.1111, 0.4444)	13.00
9	(1.0000, 0.6667, −0.1667, 0.1111, 0.4444)	13.00

The sequence settles to the vector (1.0000, 0.6667, −0.1111, 0.4444), and the sequence of scaling factors to 13.00. The components of the vector are recognizable as decimal approximations to rationals. Changing the decimals to equivalent rationals and multiplying by 18 to eliminate the fractions gives us

$$\mathbf{u} = \begin{bmatrix} 1.0000 \\ 0.6667 \\ -0.1667 \\ 0.1111 \\ 0.4444 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2/3 \\ -1/6 \\ 1/9 \\ 4/9 \end{bmatrix} \Rightarrow \mathbf{u} = \begin{bmatrix} 18 \\ 12 \\ -3 \\ 2 \\ 8 \end{bmatrix}$$

▶ We express the entries as fractions and multiply by 18 to make it easier to check our answer. This would not be done in applications. Instead, typically we would compute enough iterations to achieve a desired accuracy and take the last vector in the sequence as the eigenvector.

To test our answer, we calculate

$$A\mathbf{u} = \begin{bmatrix} 245 & -254 & -252 & -46 & -224 \\ 161 & -168 & -174 & -32 & -148 \\ -39 & 40 & 45 & 7 & 38 \\ 27 & -28 & -32 & -6 & -26 \\ 110 & -113 & -110 & -21 & -101 \end{bmatrix} \begin{bmatrix} 18 \\ 12 \\ -3 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} 234 \\ 156 \\ -39 \\ 26 \\ 104 \end{bmatrix} = 13 \begin{bmatrix} 18 \\ 12 \\ -3 \\ 2 \\ 8 \end{bmatrix}$$

confirming that $\mathbf{u}$ is an eigenvector and that $\lambda = 13$ is the associated eigenvalue.

The Power Method will frequently find an eigenvalue and associated eigenvector. But which ones? To find out, first suppose that a matrix A has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots |\lambda_n|$$

Dominant Eigenvalue

In this case λ_1 is called the **dominant eigenvalue** of A.

The next theorem tells when the Power Method will converge to an eigenvector associated with the dominant eigenvalue. The proof of this theorem is given at the end of the section.

THEOREM 6.17 ▶

Let A be an $n \times n$ matrix with linearly independent eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ and associated eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, where λ_1 is dominant. Suppose that

$$x_0=c_1u_1+\cdots+c_nu_n$$

where $c_1 \neq 0$. Then multiples of $A\mathbf{x}_0, A^2\mathbf{x}_0, \cdots$ converge to a scalar multiple of the eigenvector $\mathbf{u}_1$.

The Shifted Power Method

The Power Method can be used to find eigenvectors other than those associated with the dominant eigenvalue. Suppose that λ_1 is the dominant eigenvalue of A, and then let λ be the dominant eigenvalue of $A - \lambda_1 I_n$ with associated eigenvector $\mathbf{u}_2$. Then

$$(A-\lambda_1 I_n)u_2=\lambda u_2 \Rightarrow Au_2=(\lambda+\lambda_1)u_2$$

Shifted Power Method

Thus, if $\lambda_2 = \lambda + \lambda_1$, then λ_2 is an eigenvalue of A. Furthermore, since λ is the dominant eigenvalue of A − and $\lambda_1\ I_n$ and $\lambda = \lambda_2 - \lambda_1$, it follows that λ_2 is the eigenvalue of A that is farthest from λ_1. (Why?) Hence applying the Power Method to $A - \lambda_1 I_n$ will produce another eigenvalue and eigenvector of A. This is called the **Shifted Power Method** and is illustrated in the next example.

Example 2

For the matrix A given in Example 1, find the eigenvalue (and an associated eigenvector) farthest from $\lambda_1 = 13$.

Solution To find λ_2, the eigenvalue of A farthest from $\lambda_1 = 13$, we first apply the Power Method to the matrix

B=A−13I5=
[232−254−252−46−224161−181−174−32−148−39403273827−28−32−19−26110−113−110−21−114]

As in Example 1, we start with $\mathbf{x}_0 = (1, 1, 1, 1, 1)$. In this case, the convergence is slower, so we report only every 10th iteration in Table 4.

Table 4 **The Shifted Power Method Applied to $B = A - 13I_5$**

k	$\mathbf{x}_k$	s_k
10	(1.0000, 0.7033, −0.1483, 0.1483, 0.4450)	−15.77
20	(1.0000, 0.7078, −0.1461, 0.1461, 0.4383)	−15.86
30	(1.0000, 0.7107, −0.1447, 0.1447, 0.4340)	−15.92
40	(1.0000, 0.7123, −0.1438, 0.1438, 0.4315)	−15.96
50	(1.0000, 0.7132, −0.1434, 0.1434, 0.4302)	−15.98
60	(1.0000, 0.7137, −0.1431, 0.1431, 0.4294)	−15.99
70	(1.0000, 0.7140, −0.1430, 0.1430, 0.4290)	−15.99
80	(1.0000, 0.7141, −0.1429, 0.1429, 0.4288)	−16.00
90	(1.0000, 0.7142, −0.1429, 0.1429, 0.4287)	−16.00
100	(1.0000, 0.7142, −0.1429, 0.1429, 0.4286)	−16.00

▶ Carrying more decimal places in \Example 2 gives (after 250 iterations)

$$u=[1.000000000.71428571−0.142857140.142857140.42857143]$$

Then

$$Au=[−3.00000000−2.142857140.42857143−0.42857143−1.28571428]=−3[1.000000000.71428571−0.142857140.142857140.42857143]=−3u$$

We see that $s_k \to -16$, which implies that $\lambda = -16$ is an eigenvalue for B. Therefore $\lambda_2 = -16 + 13 = -3$ is the eigenvalue of A farthest from $\lambda_1 = 13$. Table 4 shows $\mathbf{u} = (1.000, 0.7142, -0.1429, 0.1429, 0.4286)$ is the associated eigenvector. We check this by computing

$$Au=[245−254−252−46−224161−168−174−32−148−39404573827−28−32−6−26110−113−110−21−101] [1.00000.7142−0.14290.14290.4286]≈[−2.9758−2.12660.4246−0.4258−1.2751]≈−3[0.99190.7089−0.14150.14190.4250]≈−3u$$

The approximations are a bit rough, but they can be refined by carrying more decimal places and computing additional iterations (see margin).

The Inverse Power Method

For the matrix A in Examples 1 and 2, we have found eigenvalues $\lambda_1 = 13$ and $\lambda_2 = -3$, and we also know that these are the positive and negative eigenvalues farthest from 0. Here we develop a method for finding the eigenvalue λ_3 that is *closest* to 0. Exercise 57 in Section 6.1 showed that if A is an invertible matrix with eigenvalue λ and associated eigenvector $\mathbf{u}$, then λ^{-1} is an eigenvalue of A^{-1} with associated eigenvector $\mathbf{u}$. Thus, in particular, if λ is the *largest* (in absolute value) eigenvalue of A^{-1}, then λ^{-1} must be the *smallest* (in absolute value) eigenvalue of A. (Why?) Therefore, applying the Power Method to A^{-1} will yield the smallest eigenvalue (and corresponding eigenvector) of A.

Example 3

For the matrix A in Example 1, find the smallest eigenvalue (in absolute value) and an associated eigenvector.

Solution Here we apply the Power Method to

$$A−1=178[−646698−420−1430620−461487−306−98844899−11270221−86−2740−64−14314−290329−178−663264]$$

generating a sequence of vectors of the form

$$xk+1=1skA−1xk \qquad (4)$$

▶ For large matrices it can be difficult to compute A^{-1} accurately due to round-off error. This can be avoided by multiplying each side of (4) by A, yielding the sequence of linear systems

$$Axk+1=1skxk$$

Applying LU factorization (or a related method) to A can greatly improve computational efficiency when solving these systems.

For a change of pace (and another reason to be discussed later), let's take the initial vector to be $\mathbf{x}_0 = (1, 2, 3, 4, 5)$. Table 5 gives the results of every other iteration.

Table 5 **The Inverse Power Method Applied to A**

k	$\mathbf{x}_k$	s_k
2	(1.0000, 0.6842, −0.1565, 0.1178, 0.4734)	0.2866
4	(1.0000, 0.6887, −0.1556, 0.0921, 0.4669)	0.7421
6	(1.0000, 0.6912, −0.1544, 0.0811, 0.4632)	0.9297
8	(1.0000, 0.6920, −0.1540, 0.0780, 0.4620)	0.9823
10	(1.0000, 0.6922, −0.1539, 0.0772, 0.4616)	0.9956
12	(1.0000, 0.6923, −0.1539, 0.0770, 0.4616)	0.9989
14	(1.0000, 0.6923, −0.1538, 0.0769, 0.4615)	0.9997
16	(1.0000, 0.6923, −0.1538, 0.0769, 0.4615)	0.9999

The output suggests that A^{-1} has eigenvalue $\lambda = 1$, so that $\lambda_3 = \lambda^{-1} = 1$ is an eigenvalue of A. We check the vector $\mathbf{u}$ = (1.0000, 0.6923, −0.1538, 0.0769, 0.4615) from Table 5 by computing

Au = [245−254−252−46−224161−168−174−32−148−39404573827−28−32−6−26110−113−110−21−10 1] [1.00000.6923−0.15380.07690.4615]≈[1.00000.6920−0.15370.07680.4617]≈ u

Thus $\mathbf{u}$ is an eigenvector associated with eigenvalue $\lambda_3 = 1$.

The Shifted Inverse Power Method

So far we have found three eigenvalues (and associated eigenvectors) for A,

- The largest (in absolute value) λ_1.
- The eigenvalue λ_2 that is farthest from λ_1.
- The eigenvalue λ_3 that is closest to the origin.

Another way to look for eigenvalues is to start with the matrix $B = A - cI_n$ for some scalar c. Applying the Inverse Power Method to B will find the eigenvalue λ of B that is closest to the origin. Since $B = A - cI_n$, $\lambda + c$ is the eigenvalue of A that is closest to c.

Example 4

For the matrix A given in Example 1, find the eigenvalue that is closest to $c = 4$.

Solution We start by setting

B=A−4I5=
[241−254−252−46−224161−172−174−32−148−39404173827−28−32−10−26110−113−110−21−105]

Now we apply the Inverse Power Method to B, starting out with the vector $\mathbf{x}_0$ = (5, 4, 3, 2, 1). Table 6 includes every fifth iteration.

Table 6 **The Shifted Inverse Power Method (Example 4)**

k	$\mathbf{x}_k$	s_k
5	(1.0000, 0.5363, −0.0292, 0.0141, 0.4931)	0.9416
10	(1.0000, 0.4936, 0.0051, −0.0026, 0.5013)	0.4616
15	(1.0000, 0.5008, −0.0006, 0.0003, 0.4998)	0.5053
20	(1.0000, 0.4999, 0.0001, 0.0000, 0.5000)	0.4993

k	$\mathbf{x}_k$	s_k
25	(1.0000, 0.5000, 0.0000, 0.0000, 0.5000)	0.5001
30	(1.0000, 0.5000, 0.0000, 0.0000, 0.5000)	0.5000
35	(1.0000, 0.5000, 0.0000, 0.0000, 0.5000)	0.5000

We can see that $\lambda = 0.5$ is an eigenvalue for B^{-1}, so that $\lambda = 2$ is an eigenvalue for B. By shifting back, we find that $\lambda_4 = 2 + 4 = 6$ is an eigenvalue for A, with associated eigenvector $\mathbf{u} = (1, 0.5, 0, 0, 0.5)$. We check this by computing

Au=
[245−254−252−46−224161−168−174−32−148−39404573827−28−32−6−26110−113−110−21−101] [1.00.5000.5]=[63003]=6u

Applying the Shifted Inverse Power Method with different choices of c can yield other eigenvalues and eigenvectors. It is typically not efficient to do this for random values of c, but in some applications rough estimates of the eigenvalues are known. In these instances, the Shifted Inverse Power Method can turn estimates into accurate approximations.

COMPUTATIONAL COMMENTS

Some general remarks about the Power Method and related techniques discussed in this section:

- Similar to the approximation methods for finding solutions to linear systems given in Section 1.4, the Power Method is not overly sensitive to round-off error. In fact, each successive vector can be viewed as a starting point for the algorithm, so even if an error occurs, it typically will be corrected.
- The rate of convergence has differed in the examples we have considered. With the Power Method and its relatives, the larger the dominant eigenvalue is relative to the other eigenvalues, the faster the rate of convergence.
- The Power Method is guaranteed to work only on matrices whose eigenvectors span $\mathbf{R}^n$. However, in practice this method often also will work on other matrices, although convergence may be slower.
- Earlier we stated that the choice of starting vector $\mathbf{x}_0$ does not matter, and in virtually all cases this will be true. However it is possible to get misleading results from an unlucky choice of $\mathbf{x}_0$. For instance, in Example 3 we used $\mathbf{x}_0 = (1, 2, 3, 4, 5)$ in place of (1, 1, 1, 1, 1), claiming at the time that this was done for “a change of pace.” However, there was another reason. Table 7 gives the results when starting with $\mathbf{x}_0 = (1, 1, 1, 1, 1)$.

Table 7 New Application of the Inverse Power Method

k	s_k	$\mathbf{x}_k$
5	−0.4559	(1.0000, 0.6758, −0.1624, 0.1624, 0.4866)
10	−0.4928	(1.0000, 0.6681, −0.1660, 0.1660, 0.4979)
15	−0.4990	(1.0000, 0.6669, −0.1666, 0.1666, 0.4997)
20	−0.4999	(1.0000, 0.6667, −0.1667, 0.1667, 0.5000)
25	−0.5000	(1.0000, 0.6667, −0.1667, 0.1667, 0.5000)
30	−0.5000	(1.0000, 0.6667, −0.1667, 0.1667, 0.5000)

This gives us $\lambda = -2$, which is an eigenvalue of A but not the eigenvalue closest to 0. We did not find $\lambda = 1$ because this choice of $\mathbf{x}_0$ happens to be in the span of the eigenvectors *not* associated with $\lambda = 1$. If we look at the statement of Theorem 6.17 again, we see that one of the conditions is violated.

In most applications, the entries of a matrix A are decimals and are subject to some degree of rounding. The likelihood of such an unlucky choice of $\mathbf{x}_0$ happening in practice is small.

- In cases where the eigenvalues satisfy

$$|\lambda_1| = |\lambda_2| = \cdots = |\lambda_k| > |\lambda_{k+1}| \geq \cdots \geq |\lambda_n|$$

and $\lambda_1 = \lambda_2 = \cdots = \lambda_k$, the Power Method will still work fine. However, if (for example) instead $\lambda_1 = -\lambda_2$, then the Power Method can produce strange results (see Exercises 39–40).

Proof of Theorem 6.17

Proof To understand why the Power Method works, suppose that we have an $n \times n$ matrix A that has eigenvalues $\lambda_1, \cdots, \lambda_n$, such that

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n| \qquad (5)$$

Assume that the associated eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ form a basis for $\mathbf{R}^n$. If $\mathbf{x}_0$ is an arbitrary vector, then there exist scalars $c_1, \ldots, c_n$ such that

$$x_0 = c_1u_1 + c_2u_2 + g + c_nu_n$$

Since $A^ku_j = \lambda_j^k u_j$ for each eigenvector $\mathbf{u}_j$ and each positive integer k (see Exercise 60 in Section 6.1), if we form the product $A^k \mathbf{x}_0$, we get

$$A^kx_0 = A^k(c_1u_1 + c_2u_2 + \cdots + c_nu_n) = c_1\lambda_1^ku_1 + c_2\lambda_2^ku_2 + \cdots + c_n\lambda_n^ku_n$$

Next, divide both sides by λ_1^k. (This is similar to our scaling in each step of the Power Method.) This gives us

$$\left(\frac{1}{\lambda_1^k}\right)A^kx_0 = c_1x_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k u_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k u_n$$

By (5), as k gets large, each of $\left(\frac{\lambda_2}{\lambda_1}\right)^k, \cdots \left(\frac{\lambda_n}{\lambda_1}\right)^k$ gets smaller. Hence as $k \to \infty$,

$$\left(\frac{1}{\lambda_1^k}\right)A^kx_0 \to c_1u_1$$

as claimed and observed in our examples. ■ ■

Note that the larger $|\lambda_1|$ is relative to $|\lambda_2|, \cdots, |\lambda_n|$, the faster $\left(\frac{\lambda_2}{\lambda_1}\right)^k, \cdots, \left(\frac{\lambda_n}{\lambda_1}\right)^k$ converge to 0. This is why the Power Method converges more rapidly when the dominant eigenvalue is much larger than the other eigenvalues.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Compute the first two iterations of the Power Method without scaling, starting with the given $\mathbf{x}_0$.

(a) $A = \begin{bmatrix} 3 & -1 \\ 1 & 2 \end{bmatrix}, x_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} 2 & 0 & 3 \\ 1 & 1 & 4 \\ 0 & -1 & -7 \end{bmatrix}, x_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

2. Compute the first two iterations of the Power Method with scaling, starting with the given $\mathbf{x}_0$. Round any numerical values to two decimal places.
   (a) A=[11−21],x0=[01]
   (b) A=[1−1−2203−121],x0=[001]

EXERCISES

Exercises 1–6: Compute the first three iterations of the Power Method without scaling, starting with the given $\mathbf{x}_0$.

1. A=[1−315],x0[10]
2. A=[−243−1],x0[01]
3. A=[6−374154−39],x0=[100]
4. A=[−25−7−211−13−25−7],x0=[110]
5. A=[5−122222−15],x0=[10−1]
6. A=[−12−7−1022−1022],x0=[001]

Exercises 7–12: Compute the first two iterations of the Power Method with scaling, starting with the given $\mathbf{x}_0$. Round any numerical values to two decimal places.

7. A=[2−101],x0[01]
8. A=[3120],x0=[−11]
9. A=[−1021100−21],x0=[010]
10. A=[−21103−2200],x0=[−110]
11. A=[100−1302−11],x0=[01−1]
12. A=[02−1021200],x0=[−111]

Exercises 13–18: The eigenvalues of a 3 × 3 matrix *A* are given. Determine if it is assured that the Power Method will converge to an eigenvector and eigenvalue, and if so, identify the eigenvalue.

13. λ1=5,λ2=−2,λ3=7
14. λ1=3,λ2=−4,λ3=0
15. λ1=−6,λ2=2,λ3=2
16. λ1=5,λ2=5,λ3=4
17. λ1=−4,λ2=4,λ3=6
18. λ1=3,λ2=−3,λ3=2

Exercises 19–22: The given *λ* is the dominant eigenvalue for *A*. To which matrix *B* would you apply the Power Method in order to find the eigenvalue that is farthest from *λ*?

19. A=[1232],λ=4
20. A=[−1023],λ=3
21. A=[−12−7−1022−1022],λ=9
22. A=[5−122222−15],λ=6

Exercises 23–26: To which matrix *B* would you apply the Inverse Power Method in order to find the eigenvalue that is closest to *c*?

23. A=[−3152],c=4
24. A=[1234],c=−5
25. A=[314159261],c=−1
26. A=[271828182],c=3

27. Below is the output resulting from applying the Inverse Power Method to a matrix A. Identify the eigenvalue and eigenvector.

k	$\mathbf{x}_k$	s_k
5	(1.0000, 0.5363, −0.0292)	0.4415
10	(1.0000, 0.4837, −0.0026)	0.3623
15	(1.0000, 0.5091, −0.0006)	0.2503
20	(1.0000, 0.4997, −0.0001)	0.2501
25	(1.0000, 0.5000, 0.0000)	0.2500

28. Below is the output resulting from applying the Shifted Inverse Power Method to a matrix A with c = 3. Identify the eigenvalue and eigenvector.

k	$\mathbf{x}_k$	s_k
2	(0.3577, 0.0971, 1.0000)	0.5102
4	(0.4697, 0.1021, 1.0000)	0.5063
6	(0.4925, 0.1007, 1.0000)	0.5021
8	(0.4997, 0.1002, 1.0000)	0.5003
10	(0.5000, 0.1000, 1.0000)	0.5000

FIND AN EXAMPLE Exercises 29–34: Find an example that meets the given specifications.

29. A 2 × 2 matrix A and an initial vector $\mathbf{x}_0$ such that the Power Method converges immediately. That is, $\mathbf{x}_0 = \mathbf{x}_1 = \cdots$.

30. A 3 × 3 matrix A and an initial vector $\mathbf{x}_0$ such that the Power Method converges immediately. That is, $\mathbf{x}_0 = \mathbf{x}_1 = \cdots$.

31. A 2 × 2 matrix A and an initial vector $\mathbf{x}_0$ such that the Power Method alternates between two different vectors. Thus $\mathbf{x}_0 = \mathbf{x}_2 =$ and $\mathbf{x}_1 = \mathbf{x}_3 = \cdots$, but $\mathbf{x}_0 \neq \mathbf{x}_1$.

32. A 3 × 3 matrix A and an initial vector $\mathbf{x}_0$ such that the Power Method alternates between two different vectors. Thus $\mathbf{x}_0 = \mathbf{x}_2 = \cdots$ and $\mathbf{x}_1 = \mathbf{x}_3 = \cdots$, but $\mathbf{x}_0 \neq \mathbf{x}_1$

33. A 2 × 2 matrix A and an initial vector $\mathbf{x}_0$ such that the Power Method without scaling alternates between three different vectors. Thus $\mathbf{x}_0 = \mathbf{x}_3 = \cdots$, $\mathbf{x}_1 = \mathbf{x}_4 = \cdots$, and $\mathbf{x}_2 = \mathbf{x}_5 = \cdots$, with $\mathbf{x}_0$, $\mathbf{x}_1$, and $\mathbf{x}_2$ distinct.

34. A 3 × 3 matrix A and an initial vector $\mathbf{x}_0$ such that the Power Method alternates between three different vectors. Thus $\mathbf{x}_0 = \mathbf{x}_3 = \cdots$, $\mathbf{x}_1 = \mathbf{x}_4 = \cdots$, and $\mathbf{x}_2 = \mathbf{x}_5 = \cdots$, with $\mathbf{x}_0$, $\mathbf{x}_1$, and $\mathbf{x}_2$ distinct.

TRUE OR FALSE Exercises 35–38: Determine if the statement is true or false, and justify your answer.

35.

(a) The Power Method requires a dominant eigenvalue to converge.

(b) The Power Method and Inverse Power Method cannot be applied to the same matrix.

36.

(a) If a square matrix A has a dominant eigenvalue, then the Power Method will converge.

(b) The Power Method is generally sensitive to round-off error.

37.

(a) The Inverse Power Method can only be applied to invertible matrices.

(b) If the Power Method converges, then it will converge to the same eigenvector for any initial vector $\mathbf{x}_0$.

38.

(a) Typically, the closer an initial vector $\mathbf{x}_0$ is to a dominant eigenvector $\mathbf{u}$, the faster the Power Method will converge.

(b) If $\lambda_1 = \lambda_2$ are the two largest eigenvalues of a matrix A, then the Power Method will not converge.

39.

(a) For the matrix A and vector $\mathbf{x}_0$, compute the first four iterations of the Power Method, and then explain the behavior that you observe.

A=[110−1],x0[01]

40. For the matrix A and vector $\mathbf{x}_0$, compute the first four iterations of the Power Method with scaling, and then explain the behavior that you observe.

A=[−1012],x0[10]

41. For the matrix A, $\lambda = 2$ is the largest eigenvalue. Use the given value of $\mathbf{x}_0$ to generate enough iterations of the Power Method to estimate an eigenvalue. Explain the results that you get.

A=[1022],x0=[−12]

42. For the matrix A, $\lambda = 3$ is the largest eigenvalue. Use the given value of $\mathbf{x}_0$ to generate enough iterations of the Power Method to estimate an eigenvalue. Explain the results that you get.

A=[1221],x0=[−11]

Exercises 43–48: Compute the first six iterations of the Power Method without scaling, starting with the given $\mathbf{x}_0$.

43. A=[1−315],x0=[10]

44. A=[−243−1],x0=[01]

45. A=[6−374154−19],x0=[100]

46. A=[−25−7−211−13−25−7],x0=[110]

47. A=[5−122222−15],x0=[10−1]

48. A=[−12−7−1022−1022],x0=[001]

Exercises 49–54: Compute as many iterations of the Power Method with scaling as are needed to estimate an eigenvalue and eigenvector for A, starting with the given $\mathbf{x}_0$.

49. A=[2−101],x0=[01]

50. A=[3120],x0=[−11]

51. A=[−11313−22−21],x0=[010]

52. A=[−21103−2200],x0=[−110]

53. A=[100−1302−11],x0=[01−1]

54. A=[02−1021200],x0=[−111]

SUPPLEMENTARY EXERCISES

Exercises 1–4: Find a basis for the eigenspace of A associated with λ.

1. $A=[-22-65],\lambda=2$

2. $A=[10-618-11],\lambda=1$

3. $A=[-5-106611-658-4],\lambda=-1$

4. $A=[2-11-48-10-47-9],\lambda=-2$

Exercises 5–8: Find the eigenvalues and associated eigenvectors for A.

5. $A=[01-23]$

6. $A=[-56-34]$

7. $A=[1-22-1-11-100]$

8. $A=[0-10010-1-11]$

Exercises 9–12: Compute the first three iterations of the Power Method without scaling, starting with $\mathbf{x}_0$.

9. $A=[1211],x0=[10]$

10. $A=[1512],x0=[01]$

11. $A=[102-12-100-1],x0=[100]$

12. $A=[10021-100-1],x0=[110]$

Exercises 13–16: Compute as many iterations as required of the Power Method with scaling to estimate an eigenvalue and eigenvector for A, starting with $\mathbf{x}_0$.

13. $A=[11-128-9],x0=[10]$

14. $A=[-1630-917],x0=[01]$

15. A=[3−22−23−2−22−1],x0=[100]

16. A=[02−2−16−7−14−5],x0=[010]

Exercises 17–20: Diagonalize *A* if possible.

17. A=[−410−37]

18. A=[−52−125]

19. A=[−584−232−352]

20. A=[−522−201−843]

Exercises 21–22: Find a matrix *A* with the given eigenvalues and corresponding eigenvectors.

21. λ1=2⇒{[13]},λ2=−1⇒{[2−1]}

22. λ1=0⇒{[210]},λ2=−2⇒{[1−11],[0−12]}

Exercises 23–24: Compute the given expressions for $z_1 = 3 + 2i$ and $z_2 = 4 - i$.

23.

- (a) $z_1 + z_2$
- (b) $3z_1 + 2z_2$
- (c) $z_2 - 2z_1$
- (d) $z_1 z_2$

24.

- (a) $z_1 + 2z_2$
- (b) $z_1 - 2z_2$
- (c) $z_2 + 4z_1$
- (d) z12

Exercises 25–26: Find the eigenvalues and a basis for each eigenspace for *A*.

25. A=[1−141]

26. A=[31−13]

Exercises 27–28: Determine the rotation and dilation for *A*.

27. $A=[21-12]$

28. $A=[14-41]$

Exercises 29–32: A system of linear differential equations $\mathbf{y}'=A\mathbf{y}$ has the given eigenvalues and eigenspace bases. Find the general solution for the system.

29. $\lambda1=-1\Rightarrow\{[1-1]\},\lambda2=2\Rightarrow\{[-31]\}$

30. $\lambda1=0\Rightarrow\{[-23-1]\},\lambda2=-1\Rightarrow\{[103],[-13-2]\}$

31. $\lambda1=1+i\Rightarrow\{[2i-i]\},\lambda2=1-i\Rightarrow\{[-2ii]\}$

32. $\lambda1=1\Rightarrow\{[02-1]\},\lambda2=1+2i\Rightarrow\{[1+i-i3+2i]\}\lambda3=1-2i\Rightarrow\{[1-ii3-2i]\}$

Exercises 33–36: Find the general solution to the system of linear differential equations.

33. $y1'=-5y1+6y2y2'=-3y1+5y2$

34. $y1'=-4y1+10y2y2'=-3y1+\ \ 7y2$

35. $y1'=-5y1+8y2+4y3y2'=-2y1+3y2+2y3y3'=-3y1+5y2+2y3$

36. $y1'=-5y1+2y2+2y3y2'=-2y1\ \ +y3y3'=-8y1+4y2+3y3$

CHAPTER 7

Vector Spaces

Stephen Saks/Lonely Planet Images/Getty Images

The Annapolis Tidal Station is the only tidal power plant in North America, and one of only a few tidal power plants in the entire world. It is located in Nova Scotia, Canada, by the Bay of Fundy, which is known for having the largest tidal range in the world. This geographic location is advantageous since these power plants harness the power of the tides to generate electricity. Tidal power has the benefit of being more predictable than wind or solar power, but is sometimes criticized for effects on marine life.

Over the first six chapters we have focused our attention on understanding the structure of vectors in Euclidean space. In this chapter we adapt these results to a more general setting, where we adopt a broader notion of vectors and the spaces containing them.

In Section 7.1 we generalize the definition of *vector* and define a *vector space*. Section 7.2 describes how the concepts of span and linear independence carry over from Euclidean space to vector spaces, and in Section 7.3 we revisit the topics of basis and dimension.

This chapter is relatively brief because all of the material has analogs among the concepts that we developed for Euclidean space $\mathbf{R}^n$. References to comparable earlier definitions and theorems are provided to reinforce connections. As you read this chapter, think about how the concepts presented match up with those from Euclidean space.

▶ The order of Chapter 7 and Chapter 8 can be reversed if preferred.

7.1 Vector Spaces and Subspaces

In this section we describe how to expand the concept of Euclidean space and subspaces from $\mathbf{R}^n$ to a more general setting. To get us started, let $\mathbf{P}^2$ be the set of all polynomials with real coefficients that have degree 2 or less. A typical element of $\mathbf{P}^2$ has the form p(x)=a2x2+a1x+a0, where a0, a1, and a2 are real numbers. Let's compare $\mathbf{R}^3$ and $\mathbf{P}^2$.

- If p(x)=a2x2+a1x+a0 and q(x)=b2x2+b1x+b0 are two polynomials in $\mathbf{P}^2$, then the sum is

  p(x)+q(x)=(a2+b2)x2+(a1+b1)x+(a0+b0)

  When adding polynomials, we add together the coefficients of like terms. This is similar to the componentwise addition of elements in $\mathbf{R}^3$.
- If c is a real number and p(x) is as above, then

  cp(x)=(ca2)x2+(ca1)x+(ca0)

  Scalar multiplication of polynomials distributes across terms, similar to how scalar multiples distribute across components in $\mathbf{R}^3$.
- Just as $\mathbf{R}^3$ is closed under addition and scalar multiplication, so is $\mathbf{P}^2$. The sum of two polynomials in $\mathbf{P}^2$ has degree no greater than 2, as is the degree of the scalar multiple of a polynomial in $\mathbf{P}^2$. (Note that these operations might *decrease* the degree but cannot *increase* it.)
- The zero polynomial z(x)=0 satisfies p(x)+z(x)=p(x) for every polynomial in $\mathbf{P}^2$, so that z(x) plays the same role as 0=[000] does in $\mathbf{R}^3$.
- For every polynomial p(x) in $\mathbf{P}^2$, there is another polynomial

$$-p(x) = -a_2x^2 - a_1x - a_0$$

such that $p(x)+(-p(x))=0=z(x)$. This is also true in $\mathbf{R}^3$, where for each $[a\ b\ c]$ there exists $[-a\ -b\ -c]$ such that $[a\ b\ c]+[-a\ -b\ -c]=[0\ 0\ 0]=\mathbf{0}$.

- The definitions of addition and scalar multiplication on $\mathbf{P}^2$ satisfy the same distributive and associative laws as those of $\mathbf{R}^3$ given in Theorem 2.3 in Section 2.1.

Although they look different, $\mathbf{R}^3$ and $\mathbf{P}^2$ have many similar features. Theorem 2.3 in Section 2.1 lists the algebraic properties of elements in $\mathbf{R}^n$. The preceding discussion shows that the polynomials in $\mathbf{P}^2$ have similar properties. Other sets of mathematical objects, such as matrices and continuous functions, also possess these properties. Theorem 2.3 serves as a guide for our broader definition of a vector space.

DEFINITION 7.1 ▶

Vector Space, Vector

A **vector space** consists of a set V of **vectors** together with operations of addition and scalar multiplication on the vectors that satisfy each of the following:

(1) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in V, then so is $\mathbf{v}_1+\mathbf{v}_2$. (V is closed under addition.)

(2) If c is a real scalar and $\mathbf{v}$ is in V, then so is $c\mathbf{v}$. (V is closed under scalar multiplication.)

(3) There exists a **zero vector 0** in V such that $\mathbf{0}+\mathbf{v}=\mathbf{v}$ for all $\mathbf{v}$ in V.

(4) For each $\mathbf{v}$ in V there exists an **additive inverse** (or **opposite**) vector $-\mathbf{v}$ in V such that $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$ for all $\mathbf{v}$ in V.

(5) For all $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ in V and real scalars c_1 and c_2, we have

   (a) $\mathbf{v}_1+\mathbf{v}_2=\mathbf{v}_2+\mathbf{v}_1$

(b) $(v_1+v_2)+v_3=v_1+(v_2+v_3)$
(c) $c_1(v_1+v_2)=c_1v_1+c_1v_2$
(d) $(c_1+c_2)v_1=c_1v_1+c_2v_1$
(e) $(c_1c_2)v_1=c_1(c_2v_1)$
(f) $1\cdot v_1=v_1$

Definition 7.1 is technically the definition of a *real* vector space, and we could replace the real scalars with complex numbers with minimal changes. But for now real scalars will suffice, so from here on we assume that "vector space" refers to a real vector space unless stated otherwise.

Note that Euclidean space $\mathbf{R}^n$ is a vector space. $\mathbf{P}^2$, together with the usual operations of addition and scalar multiplication, also forms a vector space.

Three important points:

- To describe a vector space, we need to specify both the set of vectors and the arithmetic operations (addition and scalar multiplication) that are performed on them. The set alone is not enough.
- **Vectors are not always columns of numbers!** (For instance, in $\mathbf{P}^2$ the polynomials are the vectors.) This takes getting used to but is crucial, so say it to yourself every night as you fall asleep until it sinks in. It is fine to think of a column of numbers as an *example* of a vector, as long as you do not assume that a vector is always a column of numbers.
- The phrase "vector space" refers to any set satisfying Definition 7.1. The phrase "Euclidean space" will be used for the specific vector space $\mathbf{R}^n$ with the standard definition of addition and scalar multiplication.

Example 1

Let $V=\mathbb{R}^{2\times 2}$ denote the set of real 2×2 matrices, together with the usual definition of matrix addition and multiplication by a constant scalar. Show that $\mathbb{R}^{2\times 2}$ is a vector space.

Solution We have a clearly defined set of vectors (the real 2×2 matrices) and definitions for addition and scalar multiplication. It remains to verify that the five conditions of Definition 7.1 hold:

(1) If A and B are real 2 × 2 matrices, then we already know that $A + B$ is also a real 2 × 2 matrix. Hence $\mathbf{R}^{2\times 2}$ is closed under addition.

(2) If c is a real scalar and A is in $\mathbf{R}^{2\times 2}$, then cA is also a real 2 × 2 matrix. Thus $\mathbf{R}^{2\times 2}$ is closed under scalar multiplication.

(3) If 022=[0000] and A=[a11a12a21a22], then

022+A=[0000]+[a11a12a21a22]=[(0+a11)(0+a12)(0+a21)(0+a22)]=
[a11a12a21a22]=A

Hence 022+A=A for all real 2×2 matrices, so that 022 is the zero vector in $\mathbf{R}^{2\times 2}$.

(4) If A=[a11a12a21a22], then −A=[−a11−a12−a21−a22] satisfies

A+(−A)=[a11a12a21a22]+[−a11−a12−a21−a22]=[0000]=022

so that each vector in $\mathbf{R}^{2\times 2}$ has an additive inverse.

(5) The six conditions (a)–(f) all follow directly from properties of the real numbers. Verification is left as an exercise.

Since all the required conditions hold, the set R2×2 together with the given arithmetic operations form a vector space.

Example 2

Let $\mathbf{Q}^2$ denote the set of polynomials with real coefficients that have degree equal to 2, together with the usual definition of addition and scalar multiplication for polynomials. Is $\mathbf{Q}^2$ a vector space?

Solution The set $\mathbf{Q}^2$ satisfies some of the conditions of Definition 7.1, but it falls short on others. For instance, if q1(x)=x2 and q2(x)=5−x2, then q_1 and q_2 are both in $\mathbf{Q}^2$, but (q1+q2)(x)=q1(x)+q2(x)=5 is not, so $\mathbf{Q}^2$ is not closed under addition. Hence $\mathbf{Q}^2$ is not a vector space.

▶ $\mathbf{Q}^2$ consists of polynomials of degree 2, whereas $\mathbf{P}^2$ consists of the polynomials of degree 2 or *less*.

Before moving on to the next example, we pause to report several properties of vector spaces that are consequences of Definition 7.1.

THEOREM 7.2 ▶

Let V be a vector space and suppose that $\mathbf{v}$ is in V. Then:

(a) If $\mathbf{0}$ is a zero vector of V, then $\mathbf{v} + \mathbf{0} = \mathbf{v}$.

(b) If $-\mathbf{v}$ is an additive inverse of $\mathbf{v}$, then $-\mathbf{v} + \mathbf{v} = \mathbf{0}$.

(c) $\mathbf{v}$ has a unique additive inverse $-\mathbf{v}$.

(d) The zero vector $\mathbf{0}$ is unique.

(e) $0 \cdot \mathbf{v} = \mathbf{0}$.

(f) $(-1) \cdot \mathbf{v} = -\mathbf{v}$.

Proof These may seem obvious based upon past experience with real numbers, but remember, all we can assume about a vector space are the properties given in Definition 7.1. We give a proof of (a) and (c), and leave the rest as Exercise 45.

For (a), we know from (3) of Definition 7.1 that there exists a zero vector $\mathbf{0}$ such that

$$0+v=v$$

Property (5a) of Definition 7.1 (the Commutative Law) lets us interchange the order of addition. Doing so on the left above gives us the equation

$$v+0=v$$

proving part (a) of our theorem.

For part (c), note that (4) of Definition 7.1 states that every vector in V has at least one additive inverse. To prove that the additive inverse is unique, we suppose to the contrary that there exists a vector v in V with two additive inverses v1 and v2. Then

$$v+v1=0 \text{ and } v+v2=0$$

so that

$$v+v1=v+v2$$

Property (5a) of Definition 7.1 (the Commutative Law) lets us interchange the order of addition, so we also have

$$v1+v=v2+v$$

To cancel out v, we start by adding v1 to both sides of the equation.

$$(v1+v)+v1=(v2+v)+v1 \Rightarrow v1+(v+v1)=v2+(v+v1)$$

The grouping on the right is justified by (5b) of Definition 7.1 (the Associative Law). Since v+v1=v+v2=0, our equation simplifies to

$$v1+0=v2+0 \Rightarrow v1=v2$$

with the cancellation of the zero vector 0 justified by part (a) of this theorem. Since v1=v2, it follows that the additive inverse is unique. ■■

Example 3

Let R∞ denote the set of all infinite sequences of real numbers v=(v1, v2,…), so the elements in R∞ have an infinite number of components. Addition and scalar multiplication are defined componentwise, by

$$(v1, v2,\ldots)+(u1, u2,\ldots)=(v1+u1, v2+u2,\ldots)$$

and

$$c(v1, v2,\ldots)=(cv1, cv2,\ldots)$$

Show that R∞ is a vector space.

Solution Here R∞ looks somewhat like $\mathbf{R}^n$ but with an infinite number of components. It is not hard to see that R∞ is closed under addition and scalar multiplication, and that the zero vector is given by

$$0=(0, 0, 0,\ldots)$$

Also, if v=(v1, v2,…), then −v=(−v1, −v2,…) satisfies v+(−v)=0, so each vector in R∞ has an additive inverse. Finally, the six conditions given in (5) of Definition 7.1 are inherited from the real numbers. Hence R∞ is a vector space.

Example 4

Determine if V is a vector space, where V is the set of vectors of the form [ab] (a and b real numbers) together with addition and scalar multiplication defined by

$$[a1b2]+[a2b2]=[a1+a2b1+b2],\ \ c\,[ab]=[ca0\,]$$

Solution We answer this question by working our way through the conditions of Definition 7.1.

(1) Addition is defined as in Euclidean space, so V is closed under addition.

(2) Scalar multiplication is different from Euclidean space, but $c\mathbf{v}$ is in V for any c and v so V is closed under scalar multiplication.

(3) The zero vector 0=[00] for Euclidean space is also the zero vector of V.

(4) Addition is defined as in Euclidean space, so for every v in V there is also an additive inverse −v in V.

(5) The commutative and associative laws of addition are inherited from Euclidean space, so (a) and (b) are satisfied. The other parts of (5) involve scalar multiplication, so they should be verified more carefully. We let v1=[a1b2] and v2=[a2b2].

(c) c1 (v1+v2)=c1 ([a1b1]+[a2b2])=c1 [a1+a2b1+b2]=[c1(a1+a2)0] = [c1a1+c1a20]=[c1a10]+[c1a20]=c1 [a1b1]+c1 [a2b2] =c1v1+c1v2

(d) (c1+c2) v1=(c1+c2) [a1b1]=[(c1+c2)a10]=[c1a1+c2a10] =[c1a10]=[c2a10]=c1 [a1b1]+c2 [a1b1]=c1v1+c2v1

(e) (c1c2) v1=(c1c2) [a1b1]=[c1c2a10]=c1 [c2a10] =c1 (c2 [a1b1])=c1(c2v1)

Thus far every required condition has been satisfied. However, condition (f) is not because

$$1\cdot v1=1\ [a1b1]=[a10]\neq v1$$

when b1≠0. Thus V is not *quite* a vector space.

Example 5

Suppose that a<b are real numbers, and let C [a, b] denote the set of all real-valued continuous functions on [a, b]. For f and g in C [a, b] and a real scalar c, we define (f+g)(x)=f(x)+g(x) and (cf)(x)=c·f(x). (These are the usual pointwise definitions of addition and

scalar multiplication of functions.) Show that C [a, b] is a vector space.

Solution The set C [a, b] looks different from the other vector spaces we have seen, where the vectors had entries reminiscent of components of vectors in $\mathbf{R}^n$. Here, we have a set consisting of functions such as f(x)=sin(x) and g(x)=e−x.

To determine if C [a, b] is a vector space, we avoid being distracted by superficial appearances by focusing on the definition. First, the set C [a, b] together with addition and scalar multiplication are clearly defined. Next we need to determine if the five conditions of Definition 7.1 are all met. We verify the first three here and leave the remaining conditions as Exercise 3.

(1) Is C [a, b] closed under addition?
 If f and g are both in C [a, b], then both are continuous on the interval [a, b]. The sum of two continuous functions is also a continuous function, so that f+g is in C [a, b]. Hence C [a, b] is closed under addition.

(2) Is C [a, b] closed under scalar multiplication?
 If f is in C [a, b] and c is a scalar, then c·f(x) is continuous on [a, b], so that cf is also in C [a, b]. Thus C [a, b] is also closed under scalar multiplication.

(3) Is there a zero function?
 If z(x)=0, the identically zero function, then z(x) is continuous on [a, b] and so is in C [a, b]. Thus if f is in C [a, b], then z(x)+f(x)=f(x) for all x in [a, b]. Hence z+f=f for all f in C [a, b], so that z is the zero function.

The remaining questions that need to be answered are:

(4) Does each function have an additive inverse?

(5) Do the six commutative, distributive, and associative laws hold?

Verifying that the answer to each is yes is left as an exercise. Therefore C [a, b] is a vector space.

Below is a brief list of vector spaces. Some have already been verified as vector spaces, and others are left as exercises.

- Euclidean space $\mathbf{R}^n$ (n>0 an integer), together with the standard addition and scalar multiplication of vectors.
- $\mathbf{P}^n$ ($n \geq 1$ an integer), the set of polynomials with real coefficients and degree no greater then n, together with the usual addition and scalar multiplication of polynomials.
- Rm×n, the set of real m×n matrices together with the usual definition of matrix addition and scalar multiplication.
- **P**, the set of polynomials with real coefficients and any degree, together with the usual addition and scalar multiplication of polynomials.
- R∞, together with addition and scalar multiplication of vectors described in Example 3.
- C [a, b], the set of real-valued continuous functions on the interval [a, b], together with the usual addition and scalar multiplication of functions.
- C(R), the set of real-valued continuous functions on the real numbers R, together with the usual addition and scalar multiplication of functions.
- T(m, n), the set of linear transformations T : Rm→Rn, together with the usual addition and scalar multiplication of functions.

▶ This list is far from exhaustive. There are many other vector spaces not included.

Subspaces

Just as in Euclidean space, a subspace of a vector space can be thought of as a vector space contained within another vector space. Since a subspace inherits the properties of the parent vector space, all that is required to certify its subspace status is that it be closed under addition and scalar multiplication. The formal definition is essentially identical to the one given for Euclidean space.

DEFINITION 7.3 ▶

Subspace

A subset S of a vector space V is a **subspace** if S satisfies the following three conditions:

(a) S contains $\mathbf{0}$, the zero vector.

(b) If $\mathbf{u}$ and $\mathbf{v}$ are in S, then $\mathbf{u} + \mathbf{v}$ is also in S.

(c) If c is a scalar and $\mathbf{v}$ is in S, then $c\mathbf{v}$ is also in S.

▶ Definition 7.3 generalizes Definition 4.1 in Section 4.1.

Example 6

Let S denote the set of all polynomials p with real coefficients such that p(0)=0. Is S a subspace of P, the set of all polynomials with real coefficients?

Solution First, we note that S is a subset of P. Next, we do the same thing that we did to show that a subset of $\mathbf{R}^n$ is a subspace: We check if the three conditions of the definition hold.

(a) The zero polynomial z(x)=0 (the identically zero function) satisfies z(0)=0 and hence is in S, so S contains the zero vector.

(b) If p and q are both in S, then p(0)=0 and q(0)=0. Therefore (p+q)(0)=p(0)+q(0)=0, so that p+q is also in S.

(c) If p is in S and c is a real scalar, then (cp)(0)=c(p(0))=0. Hence cp is also in S.

Since the three conditions of Definition 7.3 hold, S is a subspace of **P**.

Example 7

Suppose that m<n are both integers. Is $\mathbf{P}^m$ a subspace of $\mathbf{P}^n$?

Solution First, we note that $\mathbf{P}^m$ is a subset of $\mathbf{P}^n$. Moreover, since $\mathbf{P}^m$ is itself a vector space, we know that it contains the zero vector, is closed under addition, and is closed under scalar multiplication. Thus the three subspace conditions are met, so that $\mathbf{P}^m$ is a subspace of $\mathbf{P}^n$.

Example 8

Suppose that m<n are both integers. Is Rm a subspace of $\mathbf{R}^n$?

Solution While it is true that Rm is itself a vector space, it is not a subset of $\mathbf{R}^n$ (the vectors in Rm have fewer components than those in $\mathbf{R}^n$) and so cannot be a subspace of $\mathbf{R}^n$.

Example 9

The **trace** of a square matrix is the sum of the diagonal terms. Suppose that S is the subset of R2×2 consisting of matrices with trace equal to 0. Is S a subspace of R2×2?

Solution By definition S is a subset of R2×2. Let's check the three conditions of Definition 7.3.

(a) The zero matrix of R2×2 is 022=[0000], which has trace 0. Hence 022 is in S.

(b) Suppose that A=[a11a12a21a22] and B=[b11b12b21b22] are both in S. Then a11+a22=0 and b11+b22=0. Since

A+B=[(a11+b11)(a12+b12)(a21+b21)(a22+b22)] the trace of A+B is equal to

$$(a11+b11)+(a22+b22)=(a11+a22)+(b11+b22)=0$$

Hence A+B is also in S.

(c) For a scalar *c* and *A* in S as in (b), we have

$$cA=[ca11ca12ca21ca22]$$

Therefore the trace of cA is

$$ca11+ca22=c(a11+a22)=0$$

which implies that cA is also in S.

Since all three conditions for a subspace are satisfied, S is a subspace.

The trivial subspaces carry over from Euclidean space.

THEOREM 7.4

Suppose that V is a vector space. Then S={0} and S=V are both subspaces of V, sometimes called the trivial subspaces.

Theorem 7.4 generalizes the result found in Example 4 of Section 4.1.

The proof is left as an exercise. The last two examples require some knowledge of calculus.

Example 10

Let C (a, b) be the set of functions continuous on (a, b), and suppose S is the subset of C (a, b) that consists of the set of differentiable functions on (a, b). Is S a subspace of C (a, b)?

Solution By definition, S is a subset of $C(a, b)$. Since $z(x)=0$ is differentiable ($z'(x)=0$), the zero vector is in S. Also, from calculus we know that sums and constant multiples of differentiable functions are differentiable, so that S is closed under addition and scalar multiplication. Hence S is a subspace of $C(a, b)$.

▶ In this chapter the proofs of many theorems are left as exercises because their Euclidean counterparts have very similar proofs given earlier in the book.

Example 11

Let $C(R)$ denote the set of functions that are continuous on all of R. Let S denote the subset of $C(R)$ consisting of functions $y(t)$ that satisfy the differential equation

$$y''(t)+y(t)=0 \tag{1}$$

Show that S is a subspace of $C(R)$.

Solution Differential equations of this type arise in the modeling of simple harmonic motion, such as a mass moving up and down while suspended by a spring (see Figure 1).

To show that S is a subspace, we start by noting that if $y(t)=0$ is the zero function, then $y''(t)=0$ and this function satisfies 1.

Next, suppose that $y1$ and $y2$ both satisfy (1). Then

$$(y1(t)+y2(t))''+(y1(t)+y2(t))=y1''(t)+y2''(t)+y1(t)+y2(t)=(y1''(t)+y1(t))+(y2''(t)+y2(t))=0$$

so $y1+y2$ is also in S.

Finally, given a scalar c and a solution $y1$ of (1), we have

$$(cy1(t))''+cy1(t)=c(y1''(t)+y1(t))=0$$

Thus cy1 is also in S, and therefore S is a subspace.

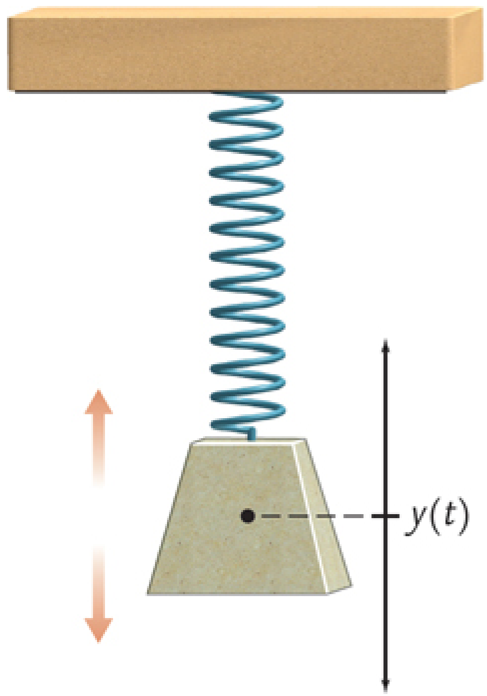


Figure 1 A mass-spring system. y(t) gives the vertical displacement at time t.

We close this section with a reminder. As mentioned earlier, often it is difficult to transition away from the notion that elements of $\mathbf{R}^n$ are the only type of "vector." As we have seen, viewed from a more general perspective, we can think of polynomials, matrices, and continuous functions as vectors. So remember: Vectors are not always columns of numbers!

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Let $\mathbf{P}^2$ be the set of polynomials with degree 2 or less, along with the usual definition for polynomial addition and scalar

multiplication. Prove that P^2 is a vector space.

2. Prove that the set of diagonal 2×2 matrices is a subspace of R2×2.
3. Let V=C(R) be the vector space of continuous functions on R. Give a proper subset S1 of V that is a subspace, and prove that it is a subspace. Then give a second proper subset S2 of V that is not a subspace, and explain why it is not a subspace.

EXERCISES

1. Complete Example 1: Verify that R2×2 with the usual definition of matrix addition and scalar multiplication satisfies the six conditions of (5) given in Definition 7.1.
2. Determine which properties of Definition 7.1 are not met by the set $\mathbf{Q}^2$ given in Example 2.

Exercises 3–8: Prove that V is a vector space.

3. V=C [a, b], the set of continuous functions defined on the interval [a, b], together with the standard pointwise definition of addition and scalar multiplication of functions. (Portions of this exercise are completed in Example 5.)
4. V=Rm×n, the set of real m×n matrices together with the usual definition of matrix addition and scalar multiplication.
5. V=Pn, the set of polynomials with real coefficients and degree no greater than *n*, together with the usual definition of polynomial addition and scalar multiplication.
6. V=P, the set of polynomials with real coefficients and any degree, together with the usual addition and scalar multiplication of polynomials.
7. V=C(R), the set of real-valued continuous functions defined on R, together with the usual pointwise addition and scalar multiplication of functions.

8. $V=T(m, n)$, the set of linear transformations $T : R^m \to R^n$, together with the usual addition and scalar multiplication of functions.

Exercises 9–12: A set V is given, together with definitions of addition and scalar multiplication. Determine if V is a vector space, and if so, prove it. If not, identify a condition of Definition 7.1 that is not satisfied.

9. V is the set of polynomials with real coefficients and degree 2 or less. Addition is defined by

$$(a_2x^2+a_1x+a_0)+(b_2x^2+b_1x+b_0)$$

$$=(a_0+b_0)x^2+(a_1+b_1)x+(a_2+b_2)$$

and scalar multiplication by

$$c(a_2x^2+a_1x+a_0)=ca_0x^2+ca_1x+ca_2$$

10. V is the set of lines in the plane through the origin, excluding the y-axis. Addition of lines is defined by adding slopes, and scalar multiplication by the scalar multiple of the slope.
11. V is the set of vectors in R^2 with the following definitions of addition and scalar multiplication:
Addition: $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}+\begin{bmatrix} a_2 \\ b_2 \end{bmatrix}=\begin{bmatrix} 0 \\ b_1+b_2 \end{bmatrix}$
Scalar multiplication: $c \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}=\begin{bmatrix} ca_1 \\ cb_1 \end{bmatrix}$
12. V is the set of vectors in R^2 with the following definitions of addition and scalar multiplication:
Addition: $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}+\begin{bmatrix} a_2 \\ b_2 \end{bmatrix}=\begin{bmatrix} 0 \\ b_1+b_2 \end{bmatrix}$
Scalar multiplication: $c \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}=\begin{bmatrix} 0 \\ cb_1 \end{bmatrix}$

Exercises 13–18: Prove that the set S is a subspace of the vector space V.

13. $V=R^{3\times3}$ and S is the set of upper triangular 3×3 matrices.
14. $V=P_5$ and S is the set of polynomials of the form $p(x)=a_4x^4+a_2x^2+a_0$.
15. $V=C(R)$ and S is the subset of functions f in V such that $f(4)=0$.

16. V=P and S is the set of all polynomials with terms of only even degree. (Thus 1, x^2, x^4, … are allowed, but not x, x^3, ….)
17. V=T(2, 2) and S is the set of linear transformations T such that

T ([x1x2])=[a1x1a2x2]

where a1 and a2 are fixed scalars.
18. V=R∞ and S consists of those vectors with a finite number of nonzero components.

Exercises 19–28: A vector space V and a subset S are given. Determine if S is a subspace of V, and if so, prove it. If not, give an example showing one of the conditions of Definition 7.3 is not satisfied.

19. V=C [−2, 2] and S=P.
20. V=C [−4, 7] and S consists of functions of the form ae^{bx} (a, b are real constants).
21. V=R∞ and S is the subset consisting of vectors where the second component is equal to zero—that is, vectorsv=(v1, 0, v3, v4,…).
22. V=R∞ and S is the subset consisting of vectors where the components are all integers.
23. V=R∞ and S is the subset of R∞ consisting of vectors where all but a finite number of components are not equal to zero.
24. V=C [−3, 3] and S is the set of real-valued functions f such that f(−1)+f(1)=0.
25. V=P4 and S is the set of real-valued functions g in P3 such that g(2)+g(3)=0.
26. V=T(4, 5) and S is the set of linear transformations that are one-to-one.
27. V=T(3, 3) and S is the set of invertible linear transformations.
28. V=C [−3, 3] and S is the set of real-valued functions h in C [−3, 3] such that h(0)=1.

Exercises 29–32: **Calculus Required** A vector space V and a subset S are given. Determine if S is a subspace of V, and if so prove it. If not explain which conditions of Definition 7.3 are not satisfied. (Assume $a<b$ as needed.)

29. $V=C(a, b)$ and $S=C^n(a, b)$, the set of functions on (a, b) with *n* continuous derivatives.

30. $V=C(\mathbb{R})$ and S is the set of all solutions to the differential equation $y'(t)-4y(t)=0$.

31. $V=C[a, b]$ and S is the set of functions g such that

$$\int_a^b g(x)\,dx=(b-a)$$

32. $V=C(\mathbb{R})$ and S is the set of functions h such that

$$\int_{-\infty}^{\infty} e^{-x^2} h(x)\,dx=0$$

FIND AN EXAMPLE Exercises 33–38: Find an example that meets the given specifications.

33. A failed vector space—that is, a set of vectors and definitions of addition and scalar multiplication that meet some but not all of the conditions of Definition 7.1.

34. A vector space *V* not given in this section.

35. A vector space *V* and a subset S that is *almost* a subspace: S contains 0 and is closed under addition, but S is not closed under scalar multiplication.

36. A vector space *V* and a subset S that is *almost* a subspace: S contains 0 and is closed under scalar multiplication, but S is not closed under addition.

37. A single set of vectors and two different definitions of addition and scalar multiplication to produce two different vector spaces V_1 and V_2.

38. A vector space *V* and an infinite sequence of subspaces $S_1 \subset S_2 \subset S_3 \subset \ldots$, with each subset proper.

TRUE OR FALSE Exercises 39–42: Determine if the statement is true or false, and justify your answer.

39.

(a) A vector space *V* consists of a set of vectors together with definitions of addition and scalar multiplication of the vectors.

(b) No two vector spaces can share the same set of vectors.

40.

(a) If v1 and v2 are in a vector space *V*, then so is v1−v2.

(b) If v1+v2 is in a subspace S, then v1 and v2 must be in S.

41.

(a) If S1 and S2 are subspaces of a vector space V, then the intersection S1∩ S2 is also a subspace of V.

(b) If S1 and S2 are subspaces of a vector space V, then the union S1∪S2 is also a subspace of V.

42.

(a) Suppose S1 and S2 are subspaces of a vector space V, and define S1+S2 to be the set of all vectors of the form s1+s2, where s1 is in S1 and s2 is in S2. Then S1+S2 is a subspace of V.

(b) A vector space V must have an infinite number of distinct elements.

43. Prove that Definition 7.3 is unchanged if condition (a) is replaced with the condition “S is nonempty.”

44. Prove Theorem 7.4: Suppose that V is a vector space. Prove that S={0} and S=V are subspaces of V.

45. Complete the proof of Theorem 7.2: Let V be a vector space and suppose that v is in V.

(a) If −v is an additive inverse of v, then −v+v=0.

(b) The zero vector 0 is unique.

(c) 0·v=0.

(d) (−1)·v=−v.

7.2 Span and Linear Independence

In this section we extend the concepts of span and linear independence from Euclidean space to vector spaces. The definitions here are similar to those for Euclidean space, so lean on your previous knowledge.

Span

Linear Combination

Just as in $\mathbf{R}^n$, a **linear combination** of a set of vectors $\{v_1, v_2, \ldots, v_m\}$ is a sum of the form

$$c_1v_1+c_2v_2+\ldots+c_mv_m \tag{1}$$

where $c_1, c_2, \ldots, c_m$ are real numbers. For instance, for the vectors $f(x)=\cos(x)$ and $g(x)=\log(x)$ in $C[1, 5]$, one possible linear combination is

$$7\cos(x)-\pi\log(x)$$

As in Euclidean space, the span of a set is defined in terms of linear combinations.

DEFINITION 7.5 ▶

Span

Let $\mathcal{V}=\{v_1, v_2,\ldots, v_m\}$ be a nonempty set of vectors in a vector space V. The **span** of this set is denoted $\text{span}\{v_1, v_2,\ldots, v_m\}$ and is defined to be the set of all linear combinations of the form

$$c_1v_1+c_2v_2+\ldots+c_mv_m$$

where $c_1, c_2, \ldots, c_m$ can be any real numbers.

If $\mathcal{V}$ consists of infinitely many vectors, then we define span($\mathcal{V}$) to be the set of all linear combinations of *finite* subsets of $\mathcal{V}$.

▶ Definition 7.5 generalizes Definition 2.5 in Section 2.2.

Example 1

Let $S=\text{span}\{x^2-2x+3, -2x^2+3x+1\}$ be a subset of $\mathbf{P}^2$. Determine if $p(x)=10x^2-17x+9$ is in S.

Solution The vector $p(x)=10x^2-17x+9$ is in S if there exist real numbers c_1 and c_2 such that

$$c_1(x^2-2x+3)+c_2(-2x^2+3x+1)=10x^2-17x+9$$

Reorganizing the left side to collect common factors, we have

$$(c_1-2c_2)x^2+(-2c_1+3c_2)x+(3c_1+c_2)=10x^2-17x+9 \qquad (2)$$

The only way that two polynomials are equal is if coefficients of like terms are equal. Thus (2) is true only if there exist c_1 and c_2 that satisfy the system

$$\begin{aligned} c_1-2c_2&=10\\ -2c_1+3c_2&=-17\\ 3c_1+c_2&=9 \end{aligned}$$

Applying our usual solution methods, we can determine that the system has the unique solution $c_1=4$ and $c_2=-3$. Hence $p(x)$ is in S.

Example 2

Let $S=\text{span}\{1, \cos(x), \cos(2x)\}$ be a subset of $C[0, \pi]$. Determine if $f(x)=\sin^2(x)$ is in S.

Solution At first glance it may not appear that $f(x)$ is in S. However, recall from trigonometry the identity

$$\sin^2(x)=\frac{1-\cos(2x)}{2}$$

Hence it follows that $\sin^2(x)$ is a linear combination of two vectors in S, with

$$\sin^2(x)=\frac{1}{2}(1)-\frac{1}{2}\cos(2x)$$

Therefore $f(x)=\sin^2(x)$ is in S.

The preceding examples both involved spanning subsets S of a vector space. In Euclidean space $\mathbf{R}^n$, such sets are subspaces. The same is true in vector spaces.

THEOREM 7.6

Suppose that $\mathcal{V}$ is a (possibly infinite) subset of a vector space V, and let $S=\text{span}(\mathcal{V})$. Then S is a subspace of V.

Theorem 7.6 generalizes Theorem 4.2 in Section 4.1.

The proof follows from verifying that the three conditions required of a subspace (Definition 7.3 in Section 7.1) are met. It is left as an exercise.

Example 3

Let S=span{[1021], [0−113], [41−21]} in R2×2 be a subspace of the vector space R2×2. Is v=[25−34] in S?

Solution In order for v to be in S, there must exist scalars c1, c2, and c3 such that

c1 [1021]+c2 [0−113]+c3 [41−21]=[25−34]⇒ [(c1+4c3)(−c2+c3)(2c1+c2−2c3)(c1+3c2+c3)]=[25−34]

For the components to be equal, c1, c2, and c3 must satisfy the linear system

c1 +4c3=2 −c2+c3=52c1+c2−2c3=−3c1+3c2+c3=4

Our standard solution methods can be used to show that this system has no solutions, so v is not in S.

As we can see from our examples, even though vector spaces can be made up of polynomials, matrices, or other objects, answering questions about spanning sets often boils down to something that we have done again and again: solving a system of linear equations. But this is not always the case.

Example 4

Can a finite set of vectors span P, the set of polynomials with real coefficients?

Solution Suppose that {f1(x), f2(x),… , fm(x)} is a set of polynomials in P, sorted by degree with

deg(f1) ≥ deg(f2) ≥ ⋯ ≥ deg(fm)

Then any linear combination

$$c_1 f_1(x)+c_2 f_2(x)+\cdots+c_m f_m(x)$$

has degree at most that of $f_1(x)$ (and possibly less, if there is cancellation). Thus, if $n=\deg(f_1)+1$, then $g(x) = x^n$ has degree greater than any linear combination of our set and hence cannot be in the span of the set. As there is nothing special about our set, this argument shows that no finite set of vectors can possibly span all of P.

The argument in Example 4 that shows that a finite set cannot span P does not apply to the infinite set $\{1, x, x^2, x^3, \ldots\}$. This infinite set does span P. Verification is left as an exercise.

Linear Independence

We now turn to the definition of linear independence. The definition is a near duplicate of the definition given in Section 2.3 for vectors in Euclidean space.

DEFINITION 7.7 ▶

Linearly Independent, Linearly Dependent

Let $\mathcal{V}=\{v_1, v_2, \ldots , v_m\}$ be a set of vectors in a vector space V. Then $\mathcal{V}$ is **linearly independent** if the equation

$$c_1v_1+c_2v_2+\cdots+c_mv_m=0$$

has only the trivial solution $c_1=\cdots=c_m=0$. The set $\mathcal{V}$ is **linearly dependent** if the equation has any nontrivial solutions.

▶ Definition 7.7 generalizes Definition 2.12 in Section 2.3.

As with span, the definition of linear independence extends to infinite sets. If $\mathcal{V}$ consists of infinitely many vectors, then $\mathcal{V}$ is linearly independent if *every finite* subset of $\mathcal{V}$ is linearly independent.

Example 5

Determine if

$$\{x^3-2x+4,\ x^3+2x^2-2,\ 2x^3+x^2-3x+5\}$$

is a linearly independent subset of P_3.

Solution Just as in Euclidean space, we start by setting

$$c_1(x^3-2x+4)+c_2(x^3+2x^2-2)+c_3(2x^3+x^2-3x+5)=0$$

and then determine the values of c_1, c_2, and c_3 that satisfy the equation. Collecting common factors gives us the equivalent equation

$$(c_1+c_2+2c_3)x^3+(2c_2+c_3)x^2+(-2c_1-3c_3)x+(4c_1-2c_2+5c_3)=0$$

Our polynomial is identically zero if and only if the coefficients are all zero. This will be true for any solution to the homogeneous system

$$\begin{aligned} c_1+c_2+2c_3&=0\\ 2c_2+c_3&=0\\ -2c_1\quad-3c_3&=0\\ 4c_1-2c_2+5c_3&=0 \end{aligned}$$

Applying our standard solution methods shows that the system has infinitely many solutions, among them $c_1=3$, $c_2=1$, and $c_3=-2$. Since a nontrivial linear combination of our vectors equals the zero vector, our set is linearly dependent.

Example 6

Determine if the subset

$$\{[102-110], [310222]\}$$

of $R2\times3$ is linearly independent.

Solution We proceed just as we did in the preceding example, by setting up the equation

$$c1\ [102-110]+c2\ [310222]=[000000]$$

Comparing the components on each side gives the linear system

$$c1+3c2=0\ \ c2=02c1\ \ =0-c1+2c2=0c1+2c2=0\ \ 2c2=0$$

We can see that this system has only the trivial solution $c1=c2=0$, so the set is linearly independent.

A few observations about linear independence carry over from Euclidean space (proofs left as exercises):

- The set $\{0, v1,\cdots, vm\}$ is linearly dependent for all vectors $v1, \cdots, vm$.
- A set of two nonzero vectors $\{v1, v2\}$ is linearly dependent if and only if one is a scalar multiple of the other. (We could solve Example 6 using this fact.)
- A set with just one vector $\{v1\}$ is linearly dependent if and only if $v1=0$.

▶ The first statement generalizes Theorem 2.13 in Section 2.3.

Example 7

Determine if $\{1, x, x2, x3,\ldots\}$ is a linearly independent subset of P, the set of all polynomials with real coefficients.

Solution Here we are working with an infinite set, so we must show that *every* finite subset is linearly independent. Let's take a typical finite subset

$$\{x^{a_1}, x^{a_2}, \ldots, x^{a_m}\}$$

where $a_1, a_2, \ldots, a_m$ are distinct nonnegative integers. Suppose that

$$c_1x^{a_1}+c_2x^{a_2}+\cdots+c_mx^{a_m}=0$$

We have the trivial solution $c_1=c_2=\cdots=c_m=0$, but are there others? The answer is no, because a polynomial is identically zero (that is, zero for all x) only if all the coefficients are zero. Since we have only the trivial solution, our subset is linearly independent. Since our subset is arbitrary, it follows that all finite subsets are linearly independent and therefore our original infinite set is linearly independent.

Example 8

Determine if $\{1, x, e^x\}$ is a linearly independent subset of C [0, 10].

Solution Just as in the other examples, we ask ourselves if there exist nontrivial scalars c_1, c_2, and c_3 such that

$$c_1(1)+c_2x+c_3e^x=0 \tag{3}$$

An equivalent formulation is

$$c_1+c_2x=-c_3e^x$$

Note that regardless of choice for c_1 and c_2, the left side is a linear equation. Since e^x is an exponential function, the only way the right side is linear is if $c_3=0$. This in turn forces $c_1=c_2=0$ as well, which shows that the only solution to our original equation is the trivial one. Therefore our set is linearly independent.

In Example 8 we used the fact that e^x is not a linear function. But suppose we did not know that? Let’s look at this example again, this time using a different approach.

Example 9

Determine if $\{1, x, e^x\}$ is a linearly independent subset of $C[0, 10]$.

Solution Suppose that c_1, c_2, and c_3 satisfy (3). Then the equation is satisfied for every value of x, and hence in particular for each of $x=0$, $x=1$, and $x=2$. Setting x equal to each of these values and plugging into (3) yields the homogeneous linear system

$$\begin{aligned} c_1 \quad\;\; + c_3 &= 0 \\ c_1 + c_2 + e c_3 &= 0 \\ c_1 + 2c_2 + e^2 c_3 &= 0 \end{aligned} \tag{4}$$

By applying our standard methods, we can verify that the system (4) has only the trivial solution $c_1=c_2=c_3=0$. Hence the only possible solution to (3) is also the trivial one, because any nontrivial solution to (3) would also satisfy (4). Therefore the set $\{1, x, e^x\}$ is linearly independent.

A note of warning: The method in Example 9 cannot be used to show that a set of functions is linearly dependent. To see why, suppose that $f(x)=-x^3+3x^2-2x$ and $g(x)=\sin(2\pi x)$. Then

$$f(0)=g(0),\ f(1)=g(1),\ f(2)=g(2)$$

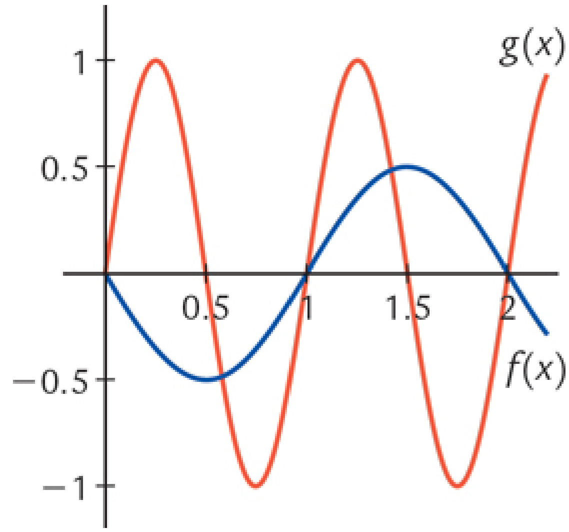


Figure 1 $f(x)=-x^3+3x^2-2x$ and $g(x)=\sin(2\pi x)$.

So if we apply the approach in Example 9, we get a linear system

$$c_1+c_2=0 \quad c_1+c_2=0 \quad c_1+c_2=0$$

Clearly this system has nontrivial solutions. However, the functions $f(x)$ and $g(x)$ are not multiples of each other (see Figure 1) and so are linearly independent. Functions as vectors are linearly dependent only if the dependence relation holds for all domain values, not just a few isolated values.

Span and Linear Independence

Sometimes it is difficult to keep straight the concepts of span and linear independence. The last two theorems in this section highlight the distinction between the two concepts.

THEOREM 7.8 ▶

The set $\mathcal{V}=\{v_1, v_2, \ldots, v_m\}$ of nonzero vectors is linearly dependent if and only if at least one vector in the set is in the span of the others.

▶ Theorem 7.8 generalizes Theorem 2.15 in Section 2.3.

The proof of Theorem 7.8 is left as an exercise.

Example 10

Determine if the set of vectors

$$\{1, \cos(x), \cos(2x), \sin^2(x)\} \tag{5}$$

is a linearly independent subset of C [0, 10].

Solution In Example 2 we showed that $\sin^2(x)$ is in span$\{1, \cos(x), \cos(2x)\}$. Hence by Theorem 7.8 the set of vectors in (5) is linearly dependent.

THEOREM 7.9 ▶

Let $\mathcal{V}=\{v_1, v_2, \ldots, v_m\}$ be a subset of a vector space V. Then:

(a) The set $\mathcal{V}$ is *linearly independent* if and only if the equation $c_1v_1+\cdots+c_mv_m=v$ has *at most* one solution for each v in V.

(b) The set $\mathcal{V}$ *spans* V if and only if the equation $c_1v_1+\cdots+c_mv_m=v$ has *at least* one solution for each v in V.

▶ Theorem 7.9 generalizes Theorem 2.11 in Section 2.2 and Theorem 2.20 in Section 2.3.

The proof is left as an exercise. Note that Theorem 7.9 applies to subspaces as well as vector spaces.

Example 11

Let S be the subspace of $T(2, 2)$ of linear transformations of the form

$$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=\begin{bmatrix}a_1x_1\\a_2x_2\end{bmatrix} \tag{6}$$

where a_1 and a_2 are scalars (see Exercise 17 of Section 7.1). Suppose that

$$T_1\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=\begin{bmatrix}x_1\\0\end{bmatrix} \text{ and } T_2\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=\begin{bmatrix}0\\x_2\end{bmatrix}$$

Show that the set $\{T_1, T_2\}$ is linearly independent and spans S.

Solution Suppose that T is a linear transformation in S and hence has the form in (6). To apply Theorem 7.9, we need to determine the number of solutions to the equation $c_1T_1(\mathbf{x})+c_2T_2(\mathbf{x})=T(\mathbf{x})$. This equation is equivalent to

$$c_1\begin{bmatrix}x_1\\0\end{bmatrix}+c_2\begin{bmatrix}0\\x_2\end{bmatrix}=\begin{bmatrix}a_1x_1\\a_2x_2\end{bmatrix}$$

For this to hold for all x_1 and x_2, we must have $c_1=a_1$ and $c_2=a_2$. Thus the equation $c_1T_1(\mathbf{x})+c_2T_2(\mathbf{x})=T(\mathbf{x})$ has exactly one solution for any a_1 and a_2. Therefore by Theorem 7.9(a) the set $\{T_1, T_2\}$ is linearly independent, and by Theorem 7.9(b) the set $\{T_1, T_2\}$ spans S.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine if the set of polynomials

$$\{x^2-2,\ 2x+1,\ 2x^2+x-3\}$$

is linearly independent and if the set spans $\mathbf{P}^2$.

2. Determine if the set of matrices

$$\{[120-1], [01-20], [1-123], [201-1]\}$$

is linearly independent and if the set spans R2×2.

EXERCISES

Exercises 1–4: Determine if v is in the subspace of $\mathbf{P}^2$ given by

$$\text{span}\{3x2+x-1, x2-3x+2\}$$

1. $v=3x2+11x-8$
2. $v=2x2-9x+7$
3. $v=10x-7$
4. $v=7x2-x$

Exercises 5–8: Determine if v is in the subspace of P3 given by

$$\text{span}\{x3+x-2, x2+2x+1, x3-x2+x\}$$

5. $v=x3+2x2-3x$
6. $v=3x2+4x$
7. $v=x2+4x+4$
8. $v=x2+2x-1$

Exercises 9–12: Determine if v is in the subspace of R2×3 given by

$$\text{span }\{[121013], [031-110]\}$$

9. $v=[-141-21-3]$
10. $v=[211115]$
11. $v=[2-5-13-16]$
12. $v=[332129]$

Exercises 13–16: Determine if v is in the subspace of R2×2 given by span{[−1341], [025−3], [1421]}

13. v=[−4355]
14. v=[233−3]
15. v=[−2−120]
16. v=[12−34]

Exercises 17–26: Determine if the subset is linearly independent in the given vector space.

17. {x2−3, 3x2+1} in $\mathbf{P}^2$
18. {2x3−x+3, −4x3+2x−6} in P3
19. {x3+2x+4, x2−x−1, x3+2x2+2} in P3
20. {x2+3x+2, x3−2x2, x3+x2−x−1} in P3
21. {[2−113], [−42−2−6]} in R2×2
22. {[−1341], [3501], [1421]} in R2×2
23. {[101214], [312033]} in R2×3
24. {[120114], [5415410], [301323]} in R3× 2
25. {sin2(x), cos2(x), 1} in C [0, π]
26. {sin(2x), cos(2x), sin(x) cos(x)} in C [0, π]

FIND AN EXAMPLE Exercises 27–32: Find an example that meets the given specifications.

27. A subset of R2×2 that spans R2×2 but is not linearly independent.
28. An infinite subset of P that is linearly independent but does not span P.
29. A vector space V and an infinite linearly independent subset $\mathcal{V}$.
30. A set of nonzero vectors that is linearly dependent and yet has a vector in the set that is *not* a linear combination of the other vectors. Explain why this does not contradict Theorem 7.8.
31. Two infinite linearly independent subsets $\mathcal{V}$1 and $\mathcal{V}$2 of R∞ such that span($\mathcal{V}$1)∩span($\mathcal{V}$2)={ 0}.

32. A subset of T(2, 2) that is linearly independent and spans T(2, 2).

TRUE OR FALSE Exercises 33–38: Determine if the statement is true or false, and justify your answer.

33.

(a) Vectors must be columns of numbers.

(b) A linearly independent set cannot have an infinite number of vectors.

34.

(a) The zero vector 0 cannot be part of a spanning set.

(b) Every vector space can be spanned by a finite number of vectors.

35.

(a) A set of vectors $\mathcal{V}$ in a vector space V can be linearly independent or can span V, but cannot do both.

(b) Suppose that f and g are linearly dependent functions in C [1, 4]. If f(1)=−3g(1), then it must be thatf(4)=−3g(4).

36.

(a) Suppose that $\mathcal{V}1\subset\mathcal{V}2$ are sets in a vector space V. If $\mathcal{V}1$ is linearly independent, then so is $\mathcal{V}2$.

(b) Let {v1,… , vk} be a linearly independent subset of a vector space V. If c≠0 is a scalar, then {cv1,… , cvk} is also linearly independent.

37.

(a) Suppose that $\mathcal{V}1\subset\mathcal{V}2$ are sets in a vector space V. If $\mathcal{V}2$ spans V, then so does $\mathcal{V}1$.

(b) Let {v1,… ,vk} be a linearly independent subset of a vector space V. For any v≠0 in V, the set {v+v1,… , v+vk} is also linearly independent.

38.

(a) If {v1, v2, v3} is a linearly independent set, then so is{v1, v2−v1, v3−v2+v1}.

(b) If $\mathcal{V}1$ and $\mathcal{V}2$ are linearly independent subsets of a vector space V and $\mathcal{V}1\cap\mathcal{V}2$ is nonempty, then $\mathcal{V}1\cap\mathcal{V}2$ is also linearly independent.

39. In Example 7 it is shown that {1, x, x2,…} is linearly independent. Prove {1, x, x2,…} also spans P.

40. Prove Theorem 7.8: A set {v1, v2,… , vm} of nonzero vectors is linearly dependent if and only if one vector in the set is in the span of the others. (Hint: This is similar to Theorem 2.15 in Section 2.3.)

41. Prove that the subset $\{0, v_1, \ldots, v_m\}$ of a vector space must be linearly dependent. (Hint: This is similar to Theorem 2.13 in Section 2.3.)

42. Prove that any set of two nonzero vectors $\{v_1, v_2\}$ is linearly dependent if and only if each is a scalar multiple of the other.

43. Prove that a set consisting of one vector $\{v_1\}$ is linearly dependent if and only if $v_1=0$.

44. Let $\mathcal{V}=\{v_1, v_2, \ldots, v_m\}$ be a subset of a vector space V. Prove Theorem 7.9 by proving each of the following:

(a) The set $\mathcal{V}$ is linearly independent if and only if the equation $c_1v_1+\ldots+c_mv_m=v$ has *at most* one solution for each v in V.

(b) The set $\mathcal{V}$ *spans* V if and only if the equation $c_1v_1+\ldots+c_mv_m=v$ has *at least* one solution for each v in V.

45. Prove that no finite subset of R^∞ can span R^∞.

46. Suppose that $\mathcal{V}_1\subset\mathcal{V}_2$ are sets in a vector space V. Prove that if $\mathcal{V}_1$ spans V, then so does $\mathcal{V}_2$.

47. Suppose that $\mathcal{V}_1\subset\mathcal{V}_2$ are nonempty sets in a vector space V. Prove that if $\mathcal{V}_2$ is linearly independent, then so is $\mathcal{V}_1$.

48. Let $v_1, \ldots, v_m$ and v be vectors in a vector space V. If v is in the span of the set $\{v_1, \ldots, v_m\}$, prove that

$$\operatorname{span}\{v, v_1, \ldots, v_m\}=\operatorname{span}\{v_1, \ldots, v_m\}$$

49. Suppose that $\mathcal{V}=\{v_1, \ldots, v_m\}$ spans a vector space V, and suppose that v is in V but not in $\mathcal{V}$. Prove that $\{v, v_1, \ldots, v_m\}$ is linearly dependent.

50. Suppose that $\mathcal{V}=\{v_1, v_2, \ldots, v_m\}$ is a linearly independent subset of a vector space V. Prove that $\{v_2, \ldots, v_m\}$ does not span V.

Exercises 51–54: If possible use the method demonstrated in Example 9 to determine if the given subset is linearly independent in the given vector space. If this is not possible, explain why.

51. $\{x, \sin(\pi x/2), e^x\}$ in $C[0, \pi]$

52. {x, sin(x), cos(x)} in C [0, π]

53. {ex, cos2(x), cos(2x), 1} in C [0, π]

54. {cos(2x), sin(2x), cos2(x), sin2(x)} in C [0, π]

7.3 Basis and Dimension

Now that we have developed the concepts of span and linear independence in the context of a vector space, we are ready to consider the notion of basis and dimension in the same setting. As with much of this chapter, the definitions of basis and dimension are essentially the same as in Euclidean space. Let's start with basis.

DEFINITION 7.10 ▶

Basis

Let $\mathcal{V}$ be a subset of a vector space V. Then $\mathcal{V}$ is a **basis** of V if $\mathcal{V}$ is linearly independent and spans V.

▶ Definition 7.10 generalizes Definition 4.8 in Section 4.2.

Example 1

Is the set $\{x^2+4x-3, x^2+1, x-2\}$ a basis for $\mathbf{P}^2$?

Solution We need to determine if the given set is linearly independent and spans $\mathbf{P}^2$. Of the two, span is generally more difficult to verify, so let's tackle that first. (We will also find out about linear independence along the way.) A typical vector in $\mathbf{P}^2$ has the form $a_2x^2+a_1x+a_0$, and for each such vector we need to know if there exist corresponding scalars c_1, c_2, and c_3 such that

$$c_1(x^2+4x-3)+c_2(x^2+1)+c_3(x-2)=a_2x^2+a_1x+a_0$$

Reorganizing the left side to collect common terms, we have

$$(c_1+c_2)x^2+(4c_1+c_3)x+(-3c_1+c_2-2c_3)=a_2x^2+a_1x+a_0$$

Comparing coefficients gives us the linear system

$$\begin{aligned} c_1+c_2 \quad &= a_2 \\ 4c_1 \quad +c_3 &= a_1 \\ -3c_1+c_2-2c_3 &= a_0 \end{aligned} \tag{1}$$

Using our standard solution methods, we find that for each choice of a_0, a_1, and a_2, the system has unique solution

$$c_1=\frac{-a_2+2a_1+a_0}{4},\ c_2=\frac{5a_2-2a_1-a_0}{4},\ c_3=a_2-a_1-a_0$$

Applying both parts of Theorem 7.9 in Section 7.2, we can conclude that our set is both linearly independent and spans $\mathbf{P}^2$. Hence the set is a basis of $\mathbf{P}^2$.

The method of solution in Example 1 suggests the following general theorem.

THEOREM 7.11

The set $\mathcal{V}=\{v_1,\ldots,v_m\}$ is a basis for a vector space V if and only if the equation

$$c_1v_1+\ldots+c_mv_m=v \tag{2}$$

has a unique solution $c_1,\ldots,c_m$ for every v in V.

Theorem 7.11 generalizes Theorem 4.9 in Section 4.2.

The proof follows from applying Theorem 7.9 in Section 7.2 and is left as an exercise.

Example 2

Verify that the set {1, x, x2,… , xn} is a basis for Pn. (This is called the **standard basis** for Pn.)

Solution Suppose that anxn+…+a1x+a0 is a typical vector in Pn. Forming a linear combination of our set and setting it equal to our typical vector produces

$$c1(1)+c2(x)+\ldots+cn+1(xn)=anxn+\ldots+a1x+a0$$

Comparing coefficients instantly shows that the unique linear combination is given by c1=a0, c2=a1, … , cn+1=an. Since this works for every vector in Pn, by Theorem 7.11 our set forms a basis for Pn.

Setting n=2 in Example 2 gives another basis for $\mathbf{P}^2$. Note that this basis has the same number of elements as the basis in Example 1, illustrating the following theorem, which carries over from Euclidean space.

THEOREM 7.12 ▶

Suppose that $\mathcal{V}1$ and $\mathcal{V}2$ are both bases of a vector space V. Then $\mathcal{V}1$ and $\mathcal{V}2$ have the same number of elements.

▶ Theorem 7.12 generalizes Theorem 4.12 in Section 4.2.

The proof is left as an exercise. The theorem also applies to vector spaces with bases that have infinitely many vectors: If one basis has infinitely many vectors, then they all do.

Example 3

Show that every basis for P has infinitely many elements.

Solution In Example 7 of Section 7.2, it is shown that the set $\{1, x, x^2, \ldots\}$ is linearly independent, and in Exercise 39 of Section 7.2 it is shown that $\{1, x, x^2, \ldots\}$ spans P. Therefore the set $\{1, x, x^2, \ldots\}$ is a basis for P, and thus by Theorem 7.12 every basis for P must have infinitely many elements.

Since every basis for a vector space has the same number of vectors, the following definition makes sense.

DEFINITION 7.13 ▶

Dimension

The **dimension** of a vector space V is equal to the number of vectors in any basis of V. If a basis of V has infinitely many vectors, then we say that the dimension is infinite.

▶ Definition 7.13 generalizes Definition 4.13 in Section 4.2.

For example, we have seen that P_n has dimension $n+1$ and P has infinite dimension. Put briefly, $\dim(P_n)=n+1$ and $\dim(P)=\infty$.

Note that a trivial vector space $V=\{0\}$ consisting only of the zero vector has no basis because it has no linearly independent subsets. We define $\dim(V)=0$ when V is a trivial vector space.

Example 4

Find the dimension of $\mathbb{R}^{2\times 2}$, the vector space of real 2×2 matrices.

Solution All we need to do is find any basis for $\mathbb{R}^{2\times 2}$ and then count the vectors. The standard basis is given by

$$\left\{\begin{bmatrix}1&0\\0&0\end{bmatrix}, \begin{bmatrix}0&1\\0&0\end{bmatrix}, \begin{bmatrix}0&0\\1&0\end{bmatrix}, \begin{bmatrix}0&0\\0&1\end{bmatrix}\right\}$$

Verification that this is a basis for $\mathbb{R}^{2\times 2}$ is left as an exercise. Hence it follows that $\dim(\mathbb{R}^{2\times 2})=4$.

A similar argument to the one in Example 4 can be used to show that $\dim(\mathbb{R}^{n\times m})=nm$ (see Exercise 46).

Since subspaces are essentially vector spaces within vector spaces, they also have dimensions.

Example 5

Find the dimension of the subspace $S=\text{span}\{T_1, T_2, T_3\}$ of $T(2, 2)$, where

$$T_1\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)=\begin{bmatrix} x_1-x_2 \\ 2x_1 \end{bmatrix},\quad T_2\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)=\begin{bmatrix} x_2 \\ 2x_1+x_2 \end{bmatrix},\quad T_3\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)=\begin{bmatrix} 2x_1-5x_2 \\ -2x_1-3x_2 \end{bmatrix}$$

Solution Let's check if the set is linearly independent. If so, then the dimension is 3 and we are done. If not, then we will identify a linear relation among the vectors that will allow us to remove one.

The zero vector in $T(2, 2)$ is the linear transformation $T_0(\mathbf{x})=\mathbf{0}$. We need to find the values of c_1, c_2, and c_3 such that

$$c_1T_1(\mathbf{x})+c_2T_2(\mathbf{x})+c_3T_3(\mathbf{x})=T_0(\mathbf{x})$$

is true for all $\mathbf{x}$ in $\mathbb{R}^2$. This is equivalent to the equation

$$c_1\begin{bmatrix} x_1-x_2 \\ 2x_1 \end{bmatrix}+c_2\begin{bmatrix} x_2 \\ 2x_1+x_2 \end{bmatrix}+c_3\begin{bmatrix} 2x_1-5x_2 \\ -2x_1-3x_2 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which in turn is equivalent to the system

$$\begin{aligned} c_1(x_1-x_2)+c_2x_2+c_3(2x_1-5x_2)&=0 \\ c_1(2x_1)+c_2(2x_1+x_2)+c_3(-2x_1-3x_2)&=0 \end{aligned}$$

Reorganizing to separate out x_1 and x_2 gives us

$$x_1(c_1+2c_3)+x_2(-c_1+c_2-5c_3)=0 \quad x_1(2c_1+2c_2-2c_3)+x_2(c_2-3c_3)=0$$

Since the system must be satisfied for all values of x_1 and x_2, we require that

$$\begin{aligned} c_1 \quad +2c_3&=0\\ -c_1+c_2-5c_3&=0\\ 2c_1+2c_2-2c_3&=0\\ c_2-3c_3&=0 \end{aligned}$$

Using our standard methods, we can show that this system has infinitely many solutions, among them $c_1=2$, $c_2=-3$, and $c_3=-1$. Therefore

$$2T_1(x)-3T_2(x)-T_3(x)=T_0(x)$$

or

$$2T_1(x)-3T_2(x)=T_3(x)$$

Since T_3 is a linear combination of T_1 and T_2, it follows (see Exercise 48 in Section 7.2) that $S=\text{span}\{T_1, T_2\}$. It is straightforward to verify that T_1 and T_2 are not multiples of each other and so are linearly independent. Hence $\{T_1, T_2\}$ is a basis for S, and therefore $\dim(S)=2$.

In Example 5, we started with a set that spanned a subspace S and removed a vector to form a basis for S. The following theorem formalizes this process, along with the process of expanding a linearly independent set to a basis.

THEOREM 7.14

Let $\mathcal{V}=\{v_1,\dots, v_m\}$ be a subset of a nontrivial finite dimensional vector space V.

(a) If $\mathcal{V}$ spans V, then either $\mathcal{V}$ is a basis for V or vectors can be removed from $\mathcal{V}$ to form a basis for V.

(b) If $\mathcal{V}$ is linearly independent, then either $\mathcal{V}$ is a basis for V or vectors can be added to $\mathcal{V}$ to form a basis for V.

▶ Theorem 7.14 generalizes Theorem 4.14 in Section 4.2.

The proof is left as an exercise. Note that Theorem 7.14 also applies to subspaces of vector spaces.

Example 6

Extend the set $\mathcal{V}$={x2+2x+1, x2+6x+3} to a basis for $\mathbf{P}^2$.

Solution We solve this by starting with a set we know spans $\mathbf{P}^2$ and then applying Theorem 7.14(b). We know that {x2, x, 1} is a basis for $\mathbf{P}^2$, so the combined set

$$\{x2+2x+1, x2+6x+3, x2, x, 1\} \tag{3}$$

spans $\mathbf{P}^2$. To find a basis that includes $\mathcal{V}$, we need to determine the dependences among the vectors in (3). We start with the equation

$$c1(x2+2x+1)+c2(x2+6x+3)+c3x2+c4x+c5=0$$

which is equivalent to

$$(c1+c2+c3)x2+(2c1+6c2+c4)x+(c1+3c2+c5)=0$$

Setting each coefficient equal to zero yields the vector equation

$$c1\ [121]+c2\ [163]+c3\ [100]+c4\ [010]+c5\ [001]+= \quad [000]$$

This is now a problem similar to Example 4 in Section 4.2. The augmented matrix and corresponding echelon form are

$$[111002601013001|000]\sim[1110002-1010001-2|000]$$

Since the leading terms appear in columns 1, 2, and 4 of the echelon matrix, it follows that the vectors associated with c_1, c_2, and c_4 are linearly independent. To see why, note that if the vectors x^2 and 1 were eliminated from (3), then we would have the linear system

$$\begin{aligned} c_1 + c_2 \quad &= 0 \\ 2c_1 + 6c_2 + c_4 &= 0 \\ c_1 + 3c_2 \quad &= 0 \end{aligned}$$

The augmented matrix and echelon form are as before, but with the third and fifth columns removed. (The row operations are the same as before.)

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

We see that the new system has a unique solution, so the set

$$\mathcal{V}_1 = \{x^2+2x+1,\ x^2+6x+3,\ x\}$$

is linearly independent. Furthermore, if $\mathcal{V}_1$ is not a basis, then by Theorem 7.14(b) we can add vectors to $\mathcal{V}_1$ to form a basis for $\mathbf{P}^2$. But this would mean that $\dim(\mathbf{P}^2) > 3$, which we know is false. Therefore $\mathcal{V}_1$ is a basis.

Dimension gives us a way to measure the size of a vector space. Thus if one vector space is contained in another, it seems reasonable that the dimension of the former would be smaller than that of latter. This was true in Euclidean space, and is also true here.

THEOREM 7.15

Let V_1 and V_2 be vector spaces with V_1 a subset of V_2, and suppose both have the same definition of addition and scalar multiplication. Then $\dim(V_1) \le \dim(V_2)$.

Theorem 7.15 is similar to Theorem 4.16 in Section 4.2.

Proof If $\dim(V2)=\infty$, then the theorem is true regardless of the dimension of V1. (Similar reasoning applies if $\dim(V1)=0$.) Now suppose that $\dim(V2)$ is finite, and let $\mathcal{V}1$ be a basis for V1. Since V1 is a subset of V2, $\mathcal{V}1$ is a linearly independent subset of V2. Thus, by Theorem 7.14(b), $\mathcal{V}1$ is a basis for V2 or can be expanded to a basis for V2. Therefore the number of vectors in $\mathcal{V}1$ is less than or equal to $\dim(V2)$. Since $\mathcal{V}1$ is a basis of V1, we conclude that $\dim(V1)\le\dim(V2)$. ■ ■

Note that in Theorem 7.15, V1 can also be considered a subspace of V2.

Example 7

Show that $\dim(C(R))=\infty$ without finding a basis for C(R).

Solution We have already shown that $\dim(P)=\infty$. Since all polynomials are continuous functions, then P is a subspace of C(R). Hence by Theorem 7.15 we have $\dim(C(R))=\infty$.

■

THEOREM 7.16 ▶

Let $\mathcal{V}=\{v1,\ldots,vm\}$ be a subset of a vector space V with $\dim(V)=n$.

(a) If $m<n$, then $\mathcal{V}$ does not span V.

(b) If $m>n$, then $\mathcal{V}$ is linearly dependent.

■

▶ Theorem 7.16 generalizes Theorem 4.17 in Section 4.2.

The proof is left as an exercise. As with previous theorems in this section, Theorem 7.16 also applies to subspaces.

Example 8

Determine by inspection which of the subsets cannot span $\mathbf{R}^{2\times 2}$ and which must be linearly dependent.

$$\mathcal{V}_1=\{[2143], [5012], [8356]\}\mathcal{V}_2=\{[5904], [7612], [4378], [6271]\}\mathcal{V}_3=\{[3690], [1438], [3201], [0166], [3255]\}$$

Solution In Example 4, we showed that $\dim(\mathbf{R}^{2\times 2})=4$. Hence by Theorem 7.16(a), $\mathcal{V}_1$ cannot span $\mathbf{R}^{2\times 2}$ because it has fewer than four vectors. By Theorem 7.16(b), $\mathcal{V}_3$ must be linearly dependent because it has more than four vectors.

Note that Theorem 7.16 cannot tell us if a set spans a vector space or is linearly independent. Hence we cannot conclude that $\mathcal{V}_2$ or $\mathcal{V}_3$ spans $\mathbf{R}^{2\times 2}$, nor can we conclude that $\mathcal{V}_1$ or $\mathcal{V}_2$ is linearly independent.

THEOREM 7.17 ▶

Let $\mathcal{V}=\{v_1,\ldots, v_m\}$ be a subset of a vector space V with $\dim(V)=m<\infty$. If $\mathcal{V}$ is linearly independent *or* spans V, then $\mathcal{V}$ is a basis for V.

▶ Theorem 7.17 generalizes Theorem 4.15 in Section 4.2.

Theorem 7.17 also applies to subspaces. This theorem tells us that if a set has the same number of vectors as the dimension of the vector space, then we need only verify either span or linear independence to determine if the set is a basis.

Example 9

Show that the set

$$\mathcal{V}=\{x^3-x^2-5x+1,\ 2x^2+3x,\ x^3+3x^2-4x+2,\ -3x^2+5\}$$

is a basis for $\mathbf{P}^3$.

Solution We have $\dim(\mathbf{P}^3)=4$ and $\mathcal{V}$ has four elements, so by Theorem 7.17 we need only show $\mathcal{V}$ is linearly independent or spans $\mathbf{P}^3$ to prove that $\mathcal{V}$ is a basis for $\mathbf{P}^3$. Let's show that $\mathcal{V}$ is linearly independent. Suppose that c_1, c_2, c_3, and c_4 are scalars such that

$$c_1(x^3-x^2-5x+1)+c_2(2x^2+3x)+c_3(x^3+3x^2-4x+2)+c_4(-3x^2+5)=0$$

Reorganizing to collect common terms and setting each coefficient equal to zero gives us

$$\begin{aligned} c_1 \quad\ \ + c_3 \quad\ \ &= 0 \\ -c_1+2c_2+3c_3-3c_4 &= 0 \\ -5c_1+3c_2-4c_3 \quad\ \ &= 0 \\ c_1 \quad\ \ +2c_3+5c_4 &= 0 \end{aligned}$$

To determine linear independence, all we need to know is if this system has any nontrivial solutions. Instead of transforming to an augmented matrix and performing row operations, let's define the coefficient matrix

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

By applying techniques from Chapter 5, we can show that $\det(C)=-59$. Since the determinant of the coefficient matrix C is nonzero, it follows from Theorem 5.7 (the Unifying Theorem, Version 7) in Section 5.2 that our linear system has only the trivial solution. Therefore $\mathcal{V}$ is a linearly independent set and is a basis for $\mathbf{P}^3$.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine if the set $\mathcal{V}$ is a basis for V.
   - (a) $\mathcal{V}$={x2+x−1, −2x+3, x2−5}, V=P2
   - (b) $\mathcal{V}$={[10−11], [0123], [1−110], [−120−1]}, V=R2×2
2. Let S be the subspace of V=R2×2 of lower triangular matrices. Find a basis for S and determine dim(S).
3. Extend {x−2, x2+1} to a basis for $\mathbf{P}^2$.

EXERCISES

Exercises 1–6: Determine by inspection if the set $\mathcal{V}$ could possibly be a basis for V. Explain your answer.

1. $\mathcal{V}$={x2+7, 3x+5}, V=P2
2. $\mathcal{V}$={x3+3, 4x2+x−1, 3x3+5, x}, V=P3
3. $\mathcal{V}$={[2104], [0315], [6221], [9002]}, V=R2×2
4. $\mathcal{V}$={[600133], [015541], [111111], [221703]}, V=R3×2
5. $\mathcal{V}$={x4+3, 4x3+x−1, 3x4+5, x2}, V=P4
6. $\mathcal{V}$={x, x3, x5, x7,…}, V=P

Exercises 7–12: Determine if the set $\mathcal{V}$ is a basis for V.

7. $\mathcal{V}$={2x2+x−3, x+1, −5}, V=P2
8. $\mathcal{V}$={x2−5x+3, 3x2−7x+5, x2−x+1}, V=P2
9. $\mathcal{V}$={[1221], [3103], [2211], [3334]}, V=R2×2
10. $\mathcal{V}$={[4321], [0123], [0021], [0001]}, V=R2×2
11. $\mathcal{V}$={(1, 0, 0, 0, 0,…), (1,−1, 0, 0, 0,…), (1,−1, 1, 0, 0,…), (1, −1, 1, −1, 0,…),…}, V=R∞
12. $\mathcal{V}$={1, x+1, x2+x+1, x3+x2+x+1,…}, V=P

Exercises 13–18: Find a basis for the subspace S and determine dim(S).

13. S is the subspace of $R^{3\times3}$ consisting of matrices with trace equal to zero. (The *trace* is the sum of the diagonal terms of a matrix.)
14. S is the subspace of $\mathbf{P}^2$ consisting of polynomials with graphs crossing the origin.
15. S is the subspace of $R^{2\times2}$ consisting of matrices with components that add to zero.
16. S is the subspace of T(2, 2) consisting of linear transformations $T : R^2 \to R^2$ such that $T(x)=ax$ for some scalar a.
17. S is the subspace of T(2, 2) consisting of linear transformations $T : R^2 \to R^2$ such that $T(v)=0$ for a specific vector v.
18. S is the subspace of P consisting of polynomials p such that $p(0)=0$.

Exercises 19–20: Determine the dimension of the subspace S. Justify your answer.

19. S is the subspace of C(R) consisting of functions f such that $f(k)=0$ for $k=0, 1, 2$.
20. S is the subspace of C(R) consisting of functions f such that $f(0)=f(1)=f(2)$.

Exercises 21–24: Extend the linearly independent set $\mathcal{V}$ to a basis for V.

21. $S=\{2x^2+1, 4x-3\}$, $V=P^2$
22. $\mathcal{V}=\{x^3, x^2+x+1, x\}$, $V=P^3$
23. $\mathcal{V}=\{[1001], [0110], [1110]\}$, $V=R^{2\times2}$
24. $\mathcal{V}=\{[101010], [010101], [101000], [000001]\}$, $V=R^{3\times2}$

Exercises 25–26: Remove vectors from $\mathcal{V}$ to yield a basis for V.

25. $\mathcal{V}=\{x+1, x+2, 2x+1\}$, $V=P^1$
26. $\mathcal{V}=\{x^2+x, x+1, x^2+1, x^2+x+1\}$, $V=P^2$

Exercises 27–30 assume some knowledge of calculus.

27. Let S be the subspace of C(R) consisting of solutions y(t) to the differential equation $y''(t)+y(t)=0$. (This equation is discussed in Example 11 of Section 7.1.) All solutions to this equation have the form

$$y(t)=c_1 \cos(t)+c_2 \sin(t).$$

Prove that $\dim(S)=2$.

28. Find a basis for the subspace S of P_4 consisting of polynomials p(x) such that $p'(x)=0$.

29. Find a basis for the subspace S of P_6 consisting of polynomials p(x) such that $p''(x)=0$.

30. Determine the dimension of the subspace S of P consisting of polynomials p such that

$$\int_{-1}^{1} p(x)\, dx=0.$$

FIND AN EXAMPLE Exercises 31–36: Find an example that meets the given specifications.

31. An infinite dimensional vector space V and a finite dimensional subspace S.

32. A vector space V and a subspace S such that $\dim(S)=5$.

33. A vector space V and subspace S such that $\dim(V)=1+\dim(S)$.

34. A vector space V and subspace S such that $\dim(V)=2(\dim(S))$.

35. A vector space V and an infinite subset $\mathcal{V}$ that is linearly independent but does not span V.

36. A vector space V and subspace S such that $\dim(V)=\dim(S)$ but $S\neq V$.

TRUE OR FALSE Exercises 37–42: Determine if the statement is true or false, and justify your answer.

37.

(a) If S is a subspace of V, then $\dim(S)<\dim(V)$.

(b) $\mathbb{R}^{4\times 3}$ and P_{11} have the same dimension.

38.

(a) The size of a vector space basis varies from one basis to another.

(b) There is no linearly independent subset $\mathcal{V}$ of P_5 containing seven elements.

39.

(a) No two vector spaces can share the same dimension.

(b) If V is a vector space with $\dim(V)=6$ and S is a subspace of V with $\dim(S)=6$, then $S=V$.

40.

(a) If V is a finite dimensional vector space, then V cannot contain an infinite linearly independent subset $\mathcal{V}$.

(b) If V_1 and V_2 are vector spaces and $\dim(V_1)<\dim(V_2)$, then $V_1\subset V_2$.

41.

(a) If $\mathcal{V}$ spans a vector space V, then vectors can be added to $\mathcal{V}$ to produce a basis for V.

(b) If V is a finite dimensional vector space, then every subspace of V must also be finite dimensional.

42.

(a) If $\{v_1,\dots,v_k\}$ is a basis for a vector space V, then so is $\{cv_1,\dots,cv_k\}$, where c is a scalar.

(b) If S_1 is a subspace of a vector space V and $\dim(S_1)=1$, then the only proper subspace of S_1 is $S_2=\{0\}$.

43. Prove that if $\{v_1, v_2,\dots,v_k\}$ is a basis for a vector space V, then so is $\{v_1, 2v_2,\dots,kv_k\}$.

44. Prove that $\dim(\mathbb{R}^\infty)=\infty$.

45. Show that

$$\{[1000], [0100], [0010], [0001]\}$$

is a basis for $\mathbb{R}^{2\times 2}$.

46. Give a basis for $\mathbb{R}^{n\times m}$, and justify that your set forms a basis. Prove that $\dim(\mathbb{R}^{n\times m})=mn$.

47. If V is a vector space with $\dim(V)=m$, prove that there exist subspaces $S_0, S_1,\dots,S_m$ of V such that $\dim(S_k)=k$.

48. Prove that $\dim(T(m, n))=mn$. (Hint: Recall that if $T : \mathbb{R}^m\to\mathbb{R}^n$ is a linear transformation, then $T(x)=Ax$ for some $n\times m$ matrix A.)

49. Prove Theorem 7.11: The set $\{v_1,\dots,v_m\}$ is a basis for a vector space V if and only if the equation

$$c_1v_1+\ldots+c_mv_m=v$$

has a unique solution $c_1,\ldots,c_m$ for every v in V. (Hint: See Theorem 7.9.)

50. Prove Theorem 7.12: Suppose that $\mathcal{V}_1$ and $\mathcal{V}_2$ are both bases of a vector space V. Prove that $\mathcal{V}_1$ and $\mathcal{V}_2$ have the same number of elements. (Note: Be sure to address the possibility that $\mathcal{V}_1$ and $\mathcal{V}_2$ both have infinitely many vectors.)

51. Let V_1 and V_2 be vector spaces with V_1 a subset of V_2.

(a) If $\dim(V_2)$ is finite, prove that $\dim(V_1)=\dim(V_2)$ if and only if $V_1=V_2$.

(b) Give an example showing that (a) need not be true if $\dim(V_2)=\infty$.

52. Prove Theorem 7.14: Let $\mathcal{V}=\{v_1,\ldots,v_m\}$ be a subset of a finite dimensional vector space V, and suppose that $\mathcal{V}$ is not a basis of V.

(a) Prove that if $\mathcal{V}$ spans V, then vectors can be removed from $\mathcal{V}$ to form a basis for V.

(b) Prove that if $\mathcal{V}$ is linearly independent, then vectors can be added to $\mathcal{V}$ to form a basis for V.

53. Prove Theorem 7.16: Let $\mathcal{V}=\{v_1,\ldots,v_m\}$ be a subset of a vector space V with $\dim(V)=n$.

(a) Prove that if $m<n$, then $\mathcal{V}$ does not span V.

(b) Prove that if $m>n$, then $\mathcal{V}$ is linearly dependent.

54. Prove Theorem 7.17: Let $\mathcal{V}=\{v_1,\ldots,v_m\}$ be a subset of a vector space V with $\dim(V)=m$. If $\mathcal{V}$ is linearly independent *or* spans V, then $\mathcal{V}$ is a basis for V.

SUPPLEMENTARY EXERCISES

Exercises 1–4: Prove that V is a vector space.

1. V=C [0, 1], the set of continuous functions defined on the interval [0,1], together with the standard pointwise definition of addition and scalar multiplication of functions.
2. V=R4×3, the set of real 4×3 matrices together with the usual definition of matrix addition and scalar multiplication.
3. V=En, the set of polynomials with real coefficients and even degree no greater than n, together with the usual definition of polynomial addition and scalar multiplication.
4. V=E, the set of polynomials with real coefficients and even degree, together with the usual addition and scalar multiplication of polynomials.

Exercises 5–8: Prove that S is a subspace of V.

5. V=C(R) and S is the set of functions of the form f(x)=ax3+b cos(x)+ce−x, where a, b, and c are any real numbers.
6. V=Rn×n and S is the set of diagonal n×n matrices.
7. V=C(R) and S is the set of continuous functions f such that $f(-3) = f(2)$.
8. V=T(2, 2) and S is the set of linear transformations such that T ([x1x2])=[0ax2] for some real number a.

Exercises 9–12: Determine if v is in the subspace S.

9. S=span{x2+3x−1, 2x2−5}, v=3x2+x−1
10. S=span{−x2+x−1, 2x2+1}, v=2x−1
11. S=span{[3−214], [−302−1]}, v=[3−320]
12. S=span{[210−121], [104−21−1]}, v=[32−1−520]

Exercises 13–16: Determine if the subset is linearly independent in the given vector space.

13. {x2+3x−1, 2x2−5, −x2+x} in $\mathbf{P}^2$.

14. {−x3+4x−1, 2x2+1, 3, x3+2x−1} in P3.

15. {[−312−5], [1−20−1], [2−132]} in R2×2.

16. {[31−1012], [2−10−22−1], [1−31−230]} in R2×3.

Exercises 17–20: Determine by inspection if the set $\mathcal{V}$ could possibly be a basis for V.

17. $\mathcal{V}$={5x2+4x−1, x2−3}, V=P2.

18. $\mathcal{V}$={5x3+4x2−x, x2+1, 4x−2, 5}, V=P3.

19. $\mathcal{V}$={[−521−3], [324−1], [01−22]}, V=R2×2.

20. $\mathcal{V}$={[−124−3], [021−1], [0023], [000−5]}, V=R2×2.

Exercises 21–24: Determine if the set $\mathcal{V}$ is a basis for V.

21. $\mathcal{V}$={x2−x−1, 2x2−3, 2x−1}, V=P2.

22. $\mathcal{V}$={x3−x, x3+1, 4x−2, 5x2}, V=P3.

23. $\mathcal{V}$={[−1210], [121−1], [−212−3], [−254−4]}, V=R2×2.

24. $\mathcal{V}$={[101−1], [031−2], [11−13], [−102−5]}, V=R2×2.

Exercises 25–28: Determine the dimension of the described subspace.

25. The set of diagonal matrices in R4×4.

26. The set of polynomials p(x) in $\mathbf{P}^2$ such that p(0)=p(2).

27. The set of polynomials p(x) in P4 such that p″(x)=0.

28. span{[−1210], [121−1], [−212−3], [−254−4]} in R2×2.

CHAPTER 8

Orthogonality

Michael Rainwater/Getty Images

Shown in the photo is a roundabout in the Pudong district of Shanghai, China. Civil engineers and city planners can use roundabouts to control the flow of traffic at intersections without the use of traffic lights, which conserves energy. Typically vehicles also spend less time idling at roundabouts than they would at traffic lights, which can lead to less pollution from vehicle exhaust. This roundabout also includes an elevated pedestrian promenade where people can safely navigate the intersection away from vehicular traffic while also taking in the surrounding sights of the city.

In this chapter we shift our focus from vector spaces back to Euclidean space Rn. Many of the topics developed here for Euclidean space are extended to vector spaces in Chapter 10.

In Section 8.1 we introduce and study the dot product, which provides an algebraic formula for determining if two vectors are perpendicular, or equivalently, *orthogonal*. The familiar notions of perpendicular, angle, and length in R2 and R3 still hold here and are extended to higher dimensions by using the dot product. Section 8.2 introduces projections of vectors and the Gram–Schmidt process, which is an algorithm for converting a linearly independent set of vectors into an orthogonal set of vectors. The remaining three sections are applications of orthogonal vectors. Section 8.3 and Section 8.4 focus on matrix factorizations, and Section 8.5 on the problem of fitting functions to data.

8.1 Dot Products and Orthogonal Sets

The *dot product* of two vectors is a form of multiplication of vectors in Rn. Unlike vector addition, which produces a new vector, the dot product of two vectors yields a scalar.

▶ The order of Chapter 7 and Chapter 8 can be reversed, if preferred.

DEFINITION 8.1 ▶

Dot Product

Suppose that

$$u=[u1⋮un] \text{ and } v=[v1⋮vn]$$

are both in Rn. Then the **dot product** of u and v is given by

$$u·v=u1v1+. . .+unvn$$

▶ An alternative way to define the dot product of u and v is with matrix multiplication, by
$u·v=uTv$

This is discussed in Exercise 69.

Example 1

Find $u·v$ for $u=[−132]$ and $v=[71−5]$.

Solution Applying Definition 8.1, we have

$$u \cdot v = (-1)(7) + (3)(1) + (2)(-5) = -14$$

It is not hard to see that $u \cdot v = v \cdot u$. Theorem 8.2 includes this and other properties of the dot product. Note the similarity to properties of arithmetic of real numbers.

THEOREM 8.2 ▶

Let u, v, and w be in R^n, and let c be a scalar. Then

(a) $u \cdot v = v \cdot u$

(b) $(u+v) \cdot w = u \cdot w + v \cdot w$

(c) $(cu) \cdot v = u \cdot (cv) = c(u \cdot v)$

(d) $u \cdot u \geq 0$, and $u \cdot u = 0$ only when $u = 0$

▶ Note that the properties given in Theorem 8.2 are similar to properties of arithmetic of the real numbers.

The proof of Theorem 8.2 is left as Exercise 60. By combining the properties (b) and (c), it can be shown (see Exercise 61) for $u_1, \ldots, u_k$ and w in R^n and scalars $c_1, \ldots, c_k$ that

$$(c_1u_1 + c_2u_2 + \cdots + c_ku_k) \cdot w = c_1(u_1 \cdot w) + c_2(u_2 \cdot w) + \cdots + c_k(u_k \cdot w) \qquad (1)$$

Example 2

Suppose that

$$u = [21-32],\ v = [-1403],\ w = [5012]$$

and $c = -3$. Show that Theorem 8.2(b) and Theorem 8.2(c) hold for these vectors and this scalar.

Solution Starting with Theorem 8.2(b), we have

$$(u+v)\cdot w=[15-35]\cdot[5012]=5+0-3+10=12$$

and

$$u\cdot w+v\cdot w=(10+0-3+4)+(-5+0+0+6)=11+1=12$$

Thus $(u+v)\cdot w=u\cdot w+v\cdot w$. For Theorem 8.2(c), we compute

$$(-3u)\cdot v=[-6-39-6]\cdot[-1403]=(6-12+0-18)=-24$$

$$u\cdot(-3v)=[21-32]\cdot[3-120-9]=(6-12+0-18)=-24$$

$$-3(u\cdot v)=-3(-2+4+0+6)=-24$$

Therefore $(-3u)\cdot v=u\cdot(-3v)=-3(u\cdot v)$.

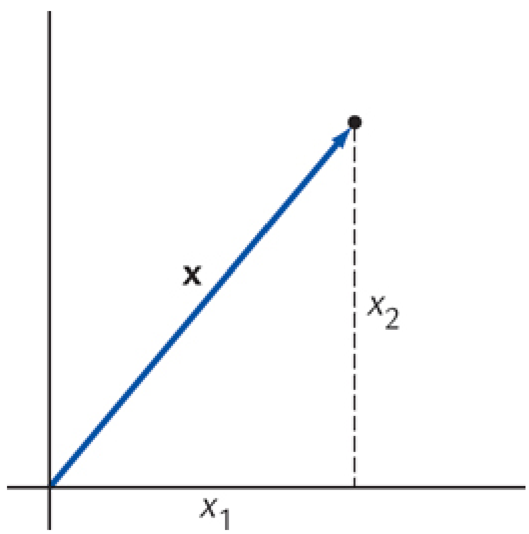


Figure 1 $x=[x1x2]$.

The dot product can be used to measure distance. Suppose that $x=[x1x2]$ is a vector in R2, as shown in Figure 1. If we denote the length of x by $\|x\|$, then from the Pythagorean Theorem we know that

$$\|x\|2=x12+x22$$

By the definition of the dot product, $x \cdot x = x12 + x22$, so that we have

$$\|x\|2 = x \cdot x \Rightarrow \|x\| = x \cdot x$$

This suggests a way to extend the definition of “length” to vectors in Rn. In this setting we generally refer to this as the *norm* of a vector.

DEFINITION 8.3 ▶

Norm of a Vector

Let x be a vector in Rn. Then the **norm** (or **length**) of x is given by

$$\|x\| = x \cdot x$$

For a scalar *c* and a vector x, it follows from Theorem 8.2c (see Exercise 62) that

$$\|cx\| = |\, c|\|x\| \tag{2}$$

Example 3

Find $\|x\|$ and $\|-5x\|$ for $x = [-314]$.

Solution We have $x \cdot x = (-3)2 + (1)2 + (4)2$, so

$$\|x\| = 9 + 1 + 16 = 26$$

By 2,

$$\|-5x\| = |-5|\, \|x\| = 526$$

Among the real numbers, we measure the distance between *r* and *s* by computing |r−s|. This serves as a model for using norms to define the distance between vectors.

DEFINITION 8.4 ▶

Distance Between Vectors

For two vectors u and v in Rn, the **distance between** u **and** v is given by ‖u−v‖.

Example 4

Compute the distance between u=[−132] and v=[71−5].

Solution We have u−v=[−827], so that

‖u−v‖=(−8)2+(2)2+(7)2=117

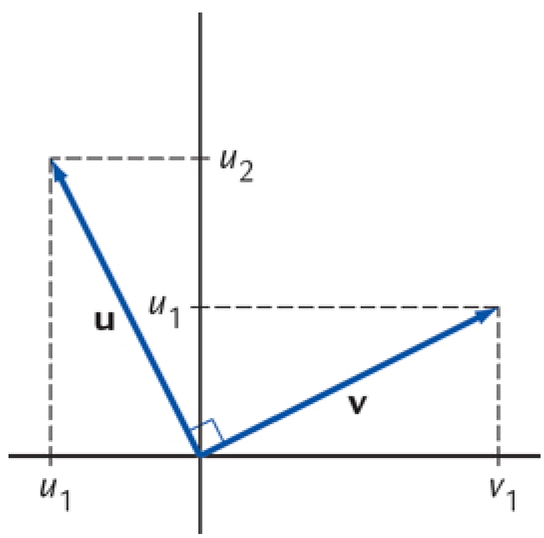


Figure 2 u and v are perpendicular vectors.

Orthogonal Vectors

Suppose that we have two vectors

$$u=[u1u2] \text{ and } v=[v1v2]$$

in R2 that have equal length and are perpendicular to each other, as shown in Figure 2. Rotating u by 90° gives us v, so that v1=u2 and v2=−u1. Therefore we have

$$u \cdot v=u1v1+u2v2=u1u2-u2u1=0$$

Even if the vectors are not the same length, after scaling we can use this argument to show that if two vectors are perpendicular, then u·v=0. (See Exercise 70 for another way to show this.) The reverse holds as well: If u·v=0, then u and v are perpendicular. The same is true in R3, which suggests using the dot product to extend the notion of perpendicular to higher dimensions. The term *orthogonal* is more commonly used and is equivalent to perpendicular.

DEFINITION 8.5 ▶

Orthogonal Vectors

Vectors u and v in Rn are **orthogonal** if u·v=0.

Example 5

Determine if any pair among u, v, and w is orthogonal.

$$u=[2-15-2],\ v=[32-40],\ w=[2964]$$

Solution We have

$$u \cdot v = (2)(3) + (-1)(2) + (5)(-4) + (-2)(0) = -16 \Rightarrow \text{Not Orthogonal} \quad u \cdot w = (2)(2) + (-1)(9) + (5)(6) + (-2)(4) = 17 \Rightarrow \text{Not Orthogonal} \quad v \cdot w = (3)(2) + (2)(9) + (-4)(6) + (0)(4) = 0 \Rightarrow \text{Orthogonal}$$

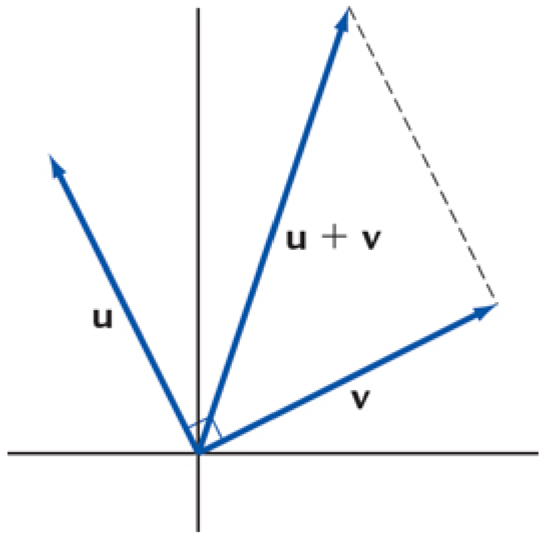


Figure 3 The vectors u+v and v together with the dashed line (which is u translated) form a right triangle.

Figure 3 shows the orthogonal vectors u and v from Figure 2, together with u+v. Note that the vectors v and u+v together with the dashed line (which is u translated) form a right triangle. Hence, by the Pythagorean Theorem, we expect

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2$$

This formulation of the Pythagorean Theorem extends to R^n and is true exactly when u and v are orthogonal—that is, when $u \cdot v = 0$.

THEOREM 8.6 ▸

(PYTHAGOREAN THEOREM) Suppose that u and v are in R^n. Then

$$\|\mathbf{u}+\mathbf{v}\|^2=\|\mathbf{u}\|^2+\|\mathbf{v}\|^2 \text{ if and only if } \mathbf{u}\cdot\mathbf{v}=0$$

Proof It is possible to verify the theorem in $\mathbb{R}^2$ using geometric arguments, but that approach is hard to extend to $\mathbb{R}^n$. Instead, we use an algebraic argument that works for any dimension. We have

$$\|\mathbf{u}+\mathbf{v}\|^2=(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) =\mathbf{u}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v} =\|\mathbf{u}\|^2+\|\mathbf{v}\|^2+2(\mathbf{u}\cdot\mathbf{v})$$

Thus the equality

$$\|\mathbf{u}+\mathbf{v}\|^2=\|\mathbf{u}\|^2+\|\mathbf{v}\|^2$$

holds exactly when $\mathbf{u}\cdot\mathbf{v}=0$. ■

Example 6

Verify the Pythagorean Theorem for the vectors $\mathbf{v}$ and $\mathbf{w}$ in Example 5.

Solution In Example 5 we showed that $\mathbf{v}\cdot\mathbf{w}=0$, so by the Pythagorean Theorem we expect that $\|\mathbf{v}+\mathbf{w}\|^2=\|\mathbf{v}\|^2+\|\mathbf{w}\|^2$. Computing each term, we find

$$\|\mathbf{v}+\mathbf{w}\|^2=\left\|\begin{bmatrix}5\\11\\2\\4\end{bmatrix}\right\|^2=5^2+11^2+2^2+4^2=166$$
$$\|\mathbf{v}\|^2=3^2+2^2+(-4)^2+0^2=29$$
$$\|\mathbf{w}\|^2=2^2+9^2+6^2+4^2=137$$

Since $29+137=166$, we have $\|\mathbf{v}+\mathbf{w}\|^2=\|\mathbf{v}\|^2+\|\mathbf{w}\|^2$.

The Pythagorean Theorem applies to orthogonal vectors. There are two inequalities that apply to any pair of vectors in $\mathbb{R}^n$.

THEOREM 8.7 ▶

(CAUCHY–SCHWARZ INEQUALITY) If u and v are in Rn, then

$$|u\cdot v| \le \| u\|\| v\| \qquad (3)$$

Proof If u=0 or v=0 then both sides of (3) are zero and the inequality holds. Now assume that both u and v are nonzero vectors. Then we have

$$0\le\|u\|u\|2-v\|v\|2\|2 =\|u\|u\|\|2+\|v\|v\|\|2-2(u\|u\|\cdot v\|v\|) =2\text{-}2u\cdot v\|u\|\|v\|$$

Dividing by 2 on both sides and reorganizing gives us

$$u\cdot v\le\| u\|\| v\| \qquad (4)$$

Inequality (4) holds if we replace u with −u, which yields

$$-u\cdot v\le\|-u\|\| v\| \Rightarrow -(u\cdot v)\le\| u\|\| v\| \qquad (5)$$

Combining (4) and (5) shows that (3) holds and completes the proof. ■■

Example 7

Verify the Cauchy–Schwarz inequality for the vectors

$$u=[13-2] \quad \text{and} \quad v=[2-54]$$

Solution We have

$$u \cdot v = (1)(2)+(3)(-5)+(-2)(4)=21 \quad \| u\|=12+32+(-2)2=14 \quad \| v\|=22+(-5)2+42=45$$

Thus $|u\cdot v|=21$ and $\| u\|\| v\|=1445\approx 25.10$, so the Cauchy–Schwarz inequality (3) is verified for u and v.

The second inequality resembles the Pythagorean Theorem, but it is not the same and it applies to all vectors instead of just orthogonal pairs.

THEOREM 8.8

(TRIANGLE INEQUALITY) If u and v are in Rn, then

$$\|u+v\|\leq\| u\|+\| v\| \qquad (6)$$

Geometrically we can think of u, v, and u+v as the three sides of a triangle. The Triangle inequality says that the sum of the lengths of two sides u and v must be greater than or equal to the length of the third side u+v.

Proof We can use the Cauchy–Schwarz inequality to show that (6) holds.

$$\|u+v\|2=(u+v)\cdot(u+v) =\| u\|2+\| v\|2+2(u\cdot v) \leq\| u\|2+\| v\|2+2\| u\|\| v\|(\text{by Cauchy–Schwarz}) =(\| u\|+\| v\|)2$$

Taking square roots on both sides gives us (6) and completes the proof.

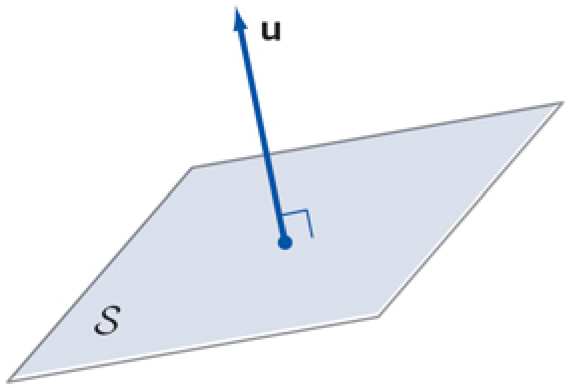


Figure 4 A subspace *S* and orthogonal vector u.

Example 8

Verify the Triangle inequality for u and v in Example 7.

Solution We have

$$u+v=[3-22]\Rightarrow \|u+v\|=32+(-2)2+22=17.$$

We already know that $\|u\|=14$ and $\|v\|=45$, so $\|u\|+\|v\|=14+45$ which is greater than 17. Therefore the Triangle inequality is verified for u and v.

Orthogonal Subspaces

Now that we are acquainted with orthogonal vectors, let's turn to the matter of how sets and vectors can be orthogonal.

DEFINITION 8.9 ▸

Orthogonal Complement

Let *S* be a subspace of Rn. A vector u is **orthogonal** to *S* if u·s=0 for every vector s in *S*. The set of all such vectors u is called the **orthogonal complement** of S and is denoted by $S^{\perp}$.

Figure 4 shows an example in R3, where we have a subspace S and an orthogonal vector u. Note that the orthogonal complement $S^{\perp}$ consists precisely of the multiples of u, so that $S^{\perp}$ is also a subspace of R3. This is true for the orthogonal complement of any subspace.

THEOREM 8.10 ▶

If S is a subspace of Rn, then so is $S^{\perp}$.

▶ $S^{\perp}$ typically is said aloud as "S-perpendicular" or just "S-perp."

Proof Recall from Definition 4.1, Section 4.1, that a subspace must satisfy three conditions.
(a) Since 0·s=0 for all s in S, it follows that $S^{\perp}$ contains 0.
(b) Suppose that u1 and u2 are in $S^{\perp}$. For any s in S, we have

$$(u1+u2)\cdot s=u1\cdot s+u2\cdot s=0+0=0$$

Hence u1+u2 is in $S^{\perp}$, so that $S^{\perp}$ is closed under addition.

(c) If c is a scalar and u is in $S^{\perp}$, then for any s in S we have

$$(cu)\cdot s=c(u\cdot s)=c(0)=0$$

Therefore cu is also in $S^{\perp}$ and hence $S^{\perp}$ is closed under scalar multiplication.

Since parts (a)–(c) of Definition 4.1 hold, $S^{\perp}$ is a subspace. ■■

How do we find $S^{\perp}$? Since a nonzero subspace S contains an infinite number of vectors, it appears that determining if u is in $S^{\perp}$ will require checking that $u \cdot s = 0$ for every s in S. Fortunately, there is another option. Theorem 8.11 shows that we need only check the basis vectors.

THEOREM 8.11 ▶

Let $V = \{s_1, \ldots, s_k\}$ be a basis for a subspace S and u be a vector. Then

$$u \cdot s_1 = 0, u \cdot s_2 = 0, \ldots, u \cdot s_k = 0$$

if and only if u is in $S^{\perp}$.

Proof First, suppose that $u \cdot s_1 = 0, \ldots, u \cdot s_k = 0$. If s is in S, then since V is a basis there exist unique scalars $c_1, \ldots, c_k$ such that

$$s = c_1 s_1 + \cdots + c_k s_k$$

Therefore

$$u \cdot s = u \cdot (c_1 s_1 + \cdots + c_k s_k) = c_1 (u \cdot s_1) + \cdots + c_k (u \cdot s_k) = 0$$

and so u is in $S^{\perp}$.

The reverse direction of the proof is easier. If u is in $S^{\perp}$, then since each of $s_1, \ldots, s_k$ are in S, it follows that $u \cdot s_1 = 0, \ldots, u \cdot s_k = 0$. ■■

Example 9

Let

$$s1=[2-130], s2=[14-2-1], \text{ and } u=[3605]$$

Suppose that S=span{s1, s2}. Determine if u is in $S^{\perp}$, and find a basis for $S^{\perp}$.

Solution To check if u is in $S^{\perp}$, by Theorem 8.11 we need only determine if s1·u=0 and s2·u=0.

$$s1\cdot u = 6-6+0+0=0 s2\cdot u = 3+24+0-5=22$$

Since s2·u=0, we know that u is not in $S^{\perp}$. To find a basis for $S^{\perp}$, we start by forming the matrix

$$A=[2-13014-2-1]$$

Note that the *rows* of *A* are made up of the components of s1 and s2, so that if u is in R4, then

$$Au=[s1\cdot us2\cdot u]$$

Thus u is in $S^{\perp}$ exactly when Au=0. To solve this equation we use our standard procedure for determining the general solution of a linear system; we find that

$$u=c1[-10790]+c2\ [1209]$$

The two vectors in the general solution give a basis for $S^{\perp}$, so we have

$$S^{\perp}=\text{span}\{[-10790], [1209]\}$$

Orthogonal Sets

We have defined what it means for a vector to be orthogonal to another vector and to a set of vectors. Our next step is to define what it means for a set of vectors to be orthogonal.

DEFINITION 8.12 ▶

Orthogonal Set

A set of vectors *V* in Rn form an **orthogonal set** if vi·vj=0 for all vi and vj in *V* with i=j.

Example 10

Show that {v1, v2, v3} is an orthogonal set, where

$$v1=[14-1],v2=[11-17],v3=[3-2-5]$$

Solution We show that {v1, v2, v3} is an orthogonal set by showing that each distinct pair of vectors is orthogonal.

$$v1\cdot v2=11-4-7=0v1\cdot v3=3-8+5=0v2\cdot v3=33+2-35=0$$

Since the dot products are all zero, the set is orthogonal.

A useful feature of orthogonal sets of nonzero vectors is that they are linearly independent.

THEOREM 8.13 ▶

An orthogonal set of nonzero vectors is linearly independent.

Proof Let {v1 ,… , vk} be a set of nonzero orthogonal vectors. Suppose that the linear combination

$$c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k=\mathbf{0} \tag{7}$$

To show that $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ is a linearly independent set, we need to show that (7) holds only when $c_1=\cdots=c_k=0$. To do so, we start by noting that since $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2,\ldots,\mathbf{v}_k$,

$$\begin{aligned}\mathbf{v}_1\cdot(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k)&=c_1(\mathbf{v}_1\cdot\mathbf{v}_1)+c_2(\mathbf{v}_1\cdot\mathbf{v}_2)+\cdots\\ &\quad+c_k(\mathbf{v}_1\cdot\mathbf{v}_k) =c_1\|\mathbf{v}_1\|^2+c_2(0)+\cdots+c_k(0)=c_1\|\mathbf{v}_1\|^2\end{aligned}$$

Because $c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k=\mathbf{0}$, it follows that $\mathbf{v}_1\cdot(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k)=0$, and thus

$$c_1\|\mathbf{v}_1\|^2=0$$

Since $\mathbf{v}_1\neq\mathbf{0}$ we know that $\|\mathbf{v}_1\|^2\neq 0$ and so it must be that $c_1=0$. Repeating this argument with each of $\mathbf{v}_2,\ldots,\mathbf{v}_k$ shows that $c_2=0,\ldots,c_k=0$, and hence $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ is a linearly independent set. ■ ■

Orthogonal Basis

A set of orthogonal vectors that forms a basis is called an **orthogonal basis**. Orthogonal sets are particularly useful for forming a basis, in part because orthogonal vectors are automatically linearly independent, and also because the dot product can be used to form linear combinations.

THEOREM 8.14 ▶

Let S be a subspace with an orthogonal basis $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$. Then any vector $\mathbf{s}$ in S can be expressed as

$$\mathbf{s}=c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k$$

where $c_i=\frac{\mathbf{v}_i\cdot\mathbf{s}}{\|\mathbf{v}_i\|^2}$ for $i=1,\ldots,k$.

Proof The proof is similar to that of Theorem 8.13. We know that there exist unique scalars $c_1, \ldots, c_k$ such that

$$s = c_1 v_1 + \cdots + c_k v_k$$

Taking the dot product of s_1 with s, we have

$$v_1 \cdot s = v_1 \cdot (c_1 v_1 + \cdots + c_k v_k) = c_1 (v_1 \cdot v_1) + c_2 (v_1 \cdot v_2) + \cdots + c_k (v_1 \cdot v_k) = c_1 (v_1 \cdot v_1) + c_2(0) + \cdots + c_k(0) = c_1 (v_1 \cdot v_1) = c_1 \| v_1 \|^2$$

Solving for c_1, we find that

$$c_1 = \frac{v_1 \cdot s}{\| v_1 \|^2}$$

A similar argument gives the formulas for $c_2, \ldots, c_k$. ■

Example 11

Verify that the set

$$v_1 = [-2\ 1\ -1], v_2 = [1\ -1\ -3], v_3 = [4\ 7\ -1]$$

forms an orthogonal basis for R^3, and express $s = [3\ -1\ 5]$ as a linear combination of v_1, v_2, and v_3.

Solution We verify orthogonality by computing dot products, which are

$$v_1 \cdot v_2 = -2 - 1 + 3 = 0 \quad v_1 \cdot v_3 = -8 + 7 + 1 = 0 \quad v_2 \cdot v_3 = 4 - 7 + 3 = 0$$

By Theorem 8.13 the vectors are linearly independent, and since there are three of them in R^3, the set also spans and hence is a basis for R^3.

To express s as a linear combination of v_1, v_2, and v_3, we apply Theorem 8.14, starting with the computations

$c1=v1 \cdot s\|v1\|2=-6-1-54+1+1=-126=-2c2=v2 \cdot s\|v2\|2=3+1-151+1+9=-1111=-1c3=v3 \cdot s\|v3\|2=12-7-516+49+1=066=0$

Therefore we have $s=-2v1-v2$.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Let

$$u1=[2-103],\ u2=[1-142],\ u3=[110-3]$$

(a) Compute $u1 \cdot u3$ and $-u2 \cdot (2u1)$.

(b) Compute the distance between u2 and $-2u3$.

2. Find all values of *a* and *b* (if any) so that u1, u2, and u3 form an orthogonal set.

$$u1=[2a-1],u2=[120],u3=[b-13],$$

3. If u1 and u2 are orthogonal vectors with $\|u1\|=4$ and $\|u2\|=1$, then find $\|2u1+3u2\|$.

4. Verify the Pythagorean Theorem, Cauchy–Schwarz inequality, and Triangle inequality for

$$u1=[31-1],u2=[125]$$

5. Find a basis for $S^{\perp}$ if $S=\text{span}\{[10-2]\}$.

EXERCISES

Exercises 1–8 refer to the vectors u1 to u8.

$$u1=[-312],u2=[111],u3=[20-1],u4=[1-32],$$

$$u5=[211],u6=[03-1],u7=[3-4-2],u8=[-1-13]$$

1. Compute the following dot products.

(a) $u1 \cdot u5$

(b) $u3 \cdot (-3u2)$

(c) $u4 \cdot u7$

(d) $2u4 \cdot u7$

2. Compute the following dot products.

(a) $3u7 \cdot u3$

(b) $u1 \cdot u1$

(c) $u2 \cdot (-2u5)$

(d) $2u2 \cdot (-u5)$

3. Compute the norms of the given vectors.

(a) $u7$

(b) $-u7$

(c) $2u5$

(d) $-3u5$

4. Compute the norms of the given vectors.

(a) $u8$

(b) $3u8$

(c) $-u2$

(d) $-2u2$

5. Compute the distance between the given vectors.

(a) $u1$ and $u2$

(b) $u3$ and $u8$

(c) $2u6$ and $-u3$

(d) $-3u2$ and $2u5$

6. Compute the distance between the given vectors.

(a) $u5$ and $u1$

(b) $u2$ and $u8$

(c) $-2u3$ and $u8$

(d) $4u2$ and $-2u6$

7. Determine if the given vectors are orthogonal.

(a) $u1$ and $u3$

(b) $u3$ and $u4$

(c) u2 and u5
(d) u1 and u8

8. Determine if the given vectors are orthogonal.

(a) u2 and u3
(b) u1 and u2
(c) u8 and u5
(d) u3 and u6

Exercises 9–12: Find all values of *a* so that u and v are orthogonal.

9. u=[a2−3], v=[4a3]

10. u=[−1a5], v=[7a−2]

11. u=[2a−3−1], v=[−546a]

12. u=[1−5a0], v=[−41a−2]

Exercises 13–16: Determine if the given vectors form an orthogonal set.

13. u1=[1−2], u2=[43]

14. u1=[123], u2=[5−41], u3=[11−1]

15. u1=[22−1], u2=[−51316], u3=[5−42]

16. u1=[120−1], u2=[5249], u3=[−22−32]

Exercises 17–18: Find all values of *a* (if any) so that the given vectors form an orthogonal set.

17. u1=[−102], u2=[432], u3=[6a3]

18. u1=[1−32−1], u2=[4210], u3=[−10a7]

Exercises 19–20: Find all values of *a* and *b* (if any) so that the given vectors form an orthogonal set.

19. u1=[21−1], u2=[3−42], u3=[2ab]

20. u1=[1−361], u2=[21a−5], u3=[0−43b]

Exercises 21–24: Verify that the Pythagorean Theorem holds for the given orthogonal vectors.

21. u1=[3−1], u2=[13]

22. u1=[68], u2=[−43]

23. u1=[2−31], u2=[431]

24. u1=[5−242], u2=[293−2]

25. Suppose u1 and u2 are orthogonal vectors, with ‖u1‖=2 and ‖u2‖=5. Find ‖3u1+4u2‖.

26. Suppose u1 and u2 are orthogonal vectors, with ‖u1‖=3 and ‖u2‖=4. Find ‖2u1−u2‖.

Exercises 27–30: Verify the Cauchy–Schwarz inequality and the Triangle inequality for u and v.

27. u=[2−1], v=[42]

28. u=[30], v=[51]

29. u=[22−2], v=[−143]

30. u=[2031], v=[0−132]

Exercises 31–32: Determine if u is orthogonal to the subspace *S*.

31. u=[2−31],S=span{[12−1], [3−11]}

32. u=[0101],S=span{[1101], [1010], [1111]}

Exercises 33–36: Find a basis for $S^{\perp}$ for the subspace *S*.

33. S=span{[1−3]}

34. S=span{[25]}

35. S=span{[11−2]}

36. S=span{[−121], [2−32]}

Exercises 37–38: Show that the given basis for *S* is orthogonal, and then write s as a linear combination of the basis vectors.

37. S=span{[110], [1−14]}, s=[12−2]

38. S=span{[101−1], [3314],[−231−1]}, s=[1001]

FIND AN EXAMPLE Exercises 39–48: Find an example that meets the given specifications.

39. Two vectors u and v such that u·v=12.

40. A vector that is orthogonal to both [10] and [01].

41. A vector u such that ‖u‖=1 and u is orthogonal to [−21].

42. An orthogonal basis for R2 that includes [34].

43. Two linearly independent vectors that are both orthogonal to [20−1].

44. Two vectors in R2 that are orthogonal but do not span R2.

45. A subspace S of R3 that has dim($S^{\perp}$) = 2.

46. Two vectors u and v such that u, v, and [111] form an orthogonal set.

47. Three vectors in R3 that form an orthogonal set but not an orthogonal basis.

48. A subspace S of R4 such that dim(S) = dim($S^{\perp}$) = 2.

TRUE OR FALSE Exercises 49–54: Determine if the statement is true or false, and justify your answer.

49.

(a) If ‖u−v‖=3, then the distance between 2u and 2v is 12.

(b) If u and v have nonnegative entries, then u·v≥0.

50.

(a) ‖u+v‖=‖u‖+‖v‖ for all u and v in Rn.

(b) Suppose that {s1, s2, s3} is an orthogonal set and that c1, c2, and c3 are scalars. Then {c1s1, c2s2, c3s3} is also an orthogonal set.

51.

(a) If S is a one-dimensional subspace of R2, then so is $S^{\perp}$.

(b) If A is an n×n matrix and u is in Rn, then ‖u‖≤‖Au‖.

52.

(a) If u1·u2=0 and u2·u3=0, then u1·u3=0.

(b) If A is an n×n matrix with orthogonal columns, then ATA is a diagonal matrix.

53.

(a) If u and v are orthogonal, then the distance between u and v is ‖u‖2+‖v‖2.

(b) If ‖u−v‖=‖u+v‖, then u and v are orthogonal.

54.

(a) If $A=[a_1\ a_2]$ and $S=\text{span}\{a_1, a_2\}$, then $S^\perp=\text{null}(A)$.

(b) Even if S is merely a nonempty subset of $\mathbb{R}^n$, the orthogonal complement $S^\perp$ is still a subspace.

55. Prove that Theorem 8.11 is true even if the set $S=\{s_1,\ldots, s_k\}$ only spans the subspace S instead of being a basis for S.

56. Prove that the zero vector 0 in $\mathbb{R}^n$ is orthogonal to all vectors in $\mathbb{R}^n$.

57. Prove that the standard basis $\{e_1,\ldots, e_n\}$ of $\mathbb{R}^n$ is an orthogonal basis.

58. Prove that if u_1 and u_2 are both orthogonal to v, then so is u_1+u_2.

59. Prove that if c_1 and c_2 are scalars and u_1 and u_2 are vectors, then $(c_1u_1)\cdot(c_2u_2)=c_1c_2(u_1\cdot u_2)$.

60. Let u, v, and w be in $\mathbb{R}^n$, and let c be a scalar. Prove each part of Theorem 8.2.

(a) $u\cdot v=v\cdot u$

(b) $(u+v)\cdot w=u\cdot w+v\cdot w$

(c) $(cu)\cdot v=u\cdot(cv)=c(u\cdot v)$

(d) $u\cdot u\geq 0$, and $u\cdot u=0$ only when $u=0$

61. Use the properties of Theorem 8.2 to prove equation (1), which says that

$$(c_1u_1+\cdots+c_ku_k)\cdot w=c_1(u_1\cdot w)+\cdots+c_k(u_k\cdot w)$$

62. Prove equation (2), which says that $\|cx\|=|c|\|x\|$ for a scalar c and vector x.

63. Prove that if $u\neq 0$ and $v=\frac{1}{\|u\|}u$, then $\|v\|=1$.

64. Let u be a vector in $\mathbb{R}^n$, and then define $T_u : \mathbb{R}^n\to\mathbb{R}$ by $T_u(v)=u\cdot v$. Show that T_u is a linear transformation.

65. Let S be a subspace. Prove that $S \cap S^\perp=\{0\}$.

66. Prove that

$$\|u+v\|^2+\|u-v\|^2=2(\|u\|^2+\|v\|^2).$$

67. For a matrix A, show that $(\text{col}(A))^\perp=\text{null}(A^T)$.

68. Prove that if *S* is a subspace, then $S=(S^{\perp})^{\perp}$.

69. Let u and v be in R^n and *A* be an $n\times n$ matrix.

(a) Explain why $u\cdot v$ and $u^T v$ are essentially the same.

(b) Show that $(Au)\cdot v=u\cdot(A^T v)$.

70. In this problem we show that if u and v are perpendicular vectors in R^2, then $u\cdot v=0$. The method of proof is different from the one given at the start of the subsection “Orthogonal Vectors.”

(a) Given nonzero vectors u and v orthogonal in R^2, show that $\|u\|$, $\|v\|$, and $\|u-v\|$ are the lengths of the sides of a right triangle.

(b) Use the version of the Pythagorean Theorem you learned in high school geometry to show that

$$\|u\|^2+\|v\|^2=\|u-v\|^2$$

(c) Write the equation in (b) in terms of dot products and then simplify to show that $u\cdot v=0$.

Exercises 71–72: Let

$$u_1=[3-1502],\ u_2=[74028],\ u_3=[03-44-3]$$

71. Compute each of the following:

(a) $u_2\cdot u_3$

(b) $\|u_1\|$

(c) $\|2u_1+5u_3\|$

(d) $\|3u_1-4u_2-u_3\|$

72. Compute each of the following:

(a) $u_2\cdot u_1$

(b) $\|u_3\|$

(c) $\|3u_2+4u_3\|$

(d) $\|-2u_1+5u_2-3u_3\|$

Exercises 73–74: Find a basis for $S^{\perp}$.

73. $S=\text{span}\{[2-135],\ [0174]\}$

74. $S=\text{span}\{[6025-1],\ [5308-6]\}$

8.2 Projection and the Gram–Schmidt Process

In Section 8.1 we developed the idea of an orthogonal basis. There are numerous applications of orthogonal bases, some explored in the other sections of this chapter. Of course, not every basis is an orthogonal basis. For example, suppose that

$$S=\text{span}\{[1011],\ [0203],\ [-3-115]\}$$

It can be shown that this basis for *S* is not orthogonal. However, we can construct an orthogonal basis for *S* from this basis. The key to understanding how this is done is vector projection, so we start there.

Projection onto Vectors

Suppose that we have two nonzero vectors **u** and **v** in R2, as shown in Figure 1a. Draw the line perpendicular to **v** that passes through the tip of **u** (dashed in Figure 1b). The projection of **u** onto **v** (denoted projvu) is the vector parallel to **v** with tip at the intersection of **v** and the dashed perpendicular line (Figure 1c). We can think of projvu as the component of **u** in the direction of **v**.

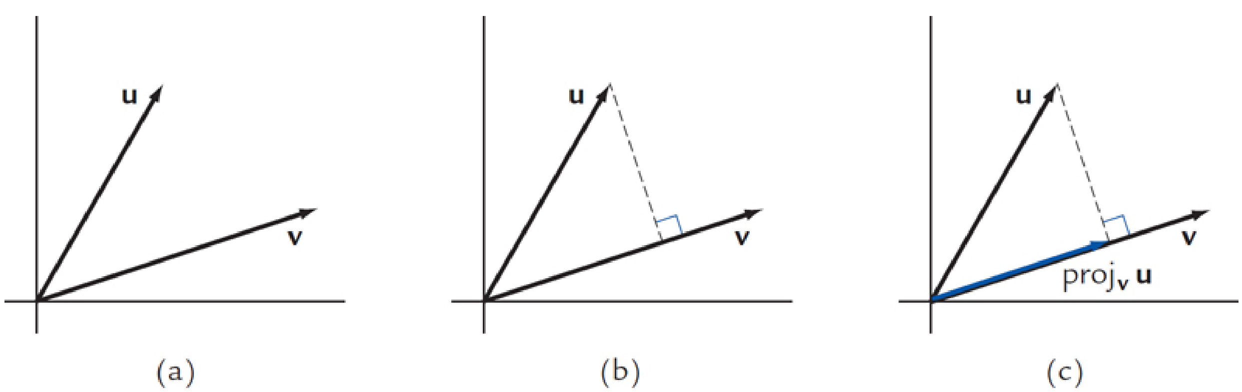


Figure 1 (a) Vectors **u** and **v** in R2; (b) The line perpendicular to **v** that passes through the tip of **u** (dashed); (c) The projection of **u** onto **v**.

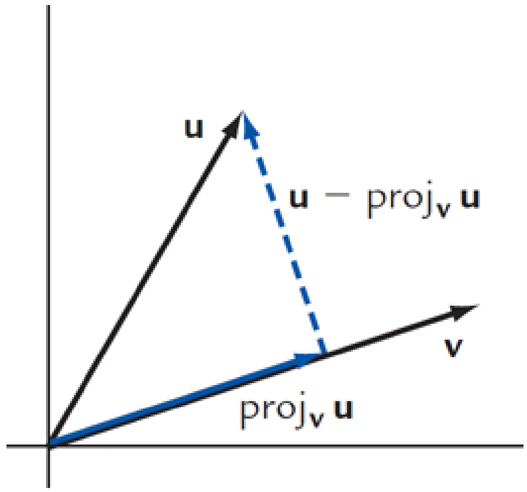


Figure 2 $\mathbf{v}$ and $\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}$ are orthogonal.

Our next step is to develop a formula for $\text{proj}_{\mathbf{v}}\mathbf{u}$. Note that the vectors $\mathbf{v}$ and $\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}$ are orthogonal to each other (see Figure 2). Therefore

$$\mathbf{v}\cdot(\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u})=0 \tag{1}$$

Since $\text{proj}_{\mathbf{v}}\mathbf{u}$ is parallel to $\mathbf{v}$, there exists a scalar c such that $\text{proj}_{\mathbf{v}}\mathbf{u}=c\mathbf{v}$. We can find a formula for c by substituting into (1) and solving for c.

$$\mathbf{v}\cdot(\mathbf{u}-c\mathbf{v})=0 \Rightarrow \mathbf{v}\cdot\mathbf{u}-c(\mathbf{v}\cdot\mathbf{v})=0 \Rightarrow c=\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{v}\cdot\mathbf{v}}=\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}$$

Thus we have

$$\text{proj}_{\mathbf{v}}\mathbf{u}=c\mathbf{v}=\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\,\mathbf{v}$$

Although this formula was developed in $\mathbb{R}^2$, it can also be evaluated for vectors in $\mathbb{R}^n$, so we use it to generalize projection to any dimension.

DEFINITION 8.15 ▶

Projection Onto a Vector

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$, with $\mathbf{v}$ nonzero. Then the **projection of** $\mathbf{u}$ **onto** $\mathbf{v}$ is given by

$$\mathrm{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\,\mathbf{v} \tag{2}$$

Example 1

Find $\mathrm{proj}_{\mathbf{v}}\mathbf{u}$ for

$$\mathbf{u} = \begin{bmatrix} 7 \\ 14 \\ 4 \end{bmatrix} \quad\text{and}\quad \mathbf{v} = \begin{bmatrix} -2 \\ 5 \\ 1 \end{bmatrix}$$

Solution Applying formula (2), we have

$$\mathrm{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\,\mathbf{v} = \frac{(-14+70+4)}{(4+25+1)}\begin{bmatrix} -2 \\ 5 \\ 1 \end{bmatrix} = \frac{60}{30}\begin{bmatrix} -2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 10 \\ 2 \end{bmatrix}$$

Projections onto vectors have several important properties that are summarized in the next theorem.

THEOREM 8.16

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$ ($\mathbf{v}$ nonzero) and c be a nonzero scalar. Then

(a) $\mathrm{proj}_{\mathbf{v}}\mathbf{u}$ is in span$\{\mathbf{v}\}$.

(b) $\mathbf{u}-\mathrm{proj}_{\mathbf{v}}\mathbf{u}$ is orthogonal to $\mathbf{v}$.

(c) if $\mathbf{u}$ is in span$\{\mathbf{v}\}$, then $\mathbf{u}=\mathrm{proj}_{\mathbf{v}}\mathbf{u}$.

(d) $\mathrm{proj}_{\mathbf{v}}\mathbf{u}=\mathrm{proj}_{c\mathbf{v}}\mathbf{u}$.

Proof We take each part in turn.

(a) Since $\mathrm{proj}_{\mathbf{v}}\mathbf{u}=\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\mathbf{v}$ is multiple of $\mathbf{v}$, $\mathrm{proj}_{\mathbf{v}}\mathbf{u}$ must be in span$\{\mathbf{v}\}$.

(b) We verify that $\mathbf{v}$ is orthogonal to $\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}$ by computing the dot product.

$$\mathbf{v}\cdot(\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u})=\mathbf{v}\cdot\left(\mathbf{u}-\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\mathbf{v}\right)=\mathbf{v}\cdot\mathbf{u}-\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}(\mathbf{v}\cdot\mathbf{v})=\mathbf{v}\cdot\mathbf{u}-\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{u}-\mathbf{v}\cdot\mathbf{u}=0$$

(c) If $\mathbf{u}$ is in span$\{\mathbf{v}\}$, then there exists a constant c such that $\mathbf{u}=c\mathbf{v}$. Hence

$$\text{proj}_{\mathbf{v}}\mathbf{u}=\frac{\mathbf{v}\cdot(c\mathbf{v})}{\|\mathbf{v}\|^2}\mathbf{v}=c\frac{\mathbf{v}\cdot\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v}=c\mathbf{v}=\mathbf{u}$$

(d) For any nonzero scalar c,

$$\text{proj}_{c\mathbf{v}}\mathbf{u}=\frac{(c\mathbf{v})\cdot\mathbf{u}}{\|c\mathbf{v}\|^2}(c\mathbf{v})=\frac{c^2}{|c|^2}\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\mathbf{v}=\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{v}\|^2}\mathbf{v}=\text{proj}_{\mathbf{v}}\mathbf{u}$$ ■ ■

Projections onto Subspaces

We can extend the idea of projecting onto a vector to projecting onto subspaces. In a sense, we have already taken a step in this direction. Since Theorem 8.16d shows $\text{proj}_{\mathbf{v}}\mathbf{u}=\text{proj}_{c\mathbf{v}}\mathbf{u}$, we can think of $\text{proj}_{\mathbf{v}}\mathbf{u}$ as projecting $\mathbf{u}$ onto the subspace span$\{\mathbf{v}\}$, the line through the origin in the direction of $\mathbf{v}$.

Given a nonzero subspace S and a vector $\mathbf{u}$, we denote the projection of $\mathbf{u}$ onto S by $\text{proj}_S\mathbf{u}$. Although we do not yet have a definition for $\text{proj}_S\mathbf{u}$, its properties should be analogous to those for $\text{proj}_{\mathbf{v}}\mathbf{u}$ given in Theorem 8.16. In particular, if $\mathbf{u}$ is in S, then we want $\text{proj}_S\mathbf{u}=\mathbf{u}$. At this point, it helps to recall Theorem 8.14 in Section 8.1, which says that if $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ is an orthogonal basis for S, then any $\mathbf{s}$ in S can be expressed

$$\mathbf{s}=\frac{\mathbf{v}_1\cdot\mathbf{s}}{\|\mathbf{v}_1\|^2}\mathbf{v}_1+\frac{\mathbf{v}_2\cdot\mathbf{s}}{\|\mathbf{v}_2\|^2}\mathbf{v}_2+\cdots+\frac{\mathbf{v}_k\cdot\mathbf{s}}{\|\mathbf{v}_k\|^2}\mathbf{v}_k \qquad (3)$$

Given our requirements for $\text{proj}_S\mathbf{u}$, then $\text{proj}_S\mathbf{u}$ must equal the right side of (3) when $\mathbf{u}$ is in S. Moreover, the right side of (3) is also equal to the sum of the projection of $\mathbf{u}$ onto each of $\mathbf{v}_1,\ldots,\mathbf{v}_k$. All of this suggests the following definition.

DEFINITION 8.17 ▶

Projection Onto a Subspace

Let S be a nonzero subspace with orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$. Then the **projection** of $\mathbf{u}$ **onto** S is given by

$$\text{proj}_S \mathbf{u} = \frac{\mathbf{v}_1 \cdot \mathbf{u}}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 + \frac{\mathbf{v}_2 \cdot \mathbf{u}}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 + \cdots + \frac{\mathbf{v}_k \cdot \mathbf{u}}{\|\mathbf{v}_k\|^2} \mathbf{v}_k \qquad (4)$$

See Figure 3 for a graphical depiction of projection onto a plane.

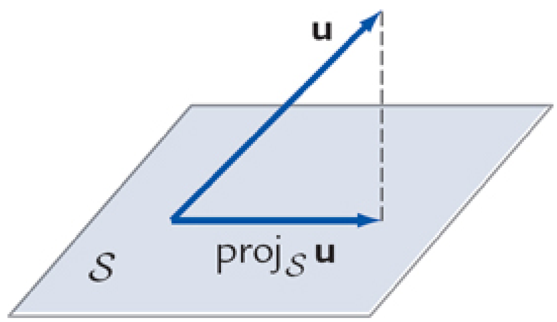


Figure 3 Projection of a vector $\mathbf{u}$ in $\mathbf{R}^3$ onto a two-dimensional subspace S.

Example 2

Find $\text{proj}_S \mathbf{u}$ for $S = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, where

$$\mathbf{u} = \begin{bmatrix} 18 \\ -20 \\ 10 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}$$

Solution The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal, so we may apply (4). We have

$$\text{proj}_S \mathbf{u} = \frac{\mathbf{v}_1 \cdot \mathbf{u}}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 + \frac{\mathbf{v}_2 \cdot \mathbf{u}}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 = \frac{42}{42} \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix} + \frac{34}{17} \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \\ -1 \end{bmatrix}$$

Regarding Definition 8.17:

- If S=span{ v} is a one-dimensional subspace, then (4) reduces to the formula for projvu.
- We can express projSu by

projSu=projv1u+projv2u+⋯+projvku

- The basis {v1,… , vk} for *S* must be orthogonal in order to apply the formula for projSu.

The next theorem shows that projSu does not depend on the choice of orthogonal basis for *S*.

THEOREM 8.18 ▶

Let *S* be a nonzero subspace of Rn with orthogonal basis {v1, … , vk}, and let u be a vector in Rn. Then

(a) projSu is in *S*.

(b) u−projSu is orthogonal to *S*.

(c) if u is in *S*, then u=projSu.

(d) projSu is independent of the choice of orthogonal basis for *S*.

▶ This extends Theorem 8.16 to subspaces.

The proof of Theorem 8.18 is given at the end of the section.

Example 3

Let

u=[3−115], v1=[10−12], v2=[2−120], w1=[1−13−2], w2=[7−216]

It can be shown that {v1, v2} and {w1, w2} both form orthogonal bases for the same subspace *S*. Show that projSu is the same for both bases.

Solution Starting with S=span{v1, v2}, we have

projSu=v1·u‖v1‖2 v1+v2·u‖v2‖2 v2=126 [10−12]+99 [2−120]=
[4−104]

On the other hand, S=span{w1, w2} yields

projSu=w1·u‖w1‖2 w1+w2·u‖w2‖2
w2=−315 [1−13−2]+5490 [7−216]=[4−104]

Both bases produce the same projection vector, as promised by Theorem 8.18.

The Gram–Schmidt Process

Now that we know how to find projections of vectors onto subspaces, we are ready to develop a method for finding an orthogonal basis for a subspace. Let's start with a simple case, an arbitrary two-dimensional subspace S=span{s1, s2} in Rn. Our goal is to find an orthogonal basis for *S*. Let

v1=s1v2=s2−projv1s2

Then $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal by Theorem 8.16b. Moreover, since $\mathbf{v}_1 = \mathbf{s}_1$, it follows that projv1s2=projs1s2=cs1 for some nonzero scalar *c*. Therefore $\mathbf{v}_2 = \mathbf{s}_2 - c\mathbf{s}_1$, so $\mathbf{v}_1$ and $\mathbf{v}_2$ are both in *S*. By Theorem 8.13 in Section 8.1, $\mathbf{v}_1$ and $\mathbf{v}_2$ are also linearly independent. Since dim(*S*) = 2, we may conclude that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for *S*.

Example 4

Let S=span{s1, s2}, where

s1=[1−23] and s2=[35−7]

Find an orthogonal basis for *S*.

Solution By the above formulas, we define

v1=u1=[1−23]

v2=u2−projv1u2=[35−7]−(−28)14 [1−23]=[51−1]

From the above discussion we know that v1·v2=0 and that span{s1, s2}=span{v1, v2}. Thus {v1, v2} forms an orthogonal basis for *S*.

The *Gram–Schmidt process* extends the procedure illustrated in Example 4 and allows us to generate an orthogonal basis for any nonzero subspace.

THEOREM 8.19 ▶

(THE GRAM–SCHMIDT PROCESS) Let *S* be a subspace with basis {s1, s2,… , sk}. Define v1, v2,… , vk, in order, by

v1=s1v2=s2−projv1s2v3=s3−projv1s3−projv2s3v4=s4−projv1s4−projv2s4−projv3s4 ⋮ vk=sk−projv1sk−projv2sk−projvk-1sk

Then {v1, v2,… , vk} is an orthogonal basis for *S*.

▶ Jörgen Gram (1850–1916) was a Danish actuary who worked on the mathematics of accident insurance, and Erhardt Schmidt (1876–1959) was a German mathematician who taught at Berlin University.

At each step of the Gram–Schmidt process, the new vector vj is orthogonal to the subspace

$$\text{span}\{v1,\ldots, vj-1\}=\text{span}\{s1,\ldots, sj-1\}$$

so we build up our basis for *S* by adding vectors orthogonal to those already in place, ensuring an orthogonal basis at the end. A proof that the Gram–Schmidt process works can be performed by induction and is left as an exercise.

Let's return to the problem we encountered at the beginning of the section.

Example 5

Find an orthogonal basis for the subspace S=span{s1, s2, s3}, where

$$s1=[1011],s2=[0203],s3=[-3-115]$$

Solution The first step of the Gram–Schmidt process is the easiest, setting v1=s1. Moving to the next step, we let

$$v2=s2-\text{proj}v1s2=s2-v1\cdot s2\|v1\|2\ v1=[0203]-33\ [1011]=[-12-12]$$

For the last step, we have

$$v3=s3-\text{proj}v1s3-\text{proj}v2s3=s3-v1\cdot s3\|v1\|2\ v1-v2\cdot s3\|v2\|2\ v2$$

$$=[-3-115]-33\ [1011]-1010\ [-12-12]=[-3-312]$$

This gives us the orthogonal basis

$$\{[1011],\ [-12-12],\ [-3-312]\}$$

Orthonormal Bases

In Theorem 8.14 we showed that if *S* is a subspace with orthogonal basis {v1,… ,vk}, then any vector **s** in *S* can be expressed

$$s = \frac{v_1 \cdot s}{\|v_1\|^2} v_1 + \frac{v_2 \cdot s}{\|v_2\|^2} v_2 + \cdots + \frac{v_k \cdot s}{\|v_k\|^2} v_k$$

If each of the vectors v_i also satisfies $\|v_i\| = 1$, then this formula simplifies to

$$s = (v_1 \cdot s)v_1 + (v_2 \cdot s)v_2 + \cdots + (v_k \cdot s)v_k \qquad (5)$$

DEFINITION 8.20 ▶

Orthonormal Set

A set of vectors $\{w_1, \ldots, w_k\}$ is **orthonormal** if the set is orthogonal and $\|w_j\| = 1$ for each of $j = 1, 2, \ldots k$.

To obtain an orthonormal basis for a subspace $S = \text{span}\{s_1, \ldots, s_k\}$ of dimension *k*, we first use Gram–Schmidt to find an orthogonal basis $\{v_1, \ldots, v_k\}$ for *S* and then let

$$w_j = \frac{1}{\|v_j\|} v_j \quad \text{for } j = 1, 2, \ldots k$$

▶ In addition to the simplified formula (5), orthonormal sets are also used in matrix factorizations described in Section 8.3 and Section 8.4.

Normalizing

This step is called **normalizing** the vectors. Since each w_j is a multiple of v_j, the set $\{w_1, \ldots, w_k\}$ is orthogonal and $\text{span}\{v_1, \ldots, v_k\} = \text{span}\{w_1, \ldots, w_k\}$. Furthermore, as

$$\|w_j\| = \left\| \frac{1}{\|v_j\|} v_j \right\| = \frac{1}{\|v_j\|} \|v_j\| = 1$$

the set $\{w_1, \ldots, w_k\}$ is an orthonormal basis for *S*.

Example 6

Find an orthonormal basis for the subspace *S* given in Example 5.

Solution We already have the orthogonal basis

$$v1=[1011],v2=[-12-12],v3=[-3-312]$$

All that remains is to normalize each of v1, v2, and v3 by dividing by their respective lengths. Since ‖v1‖=3, ‖v2‖=10, and ‖v3‖=23, the orthonormal basis is

$$\{13\ [1011],\ 110\ [-12-12],\ 123\ [-3-312]\}$$

Example 7

Let S=span{s1, s2, s3}, where

$$s1=[12-2],s2=[10-4],s3=[520], \text{ and } s=[11-1]$$

Use the Gram–Schmidt process to find an orthonormal basis for *S*, and then write s as a linear combination of the orthonormal basis vectors.

Solution We start by finding an orthogonal basis. After setting v1=s1, we have

$$v2=s2-projv1s2=s2-s2\cdot v1\|v1\|2\ v1=[10-4]-99\ [12-2]=[0-2-2]$$

and

$$v3=s3-projv1s3-projv2s3=s3-s3\cdot v1\|v1\|2\ v1-s3\cdot v2\|v2\|2\ v2$$

$$=[520]-99\ [12-2]\ -(-4)8\ [0-2-2]=[4-11]$$

Now that we have an orthogonal basis, we obtain an orthonormal basis by normalizing each of v1, v2, and v3:

$$w1=1\|v1\|v1=13[\ 12-2\ ]w2=1\|v2\|v2=18[\ 0-2-2\ ]=12[\ 0-1-1\ ]w3=1\|v3\|v3=118[\ 4-11\ ]=132[\ 4-11\ ]$$

To write s as a linear combination of w1, w2, and w3, we apply the formula in (5). This produces

$$s=(w1\cdot s)w1+(w2\cdot s)w2+(w3\cdot s)w3 \quad =(53)\, w1+(0)w2+(23)\, w3=53\, w1+23\, w3$$

We can check that this is correct by computing

$$(53)\, w1+(23)\, w3=(53)\cdot 13\ [12-2] + (23)\cdot 132\ \ [4-11]= [11-1]=s$$

COMPUTATIONAL COMMENTS

- When implemented on a computer, the Gram–Schmidt process can suffer from significant build up of round-off error as the orthogonal vectors are computed. There is a modified version of the Gram–Schmidt process that requires more operations, but is also more numerically stable and hence is not as prone to loss of orthogonality due to round-off error.

Proof of Theorem 8.18

Proof of Theorem 8.18 Part (a) follows from the definition of projSu, and part (c) follows from Theorem 8.14. For part (b), suppose that {v1,… , vk} is an orthogonal basis for *S*. Then

$$v1\cdot(u-projsu)=v1\cdot u-v1\cdot(v1\cdot u\|v1\|2v1+v2\cdot u\|v2\|2v2+\dots+vk\cdot u\|vk\|2vk) =v1\cdot u-(v1\cdot u\|v1\|2(v1\cdot v1)+v2\cdot u\|v2\|2(v1\cdot v2)+\dots+vk\cdot u\|vk\|2(v1\cdot vk)) =v1\cdot u-(v1\cdot u\|v1\|2\|v1\|2+v2\cdot u\|v2\|2(0)+\dots+vk\cdot u\|vk\|2(0)) =v1\cdot u-v1\cdot u=0$$

The same argument can be used to show that each of

$$v2\cdot(u-projSu)=0,\ \dots\ vk\cdot(u-projSu)=0$$

Thus, by Theorem 8.11, $u - \text{proj}_S u$ is orthogonal to S.

To verify part (d), suppose that $\{\tilde{v}_1, \ldots, \tilde{v}_k\}$ is another orthogonal basis for S, and let $\text{proj}_{S_v} u$ and $\text{proj}_{S_{\tilde{v}}} u$ denote the projections for bases $\{v_1, \ldots, v_k\}$ and $\{\tilde{v}_1, \ldots, \tilde{v}_k\}$, respectively. By part (b), both $u - \text{proj}_{S_v} u$ and $u - \text{proj}_{S_{\tilde{v}}} u$ are in $S^\perp$, and since $S^\perp$ is a subspace, the difference

$$(u - \text{proj}_{S_v} u) - (u - \text{proj}_{S_{\tilde{v}}} u) = \text{proj}_{S_{\tilde{v}}} u - \text{proj}_{S_v} u$$

is in $S^\perp$. But by part (a), both $\text{proj}_{S_{\tilde{v}}} u$ and $\text{proj}_{S_v} u$ are also in the subspace S, so that $\text{proj}_{S_{\tilde{v}}} u - \text{proj}_{S_v} u$ is as well. However, $S \cap S^\perp = \{0\}$ (see Exercise 65 of Section 8.1), which implies $\text{proj}_{S_{\tilde{v}}} u - \text{proj}_{S_v} u = 0$. Hence $\text{proj}_{S_{\tilde{v}}} u = \text{proj}_{S_v} u$, and therefore $\text{proj}_S u$ is independent of choice of basis for S. ■ ■

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find $\text{proj}_{u_1} u_2$ for

$$u_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, u_2 = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

2. Find $\text{proj}_S u_3$ for

$$S = \text{span}\left\{\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}\right\}, u_3 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$$

3. Apply the Gram–Schmidt process to find an orthogonal basis for S.

$$S = \text{span}\left\{\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 2 \\ -1 \end{bmatrix}\right\}$$

4. Apply the Gram–Schmidt process to find an orthonormal basis for S.

$$S = \text{span}\left\{\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\right\}$$

EXERCISES

Exercises 1–6: Refer to the vectors given below.

u1=[−312], u2=[111], u3=[20−1], u4=[1−32]

u5=[211], u6=[03−1], u7=[3−4−2], u8=[−1−13]

1. Compute the following projections.
 (a) proju3u2
 (b) proju1u2

2. Compute the following projections.
 (a) proju5u1
 (b) proju5u8

3. Compute projSu2, where S=span{u3, u4}.

4. Compute projSu8, where S=span{u5, u7}.

5. Normalize the given vectors.
 (a) u1
 (b) u4

6. Normalize the given vectors.
 (a) u3
 (b) u6

Exercises 7–14: Apply the Gram–Schmidt process to find an orthogonal basis for *S*.

7. S=span{[13], [42]}

8. S=span{[2−1], [43]}

9. S=span{[−221], [34−2]}

10. S=span{[10−2], [133]}

11. S=span{[1−101], [4120]}

12. S=span{[−1 012], [420−1]}

13. S=span{[−101],[341],[416]}

14. S=span{[110−1],[1301],[4220]}

Exercises 15–22: Find projSu.

15. *S* = subspace in Exercise 7; u=[11].

16. *S* = subspace in Exercise 8; u=[−11].

17. *S* = subspace in Exercise 9; u=[102].

18. *S* = subspace in Exercise 10; u=[111].

19. *S* = subspace in Exercise 11; u=[1−101].

20. *S* = subspace in Exercise 12; u=[0110].

21. *S* = subspace in Exercise 13; u=[102].

22. *S* = subspace in Exercise 14; u=[1010].

Exercises 23–30: Find an orthonormal basis for *S*.

23. *S*= subspace in Exercise 7.

24. *S*= subspace in Exercise 8.

25. *S*= subspace in Exercise 9.

26. *S*= subspace in Exercise 10.

27. *S*= subspace in Exercise 11.

28. *S*= subspace in Exercise 12.

29. *S*= subspace in Exercise 13.

30. *S*= subspace in Exercise 14.

FIND AN EXAMPLE Exercises 31–36: Find an example that meets the given specifications.

31. Two vectors u and v in R2 with projvu=u.

32. Two vectors u and v in R3 with projvu=v.

33. Two vectors u and v in R2 with projvu=0.

34. A two-dimensional subspace *S* in R3 and a vector u such that projSu=u.

35. Two nonparallel vectors u and v with projvu=[12].

36. A two-dimensional subspace *S* in R3 and a vector u not in *S* such that projSu=[301].

TRUE OR FALSE Exercises 37–42: Determine if the statement is true or false, and justify your answer. Assume *S* is nontrivial and u and v are both nonzero.

37.

(a) Since u and v are both nonzero, projvu≠0.

(b) projvu points in the same direction as v.

38.

(a) Every subspace *S* of Rn has an orthonormal basis.

(b) If u is in R5 and *S* is a three-dimensional subspace of R5, then projSu is in R3.

39.

(a) If *S* is a subspace, then projSu is in *S*.

(b) If u and v are vectors, then projvu is a multiple of u.

40.

(a) If u and v are orthogonal, then projvu=0.

(b) If projSu=u, then u is in *S*.

41.

(a) If u is in *S*, then projS⊥u=0.

(b) For a vector u and a subspace *S*,

projS(projSu)=projSu

42.

(a) For vectors u and v,

proju(projvu)=u

(b) For every subspace *S* there exists a nonzero vector u such that projSu=2u.

43. Let {v1,… , vk} be the orthogonal set generated in the course of applying the Gram–Schmidt process to a basis, and define

Sj=span{v1,… , vj} for j=1,… , k

(a) Prove that if i<j, then Si is a subspace of Sj.

(b) Prove that if i<j, then Sj⊥ is a subspace of Si⊥.

44. Suppose that {u1, u2} are linearly independent and that u3 is in span{u1, u2}. Suppose further that the Gram–Schmidt process is applied to $\{u_1, u_2, u_3\}$ to generate a new set $\{v_1, v_2, v_3\}$. What is v_3? Explain your answer.

45. Suppose that u and v are nonzero vectors and that S is a subspace. Prove that if u is in S and v is in $S^\perp$, then $u+v$ is not in S or $S^\perp$.

46. Suppose that $\{w_1, \dots, w_n\}$ is an orthonormal set and that $x=c_1w_1+\cdots+c_nw_n$. Prove that

$$\|x\|^2=c_1^2+\cdots+c_n^2$$

47. Let $v\neq 0$ be a fixed vector in $\mathbb{R}^n$. Prove that $T : \mathbb{R}^n\to\mathbb{R}^n$ given by $T_v(u)=\text{proj}_v u$ is a linear transformation.

48. Let S be a nonzero subspace. Prove that $T : \mathbb{R}^n\to\mathbb{R}^n$ given by $T_S(u)=\text{proj}_S u$ is a linear transformation.

49. Here we prove that the Gram–Schmidt process works. Suppose that $\{u_1, \dots, u_k\}$ are linearly independent vectors, and that $\{v_1, \dots, v_k\}$ are the vectors generated by the Gram–Schmidt process.

(a) Use induction to show $\{v_1, \dots, v_j\}$ is an orthogonal set for $j=1, \dots, k$.

(b) Use induction to show $\text{span}\{u_1, \dots, u_j\}=\text{span}\{v_1, \dots, v_j\}$ for $j=1, \dots, k$.

(c) Explain why (a) and (b) imply that the Gram–Schmidt process yields an orthogonal basis.

50. Prove that for any nonzero vectors u and v,

$$\|u\|^2=\|\text{proj}_v u\|^2+\|u-\text{proj}_v u\|^2 \tag{6}$$

(HINT: Apply Theorem 8.16 and Theorem 8.6.)

51. Prove that for any vector u and nonzero subspace S,

$$\|u\|^2=\|\text{proj}_S u\|^2+\|u-\text{proj}_S u\|^2$$

(HINT: Apply Theorem 8.18 and Theorem 8.6.)

52. If u and v are nonzero vectors in $\mathbb{R}^n$, then the angle θ between u and v is defined in terms of the formula

$$\cos(\theta)=\frac{u\cdot v}{\|u\|\|v\|} \tag{7}$$

In this exercise, we use trigonometry to prove that this is true in $\mathbb{R}^2$. The equation in (7) is an extension from $\mathbb{R}^2$ to $\mathbb{R}^n$.

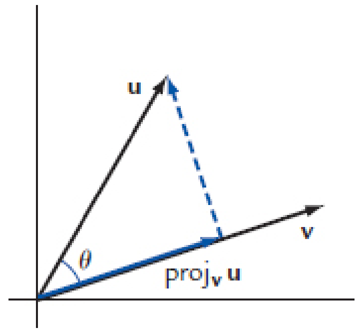


Figure 4 The angle θ between u and v.

(a) Refer to Figure 4 and use it to explain why

$$\cos(\theta)=\frac{\|\text{proj}_v u\|}{\|u\|}=\frac{|u\cdot v|}{\|u\|\|v\|}$$

(b) Explain why $|u\cdot v|=u\cdot v$ in Figure 4. (HINT: $\text{proj}_v u=cv$. Is c positive or negative?) Conclude that (7) holds for $\theta < 90°$.

(c) Draw a new diagram with u and v arranged so that $\theta > 90°$. Explain why in this case $|u\cdot v|=-u\cdot v$, and conclude that (7) also holds in this case.

(d) Complete the proof by showing that (7) holds when $\theta = 90°$

53. In this exercise we use projection to prove the Cauchy–Schwarz inequality, which states that

$$|u\cdot v|\le\|u\|\|v\| \tag{8}$$

for vectors u and v in $\mathbb{R}^n$.

(a) Prove that $\|\text{proj}_v u\|\le\|u\|$. (HINT: See (6).)

(b) Use (a) and the definition of projection to show that (8) holds.

(c) Show that $|u\cdot v|=\|u\|\|v\|$ if and only if $u=cv$. Hence there is equality in the Cauchy–Schwarz inequality exactly when u is a scalar multiple of v.

54. Here we prove that

$$u=\text{proj}_S u+\text{proj}_{S^\perp} u \tag{9}$$

for a vector u and a nontrivial subspace S.

(a) Explain why $u-\text{proj}_S u$ and $\text{proj}_{S^\perp} u$ are both in $S^\perp$, and use this to prove that

$$u-\text{proj}_S u-\text{proj}_{S^\perp} u \tag{10}$$

is in $S^\perp$.

(b) Explain why $u-\text{proj}_{S^\perp} u$ and $\text{proj}_S u$ are both in S, and use this to prove that the expression in (10) is also in S.

(c) Combine (a) and (b) to show that

$$u-\mathrm{proj}_S u-\mathrm{proj}_{S^\perp} u=0$$

and from this conclude that (9) is true.

Exercises 55–56: Find an orthonormal basis for *S*.

55. S=span{[12−4−1], [−305−2], [072−6]}

56. S=span{[2−10−23], [4−21−4−2], [5−120−4], [402−3−3]}

Exercises 57–58: Determine projSu.

57. S=span{[3−1−27], [2163]}, u=[3−1−14]

58. S=span{[21710], [3−2−504],[52114]}, u=[3−1−140]

8.3 Diagonalizing Symmetric Matrices and QR Factorization

We start this section by revisiting the problem of diagonalizing matrices. As we discovered in Section 6.2, not all square matrices can be diagonalized, and in general it is not easy to tell if a given matrix can be diagonalized. However, the situation is different if the matrix is symmetric. Let's consider an example.

Example 1

▶ Recall that a square matrix A is symmetric if $A^T = A$.

If possible, diagonalize the matrix

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

Solution The characteristic polynomial for A is

$$\det(A-\lambda I)=-\lambda^3-6\lambda^2-9\lambda-54=-(\lambda-6)(\lambda-3)(\lambda+3)$$

so the eigenvalues are $\lambda=6,\ 3,\ -3$. Since the eigenvalues each have multiplicity 1, we know that A is diagonalizable. Following our usual procedure, we find that an eigenvector associated with each eigenvalue is

$$\lambda=6 \Rightarrow \mathbf{u}_1=\begin{bmatrix}1\\ -1\\ 1\end{bmatrix},\quad \lambda=3 \Rightarrow \mathbf{u}_2=\begin{bmatrix}1\\ 2\\ 1\end{bmatrix},\quad \lambda=-3 \Rightarrow \mathbf{u}_3=\begin{bmatrix}-1\\ 0\\ 1\end{bmatrix}$$

Forming the diagonal matrix D and the matrix of eigenvectors P, we have

$$D=\begin{bmatrix}6 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & -3\end{bmatrix} \text{ and } P=\begin{bmatrix}1 & 1 & -1\\ -1 & 2 & 0\\ 1 & 1 & 1\end{bmatrix}$$

We know from previous work that eigenvectors associated with distinct eigenvalues are linearly independent. In Example 1, even more is true: The eigenvectors are orthogonal. This is not a coincidence—it always happens when A is a symmetric matrix.

THEOREM 8.21

If A is a symmetric matrix, then eigenvectors associated with distinct eigenvalues are orthogonal.

Note that $\mathbf{x}_1^T\mathbf{x}_2 = \mathbf{x}_1 \cdot \mathbf{x}_2$. (See Exercise 69 of Section 8.1.)

Proof Let A be a symmetric matrix, and suppose that $\lambda_1 \neq \lambda_2$ are distinct eigenvalues of A with associated eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively. We now compute $(A\mathbf{u}_1)^T\mathbf{u}_2$ in two different ways. First, we have

$$(A\mathbf{u}_1)^T\mathbf{u}_2 = (\lambda_1\mathbf{u}_1)^T\mathbf{u}_2 = \lambda_1\mathbf{u}_1^T\mathbf{u}_2 = \lambda_1(\mathbf{u}_1 \cdot \mathbf{u}_2)$$

Second, A symmetric means $A = A^T$. Thus

$$(A\mathbf{u}_1)^T\mathbf{u}_2 = (\mathbf{u}_1^TA^T)\mathbf{u}_2 = (\mathbf{u}_1^TA)\mathbf{u}_2 = \mathbf{u}_1^T(A\mathbf{u}_2) = \mathbf{u}_1^T(\lambda_2\mathbf{u}_2) = \lambda_2(\mathbf{u}_1^T\mathbf{u}_2) = \lambda_2(\mathbf{u}_1 \cdot \mathbf{u}_2)$$

Hence $\lambda_1(\mathbf{u}_1 \cdot \mathbf{u}_2) = \lambda_2(\mathbf{u}_1 \cdot \mathbf{u}_2)$, or equivalently,

$$(\lambda_1 - \lambda_2)(\mathbf{u}_1 \cdot \mathbf{u}_2) = 0$$

Since $\lambda_1 \neq \lambda_2$, then $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$. Therefore $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal. ■

Returning to Example 1, if we compute P^TP we find that

$$P^TP = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & -1 \\ -1 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Thus PTP is a diagonal matrix. To further simplify PTP, we redefine *P* to have normalized columns (i.e., columns that have length 1),

$$P=[1316-12-13260131612]$$

Since each column is a constant multiple of its predecessor, it is still an eigenvector associated with the same eigenvalue. Thus we could use this definition of *P* to diagonalize *A*. Although this choice of *P* is not as tidy as the previous one, it does have the nice property that

$$PT\,P=[13-1313162616-12012]\,[1316-12-13260131612]=[100010001]=I3 \qquad (1)$$

DEFINITION 8.22 ▶

Orthogonal Matrix

A square matrix *P* with orthonormal columns is called an **orthogonal matrix**.

The property in (1) holds for all orthogonal matrices.

THEOREM 8.23 ▶

If *P* is an n×n orthogonal matrix, then P−1=PT.

Proof When computing the matrix product PT P, we are just computing the dot products of the columns of *P*. The diagonal terms of PT P come from the dot product of a column with itself, with each equal to 1 because of the normality. The nondiagonal terms come from the dot products of distinct columns, and so are zero because the columns are orthogonal. Thus PTP=In, and hence P−1=PT. ■ ■

For example, since the following matrix *P* is orthogonal, we have P−1=PT.

P=[0001100001000010]⇒P−1=PT=[0100001000011000]

▶ It might be better if an “orthogonal matrix” was called an “orthonormal matrix” but “orthogonal matrix” is standard in linear algebra. There is no special name for a matrix that has orthogonal columns that are not normalized.

Orthogonally Diagonalizable Matrices

Next we show how to find matrices *D* and *P* in the special case where *P* is an orthogonal matrix.

DEFINITION 8.24 ▶

Orthogonally Diagonalizable

A square matrix *A* is **orthogonally diagonalizable** if there exists an orthogonal matrix *P* and a diagonal matrix *D* such that A=PDP−1.

Since we can write the symmetric matrix *A* in Example 1 as A=PDP−1 for

D=[60003000−3] and
P=[1316−12−13260131612]

it follows that *A* is orthogonally diagonalizable. It is not hard to show that *any* orthogonally diagonalizable matrix must be symmetric.

THEOREM 8.25 ▶

Let *A* be an orthogonally diagonalizable matrix. Then *A* is symmetric.

Proof If *A* is orthogonally diagonalizable, then there exists an orthogonal matrix *P* and diagonal matrix *D* such that $A=PDP^{-1}$. Using the fact that $P^{-1}=P^T$ (by Theorem 8.23) and $D^T=D$ (because *D* is diagonal), we have

$$A^T=(PDP^{-1})^T=(PDP^T)^T=(P^T)^T D^T P^T=PDP^T=PDP^{-1}=A$$

Since $A^T=A$, it follows that *A* is symmetric. ■ ■

Remarkably, the converse of Theorem 8.25 is also true: If *A* is a symmetric matrix, then *A* is orthogonally diagonalizable.

THEOREM 8.26 ▶

(SPECTRAL THEOREM) A matrix *A* is orthogonally diagonalizable if and only if *A* is symmetric.

A complete proof of the Spectral Theorem is not included here. Two consequences of the Spectral Theorem:

- All eigenvalues of a symmetric matrix *A* are real.
- Each eigenspace of a symmetric matrix *A* has dimension equal to the multiplicity of the associated eigenvalue.

Example 2

Orthogonally diagonalize the symmetric matrix

$$A=[13-3-33-33-1-331-3-3-1-3-3]$$

Solution Finding the characteristic polynomial of *A* (which has degree 4) and then factoring by hand is difficult, but using computer software we find that

$$\det(A-\lambda I)=\lambda 4+4\lambda 3-48\lambda 2-64\lambda+512=(\lambda+8)(\lambda+4)(\lambda-4)2$$

Thus we have eigenvalues $\lambda=-8$, $\lambda=-4$ (both multiplicity 1), and $\lambda=4$ (multiplicity 2). Our usual methods produce bases for each eigenspace,

$$\lambda=-8 \Rightarrow \{[1-111]\};\lambda=-4 \Rightarrow \{[0101]\};\lambda=4 \Rightarrow \{[-1010],\ [-2-101]\}$$

Since our goal is to orthogonally diagonalize *A*, we first need an orthogonal basis for each eigenspace. The bases for $\lambda=-8$ and $\lambda=-4$ are fine as is, but we need to apply Gram–Schmidt to the basis for $\lambda=4$. Designate

$$u1=[-1010],u2=[-2-101]$$

Setting v1=u1, the second vector v2 is given by

$$v2=u2-v1\cdot u2\|v1\|2\ v1=[-2-101]\ -22\ [-1010]=[-1-1-11]$$

▶ We know the vectors are orthogonal by Theorem 8.21, which says that for a symmetric matrix, eigenvectors associated with distinct eigenvalues are orthogonal.

This gives us four orthogonal eigenvectors

$$[1-111],[0101],\ [-1010],[-1-1-11]$$

Next, we normalize each vector,

$$[12-121212],[012012],\ [-120120],[-12-12-1212]$$

Finally, we form *D* and *P*,

$$D=[-80000-40000400004] \text{ and } P=[120-12-12-12120-1212012-121212012]$$

Example 3

Orthogonally diagonalize the matrix AT A for

A=[122002]

Solution We have

AT A=[120202] [122002]=[5228]

▶ All matrices of the form AT A are symmetric. (See Exercise 69 of Section 3.2.)

ATA is symmetric and hence by the Spectral Theorem orthogonally diagonalizable. The characteristic polynomial is λ2−13λ+36=(λ−4)(λ−9), yielding eigenvalues λ1=9 and λ2=4. The corresponding normalized eigenvectors are

λ=9⇒{[1525]};λ=4⇒{[25−15]}

Thus, if we define

D=[9004] and P=[152525−15]

then *P* is orthogonal and AT A=PDPT.

Since ATA is a symmetric matrix, we know from the Spectral Theorem that the eigenvalues are real. In fact, for symmetric matrices of this form, it turns out that the eigenvalues will be nonnegative. Since this result is handy to know in the next section, we state and prove it here.

THEOREM 8.27 ▶

If *A* is a real matrix, then AT A has nonnegative eigenvalues.

Proof Suppose that λ is an eigenvalue of ATA with associated eigenvector u. Then

$$\|Au\|^2=(Au)\cdot(Au)=(Au)^T(Au)=(u^TA^T)(Au)=u^T(A^TAu)=u^T(\lambda u)=\lambda(u^Tu)=\lambda\|u\|^2$$

In summary, $\|Au\|^2=\lambda\|u\|^2$. Since both $\|Au\|^2$ and $\|u\|^2$ are nonnegative, it must be that λ is nonnegative as well. ■ ■

QR Factorization

Diagonalizing a matrix is one type of matrix factorization. Diagonalizing is always possible when a matrix is symmetric, but it may or may not be otherwise. Here we consider another type of factorization, which applies to any matrix that has linearly independent columns.

THEOREM 8.28 ▶

(QR FACTORIZATION) Let $A=[a_1 \cdots a_m]$ be an $n\times m$ matrix with linearly independent columns. Then A can be factorized as $A=QR$, where $Q=[q_1 \cdots q_m]$ is an $n\times m$ matrix with orthonormal columns and R is an $m\times m$ upper triangular matrix with positive diagonal entries.

Proof Suppose that $\{q_1,\ldots,q_m\}$ is the orthonormal set of vectors we get by applying the Gram–Schmidt process to the set of columns $\{a_1,\ldots,a_m\}$. Now define $Q=[q_1\cdots q_m]$. From the Gram–Schmidt construction, for each $1\le k\le m$ the vector a_k is in the span of the orthonormal set $\{q_1,\ldots,q_k\}$. Hence by equation (5) in Section 8.2 we have

$$a_k=(q_1\cdot a_k)q_1+(q_2\cdot a_k)q_2+\cdots+(q_k\cdot a_k)q_k \qquad (2)$$

Now define $r_{ik}=q_i\cdot a_k$ for $1\le k\le m$ and $1\le i\le k$, and let

$$R=\begin{bmatrix} r_{11} & r_{12} & r_{13} & \ldots & r_{1n}\\ 0 & r_{22} & r_{23} & \ldots & r_{2n}\\ 0 & 0 & r_{33} & \ldots & r_{3n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & r_{mm}\end{bmatrix}$$

Since $Q=[\mathbf{q}_1 \cdots \mathbf{q}_m]$, the kth column of the product QR is equal to

$$r_{1k}\mathbf{q}_1+r_{2k}\mathbf{q}_2+\cdots+r_{kk}\mathbf{q}_k=\mathbf{a}_k$$

by (2) and the definition of r_{ik}. Since $\mathbf{a}_k$ is the kth column of *A*, it follows that $A=QR$.

Finally, since $\mathbf{a}_k$ is *not* in $\operatorname{span}\{\mathbf{q}_1,\ldots,\mathbf{q}_{k-1}\}$ (why?), it must be that $r_{kk}\neq 0$. If $r_{kk}<0$, then we replace $\mathbf{q}_k$ with $-\mathbf{q}_k$, which will make $r_{kk}>0$ while keeping the columns of *Q* orthonormal and the column space $\operatorname{col}(Q)$ unchanged. Hence we can ensure that the diagonal entries of *R* are positive. ■ ■

Before considering an example, we note that once *Q* has been found we can use matrix multiplication to compute *R*. Since $A=QR$ we have

$$Q^T A=Q^T QR=R$$

because $Q^T Q=I_m$. It also can be shown directly that the entries of Q^TA are equal to the entries of *R* (see Exercise 55).

Example 4

Find the QR factorization for

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

Solution The columns of *A* are the vectors from Examples 5–6 in Section 8.2, where we found the corresponding orthonormal set

$$\left\{\frac{1}{\sqrt{3}}\begin{bmatrix}1\\0\\1\\1\end{bmatrix},\ \frac{1}{\sqrt{10}}\begin{bmatrix}-1\\2\\-1\\2\end{bmatrix},\ \frac{1}{\sqrt{23}}\begin{bmatrix}-3\\-3\\1\\2\end{bmatrix}\right\}$$

Therefore we define

$$Q=\begin{bmatrix}\frac{1}{\sqrt{3}}&-\frac{1}{\sqrt{10}}&-\frac{3}{\sqrt{23}}\\0&\frac{2}{\sqrt{10}}&-\frac{3}{\sqrt{23}}\\\frac{1}{\sqrt{3}}&-\frac{1}{\sqrt{10}}&\frac{1}{\sqrt{23}}\\\frac{1}{\sqrt{3}}&\frac{2}{\sqrt{10}}&\frac{2}{\sqrt{23}}\end{bmatrix}$$

We find *R* by computing

R=QT A=[1301313−110210−110210−323−323123223]
[10−302−1101135]=[333010100023]

Thus we have the factorization

QR=[13−110−3230210−32313−11012313210223]
[333010100023]=[10−302−1101135]=A

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine if *A* is an orthogonal matrix.

   A=[14−1212−121214140−14]

2. Find the matrices *D* and *P* for an orthogonal diagonalization of A=[3221].
3. Verify that the eigenvalues of AT A are nonnegative.

   A=[310221]

4. Find the QR factorization for A=[2−332].

EXERCISES

Exercises 1–8: Determine if *A* is symmetric.

1. A=[1−221]
2. A=[4335]

3. $A=[321213132]$
4. $A=[201020002]$
5. $A=[3-14-143]$
6. $A=[-52-1210-10-6]$
7. $A=[17-3724-30-64-6-1]$
8. $A=[420-223450420-2501]$

Exercises 9–14: Determine if *A* is orthogonal.

9. $A=[1-221]$
10. $A=[110310310-110]$
11. $A=[-51312131213513]$
12. $A=[101010-101]$
13. $A=[121314-121314013-12]$
14. $A=[2141344211413-542-31413142]$

Exercises 15–18: The eigenvalues and corresponding eigenvectors for a symmetric matrix *A* are given. Find matrices *D* and *P* of an orthogonal diagonalization of *A*.

15. $\lambda1=2$, $u1=[12]$; $\lambda2=-3$, $u2=[-21]$
16. $\lambda1=-1$, $u1=[34]$; $\lambda2=1$, $u2=[-43]$
17. $\lambda1=0$, $u1=[111]$; $\lambda2=2$, $u2=[1-10]$; $\lambda3=-1$, $u3=[-1-12]$
18. $\lambda1=-1$, $u1=[210]$; $\lambda2=0$, $u2=[1-21]$; $\lambda3=3$, $u3=[-125]$

Exercises 19–24: The eigenvalues for the symmetric matrix *A* are given. Find the matrices *D* and *P* of an orthogonal diagonalization of *A*.

19. $A=[4221]$, $\lambda=0, 5$
20. $A=[3443]$, $\lambda=-1, 7$
21. $A=[012111210]$, $\lambda=-2, 0, 3$
22. $A=[123213330]$, $\lambda=-3, -1, 6$
23. $A=[001010100]$, $\lambda=-1, 1$
24. $A=[111111111]$, $\lambda=0, 3$

Exercises 25–28: Verify that the eigenvalues of AT A are nonnegative.

25. A=[112112]

26. A=[021021]

27. A=[021100]

28. A=[223110]

Exercises 29–32: The matrix *Q* has orthogonal columns. Find Q−1 without using row operations. (HINT: Exercise 54 could be helpful. If you use it, explain why it works.)

29. Q=[12−21]

30. Q=[455−4]

31. Q=[020101−101]

32. Q=[11521−43−11]

Exercises 33–40: Find the QR factorization for the matrix *A*.

33. A=[3−223]

34. A=[−4326]

35. A=[1432]

36. A=[24−13]

37. A=[121−351]

38. A=[0−36−392]

39. A=[−23241−2]

40. A=[1103−23]

FIND AN EXAMPLE Exercises 41–48: Find an example that meets the given specifications.

41. A 2×2 matrix *A* that has eigenvalues λ1=1 and λ2=2 and is orthogonally diagonalizable.

42. A 3×3 matrix *A* that has eigenvalues λ1=2, λ2=3, and λ3=5 and is orthogonally diagonalizable.

43. A 2×2 matrix A that is orthogonally diagonalizable, has eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$, and corresponding eigenvectors

$$u_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, u_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

44. A 3×3 matrix A that is orthogonally diagonalizable, has eigenvalues $\lambda_1 = -3$, $\lambda_2 = 0$, and $\lambda_3 = 4$, and has corresponding eigenvectors

$$u_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, u_2 = \begin{bmatrix} 4 \\ -1 \\ -2 \end{bmatrix}, u_3 = \begin{bmatrix} 2 \\ 10 \\ -1 \end{bmatrix}$$

45. A 2×2 matrix A that does not have a QR factorization.

46. Two 3×3 matrices A and B that both have a QR factorization, but $A+B$ does not.

47. A 2×2 matrix A that is diagonalizable but not orthogonally diagonalizable.

48. A 2×2 matrix A that is orthogonally diagonalizable but not invertible.

TRUE OR FALSE Exercises 49–52: Determine if the statement is true or false, and justify your answer.

49.

(a) If A is a symmetric matrix, then A is diagonalizable.

(b) If A is a square matrix, then A is diagonalizable.

50.

(a) A^TA is symmetric for any matrix A.

(b) All matrices have a QR factorization.

51.

(a) In the QR factorization of a matrix A, the matrix R has columns that span the column space of A.

(b) If A and B are orthogonal $n \times n$ matrices, then so is $A+B$.

52.

(a) If $A = QR$ is a QR factorization for a matrix A, then R is invertible.

(b) If $A^TA = AA^T$ for a square matrix A, then A is an orthogonal matrix.

53. Prove that if A is orthogonal, then $\det(A) = \pm 1$.

54. Suppose $Q = [q_1 \cdots q_n]$ has orthogonal columns. Show that

$Q−1=[1‖q1‖2\ q1\ T⋯1‖qn‖2\ qn\ T]$

55. Let *A* and *Q* be the matrices in Theorem 8.28. Prove that the entry in position (i,k) of QTA is equal to qi·ak.

56. Prove that if *A* is an orthogonal matrix, then so is AT.

57. Prove that if *A* is orthogonally diagonalizable, then so is AT.

58. Prove that if *A* and *B* are orthogonally diagonalizable matrices, then so is A+B.

59. Prove that if *A* is orthogonally diagonalizable, then so is A2.

60. Suppose that *P* is an orthogonal matrix. Show that for any vector x, ‖Px‖2=‖x‖2 and therefore any eigenvector of *P* satisfies |λ|=1.

Exercises 61–64: Find an orthogonal diagonalization for the matrix *A*.

61. A=[21310−43−45]

62. A=[−1−30−327074]

63. A=[2140132−542−130−532]

64. A=[0−832−8−127320−127−14]

Exercises 65–68: Find a QR factorization for the matrix *A*.

65. A=[−133024115]

66. A=[42015−23−31]

67. A=[114132102−114]

68. A=[2−1142303−2510]

8.4 The Singular Value Decomposition

▶ This section is optional and can be omitted without loss of continuity.

In this section we develop another type of matrix factorization that is a generalization of diagonalization. This new type of matrix factorization is called the *singular value decomposition* (SVD), and it can be applied to any type of matrix, even those that are not square. We start by developing the factorization method and then describe applications to image processing and estimating the rank of a matrix.

▶ $0_{(n-m)m}$ is the $(n-m)\times m$ matrix with all entries equal to zero.

Singular Value Decomposition

Suppose that we have an $n\times m$ matrix *A*. If $n\geq m$, then the **singular value decomposition** is the factorization of *A* as the product $A=U\Sigma V^T$, where

- *U* is an $n\times n$ orthogonal matrix.
- Σ is an $n\times m$ matrix of the form $\Sigma=[D\ 0_{(n-m)m}]$, where *D* is a diagonal matrix with

  $$D=\begin{bmatrix}\sigma_1 & 0 & \cdots & 0\\ 0 & \sigma_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_m\end{bmatrix}$$

  and $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_m\geq 0$ are the **singular values** of *A*. The singular values are given by $\sigma_i=\lambda_i$, where λ_i is an eigenvalue of A^TA.
- *V* is an $m\times m$ orthogonal matrix.

If $n<m$, then $\Sigma=[D_{nn}\ 0_{n(m-n)}]$ with everything else the same.

THEOREM 8.29 ▶

Every $n\times m$ matrix *A* has a singular value decomposition.

The proof of this theorem is given at the end of the section. For now, let's look at an example that illustrates how we find the SVD.

Example 1

Find the SVD for the 3×2 matrix

$$A=[122002]$$

Solution We find the SVD $A=U\Sigma VT$ by applying the following sequence of steps.

1. **Orthogonally Diagonalize** AT A **to find *V*.** Since ATA is symmetric, the Spectral Theorem guarantees that it is orthogonally diagonalizable. Our matrix *A* appears in Example 3 of Section 8.3, where we showed that the eigenvalues of ATA are $\lambda 1=9$ and $\lambda 2=4$ and that the orthogonal diagonalizing matrix is

   $$V=[152525-15]$$

   This is the matrix *V* in the SVD of *A*.

2. **Find Σ.** The singular values for a matrix *A* are given by $\sigma i=\lambda i$, the square roots of the eigenvalues of ATA. Here we have $\sigma 1=9=3$ and $\sigma 2=4=2$, so that

   $$\Sigma=[300200]$$

   ▶ By Theorem 8.27, the eigenvalues of ATA are always nonnegative.

   ▶ Recall that col(A) is the column space of *A*.

3. **Find *U*.** We determine the columns of *U* in two steps, one for the columns corresponding to positive singular values, and the other for the columns that form an orthonormal basis for $(\mathrm{col}(A))^{\perp}$. (We will see why later.)

   ▶ $VT=V-1$ because *V* is an orthogonal matrix.

3a. Positive Singular Values. Our ultimate goal is to find *U* so that A=UΣVT, or equivalently, AV=UΣ. Note that the ith column of AV is Avi, while the ith column of UΣ is σiui. Thus, for AV=UΣ, we must have Avi=σiui. When σi>0, we arrange for this by defining

$$ui=1σiAvi$$

In this example, we have

$$u1=1σ1\ Av1=13\ [122002]\ [1525]=135\ [524]$$

$$u2=1σ2\ Av2=12\ [122002]\ [25−15]=15\ [02−1]$$

Two important observations:
(a) u1 and u2 are orthonormal.
(b) span{u1, u2}=col(A), the column space of *A*.
This is not a coincidence, but rather a consequence of our method for finding *U*. We will show why in our proof of Theorem 8.29.

3b. Filling Out *U*. We now have the first two columns of *U*. Since the third row of Σ consists of zeros, the product U Σ will be the same regardless of our choice of u3. We want *U* to be an orthogonal matrix, which we can accomplish by extending span{u1, u2} to an orthonormal basis for R3. Since span{u1, u2}=col(A), we proceed by finding an orthonormal basis for $(\mathrm{col}(A))^{\perp}$, the orthogonal complement of the column space of *A*. We do this by noting that $(\mathrm{col}(A))^{\perp} = \mathrm{null}(A^{T})$ (see Exercise 67 of Section 8.1). The null space of AT is equal to the set of solutions to ATx=0. The augmented matrix and echelon form are

$$[12002020]\sim[12000−210]$$

Back substitution and normalizing the solution yield

$$u3=[−231323]$$

The vector u3 gives the final column of *U*. We have

$$U=[u1u2u3]=[5350−232352513435−1523]$$

We can check our work by computing

$$UΣVT=[5350−232352513435−1523]\ [300200]\ [152525−15]=[122002]=A$$

This procedure will lead to the SVD for any matrix *A*. We find Σ and *V* by orthogonally diagonalizing ATA, so we can see that this is always possible. The development of *U* is not as transparent. The key subtle fact that makes our procedure work is that if σ1,… , σk are the *positive* singular values of an $n\times m$ matrix A with associated orthonormal eigenvectors $\mathbf{v}_1,\ldots,\mathbf{v}_k$ of A^TA, then the set

$$\mathbf{u}_i=\frac{1}{\sigma_i}A\mathbf{v}_i,\quad 1\le i\le k$$

forms an orthonormal basis for $\mathrm{col}(A)$. Finding an orthonormal basis for $(\mathrm{col}(A))^{\perp}=\mathrm{null}(A^T)$ allows us to extend the set $\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}$ to an orthonormal basis for $\mathbb{R}^n$ and gives the remaining columns $\mathbf{u}_{k+1},\ldots,\mathbf{u}_n$ of U. Since rows $k+1,\ldots,n$ of Σ are made up of zeros, the product $U\Sigma$ is independent of $\mathbf{u}_{k+1},\ldots,\mathbf{u}_n$, and the definition of $\mathbf{u}_1,\ldots,\mathbf{u}_k$ ensures that $AV=U\Sigma$.

Our example covered the case of an $n\times m$ matrix A where $n>m$. The same procedure can be used if $n=m$, but suppose that $n<m$? In this case, we can take transposes. Let $B=A^T$, and suppose $B=U\Sigma V^T$ is the SVD. Since $A=B^T$, we have $A=(U\Sigma V^T)^T=V\Sigma^TU^T$, the form required by the SVD.

▶ Since U and V are orthogonal, so are U^T and V^T (see Exercise 56 of Section 8.3).

Example 2

Find the SVD for the matrix

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

Solution We start by setting

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

Now we apply our algorithm to find the SVD of B.

1. **Orthogonally Diagonalize** B^TB **to find** V. We have

$$B^TB=\begin{bmatrix}6 & 3\\ 3 & 6\end{bmatrix}$$

which has eigenvalues and eigenvectors

$$\lambda_1=9\Rightarrow \mathbf{v}_1=\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix};\quad \lambda_2=3\Rightarrow \mathbf{v}_2=\begin{bmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$$

Thus the orthogonal diagonalizing matrix is

V=[12−121212]

2. **Find** Σ. The singular values for a matrix *B* are $\sigma_1=3$ and $\sigma_2=3$. Hence

Σ=[30030000]

3. **Find** *U*. As before, determining the columns of *U* is performed in two steps.

   3a. Positive Singular Values. For these we have ui=1σiBvi, so that

   u1=1σ1B v1=13 [12201101] [1212]=132 [3221]

   u2=1σ2B v2=13 [12201101] [−1212]=16 [1−201]

   3b. Filling Out *U*. We need two columns to complete *U*, and we get them from an orthonormal basis for $(\mathrm{col}(B))^{\perp}$=null(BT). Solving BTx=0 using our standard procedure and then applying the Gram–Schmidt process gives us

   u3=13 [01−22],u4=13 [−1011]

Combining the four vectors into *U* yields

U=[u1u2u3u4]=[12160−13232−261302320−2313132162313]

Since A=BT, we have A=(UΣVT)T=VΣTUT. Checking the calculations, we have

VΣTUT=[12−121212] [30000300] [1223223213216−26016013−2323−1301313] =[12102011]=A

Image Compression

SVDs can be used to store and transfer digital images efficiently. A digital black-and-white photo can be stored in matrix form, with each entry representing the gray level (the proportion of black to white) for a particular pixel. To simplify the discussion, let's assume that we have a square n×n matrix *A* made up of nonnegative entries that are the gray levels for a photo. Such a matrix has n2 entries, which grows quickly with *n* and can have significant implications for storage and transmission of digital images.

To improve efficiency we can take advantage of the fact that pixels near one another in a digital photo frequently have similar gray levels. Hence there can be a lot of redundant information in the image matrix, so that it may be possible to represent the image using much less storage space while still retaining the essential elements. One way to do this is to use the singular value decomposition of the image matrix $A=U\Sigma V^T$. We can use the *outer product expansion* (see Exercise 34) to express

$$A=\sigma_1 u_1 v_1^T+\sigma_2 u_2 v_2^T+\cdots+\sigma_n u_n v_n^T$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ are the singular values. The terms with the largest singular values often contain most of the "information" in an image, while those associated with the smallest singular values frequently contribute relatively little. We can sometimes discard many—or even most—of the terms and still have a good approximation of the original image, by just taking the first *k* terms,

$$A_k=\sigma_1 u_1 v_1^T+\sigma_2 u_2 v_2^T+\cdots+\sigma_k u_k v_k^T$$

Figure 1 A photo of Abraham Lincoln processed with (from left to right) 2, 7, 14, and 28 of the 273 singular values. *Library of Congress*

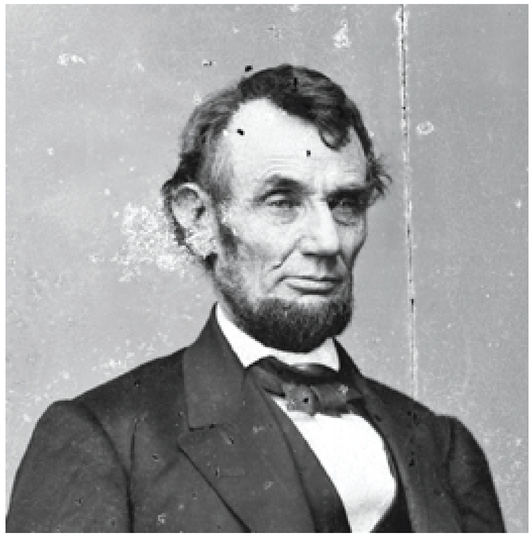


Figure 2 Abraham Lincoln image. *Library of Congress*

▶ For simplicity in notation we consider only square matrices, but the discussion can be extended to nonsquare matrices.

Figure 1 shows the results of using 2, 7, 14, and 28 of the original 273 singular values from a famous photo of a famous American (see Figure 2 for the original). Although the image using 28 singular values requires only about 20% of the storage capacity of the original image, it is still fairly good.

Estimating the Rank of a Matrix

Machine ∈

In most applications, matrix calculations are carried out on a computer. Unfortunately, the finite precision arithmetic employed by computers can sometimes lead to subtle but critical round-off of matrix entries, which can make it very difficult to determine that rank of the matrix. For example, an $n \times n$ matrix A might have true rank n but appear to the computer to have a lower rank, which we refer to as the *numerical rank* of the matrix. Here we briefly describe how to use singular values to find the numerical rank of a matrix.

A computer's sensitivity to round-off depends on the degree of precision used in storing numerical values. The *machine* ε provides a measure of this sensitivity. Roughly speaking, the machine ε gives an upper bound on the relative error that can occur when representing a number in the computer's floating point memory.

There are different ways to define the numerical rank of an n×n matrix *A*. One method employs the singular values σ1≥ σ2≥ ⋯ ≥ σn. Using a machine ε as an upper bound of the relative error, we let

$$b=\sigma_1 \cdot \varepsilon \cdot n$$

Now define *k* to be the largest integer such that σk≥b. Then *k* is the numerical rank of *A*.

Example 3

▶ The choice of $\in = 10^{-10}$ in this example is arbitrary, and could be different depending on the application.

Suppose that A is a 4 × 4 matrix with singular values $\sigma_1 = 3$, $\sigma_2 = 1$, $\sigma_3 = 10^{-8}$, and $\sigma_4 = 10^{-9}$. If we have $\in = 10^{-10}$, what is the numerical rank of A?

Solution We have the bound

$$b = 3 \cdot 10^{-10} \cdot 4 = 1.2 \times 10^{-9}$$

Since $\sigma_3 \geq b$ but $\sigma_4 < b$, it follows that the numerical rank of *A* is 3.

Proof of Theorem 8.29

We have verified most of the elements required to show that an SVD always exists. All that remains is to show that the key fact mentioned earlier is true: If σ1,… , σk are the positive singular values of an n×m matrix *A* with associated orthonormal eigenvectors (of ATA) v1, … , vk, then the set

$$u_i=\frac{1}{\sigma_i}Av_i, 1\leq i\leq k$$

forms an orthonormal basis for col(A). Once this is established, the Rank–Nullity Theorem and Gram–Schmidt process ensure the existence of an orthonormal basis $\{u_{k+1},\ldots, u_n\}$ for $(\mathrm{col}(A))^{\perp}=\mathrm{null}(A^T)$, so we are assured that the required orthogonal matrix *U* can be formed.

First, note that for $1\le i, j\le k$ we have

$$u_i\cdot u_j=\left(\frac{1}{\sigma_i}Av_i\right)\cdot\left(\frac{1}{\sigma_j}Av_j\right)=\frac{1}{\sigma_i\sigma_j}(Av_i)^T(Av_j)=\frac{1}{\sigma_i\sigma_j}v_i^T(A^TAv_j)=\frac{1}{\sigma_i\sigma_j}v_i^T(\lambda_j v_j)=\frac{\lambda_j}{\sigma_i\sigma_j}(v_i\cdot v_j)$$

If $i\ne j$, then $v_i\cdot v_j=0$ and so $u_i\cdot u_j=0$. On the other hand, if $i=j$, then $v_i\cdot v_i=\frac{\lambda_i}{\sigma_i^2}=1$ and hence $u_i\cdot u_i=1$. Thus $\{u_1,\cdots, u_k\}$ are orthonormal and therefore also linearly independent.

Next, since $u_i=\frac{1}{\sigma_i}Av_i$, each u_i $(1\le i\le k)$ is in col(A). So if we can show that $\dim(\mathrm{col}(A))=k$, we are done because $\{u_1,\cdots, u_k\}$ is an orthonormal basis for col(A). To see why this is true, we note the following:

(a) $\mathrm{rank}(A^TA)=\mathrm{rank}(A)$. (See Exercise 35.)

(b) The eigenvectors $\{v_1,\ldots, v_k\}$ associated with the nonzero eigenvalues of A^TA form a basis for col(A). (See Exercise 36.)

Thus, if we have positive singular values $\sigma_1,\ldots, \sigma_k$, then $\mathrm{rank}(A^TA)=k$, which implies that $\mathrm{rank}(A)=k$. Hence $\dim(\mathrm{col}(A))=k$, completing the proof.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the singular values for A=[31−1102].
2. Find the singular value decomposition for A=[10−11−12].
3. Determine the numerical rank of a 4×4 matrix *A* with singular values $\sigma_1=22$, $\sigma_2=7$, $\sigma_3=10^{-5}$, $\sigma_4=10^{-8}$; $\varepsilon=10^{-7}$.

EXERCISES

Exercises 1–8: Find the singular values for *A*.

1. $A=[12-12]$
2. $A=[12-22]$
3. $A=[3-1-13]$
4. $A=[2-241]$
5. $A=[12022-1]$
6. $A=[31-1012]$
7. $A=[12101-1]$
8. $A=[-13211-1]$

Exercises 9–16: Find a singular value decomposition for *A*.

9. $A=[1221]$
10. $A=[22-21]$
11. $A=[21-1310]$
12. $A=[1-22102]$
13. $A=[-110221]$
14. $A=[1202-1-2]$
15. $A=[22101-101]$
16. $A=[133212-11]$

Exercises 17–20: Determine the numerical rank of the matrix *A*.

17. *A* is a 3×3 matrix with singular values $\sigma1=10$, $\sigma2=6$, $\sigma3=10-8$; $\varepsilon=10-9$.
18. *A* is a 4×4 matrix with singular values $\sigma1=5$, $\sigma2=3$, $\sigma3=10-7$, $\sigma4=10-8$; $\varepsilon=10-8$.
19. *A* is a 4×4 matrix with singular values $\sigma1=12$, $\sigma2=4$, $\sigma3=10-6$, $\sigma4=10-9$; $\varepsilon=10-7$.
20. *A* is a 5×5 matrix with singular values $\sigma1=15$, $\sigma2=8$, $\sigma3=10-6$, $\sigma4=10-8$, $\sigma5=10-9$; $\varepsilon=10-7$.

TRUE OR FALSE Exercises 21–24: Determine if the statement is true or false, and justify your answer.

21.

(a) For an $n\times m$ matrix *A*, both *A* and A^T have the same singular values.

(b) The singular values of *A* are equal to the square root of the eigenvalues of *A*.

22.

(a) If *A* is an $n\times m$ matrix, then *A* has a singular value decomposition only if $n>m$.

(b) The singular values of a matrix *A* are all positive.

23.

(a) If *A* is an invertible matrix with singular value σ, then A^{-1} has singular value σ^{-1}.

(b) If *A* is a square matrix, then the singular value decomposition of *A* is the same as the diagonalization of *A*.

24.

(a) If *A* is a square matrix, then $|\det(A)|$ is equal to the product of the singular values of *A*.

(b) The largest singular value of an orthogonal matrix is 1.

25. For the matrix in Exercise 11, compute $\sigma_1 u_1 v_1^T$ and $\sigma_1 u_1 v_1^T+\sigma_2 u_2 v_2^T$, and compare your results to the original matrix.

26. For the matrix in Exercise 12, compute $\sigma_1 u_1 v_1^T$ and $\sigma_1 u_1 v_1^T+\sigma_2 u_2 v_2^T$, and compare your results to the original matrix.

27. For the matrix in Exercise 15, compute $\sigma_1 u_1 v_1^T$ and $\sigma_1 u_1 v_1^T+\sigma_2 u_2 v_2^T$, and compare your results to the original matrix.

28. Prove that if *A* is a symmetric matrix with eigenvalue λ, then *A* has singular value $|\lambda|$.

29. Prove that the positive singular values of *A* and A^T are the same.

30. Prove that if σ is a singular value of *A*, then there exists a nonzero vector $\mathbf{x}$ such that

$$\sigma=\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$$

Exercises 31–34: Assume that *A* is an $n\times n$ matrix with SVD $A=U\Sigma V^T$.

31. If *A* is invertible, find a SVD of A^{-1}.

32. Prove that the columns of *U* are eigenvectors of AA^T.

33. If *P* is an orthogonal $n\times n$ matrix, prove that PA has the same singular values as *A*.

34. Let $U=[u_1 \cdots u_n]$ and $V=[v_1 \cdots v_n]$.

(a) If $n=2$, show that $A=\sigma_1 u_1 v_1^T+\sigma_2 u_2 v_2^T$.

(b) For $n\geq 1$, show that $A=\sigma_1 u_1 v_1^T+\cdots+\sigma_n u_n v_n^T$.

35. Suppose that *A* is an $m\times n$ matrix. Prove that $\text{rank}(A^TA)=\text{rank}(A)$ by verifying each of the following:

(a) Show that if x is a solution to $Ax=0$, then x is a solution to $A^TAx=0$.

(b) Suppose that x satisfies $A^TAx=0$. Show that Ax is in $(\text{col}(A))^{\perp}$. Since Ax is also in $\text{col}(A)$ (justify!), show that this means that $Ax=0$.

(c) Combine (a) and (b) to show that $\text{nullity}(A^TA)=\text{nullity}(A)$, and then apply the Rank–Nullity Theorem to conclude that $\text{rank}(A^TA)=\text{rank}(A)$.

36. Prove that the orthogonal eigenvectors $\{v_1,\ldots,v_k\}$ associated with the nonzero eigenvalues of A^TA form a basis for $\text{col}(A)$ by verifying each of the following:

(a) Apply Exercise 35 to show that $\dim(\text{col}(A^TA))=\dim(\text{col}(A))$.

(b) Apply the Spectral Theorem to explain why the orthogonal eigenvectors $\{v_1,\ldots,v_k\}$ associated with the nonzero eigenvalues of A^TA form a basis for $\text{col}(A^TA)$.

(c) Combine (a) and (b) to reach the desired conclusion.

C Exercises 37–40: Find a singular value decomposition for *A* by following the steps illustrated in this section, by using computer software to assist with finding the required eigenvalues, eigenvectors, and orthogonal bases.

37. $A=[35-12]$

38. $A=[2-16-302]$

39. $A=[-5021-13042]$

40. $A=[23-101213-2-113]$

8.5 Least Squares Regression

A problem that arises in a wide variety of disciplines is that of finding algebraic formulas to describe data. A simple example involving the relationship between barometric pressure and the boiling point of water is described below.

▶ This section is optional. However, least squares regression is revisited in optional Section 10.3.

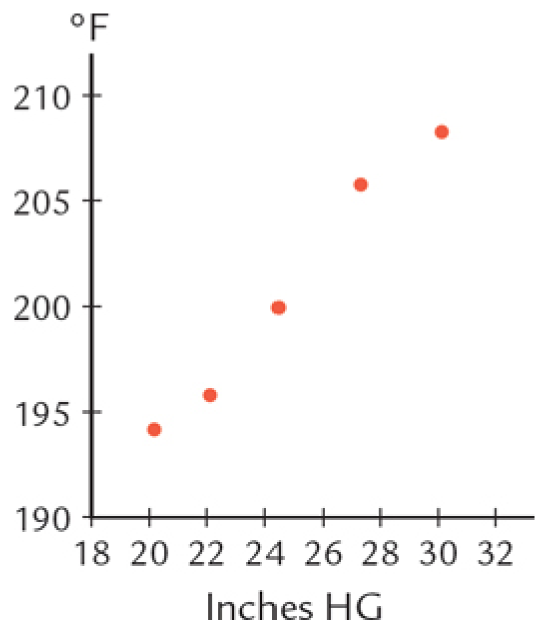


Figure 1 Scatter plot of pressure against boiling point.

Example 1

The boiling point of water is known to vary depending on the barometric pressure. To determine the relationship between boiling point and pressure, boiling points were found experimentally at several different barometric pressures. The results are summarized in Table 1 and the data plotted in Figure 1. Find a linear equation of the form T=c0+c1 P that will allow us to make accurate predictions of boiling point *T* for a given barometric pressure *P*.

Table 1 **Boiling Point of Water at Different Barometric Pressures**

Barometric Pressure (inches HG)	20.2	22.1	24.5	27.3	30.1
Boiling Point (°F)	195	197	202	209	212

Solution To find the coefficients c0 and c1, we could try plugging *T* and *P* for each data point into T=c0+c1P, yielding the linear system

$$c0+20.2c1=195c0+22.1c1=197c0+24.5c1=202c0+27.3c1=209c0+30.1c1=212 \quad (1)$$

Unfortunately, this system cannot have any solutions. If it did, then our points would lie exactly on a line, but we see in Figure 1 that they do not. However, since the points all lie close to a line, the system (1) “almost” has a solution.

In this section we develop an approximation method that gives us a way to change a linear system that has no solutions into a new system that has a solution. Our method is such that we change the system as little as possible, so that the solution to the new system can serve as an approximate solution for the original system. We will return to this example when we have the tools we need to find an answer.

Example 1 suggests a more general problem—namely, that of finding an “approximate” solution to a linear system

$$Ax=y \quad (2)$$

that has no solutions. We solve this problem by changing the vector y in (2) into a new vector y^ such that

$$Ax=y\hat{} \quad (3)$$

has a solution. In order for (3) to have a solution, we must select y^ from among the vectors in col(A), the column space of *A*. We want the systems (2) and (3) to be as similar as possible, so we choose $\hat{\mathbf{y}}$ in order to minimize

$$\|\mathbf{y}-\hat{\mathbf{y}}\|^2=(y_1-\hat{y}_1)^2+(y_2-\hat{y}_2)^2+\cdots+(y_n-\hat{y}_n)^2 \tag{4}$$

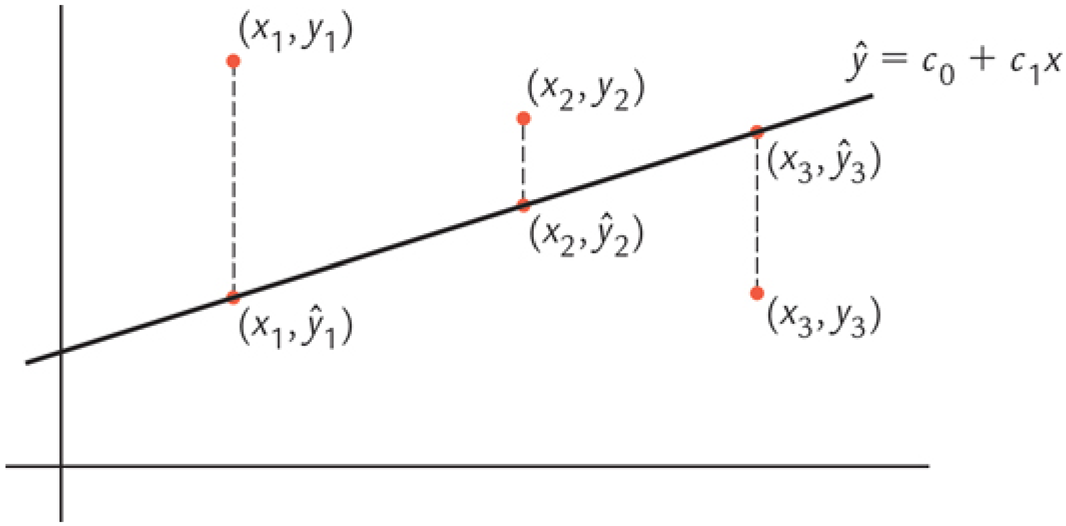


Figure 2 The data points are (x_i, y_i), and $(x_i, \hat{y}_i)$ are the corresponding points on the line $\hat{y}=c_0+c_1x$. The sum of the squares of the lengths of the dotted lines equals the expression (4). The coefficients c_0 and c_1 are chosen to make (4) as small as possible.

▶ The vector $\hat{\mathbf{y}}$ that minimizes (4) also minimizes $\|\mathbf{y}-\hat{\mathbf{y}}\|$. Squaring the distance simplifies calculations by eliminating the square root in the norm formula.

Thus we want to find $\hat{\mathbf{y}}$ as close as possible to $\mathbf{y}$, subject to the constraint that $\hat{\mathbf{y}}$ is in col(A). This is depicted in Figure 2.

To find $\hat{\mathbf{y}}$, we use projections developed in Section 8.2. A key property of projections is contained in the next theorem. A proof is given at the end of the section.

THEOREM 8.30 ▶

Let $\mathbf{y}$ be a vector and S a subspace. Then the vector closest to $\mathbf{y}$ in S is given by $\hat{\mathbf{y}}=\text{proj}_S\mathbf{y}$.

To sum up, given a linear system $A\mathbf{x}=\mathbf{y}$ that has no solutions, we find an approximate solution by solving $A\mathbf{x}=\hat{\mathbf{y}}$, where $\hat{\mathbf{y}}=\text{proj}_S\mathbf{y}$ and $S=\text{col}(A)$. This approach is called *Least Squares Regression* (or *Linear Regression*), and a solution $\hat{\mathbf{x}}$ to $A\mathbf{x}=\hat{\mathbf{y}}$ is called a **least squares solution**.

Least Squares Solution

Example 2

Find the least squares solution $\hat{\mathbf{x}}$ to $A\mathbf{x}=\mathbf{y}$, where

$$A=[\mathbf{a}_1\ \mathbf{a}_2]=\begin{bmatrix}1 & 4\\ -3 & 3\\ 5 & 1\end{bmatrix} \text{ and } \mathbf{y}=\begin{bmatrix}-16\\ 28\\ 6\end{bmatrix}$$

Solution The first step is to find $\hat{\mathbf{y}}$. If $S=\text{col}(A)$, then by Theorem 8.30 the vector $\hat{\mathbf{y}}$ in S that is closest to $\mathbf{y}$ is $\hat{\mathbf{y}}=\text{proj}_S\mathbf{y}$. Since the columns of A are orthogonal, we can apply the projection formula in Definition 8.17 in Section 8.2 to compute

$$\hat{\mathbf{y}}=\text{proj}_S\mathbf{y}=\frac{\mathbf{a}_1\cdot\mathbf{y}}{\|\mathbf{a}_1\|^2}\mathbf{a}_1+\frac{\mathbf{a}_2\cdot\mathbf{y}}{\|\mathbf{a}_2\|^2}\mathbf{a}_2=\frac{-70}{35}\begin{bmatrix}1\\ -3\\ 5\end{bmatrix}+\frac{26}{26}\begin{bmatrix}4\\ 3\\ 1\end{bmatrix}=\begin{bmatrix}2\\ 9\\ -9\end{bmatrix}$$

We find the least squares solution $\hat{\mathbf{x}}$ by solving the system $A\mathbf{x}=\hat{\mathbf{y}}$, which is

$$\begin{aligned} x_1+4x_2&=2\\ -3x_1+3x_2&=9\\ 5x_1+x_2&=-9\end{aligned}$$

Using our usual solution methods, we find that $\hat{\mathbf{x}}=\begin{bmatrix}-2\\ 1\end{bmatrix}$.

An alternative definition of least squares solution is given below.

DEFINITION 8.31 ▶

Least Squares Solution

If A is an $n \times m$ matrix and $\mathbf{y}$ is in $\mathbf{R}^n$, then a **least squares solution** to $A\mathbf{x} = \mathbf{y}$ is a vector $\hat{\mathbf{x}}$ in $\mathbf{R}^m$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{y}\| \leq \|A\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}$ in $\mathbf{R}^m$.

If $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}_0$, then $\hat{\mathbf{x}} = \mathbf{x}_0$. If A has linearly independent columns, then $\hat{\mathbf{x}}$ will be unique. If not, then there are infinitely many least squares solutions $\hat{\mathbf{x}}$.

The solution method using projection demonstrated in Example 2 requires an orthogonal basis for $S = \mathrm{col}(A)$. The following theorem is more convenient to use because it does not have this requirement.

THEOREM 8.32

Normal Equations

The set of least squares solutions to $A\mathbf{x} = \mathbf{y}$ is equal to the set of solutions to the system

$$A^T A\mathbf{x} = A^T\mathbf{y} \tag{5}$$

The equations in (5) are called the **normal equations** for $A\mathbf{x} = \mathbf{y}$. If A has linearly independent columns, then there is a unique least squares solution given by

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y} \tag{6}$$

Otherwise, there are infinitely many least squares solutions.

Proof Starting with (5), suppose that $\hat{\mathbf{x}}$ is a solution to $A\mathbf{x} = \hat{\mathbf{y}}$, where $\hat{\mathbf{y}} = \mathrm{proj}_S \mathbf{y}$ and $S = \mathrm{col}(A)$. By Theorem 8.18 in Section 8.2, $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathrm{proj}_S \mathbf{y}$ is in $S^\perp$. Since $S^\perp = (\mathrm{col}(A))^\perp = \mathrm{null}(A^T)$ (see Exercise 67 of Section 8.1), it follows that $A^T(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{0}$. As $A\hat{\mathbf{x}} = \hat{\mathbf{y}}$, we have

$$A^T(\mathbf{y}-A\hat{\mathbf{x}})=\mathbf{0} \;\Rightarrow\; A^TA\hat{\mathbf{x}}=A^T\mathbf{y}$$

The reasoning also works in the reverse direction, completing the proof of (5).

To prove (6), note that A^TA is invertible if and only if *A* has linearly independent columns (see Exercise 29). Hence (5) has a unique solution if *A* has linearly independent columns, with the solution given by (6). Otherwise, A^TA is not invertible, and (5) has infinitely many solutions. ■ ■

▶ When *A* has linearly independent columns, the formula for $\hat{\mathbf{x}}$ in (6) can be applied. However, for large data sets numerical issues can arise in calculating $(A^TA)^{-1}$ that may make using (5) attractive.

Example 3

Complete Example 1 by finding the coefficients c_0 and c_1 for the line $T=c_0+c_1P$ that best fits the data in Table 1.

Solution We need to find the least squares solution $\hat{\mathbf{c}}$ for the linear system (1), which is equivalent to $A\mathbf{c}=\mathbf{t}$, where

$$A=\begin{bmatrix}1 & 120.2\\ 1 & 122.1\\ 1 & 124.5\\ 1 & 127.3\\ 1 & 130.1\end{bmatrix},\quad \mathbf{c}=\begin{bmatrix}c_0\\ c_1\end{bmatrix},\quad \text{and}\quad \mathbf{t}=\begin{bmatrix}195\\ 197\\ 202\\ 209\\ 212\end{bmatrix}$$

Although the notation is different than in our general development of least squares solutions, the method of solution is the same. Since the columns of *A* can be seen to be linearly independent, we have

$$\hat{\mathbf{c}}=(A^TA)^{-1}A^T\mathbf{t}=\begin{bmatrix}157.17\\ 1.845\end{bmatrix}$$

Therefore the equation that best fits the data is $T=157.17+1.845P$. A graph of the data and line is shown in Figure 3.

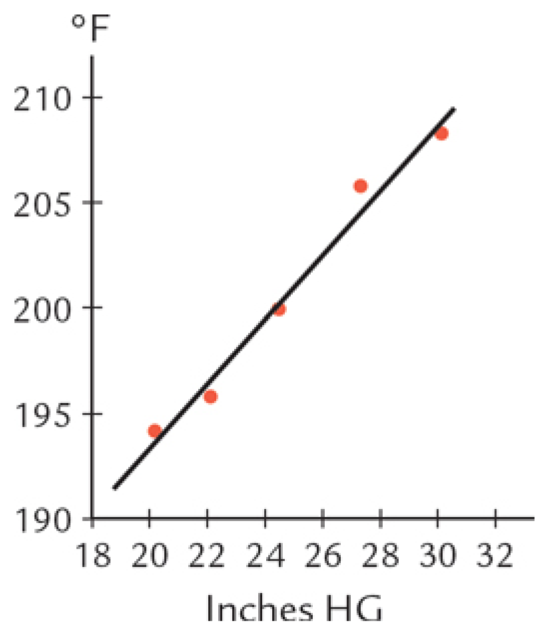


Figure 3 The data and linear equation from Example 3.

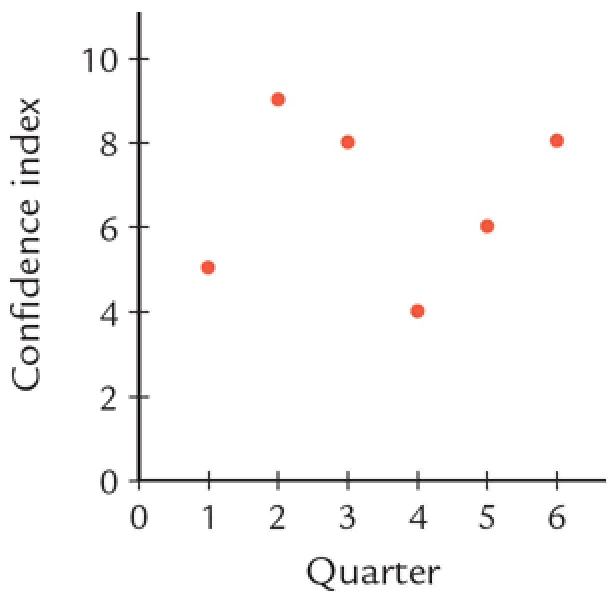


Figure 4 Data for the consumer confidence index in Example 4.

Fitting Functions to Data

Example 4

An economist conducts quarterly surveys to measure consumer confidence. The confidence indices for six consecutive quarters are given in Table 2. Find a cubic polynomial that approximates the data.

Table 2 **Quarterly Consumer Confidence Index Data**

Quarter	1	2	3	4	5	6
Confidence Index	5	9	8	4	6	8

Solution The data set is displayed in Figure 4. We want to fit to it a cubic polynomial of the form

$$g(t)=c_0+c_1t+c_2t^2+c_3t^3$$

For each data point, we plug the quarter number into g and set the result equal to the corresponding index. We get the system of equations

$$\begin{aligned}
(1,5) &\Rightarrow c_0+c_1+c_2+c_3=5\\
(2,9) &\Rightarrow c_0+2c_1+4c_2+8c_3=9\\
(3,8) &\Rightarrow c_0+3c_1+9c_2+27c_3=8\\
(4,4) &\Rightarrow c_0+4c_1+16c_2+64c_3=4\\
(5,6) &\Rightarrow c_0+5c_1+25c_2+125c_3=6\\
(6,8) &\Rightarrow c_0+6c_1+36c_2+216c_3=8
\end{aligned}$$

The system is equivalent to $A\mathbf{c}=\mathbf{y}$, where

$$A=\begin{bmatrix}1&1&1&1\\1&2&4&8\\1&3&9&27\\1&4&16&64\\1&5&25&125\\1&6&36&216\end{bmatrix},\quad \mathbf{c}=\begin{bmatrix}c_0\\c_1\\c_2\\c_3\end{bmatrix},\quad \mathbf{y}=\begin{bmatrix}5\\9\\8\\4\\6\\8\end{bmatrix}$$

The columns of A can be verified to be linearly independent, so that by Theorem 8.32 we have (some rounding is included)

$$\hat{\mathbf{c}}=(A^TA)^{-1}A^T\mathbf{y}=\begin{bmatrix}-5.33\\15.07\\-5.02\\0.48\end{bmatrix}$$

Hence the best-fitting cubic polynomial is

$$g(t)=-5.33+15.07t-5.02t^2+0.48t^3$$

A plot of g(t) together with the data is shown in Figure 5.

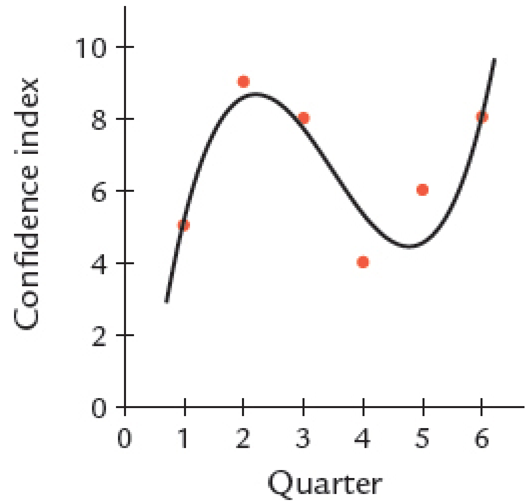


Figure 5 g(t) and the consumer confidence index data.

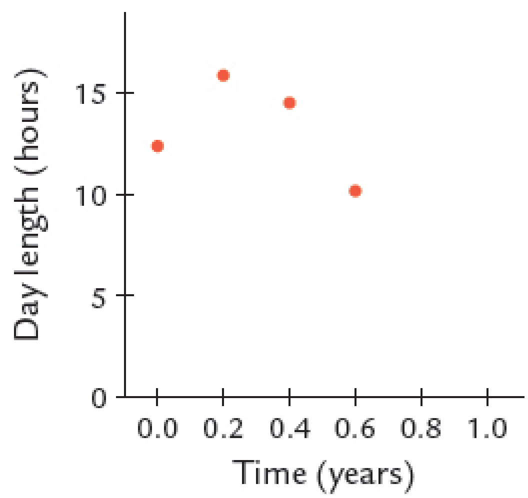


Figure 6 Length of day data. The time t=0 corresponds to March 20, the vernal equinox.

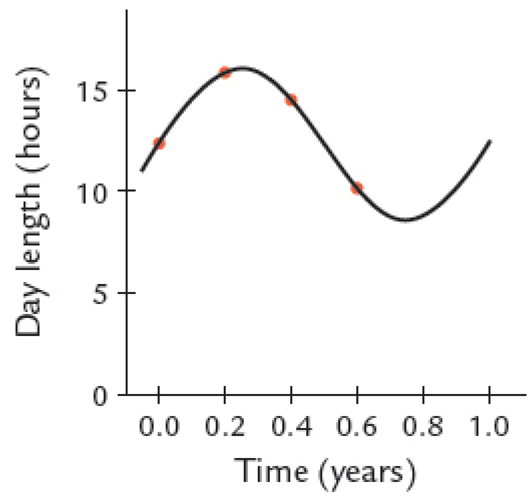


Figure 7 Length of day data and the graph of L(t).

Example 5

The times from sunrise to sunset in Vancouver, BC, on selected days in 2010 are given in Table 3 and plotted in Figure 6. The length of day can be modeled by a function of the form $L(t)=c1+c2\sin(2\pi t)$, where *t* is time in years. Find the coefficients c1 and c2 that will give the best fit to the data.

Table 3 **The Time from Sunrise to Sunset on Selected Days in Vancouver, BC**

Date	Mar 20	June 1	Aug 12	Oct 25
Day Length (hours)	12.17	16.23	12.12	8.18

Solution If we let t=0 correspond to March 20, then the remaining dates occur at t=0.2, t=0.4, and t=0.6, respectively. Evaluating L(t) at each of the four times, we obtain the system

$$(0,\ 12.17) \Rightarrow c1+c2\ \sin(0)=c1\ \ =12.17(0.2,\ 15.95) \Rightarrow$$
$$c1+c2\ \sin(0.4\pi)=c1+0.951c2=15.95(0.4,\ 14.55) \Rightarrow$$
$$c1+c2\ \sin(0.8\pi)=c1+0.588c2=14.55(0.6,\ 10.22) \Rightarrow$$
$$c1+c2\ \sin(1.2\pi)=c1-0.588c2=10.22$$

The equivalent system is Ac=y, where

$$A=[1010.95110.5881-0.588],c=[c1c2],y=[12.1715.9514.5510.22]$$

The columns of *A* are linearly independent, so that by Theorem 8.32 we have (some rounding is included)

$$\hat{c}=(ATA)-1ATy=[12.333.75]$$

Therefore the best fit is given by L(t)=12.33+3.75 sin(2πt). A graph of the data and L(t) are shown in Figure 7.

Planetary Orbits Revisited

In Section 1.3 we considered the problem of finding a model that predicts the orbital period (time required to circle the sun) for a planet based on the planet's distance from the sun. We start with the model

$$p=adb$$

where *P* is the orbital period, *d* is the distance from the sun, and *a* and *b* are constants we estimate from the data. To make the model linear, we take the logarithm on both sides of the equation and set a1=ln(a), giving us

$$\ln(p)=a1+b \ln(d)$$

Our goal here is to find values for a1 and *b*. Previously, we did not have the tools to simultaneously incorporate all of our data into the model, because doing so would have resulted in a system with no solutions. However, now we can find a least squares solution to the system that does incorporate all of the data.

Since the model involves ln(p) and ln(d), as a first step we need the logarithms of the periods and distances. This is given in Table 4.

Table 4 Planetary Orbital Distances and Periods

Planet	Distance (×106 km)	Orbital Period (Days)	ln(d)	ln(p)
Mercury	57.9	88	4.059	4.477
Venus	108.2	224.7	4.684	5.415
Earth	149.6	365.2	5.008	5.900
Mars	227.9	687	5.429	6.532
Jupiter	778.6	4331	6.657	8.374
Saturn	1433.5	10,747	7.268	9.282
Uranus	2871.5	30,589	7.963	10.328
Neptune	4495.1	59,800	8.411	10.999

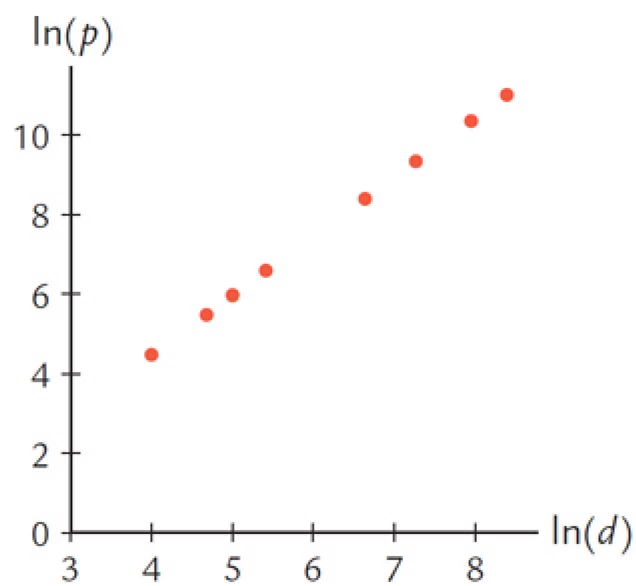


Figure 8 Scatter plot of data (ln(d), ln(p)).

The graph in Figure 8 shows a plot of the points (ln(d), ln(p)). The points lie very close to a line, suggesting that we are on the right track. Substituting each of the points into the equation ln(p)=a1+b ln(d) yields the system of equations Ax=y, where

A=[14.05914.68415.00815.42916.65717.26817.96318.411],x=[a1b], y=[4.4775.4155.9006.5328.3749.28210.32810.999]

Since A has linearly independent columns, we can use (6) to compute

$$\hat{x} = (A^T A)^{-1} A^T y = \begin{bmatrix} -1.60322 \\ 1.49827 \end{bmatrix}$$

Therefore $a_1 = -1.60322$ and $b = 1.49827$. Since $a_1 = \ln(a)$, it follows that $a = e^{-1.60322} = 0.201247$. Thus our model is

$$p = 0.201247 d^{1.49827}$$

This is consistent with Kepler's third law of motion, which predicts that the exponent should be 3/2. Table 5 gives the predicted values for the orbital periods. Note that the model provides generally better predictions than those we came up with in Section 1.3.

Table 5 **Orbital Distances, Periods, and Predicted Periods**

Planet	Distance ($\times 10^6$ km)	Orbital Period (Days)	Predicted Period
Mercury	57.9	88	88.0
Venus	108.2	224.7	224.7
Earth	149.6	365.2	365.1
Mars	227.9	687	685.9
Jupiter	778.6	4331	4322.2
Saturn	1433.5	10,747	10,786.1
Uranus	2871.5	30,589	30,543.0
Neptune	4495.1	59,800	59,775.1

Proof of Theorem 8.30

Proof of Theorem 8.30 Let s be a vector in S, and let $\hat{y} = \text{proj}_S y$. Then by Theorem 8.18 in Section 8.2, $y - \hat{y}$ is in $S^{\perp}$. Also, both $\hat{y}$ and s are in S, so $\hat{y} - s$ is in S because S is a subspace. Therefore $y - \hat{y}$ and $\hat{y} - s$ are orthogonal, so by Theorem 8.6 in Section 8.1 (the Pythagorean Theorem) we have

$$\|y - s\|^2 = \|(y - \hat{y}) + (\hat{y} - s)\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - s\|^2$$

Since ‖y^−s‖2≥0, we may conclude that ‖y−s‖2≥‖y−y^‖2 for all s in *S*. Therefore no vector in *S* is closer to y than y^, so y^=projSy is the vector in *S* that is closest to y. Furthermore, there is equality only when y^=s, so y^ is the unique closest point.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the vector in the subspace *S* closest to y.

(a) y=[21], S=span{[1−4]}

(b) y=[10−1], S=span{[12−1], [−31−1]}

2. Find the normal equations for the system

3x1−2x2=12x1+x2=−2x1−3x2=3

3. Find the equation for the line that best fits the points (−1, 1.7), (0, 1.2), (1, 0.5).

EXERCISES

Exercises 1–4: Find the vector in the subspace *S* closest to y.

1. y=[12], S=span{[1−1]}

2. y=[2−3], S=span{[31]}

3. y=[1−12], S=span{[13−2], [5−11]}

4. y=[301], S=span{[02−1], [4−1−2]}

Exercises 5–8: Find the normal equations for the given system.

5. 2x1−x2=4x1+2x2=33x1−x2=4

6. 4x1+2x2=3−2x1−x2=13x1+2x2=0

7. x1+x2−x3=22x1−x2+2x3=−1x1+4x2−5x3=6

8. −x1−2x2+2x3=−32x1+x2−3x3=8−x1−5x2+3x3=0

Exercises 9–12: Find the least squares solution(s) for the given system.

9. −x1−x2=32x1+3x2=−1−3x1+2x2=2

10. x1+2x2=−13x1−2x2=1−x1−3x2=−2

11. 2x1−x2−x3=1−x1+x2+3x3=−13x1−2x2−4x3=3

12. 3x1+2x2−x3=−22x1+3x2−x3=1−x1−4x2+x3=−3

13. Find the normal equations for the parabolas that best fit the points (0, 1) and (2, 5), and explain why the system should have infinitely many solutions.

14. Find the normal equations for the cubic polynomials that best fit the points (0, 1), (1, 4), and (3, −1), and explain why the system should have infinitely many solutions.

FIND AN EXAMPLE Exercises 15–20: Find an example that meets the given specifications.

15. A linear system with three equations and two variables that has no solutions and a unique least squares solution.

16. A linear system with four equations and three variables that has no solutions and a unique least squares solution.

17. A linear system with four equations and two variables that has no solutions and infinitely many least squares solutions.

18. A linear system with four equations and three variables that has no solutions and infinitely many least squares solutions.

19. A linear system with three equations and three variables that has a unique solution and a unique least squares solution.

20. A linear system with four equations and three variables that has a unique solution and a unique least squares solution.

TRUE OR FALSE Exercises 21–24: Determine if the statement is true or false, and justify your answer.

21.

(a) A least squares solution can be found only for a linear system that has more equations than variables.

(b) If a linear system has infinitely many solutions, then it also has infinitely many least squares solutions.

22.

(a) If *A* is an n×m matrix, then any least squares solution of Ax=y must be in Rm.

(b) The linear system Ax=y has a unique least squares solution if the columns of *A* are linearly independent.

23.

(a) A least squares solution to Ax=y is a vector x^ such that Ax^ is as close as possible to y.

(b) The system of normal equations for the linear system Ax=y has solutions if and only if Ax=y has solutions.

24.

(a) If x^1 and x^2 are least squares solutions of Ax=y, then so is x^1+x^2.

(b) A linear system must be inconsistent in order for there to be infinitely many least squares solutions.

25. Prove that if the matrix *A* has orthogonal columns, then Ax=y has a unique least squares solution.

26. Suppose that *A* is a nonzero matrix and S=col(A). Prove that if Ax=y has a solution, then y=projSy.

27. Prove that if *A* is an orthogonal matrix, then any least squares solution of Ax=y is a linear combination of the rows of *A*.

28. Prove that if x^ is a least squares solution of Ax=y and x0 is in null(ATA), then x^+x0 is also a least squares solution ofAx=y.

29. For a matrix *A*, prove that ATA is invertible if and only if *A* has linearly independent columns. (HINT: See Exercise 35 of Section 8.4.)

30. Prove that if *A* has orthonormal columns, then x^=ATy is the unique least squares solution to Ax=y.

Exercises 31–34: Find the equation for the line that best fits the given data.

31. (−2, 1.3), (0, 1.8), (1, 3)

32. (−3, −1.6), (−1, −1.9), (1, −2.5)

33. (−2, 2.0), (−1, 1.7), (1, 2.6), (3, 2.1)

34. (1, 3.1), (2, 2.6), (4, 1.9), (5, 2.1)

Exercises 35–38: Find the equation for the parabola that best fits the given data.

35. (−2, 3), (−1, 2), (1, 2.1), (2, 3.4)

36. (−3, −1), (−1, 2), (1, 4), (2, 1)

37. (−2, 0), (−1, 1.5), (0, 2.5), (1, 1.3), (2, −0.2)

38. (0, 4), (1, 3), (2, 1), (3, 2), (4, 5)

Exercises 39–42: Find constants *a* and *b* so that the model y=aebx best fits the given data.

39. (−1, 0.3), (0, 1.3), (1, 3.1), (2, 5.7)

40. (1, 2.1), (3, 3.2), (4, 3.9), (6, 5.8), (9, 10.8)

41. (2, 11.3), (4, 8.2), (5, 7.1), (7, 5.3), (10, 3.2)

42. (−1, 5.1), (0, 1.9), (1, 0.8), (2, 0.3)

Exercises 43–49: Find constants *a* and *b* so that the model y=axb best fits the given data.

43. (2, 5.4), (4, 13.5), (5, 17.6), (7, 26.0), (9, 40.2)

44. (1, 4.0), (3, 25.2), (4, 42.1), (6, 78.3), (8, 130.4)

45. (2, 26.1), (3, 21.7), (5, 15.8), (7, 12.7), (10, 11.2)

46. (3, 24.1), (4, 18.2), (6, 15.1), (8, 11.9), (9, 10.9)

47. Apply least squares regression to the data for the planets Venus, Earth, and Mars to develop a model to predict orbital period from distance.

48. Apply least squares regression to the data for the planets Mercury, Earth, Jupiter, and Uranus to develop a model to predict orbital period from distance.

49. On January 10, 2010, Nasr Al Niyadi and Omar Al Hegelan parachuted off a platform suspended from a crane attached to the 160th floor of the Burj Khalifa in Dubai, the tallest building in the world. Suppose that an anvil was dropped from the same platform, located 2205 feet above the ground. Measurements of the elevation of the anvil *t* seconds after release are given in the table.

t	**Elevation**
1	2185 ft
2	2140 ft
3	2055 ft
4	1943 ft

Find the quadratic polynomial that best fits the data, and use it to predict how long it will take for the anvil to hit the ground.

50. Warren invests $100,000 into a fund that combines stocks and bonds. The return varies from year to year. The balance at 5-year intervals is given in the table.

t	**Balance (×1000)**
5	142
10	230
15	314
20	483

Find constants *a* and *b* such that the model y=aebt best fits the data. Use your model to predict when the investment fund balance will reach $1 million.

51. The isotope Polonium-218 is unstable and subject to rapid radioactive decay. The quantity of a sample is measured at various times, with the results in the table.

***t* (min)**	**Mass (g)**
2	1.50
4	0.97
6	0.57
8	0.41

Find constants *a* and *b* such that the model y=aebt best fits the data. Use your model to predict the initial size of the sample and the amount that will be present at t=15.

52. Measurements of CO_2 in the atmosphere have been taken regularly over the last 50 years at the Mauna Loa Observatory in Hawaii. In addition to a general upward trend, the CO_2 data also has an annual cyclic behavior. The table has the monthly measurements (in parts per million) for 2009.

Month	CO2	Month	CO2
1	386.92	7	387.74
2	387.41	8	385.91
3	388.77	9	384.77
4	389.46	10	384.38
5	390.18	11	385.99
6	389.43	12	387.27

Find constants *a*, *b*, and *c* such that the model $y=a+bt+c\sin(t\pi/6)$ best fits the data, where *t* is time in months. Use your model to predict the CO_2 level in January 2020.

SUPPLEMENTARY EXERCISES

Exercises 1–12: Refer to the following vectors:

u1=[3−10], u2=[271], u3=[11−2], u4=[13−2], u5=[042], u6=[5−24]

1. Compute the following dot prodicts:
   (a) u1·u3
   (b) u4·u5
   (c) u2·(−u6)
   (d) 2u5·(−3u2)
2. Compute the following dot prodicts:
   (a) u4·u3
   (b) u2·u5
   (c) 3u6·(−u4)
   (d) −2u2·(−2u3)
3. Compute the norms of the given vectors.
   (a) u1
   (b) −u4
   (c) 3u5
   (d) u3+u6
4. Compute the norms of the given vectors.
   (a) u5
   (b) −u2
   (c) −2u6
   (d) u4+3u2
5. Determine the distance between the given vectors.
   (a) u1 and u3
   (b) u2 and u5
   (c) u1 and u4
   (d) u3 and u6

6. Determine the distance between the given vectors.

(a) u6 and u2
(b) u2 and u4
(c) u5 and u1
(d) u4 and u3

7. Determine if the given vectors are orthogonal.

(a) u1 and u6
(b) u2 and u4
(c) u4 and u1
(d) u5 and u3

8. Determine if the given vectors are orthogonal.

(a) u1 and u2
(b) u2 and u4
(c) u5 and u1
(d) u6 and u2

9. Compute the following projections.

(a) proju2u3
(b) proju1u6
(c) proju5u1
(d) proju4u2

10. Compute the following projections.

(a) projSu3, where S=span{u1, u4}.
(b) projSu5, where S=span{u2, u6}.
(c) projSu6, where S=span{u3, u5}.
(d) projSu2, where S=span{u1, u4}.

11. Normalize the given vectors.

(a) u1
(b) u2
(c) u4
(d) u6

12. Apply Gram–Schmidt to find an orthogonal basis for *S*.

(a) S=span{u1, u2}.
(b) S=span{u2, u3}.
(c) S=span{u3, u4}.
(d) S=span{u5, u6}.

Exercises 13–18: Determine if *A* is symmetric and if *A* is orthogonal.

13. $A=[2-3-34]$

14. $A=[0110]$

15. $A=[121212-12]$

16. $A=[121212-12]$

17. $A=[111010111]$

18. $A=[131212130-1213-120]$

Exercises 19–22: Orthogonally diagonalize *A*, given the provided eigenvalues.

19. $A=[1-2-24]$, $\lambda=0,\ 5$

20. $A=[4-3-34]$, $\lambda=1,\ 7$

21. $A=[210111012]$, $\lambda=0,\ 2,\ 3$

22. $A=[112111211]$, $\lambda=0,\ -1,\ 3$

Exercises 23–26: Find the QR factorization of *A*.

23. $A=[4324]$

24. $A=[5231]$

25. $A=[211003]$

26. $A=[524137]$

Exercises 27–32: Find the singular value decomposition of *A*.

27. $A=[1331]$

28. $A=[33-31]$

29. $A=[210-110]$

30. $A=[-102103]$

31. $A=[110201]$

32. $A=[211-101]$

Exercises 33–36: Find the vector in *S* that is closest to y.

33. $y=[32], S=\text{span}\{[1-2]\}$

34. $y=[1-3], S=\text{span}\{[34]\}$

35. y=[011], S=span{[12−2], [4−11]}

36. y=[3−21], S=span{[221], [0−12]}

Exercises 37–40: Find the normal equations for the given system.

37. −x1−2x2=1x1+x2=22x1−x2=−1

38. 3x1+x2=−1x1−2x2=12x1+x2=3

39. x1+2x2−x3=4−x1−3x2+x3=−22x1+3x2−4x3=5

40. −3x1−2x2+2x3=−1x1+3x2−2x3=7−x1−5x2+3x3=2

Exercises 41–44: Find the least squares solution(s) for the given system.

41. x1−2x2=1−2x1+x2=−4−x1+x2=3

42. 2x1+x2=1−2x1−2x2=−3−x1−x2=2

43. x1−3x2−2x3=−1−x1+5x2+3x3=23x1−2x2−x3=−3

44. x1+x2−2x3=−13x1+3x2−3x3=0−4x1−2x2+2x3=−2

CHAPTER 9

Linear Transformations

Monty Rakusen/Cultura/Getty Images

Plants and other organic materials can be converted to biofuels such as ethanol. Unlike fossil fuels, biofuels are renewable because more plants can be grown. Sometimes crops such as corn are used to create biofuels. There are criticisms of biofuels, including whether crops that could be used as food should be used for producing fuels and whether the process of growing and processing crops into fuel actually requires more energy than they provide. One possible solution is to use grasses like Miscanthus, pictured here. Miscanthus is a quick-growing grass than can be used to produce ethanol or burned to produce heat and steam for power turbines.

In Chapter 3 we defined and studied the properties of linear transformations in Euclidean space Rn. In Chapter 7 we developed the concept of a vector space. In this chapter we combine these by extending the definition of linear transformations to vector spaces.

Section 9.1 introduces the definition and basic properties of a linear transformation in the context of a vector space. Most of the definitions, such as one-to-one, onto, kernel, and range, carry over almost word-for-word from Rn. Section 9.2 focuses on a special type of linear transformation called an isomorphism and develops methods for determining when two different vector spaces have the same essential structure. In Section 9.3 we establish matrix representations for linear transformations that are similar to those for linear transformations in Euclidean space. Section 9.4 considers similar matrices, which is a way to group matrices based on their relationship to a linear transformation.

9.1 Definition and Properties

In Section 3.1 we gave the definition for a linear transformation $T : R^m \to R^n$. Recall that such a function preserves linear combinations by satisfying

$$T(c_1v_1+c_2v_2)=c_1T(v_1)+c_2T(v_2)$$

In Chapter 7 we defined the vector space, an extension of Euclidean space that allows vectors to be polynomials, matrices, continuous functions, and other types of mathematical objects.

We start this section by extending the definition of linear transformation to allow domains and codomains that are vector spaces.

DEFINITION 9.1 ▶

Linear Transformation

Let *V* and *W* be vector spaces. Then $T : V \to W$ is a **linear transformation** if for all v_1 and v_2 in *V* and all real scalars *c*, the function *t* satisfies

(a) $T(v_1+v_2)=T(v_1)+T(v_2)$

(b) $T(cv_1)=cT(v_1)$

▶ Definition 9.1 generalizes Definition 3.1 in Section 3.1.

Domain, Codomain

For a linear transformation $T : V \to W$, the vector space *V* is the **domain** and the vector space *W* is the **codomain**.

Example 1

Let $T : R^2 \to P_2$ be given by

$$T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = a_1x^2 + (a_1 - a_2)x - a_2.$$

Show that *T* is a linear transformation.

▶ Recall that P_n denotes the vector space of polynomials with real coefficients that have degree *n* or less.

Solution We need to verify that the two conditions given in Definition 9.1 are satisfied. Suppose that $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. Starting with condition (a), we have

$$\begin{aligned} T(\mathbf{a} + \mathbf{b}) &= T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \end{bmatrix}\right) \\ &= (a_1 + b_1)x^2 + ((a_1 + b_1) - (a_2 + b_2))x - (a_2 + b_2) \\ &= (a_1x^2 + (a_1 - a_2)x - a_2) + (b_1x^2 + (b_1 - b_2)x - b_2) \\ &= T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) + T\left(\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right) = T(\mathbf{a}) + T(\mathbf{b}) \end{aligned}$$

Thus condition (a) of Definition 9.1 is satisfied. For condition (b), let *c* be a real scalar. Then we have

$$\begin{aligned} T(c\mathbf{a}) &= T\left(c\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} ca_1 \\ ca_2 \end{bmatrix}\right) \\ &= ca_1x^2 + (ca_1 - ca_2)x - ca_2 \\ &= c(a_1x^2 + (a_1 - a_2)x - a_2) = cT\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = cT(\mathbf{a}) \end{aligned}$$

Hence condition (b) of Definition 9.1 is also satisfied, so *T* is a linear transformation.

Before moving on to the next example, we report the following useful theorem that wraps the two conditions of Definition 9.1 into a single package.

THEOREM 9.2 ▶

$T : V \to W$ is a linear transformation if and only if

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) \qquad (1)$$

for all vectors v_1 and v_2 in V and real scalars c_1 and c_2.

▶ On the left of (1) we have the transformation of a linear combination of two vectors, and on the right we have the same linear combination of the images of the vectors. Hence linear transformations preserve linear combinations.

The proof follows from applications of Definition 9.1 and is left to Exercise 54. Some consequences of Theorem 9.2:

- The expression (1) can be extended to

$$T(c_1v_1+c_2v_2+\cdots+c_mv_m)=c_1T(v_1)+c_2T(v_2)+\cdots+c_mT(v_m)$$

  where $v_1,\ldots,v_m$ are in V and $c_1,\ldots,c_m$ are real scalars (see Exercise 50).
- If 0_V denotes the zero vector in V and 0_W the zero vector in W, then $T(0_V)=0_W$ (see Exercise 51).
- For any v in V, we have $T(-v)=-T(v)$ (see Exercise 52).

Example 2

Let $T : P_2\to P_4$ be given by

$$T(p(x))=x^2p(x)$$

Show that T is a linear transformation.

Solution Here we apply Theorem 9.2, so that only one condition needs to be verified.

$$T(c_1p_1(x)+c_2p_2(x))=x^2(c_1p_1(x)+c_2p_2(x)) =c_1x^2p_1(x)+c_2x^2p_2(x) =c_1T(p_1(x))+c_2T(p_2(x))$$

Therefore Theorem 9.2 is satisfied, so T is a linear transformation.

Example 3

Suppose that $T : C[0, 1] \to C[0, 1]$ is defined by

$$T(f(x))=(f(x))^2$$

Prove that *T* is *not* a linear transformation.

Solution All we need to do is show that Definition 9.1 fails to hold for at least one scalar or vector. Given a scalar *c* and continuous function *f*, we have

$$T(cf(x))=(cf(x))^2=c^2(f(x))^2=c^2T(f(x))$$

Thus if $c = 2$, then

$$T(2f(x))=4T(f(x))\neq 2T(f(x))$$

Condition (b) of Definition 9.1 is violated, so *T* is not a linear transformation.

Image, Range, and Kernel

The definitions and notation for image, range, and kernel carry over essentially unchanged from Euclidean space.

Image

- If v is a vector in *V*, then T(v) is the **image of** v **under** *T*.
- If *S* is a subspace of *V*, then T(S) denotes the subset of *W* consisting of all images of elements of *S*.

Range

- The **range** of *T* is denoted range(T) and is the subset of *W* consisting of all images of elements of *V* (also can be written T(V)).

Kernel

- The **kernel** of *T* is denoted ker(T) and is the set of all elements v in *V* such that T(v)=0W.

Note that ker(T) is a subset of *V*, while range(T) is a subset of *W*.

Recall that a subspace *S* is a subset closed under linear combinations. Since linear transformations preserve linear combinations, it makes sense that the image *T*(*S*) is also a subspace, as shown in the next theorem.

THEOREM 9.3

Let T : V→W be a linear transformation. Then ker(T) is a subspace of *V*. If *S* is a subspace of *V*, then *T*(*S*) is a subspace of *W*.

Theorem 9.3 generalizes Theorem 4.5 in Section 4.1.

Proof We leave the proof that ker(T) is a subspace of *V* as Exercise 53 and prove that T(S) is a subspace of *W* here. Recall that we must verify the three conditions required of a subspace.

(a) As T(0V)=0W and 0V must be in *S*, it follows that 0W is in T(S).

(b) Suppose that w1 and w2 are both in T(S). Then there exist vectors v1 and v2 in *S* such that T(v1)=w1 and T(v2)=w2. Since v1+v2 must be in *S* and

$$T(v1+v2)=T(v1)+T(v2)=w1+w2$$

it follows that w1+w2 is also in T(S). Thus *S* is closed under addition.

(c) Suppose that w is in T(S) and that *c* is a scalar. Then there exists a vector v in *S* such that T(v)=w. As cv is also in *S* and

$$T(cv)=cT(v)=cw$$

we have cw in T(S). Hence T(S) is also closed under scalar multiplication.

Since (a), (b), and (c) are all satisfied, T(S) is a subspace of *W*.

Example 4

Let $T : R^{2\times 2} \to R$ be given by

$$T(A) = tr(A)$$

where tr(A) denotes the trace of *A*. Determine range(T) and ker(T).

Solution Recall that tr(A) is the sum of the diagonal elements of a square matrix *A*, so that

$$tr\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11} + a_{12}$$

Then *T* is a linear transformation. (Verifying this is left as Exercise 9.) Note that for any real number *r*, we have

$$tr\left(\begin{bmatrix} r & 0 \\ 0 & 0 \end{bmatrix}\right) = r$$

Therefore every real number is the image of an element of $R^{2\times 2}$, so we may conclude that $range(T) = R$.

Next, $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is in ker(T) if

$$tr(A) = a_{11} + a_{22} = 0$$

or $a_{22} = -a_{11}$. Therefore ker(T) is the set of all 2×2 real matrices of the form

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & -a_{11} \end{bmatrix}$$

Since this matrix can be written

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & -a_{11} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ 0 & -a_{11} \end{bmatrix} + \begin{bmatrix} 0 & a_{12} \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ a_{21} & 0 \end{bmatrix}$$

a basis for ker(T) (which we already know is a subspace) is given by

$$\{[1\,0\,0\,-1],\ [0\,1\,0\,0],\ [0\,0\,1\,0]\}$$

One-to-One and Onto Linear Transformations

The definitions of one-to-one and onto for linear transformations carry over almost word-for-word from Chapter 3.

DEFINITION 9.4 ▶

One-to-One, Onto

Let $T : V \to W$ be a linear transformation. Then

(a) T is **one-to-one** if for each w in W there is *at most* one v in V such that $T(v)=w$.

(b) T is **onto** if for each w in W there is *at least* one v in V such that $T(v)=w$.

▶ Definition 9.4 generalizes Definition 3.4 in Section 3.1.

One way to determine if a linear transformation is one-to-one is to find ker(T) and then apply the next theorem.

THEOREM 9.5 ▶

Let $T : V \to W$ be a linear transformation. Then T is one-to-one if and only if $T(v)=0_W$ has only the trivial solution $v=0_V$.

▶ Theorem 9.5 generalizes Theorem 3.5 in Section 3.1.

The proof is similar to that of Theorem 3.5 in Section 3.1 and is left as Exercise 55.

Example 5

Let $T: \mathbb{R}_{2\times 2} \to P_2$ be given by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a-d)x^2 - bx + c$$

Then *T* is a linear transformation (see Exercise 8). Determine if *T* is onto or one-to-one.

Solution A typical element of P_2 has the form $h(x) = ex^2 + fx + g$. Thus, for *T* to be onto, we need to be able to find a solution to

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a-d)x^2 - bx + c = ex^2 + fx + g$$

Comparing coefficients yields the linear system

$$\begin{aligned} a - d &= e \\ -b &= f \\ c &= g \end{aligned}$$

This system has infinitely many solutions, among them

$$a = e, \quad b = -f, \quad c = g, \quad d = 0$$

Hence *T* is onto. Moreover, for the special case where $e = f = g = 0$, a solution to the system is $a = d = 1$, $b = c = 0$. Therefore

$$T\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = 0$$

so that $\ker(T)$ is nontrivial. Hence *T* is not one-to-one by Theorem 9.5.

The next theorem shows how a linear transformation $T : V \to W$ can relate linearly independent sets in *V* and *W*.

THEOREM 9.6

Let $T : V \to W$ be a linear transformation. Suppose that $\mathcal{V}=\{v_1, \ldots, v_m\}$ is a subset of *V*, $\mathcal{W}=\{w_1, \ldots, w_m\}$ is a subset of *W*, and $T(v_i)=w_i$ for $i=1, \ldots, m$. If $\mathcal{W}$ is linearly independent, then so is $\mathcal{V}$.

Proof Suppose that

$$c_1v_1+\cdots+c_mv_m=0_V \tag{2}$$

Since

$$T(c_1v_1+\cdots+c_mv_m)=c_1T(v_1)+\cdots+c_mT(v_m)=c_1w_1+\cdots+c_mw_m$$

and $T(0_V)=0_W$, we have

$$c_1w_1+\cdots+c_mw_m=0_W$$

As $\mathcal{W}$ is a linearly independent set, it must be that $c_1=\cdots=c_m=0$, so that (2) has only the trivial solution. Hence $\mathcal{V}$ is also linearly independent. ■ ■

Note that the reverse is not always true. Just because $\mathcal{V}$ is linearly independent does *not* guarantee that $\mathcal{W}$ is linearly independent (see Exercise 37). However, if *T* is one-to-one and $\mathcal{V}$ is a linearly independent set, then that is enough to ensure that $\mathcal{W}$ is also linearly independent (see Exercise 57).

Example 6

Let $T : C[0, 1] \to \mathbb{R}^2$ be defined by

$$T(f)=\begin{bmatrix} f(0) \\ f(1) \end{bmatrix}$$

Use T to prove that the set $\{\cos(x\pi/2), \sin(x\pi/2)\}$ is linearly independent.

Solution It is shown in Exercise 6 that T is a linear transformation. Next, note that

$$T(\cos(x\pi/2))=\begin{bmatrix} \cos(0) \\ \cos(\pi/2) \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$T(\sin(x\pi/2))=\begin{bmatrix} \sin(0) \\ \sin(\pi/2) \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Since the set $\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$ is linearly independent, then by Theorem 9.6 the set $\{\cos(x\pi/2), \sin(x\pi/2)\}$ is also linearly independent.

Our next theorem relates the dimensions of V, $\ker(T)$, and $\mathrm{range}(T)$ for a linear transformation $T : V \to W$.

THEOREM 9.7

Let $T : V \to W$ be a linear transformation, with V and W finite dimensional. Then

$$\dim(V)=\dim(\ker(T))+\dim(\mathrm{range}(T)) \qquad (3)$$

- Theorem 9.7 generalizes Theorem 4.23 in Section 4.3.
- The dimension of the kernel is called the **nullity** of the linear transformation, and the dimension of the range is called the **rank** of the linear transformation.

Proof We start by letting $\{v_1, \dots, v_k\}$ so that $\dim(\ker(T))=k$. Now extend this set to a basis $\{v_1, \dots, v_k, v_{k+1}, \dots, v_m\}$ for V. Hence

$\dim(V)=m$. For $i=k+1,\dots,m$, let $w_i=T(v_i)$. Our goal is to show that $\{w_{k+1},\dots,w_m\}$ is a basis for range(T).

Since $\{v_1,\dots,v_m\}$ is a basis for V, every vector v in V can be expressed as a unique linear combination

$$v=c_1v_1+\cdots+c_mv_m$$

for real scalars $c_1,\dots,c_m$. Therefore

$$T(v)=T(c_1v_1+\cdots+c_kv_k+c_{k+1}v_{k+1}+\cdots+c_mv_m) =c_1T(v_1)+\cdots+c_kT(v_k)+c_{k+1}T(v_{k+1})+\cdots+c_mT(v_m) =c_10_W+\cdots+c_k0_W+c_{k+1}w_{k+1}+\cdots+c_mw_m =c_{k+1}w_{k+1}+\cdots+c_mw_m$$

▶ The cases $\dim(V)=0$ and $\dim(\ker(T))=0$ are left to the reader.

As range(T) is the set of all $T(v)$, it follows that $\{w_{k+1},\dots,w_m\}$ spans range(T). Moreover, if

$$c_{k+1}w_{k+1}+\cdots+c_mw_m=0_W$$

then $c_{k+1}v_{k+1}+\cdots+c_mv_m$ is in ker(T). Since $v_{k+1},\dots,v_m$ are linearly independent and none are in ker(T), this implies $c_{k+1}=\cdots=c_m=0$. Hence $\{w_{k+1},\dots,w_m\}$ is also linearly independent and thus is a basis for range(T). Therefore $\dim(\text{range}(T))=m-k$, and so (3) is true. ■■

For example, recall $T:\mathbb{R}^{2\times 2}\to\mathbb{R}$ given by $T(A)=\text{tr}(A)$ in Example 4. There, it is shown that a basis for ker(T) is

$$\left\{\begin{bmatrix}1&0\\0&-1\end{bmatrix}, \begin{bmatrix}0&1\\0&0\end{bmatrix}, \begin{bmatrix}0&0\\1&0\end{bmatrix}\right\}$$

so that $\dim(\ker(T))=3$. Since range(T)$=\mathbb{R}$, we have $\dim(\text{range}(T))=1$. We also have $\dim(\mathbb{R}^{2\times 2})=4$, which is exactly as predicted by Theorem 9.7.

Another application of Theorem 9.7 is given in the next theorem.

THEOREM 9.8 ▶

Let $T : V \to W$ be a linear transformation, with V and W finite dimensional.

(a) If T is onto, then $\dim(V) \geq \dim(W)$.

(b) If T is one-to-one, then $\dim(V) \leq \dim(W)$.

Proof If T is onto, then $\text{range}(T) = W$. Therefore $\dim(\text{range}(T)) = \dim(W)$, so that by Theorem 9.7,

$$\dim(V) = \dim(\ker(T)) + \dim(W)$$

Since $\dim(\ker(T)) \geq 0$, it follows that $\dim(V) \geq \dim(W)$, so (a) is true.

For part (b), if T is one-to-one, then by Theorem 9.5 we have $\ker(T) = \{0_V\}$. Hence $\dim(\ker(T)) = 0$, and so by Theorem 9.7,

$$\dim(V) = \dim(\text{range}(T)) \leq \dim(W)$$

because $\text{range}(T)$ is a subset of W. Hence (b) is also true. ■■

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Let $T : V \to P_1$ be a linear transformation satisfying

$$T(v_1) = 3x - 2 \quad T(v_2) = -5x \quad T(v_3) = x + 7$$

Find $T(v_1 - 4v_2 + 2v_3)$.

2. Let $T : P_1 \to \mathbf{R}^2$ be a linear transformation satisfying

$$T(x-1) = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad T(-2x+3) = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

Find $T(-x+3)$.

3. Prove that T is a linear transformation.

(a) T : P2→R2 with T(ax2+bx+c)=[a+bb−2c]
(b) T : R2×2→P2 with T ([abcd])=ax2−dx+5b

4. Determine if *T* is a linear transformation. Justify your answer.

(a) T : P2→P3 with T(ax2+bx+c)=(ax+b)3
(b) T : R2×3→R3×2 with T(A)=AT

5. Find the kernel and range of the linear transformation *T*.

(a) T : P1→R2 with T(ax+b)=[6a−4b−9a+6b]
(b) T : P2→R with T(ax2+b+c)=a−3b+c
(c) The **identity linear transformation** T : V→V given by T(v)=v
(d) The **trivial linear transformation** T : V→W given by T(v)=0W

EXERCISES

1. Let T : V→R2 be a linear transformation satisfying

T(v1)=[12], T(v2)=[−31]

Find T(v2−2v1).

2. Let T : V→P2 be a linear transformation satisfying

T(v1)=x2+1T(v2)=−x2+3xT(v3)=4x−2

Find T(2v1+v2−3v3).

3. Let T : P2→R2 be a linear transformation satisfying

T(x2+1)=[−13], T(4x+3)=[21]

Find T(2x2−4x−1). (HINT: Express 2x2−4x−1 as a linear combination of x2+1 and 4x+3.)

4. Let T : R2×2→P2 be a linear transformation satisfying

T ([1−2−13])=x2−x+3

T ([241−1])=3x2+4x−1

Find T ([72−17]). (HINT: Write [72−17] as a linear combination of [1−2−13] and [241−1].)

Exercises 5–10: Prove that the given function is a linear transformation.

5. $T : P2 \to P2$ with

$$T(ax2+bx+c)=cx2+bx+a$$

6. $T : C\,[0, 1] \to R2$ with $T(f\ )=[f\,(0)f\,(1)]$

7. $T : Pn \to C\,[0, 1]$ with $T(p(x))=exp(x)$

8. $T : R2\times2 \to P2$ with

$$T\,([abcd])=(a-d)x2-bx+c$$

9. $T : R2\times2 \to R$ with $T(A)=tr(A)$ (the trace of *A*).

10. $T : V \to W$ with $T(v)=0W$ for all v.

Exercises 11–22: Determine if the given function is a linear transformation. Be sure to completely justify your answer.

11. $T : Rn \to Rn$ with $T(v)=-4v$

12. $T : R2 \to R2$ with

$$T\,([ab])=-[ab]+[32]$$

13. $T : R2 \to R2$ with $T\,([ab])=[ba]$

14. $T : Rn \to Rn$ with $T(v)=[0\vdots0]$

15. $T : P2 \to R2$ with $T(ax2+bx+c)=[a-bb+c]$

16. $T : R2\times2 \to R$ with $T(A)=det(A)$

17. $T : Rn\times n \to Rn\times n$ with $T(A)=AT$

18. $T : C\,[0, \pi] \to R$ with $T(f)=f(2)$

19. $T : Rn\times n \to R$ with $T(A)=tr(A)$

20. $T : C\,[0, 1] \to C\,[0, 1]$ with $T(f)=x+f(x)$

21. $T : R3\times2 \to R2\times2$ with $T(A)=ATA$

22. $T : R2 \to C\,[0, 1]$ with $T\,([ab])=aebx$

Exercises 23–26: Describe the kernel and range of the given linear transformation.

23. $T : P1 \to R$ with $T(ax+b)=a-b$

24. $T : P_1 \to P_2$ with $T(f) = xf(x)$
25. $T : P_2 \to R^{2\times 2}$ with

$$T(ax^2+bx+c) = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

26. $T : C[0, 1] \to P_2$ with

$$T(f) = f(0)x^2 + f(1)$$

Exercises 27–30: Determine if the given linear transformation is onto and/or one-to-one.

27. $T : P_2 \to R^2$ with $T(f) = \begin{bmatrix} f(1) \\ f(2) \end{bmatrix}$
28. $T : P_2 \to P_3$ with $T(f) = xf(x)$
29. $T : C[0, 1] \to R$ with $T(f) = f(1)$
30. $T : R^2 \to P_2$ with

$$T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = ax^2 + (b-a)x + (a-b)$$

FIND AN EXAMPLE Exercises 31–38: Find an example of vector spaces *V* and *W* and a function $T : V \to W$ that meets the given specifications.

31. *T* is a linear transformation that is one-to-one but not onto.
32. *T* is a linear transformation that is onto but not one-to-one.
33. *T* is a linear transformation that is neither onto nor one-to-one.
34. *T* is a linear transformation that is both onto and one-to-one.
35. *T* is a linear transformation such that $\dim(\ker(T)) = 1$ and $\dim(\text{range}(T)) = 3$.
36. *T* is a linear transformation such that $\dim(\ker(T)) = 4$ and $\dim(\text{range}(T)) = 2$.
37. *T* is a linear transformation such that, for any set of linearly independent vectors $\{v_1, \ldots, v_k\}$ the set $\{T(v_1), \ldots, T(v_k)\}$ is linearly dependent.
38. *T* satisfies condition (b) but not condition (a) of Definition 9.1.

TRUE OR FALSE Exercises 39–44: Determine if the statement is true or false, and justify your answer.

39.

(a) If $T : V \to W$ is a linear transformation, then $T(v_1 - v_2) = T(v_1) - T(v_2)$.

(b) If $T : V \to W$ is a linear transformation, then $T(v) = 0_W$ implies that $v = 0_V$.

40.

(a) If $T : V \to W$ is a linear transformation and S is a subspace of V, then $T(S)$ is a subspace of W.

(b) If $T : V \to W$ is a linear transformation, then $\dim(\ker(T)) \le \dim(\mathrm{range}(T))$.

41.

(a) There exists a linear transformation $T : V \to W$ such that $T(v_1) = w_1$ and $T(-v_1) = 2w_1$ for some v_1 in V and w_1 in W.

(b) If $T : V \to W$ is a linear transformation and $T(v) = 0$ for some nonzero vector v, then $\dim(\mathrm{range}(T)) < \dim(W)$.

42.

(a) If $T : V \to W$ is a linear transformation and $\{v_1, \ldots, v_k\}$ is a linearly independent set, then so is $\{T(v_1), \ldots, T(v_k)\}$.

(b) If $T : V \to W$ is a linear transformation and $\{v_1, \ldots, v_k\}$ is a linearly dependent set, then so is $\{T(v_1), \ldots, T(v_k)\}$.

43.

(a) If $T : \mathbb{R}^{2\times 2} \to P_6$, then is it impossible for T to be onto.

(b) If $T : P_4 \to \mathbb{R}^6$, then it is impossible for T to be one-to-one.

44.

(a) Let $T : V \to W$ be a linear transformation and w a nonzero vector in W. Then the set of all v in V such that $T(v) = w$ forms a subspace.

(b) For every pair of vector spaces V and W, it is always possible to define a linear transformation $T : V \to W$.

45. Let $T : \mathbb{R}^5 \to C[0, 1]$ be a linear transformation, and suppose that $\dim(\ker(T)) = 2$. What is $\dim(\mathrm{range}(T))$?

46. Let $T : \mathbb{R}^{4\times 3} \to P$ be a one-to-one linear transformation. What is $\dim(\mathrm{range}(T))$?

47. Prove that if $T : V \to W$ is a linear transformation with $T(v_1) = T(v_2)$, then $v_1 - v_2$ is in $\ker(T)$.

48. Prove that if $T : V \to W$ is an onto and one-to-one linear transformation, and both V and W are of finite dimension, then $\dim(V) = \dim(W)$.

49. Suppose that $T_1 : V \to W$ and $T_2 : V \to W$ are both linear transformations. Prove that $T_1 + T_2$ is also a linear transformation from V to W.

50. Prove the extended version of Theorem 9.2: If $T : V \to W$ is a linear transformation, then

$$T(c_1v_1+c_2v_2+\cdots+c_mv_m) = c_1T(v_1)+c_2T(v_2)+\cdots+c_mT(v_m),$$

where $v_1, \ldots , v_m$ are in V and $c_1, \ldots , c_m$ are real scalars.

51. Prove that if $T : V \to W$ is a linear transformation, then $T(0_V)=0_W$.

52. Prove that if $T : V \to W$ is a linear transformation, then for any v in V we have $T(-v)=-T(v)$.

53. Prove part of Theorem 9.3: Let $T : V \to W$. Then $\ker(T)$ is a subspace of V.

54. Prove Theorem 9.2: $T : V \to W$ is a linear transformation if and only if

$$T(c_1v_1+c_2v_2)=c_1T(v_1)+c_2T(v_2)$$

for all vectors v_1 and v_2 in V and real scalars c_1 and c_2.

55. Prove Theorem 9.5: Let $T : V \to W$ be a linear transformation. Then T is one-to-one if and only if $\ker(T)=\{0_V\}$. (HINT: Theorem 3.5 is similar.)

56. Suppose that $T_1 : V \to W$ and $T_2 : W \to Y$ are both linear transformations. Prove that $T_2(T_1(v))$ is also a linear transformation from V to Y.

57. Prove a partial converse of Theorem 9.6: Let $T : V \to W$ be a one-to-one linear transformation, with $\mathcal{V}=\{v_1, \ldots , v_m\}$ a subset of V, and $\mathcal{W}=\{w_1, \ldots , w_m\}$ a subset of W. Suppose that $T(v_i)=w_i$ for $i=1, \ldots , m$. If $\mathcal{V}$ is linearly independent, then so is $\mathcal{W}$.

58. Prove Theorem 9.8, but with the condition that V and W are finite dimensional removed.

Exercises 59–64: Basic knowledge of calculus is required.

59. Let $C^1(a, b)$ denote the set of functions that are continuously differentiable on the interval (a, b). Prove that

$T : C^1(a, b) \to C(a, b)$ given by

$$T(f) = f'(x)$$

is a linear transformation.

60. Let $T : C[a, b] \to \mathbb{R}$ be given by

$$T(f) = \int_a^b f(x)\, dx.$$

Prove that *T* is a linear transformation.

61. Determine if $T : P_4 \to P_2$ with $T(p) = p''(x)$ is a linear transformation.

62. Determine if $T : P_3 \to \mathbb{R}$ with

$$T(p) = \int_a^b x p(x)\, dx$$

is a linear transformation.

63. Determine if $T : P_4 \to P_5$ with $T(p) = (x^2 p(x))'$ is a linear transformation.

64. Determine if $T : P_3 \to \mathbb{R}$ with

$$T(p) = \int_a^b e^{x - x^2} p(x)\, dx$$

is a linear transformation.

9.2 Isomorphisms

At the beginning of Section 7.1, we compared the features of R3 and P2 and concluded that these two vector spaces have a lot in common. In fact, many superficially different vector spaces are essentially the same in important ways. In this section we define precisely what it means for two vector spaces to be essentially the same, and determine which vector spaces are essentially the same.

Definition of Isomorphism

Our mechanism for establishing when two vector spaces are essentially the same is through a special type of linear transformation called an *isomorphism*.

DEFINITION 9.9 ▶

Isomorphism, Isomorphic

A linear transformation $T : V \rightarrow W$ is an **isomorphism** if *T* is both one-to-one and onto. If such an isomorphism exists, then we say that *V* and *W* are **isomorphic** vector spaces.

▶ The word *isomorphism* has Greek origins and means “same structure.”

Regarding isomorphisms and isomorphic vector spaces:

- Some pairs of vector spaces are isomorphic, while others are not. For instance, R3 and P2 are isomorphic (see Example 1 below), while R2×3 and C [0, 1] are not. (We will explain why later.)

- If $T : V \to W$ is an isomorphism, then there will also exist an isomorphism $S : W \to V$. (This is developed in more detail later in this section.) Thus the notion of isomorphic is symmetric. If *V* and *W* are isomorphic, then *W* and *V* are isomorphic.
- *V* and *W* may be isomorphic even if there exists a linear transformation $T : V \to W$ that is not an isomorphism. There may be a different linear transformation that is an isomorphism.
- The requirement that $T : V \to W$ be both onto and one-to-one ensures that there is an exact correspondence between the elements of *V* and *W*, and it is called a **one-to-one correspondence**: Every v in *V* is paired up with a specific w in *W* by $T(v)=w$. Matching up elements of *V* and *W* in this manner is one part of establishing that *V* and *W* are essentially the same.
- Because *T* is a linear transformation, addition and scalar multiplication work the same between corresponding elements of *V* and *W*. For instance, if $T(v_i)=w_i$ for $i=1, 2, 3$ and $v_1+v_2=v_3$, then

$$w_1+w_2=T(v_1)+T(v_2)=T(v_1+v_2)=T(v_3)=w_3$$

  That is, $v_1+v_2=v_3$ implies that $w_1+w_2=w_3$, so that addition works the same in *V* and *W*. This principle also holds for scalar multiplication.

Since an isomorphism between two vector spaces *V* and *W* matches up vectors and preserves the respective arithmetic operations, from the standpoint of vector spaces *V* and *W* are essentially the same.

Let’s consider some examples.

Example 1

Show that $\mathbb{R}^3$ and P_2 are isomorphic.

Solution To show that two vector spaces *V* and *W* are isomorphic, we need to find an isomorphism $T : V \to W$. That is, *T* must be a linear transformation that is onto and one-to-one.

▶ To establish that *V* and *W* are isomorphic, we need only find one isomorphism $T : V \to W$.

Since our goal is to show that R^3 and P_2 are isomorphic, we choose a linear transformation $T : R^3 \to P_2$ that is as simple as possible while still meeting the requirements of an isomorphism. Frequently something obvious makes a good choice. Here we try

$$T([a_0 a_1 a_2]) = a_2x^2+a_1x+a_0$$

We have

$$T([a_0 a_1 a_2]+[b_0 b_1 b_2])=T([a_0+b_0\; a_1+b_1\; a_2+b_2]) =(a_2+b_2)x^2+(a_1+b_1)x+(a_0+b_0) =(a_2x^2+a_1x+a_0)+(b_2x^2+b_1x+b_0)=T([a_0 a_1 a_2]) +T([b_0 b_1 b_2])$$

and

$$T(c\,[a_0 a_1 a_2])=T([ca_0\; ca_1\; ca_2])=ca_2x^2+ca_1x+ca_0=c(a_2x^2+a_1x+a_0)=cT([a_0 a_1 a_2])$$

so that *T* is a linear transformation. Next suppose

$$T([a_0 a_1 a_2]) =T([b_0 b_1 b_2])$$

Then $a_2x^2+a_1x+a_0=b_2x^2+b_1x+b_0$, so we may conclude from the properties of polynomials that $a_2=b_2$, $a_1=b_1$, and $a_0=b_0$. Hence

$$[a_0 a_1 a_2]=[b_0 b_1 b_2]$$

so *T* is one-to-one. Finally, any polynomial $a_2x^2+a_1x+a_0$ satisfies

$$T[a_0 a_1 a_2] =a_2x^2+a_1x+a_0$$

so *T* is onto. Therefore *T* is an isomorphism, which proves that R^3 and P_2 are isomorphic.

It is possible for a subspace of one vector space to be isomorphic to a vector space or even another subspace.

Example 2

Show that R2 and the subspace

$$S=\text{span }\{[120], [311]\} \tag{1}$$

of R3 are isomorphic.

Solution The subspace *S* is a plane in R3 (Figure 1), so it resembles the coordinate plane R2. It seems plausible that the two would be isomorphic, but this is not a proof.

To prove that *S* and R2 are isomorphic, we need to find an isomorphism. Let T : R2→S be given by T(x)=Ax, where

$$A=[132101]$$

Then *T* is a linear transformation. Since S=col(A), the column space of *A*, it follows that *T* is onto. Furthermore, the columns of *A* are linearly independent, so *T* is one-to-one by Theorem 3.6 in Section 3.1. Therefore *T* is an isomorphism, and *S* and R2 are isomorphic.

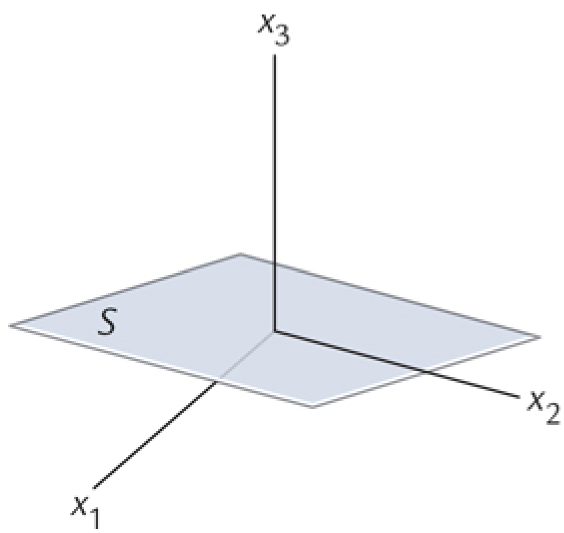


Figure 1 The subspace S in R3.

In both of the examples we have considered, the isomorphic vector spaces had the same dimension. This is not a coincidence.

THEOREM 9.10 ▶

Suppose that finite dimensional vector spaces V and W are isomorphic. Then $\dim(V)=\dim(W)$.

Proof Since V and W are isomorphic, there exists an isomorphism $T : V \to W$. Now recall from Theorem 9.8 that

- If T is onto, then $\dim(V) \geq \dim(W)$.
- If T is one-to-one, then $\dim(V) \leq \dim(W)$.

Since T is onto and one-to-one, both inequalities must hold. The only way this can happen is if $\dim(V)=\dim(W)$. ■ ■

Theorem 9.10 provides a quick and easy way to show that two vector spaces are *not* isomorphic, because

If $\dim(V) \neq \dim(W)$, then V and W are not isomorphic.

Example 3

Show that $\mathbb{R}^{3\times 2}$ and P_4 are not isomorphic.

Solution Since $\dim(\mathbb{R}^{3\times 2})=6$ and $\dim(P_4)=5$, these two vector spaces cannot be isomorphic.

Theorem 9.10 also holds for infinite-dimensional vector spaces. If *V* and *W* are isomorphic and one is infinite dimensional, then so is the other. Such vector spaces can lead to interesting and counter-intuitive results.

Example 4

Let P_e denote the set of polynomials with real coefficients and only even-powered terms, and let P be the set of all polynomials with real coefficients. Then P_e is a subspace of P. Is P_e isomorphic to P?

Solution Viewed one way, it seems unlikely that P_e is isomorphic to P. After all, P_e is a proper subspace of P, so how could a one-to-one correspondence—required of an isomorphism—possibly exist between these two sets? But intuition can be misleading when it comes to infinite-dimensional vector spaces.

Suppose that we define $T : P_e \to P$ by

$$T(a_n x^{2n}+a_{n-1}x^{2(n-1)}+\cdots+a_1x^2+a_0)=a_n x^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$$

It is not hard to verify that *T* is a linear transformation, and that *T* is also onto and one-to-one. Therefore, perhaps surprisingly, P_e and P are isomorphic.

Thus far we know that if V and W are isomorphic, then $\dim(V)=\dim(W)$. Now we consider the reverse direction. Do the dimensions of V and W tell us anything about whether or not V and W are isomorphic?

THEOREM 9.11 ▶

Let V and W be finite-dimensional vector spaces with $\dim(V)=\dim(W)=m$, where $m>0$. Suppose that

$$\mathcal{V}=\{v_1,\ldots,v_m\} \text{ and } \mathcal{W}=\{w_1,\ldots,w_m\}$$

are bases for V and W, respectively. Now define $T: V\to W$ as follows: For v in V, let $c_1,\ldots,c_m$ be such that $v=c_1v_1+\cdots+c_mv_m$. Then set

$$T(v)=T(c_1v_1+\cdots+c_mv_m)=c_1w_1+\cdots+c_mw_m$$

Then T is an isomorphism, and V and W are isomorphic vector spaces.

The proof of Theorem 9.11 is not hard, but it is a bit long and so is given at the end of the section. Here we report the significant implication of this theorem.

THEOREM 9.12 ▶

Finite dimensional vector spaces V and W are isomorphic if and only if $\dim(V)=\dim(W)$.

▶ Remember that in this chapter all vector spaces are real vector spaces.

Proof By Theorem 9.11, if two finite-dimensional vector spaces V and W have the same dimension, then they are isomorphic. The converse follows from Theorem 9.10, which tells us that two finite-dimensional isomorphic vector spaces must have the same dimension. ■ ■

If we think of dimension as giving a measure of the size of a vector space, then Theorem 9.12 tells us that size is *all* that matters when determining if two vector spaces are isomorphic. Now we can easily answer questions such as:

- Are P_7 and $R^{4\times2}$ isomorphic? (Yes. Both have dimension 8.)
- Are $R^{2\times3}$ and $C[0, 1]$ isomorphic? (No. $\dim(R^{2\times3})=6$ and $\dim(C[0, 1])=\infty$.)
- Are R^5 and $S=\text{span}\{\cos(x), \sin(x), \cos(2x), \sin(2x)\}$ in $C[0, 1]$ isomorphic? (No. $\dim(R^5)=5$ and $\dim(S)\le4$.)
- Are $R^{7\times5}$ and R^{35} isomorphic? (Yes. Both have dimension 35.)
- Are R^∞ and P isomorphic? (Maybe. We cannot tell because Theorem 9.12 does not apply to a pair of vector spaces having infinite dimension. Not every question is easily answered.)

Another consequence of Theorem 9.12 is contained in the next theorem.

THEOREM 9.13 ▶

If V is a vector space and $\dim(V)=n$, then V is isomorphic to R^n.

In a way, this brings our development of vector spaces full circle. Since all *n*-dimensional vector spaces are isomorphic to *n*-dimensional Euclidean space, it is not so surprising that our Euclidean space results carried over so readily to vector spaces.

Inverses

We now revisit the notion of inverse functions, first treated in Section 3.3. Here is a definition, updated from earlier to our current more general setting.

DEFINITION 9.14 ▶

Inverse, Invertible

A linear transformation $T: V \to W$ is **invertible** if *T* is one-to-one and onto. When *T* is invertible, the **inverse** function $T^{-1}: W \to V$ is defined by

$$T^{-1}(w) = v \text{ if and only if } T(v) = w$$

Much of the development of Section 3.3 carries over directly to vector spaces. The main points are:

- A linear transformation *T* is invertible exactly when *T* is an isomorphism.
- If a linear transformation *T* is invertible, then the inverse T^{-1} is unique.
- If $T: V \to W$ is an isomorphism, then $T^{-1}: W \to V$ is also an isomorphism.

The proofs of these properties are left as exercises. Let's consider an example.

Example 5

Let $T: P_1 \to \mathbb{R}^2$ be the linear transformation given by

$$T(p(x)) = \begin{bmatrix} p(0) \\ p(1) \end{bmatrix}$$

Show that *T* is an isomorphism by showing that *T* is one-to-one and onto, and find T^{-1}.

Solution Elements of P_1 are all polynomials of the form $p(x)=ax+b$, so that $p(0)=b$ and $p(1)=a+b$. Thus

$$T(ax+b)=\begin{bmatrix} b \\ a+b \end{bmatrix}$$

For any real numbers *c* and *d*, the vector equation

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

has unique solution $a=d-c$ and $b=c$. Therefore *T* is one-to-one and onto. Moreover, the unique solution also shows us how to define $T^{-1} : \mathbb{R}^2\to P_1$. Since $a=d-c$ and $b=c$, it follows that

$$T((d-c)x+c)=\begin{bmatrix} c \\ d \end{bmatrix} \Rightarrow T^{-1}\left(\begin{bmatrix} c \\ d \end{bmatrix}\right)=(d-c)x+c$$

We can check that this is correct by computing

$$T^{-1}(T(ax+b))=T^{-1}\left(\begin{bmatrix} b \\ a+b \end{bmatrix}\right)=((a+b)-b)x+b=ax+b$$

and

$$T\left(T^{-1}\left(\begin{bmatrix} c \\ d \end{bmatrix}\right)\right)=T((d-c)x+c)=\begin{bmatrix} c \\ (d-c)+c \end{bmatrix}=\begin{bmatrix} c \\ d \end{bmatrix}$$

Proof of Theorem 9.11

Proof First, since $\mathcal{V}=\{v_1,\dots,v_m\}$ is a basis for *V*, for any v there is exactly one set of scalars $c_1,\dots,c_m$ such that $v=c_1v_1+\cdots+c_mv_m$. Therefore *T* is actually a well-defined function.

Second, suppose that u is also in *V*, with $u=d_1v_1+\cdots+d_mv_m$. For $v=c_1v_1+\cdots+c_mv_m$ as above and scalars *a* and *b*, we have

$$T(av+bu)=T((ac_1+bd_1)v_1+\cdots+(ac_m+bd_m)v_m) =(ac_1+bd_1)w_1+\cdots+(ac_m+bd_m)w_m =a(c_1w_1+\cdots+c_mw_m)+b(d_1w_1+\cdots+d_mw_m) =aT(v)+bT(u)$$

Hence T is a linear transformation.

Third, since $\mathcal{W}=\{w_1,\dots,w_m\}$ is a basis, every element w in W can be expressed in the form $w=c_1w_1+\cdots+c_mw_m$ for unique $c_1,\dots,c_m$. For such a w, we see that $v=c_1v_1+\cdots+c_mv_m$ satisfies $T(v)=w$, and therefore T is onto.

Finally, as $\mathcal{W}$ is a basis, the only linear combination $c_1w_1+\cdots+c_mw_m=0_W$ is when $c_1=\cdots=c_m=0$. Therefore the only v in V such that $T(v)=0_W$ is $v=0_V$, which implies that T is also one-to-one. Thus T is an isomorphism, and therefore V and W are isomorphic vector spaces. ■ ■

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Based on dimensions, determine if the vector spaces are isomorphic. If dimensions cannot be used, explain why.

(a) $V=\mathbb{R}^{4\times5}$ and $W=\mathbb{P}^{19}$

(b) $V=\mathbb{R}^\infty$ and $W=C[0,1]$

2. Prove that the given function is an isomorphism.

(a) $T:\mathbb{R}^2\to\mathbb{P}^1$ with $T\left(\begin{bmatrix}a\\b\end{bmatrix}\right)=bx+(a-b)$

(b) $T:\mathbb{R}^{2\times3}\to\mathbb{R}^{3\times2}$ with $T(A)=A^T$

3. Find T^{-1} for the given isomorphism.

(a) $T:\mathbb{P}^2\to\mathbb{R}^3$ with $T(ax^2+bx+c)=\begin{bmatrix}b\\c\\a\end{bmatrix}$

(b) $T:\mathbb{R}^{2\times2}\to\mathbb{P}^3$ with $T\left(\begin{bmatrix}a&b\\c&d\end{bmatrix}\right)=bx^3+cx^2+dx-a$

EXERCISES

Exercises 1–6: Use dimensions when possible to determine if the given vector spaces are isomorphic. If not possible, explain why.

1. V=R8 and W=P9
2. V=R5×3 and W=R15
3. V=R3×6 and W=P17
4. V=R∞ and W=P20
5. V=R13 and W=C [0, 1]
6. V=R∞ and W=C [0, 1]

Exercises 7–10: Prove that *T* is an isomorphism.

7. T : R3→P2 with T([abc]) =cx2+bx+a
8. T : P1→R2 with T(p)=[p(−2)p(1)]
9. T : P3→R2×2 with T(ax3+bx2+cx+d)=[(a+b+c+d)(a+b+c)(a+b)a]
10. T : R2×2→R2×2 with T(A)=AT

Exercises 11–14: Determine if *T* is an isomorphism.

11. T : P1→R2 with T(ax+b)=[a+bb−a]
12. T : P→P with T(p)=xp(x)
13. T : C(R)→R∞ with T(f)=(f(1), f(2), f(3),…)
14. T : R2×2→P3 with

    T(A)=tr(A)x3+a11x2+a21x−a12

Exercises 15–18: Find T−1 for the given isomorphism *T*.

15. T : P1→R2 with T(ax+b)=[2ba−b]
16. T : R2×2→R2×2 with T(A)=AT
17. T : P2→P2 with T(ax2+bx+c)=cx2−bx+a
18. T : P3→R2×2 with

    T(ax3+bx2+cx+d)=[−ac−db]

19. Let *S* be the subspace of R3 given by

    S=span {[100], [010]}

    Show that T : R2→S given by

$$T\left(\begin{bmatrix}a_1\\a_2\end{bmatrix}\right)=\begin{bmatrix}a_1\\a_2\\0\end{bmatrix}$$

is an isomorphism.

20. Let *S* be the subspace of $\mathbb{R}^4$ given by

$$S=\text{span}\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\2\end{bmatrix}\right\}$$

Show that $T : \mathbb{R}^2 \to S$ given by

$$T\left(\begin{bmatrix}a_1\\a_2\end{bmatrix}\right)=\begin{bmatrix}a_1\\0\\0\\a_2\end{bmatrix}$$

is an isomorphism.

21. Let P^e be the subspace of P defined in Example 4. Show that $T : P^e \to P$ given by

$$T(a_n x^{2n}+a_{n-1}x^{2(n-1)}+\cdots+a_1x^2+a_0) =a_nx^n+a_{n-1} x^{n-1}+\cdots+a_1x+a_0$$

is an isomorphism.

22. Let P^o be the subspace of P consisting of polynomials with only odd-powered terms, and the zero polynomial. Define $T : P^o \to P$ by

$$T(a_nx^{2n+1}+a_{n-1}x^{2n-1}+\cdots+a_1x^3+a_0x) =a_nx^n+a_{n-1} x^{n-1}+\cdots+a_1x+a_0$$

and $T(0)=0$ for the zero polynomial. Is *T* a linear transformation?

FIND AN EXAMPLE Exercises 23–30: Find an example that meets the given specifications. Prove your claim.

23. An isomorphism $T : \mathbb{R}^5 \to P^4$.

24. An isomorphism $T : \mathbb{R}^{2\times 3} \to \mathbb{R}^6$.

25. An isomorphism $T : \mathbb{R}^{2\times 2} \to P^3$.

26. An isomorphism $T : \mathbb{R}^4 \to S$, where *S* is the subspace of P^6 of polynomials that have only terms with even-powered exponents.

27. A subspace of $\mathbb{R}^{2\times 3}$ that is isomorphic to P^3.

28. A subspace of P that is isomorphic to $\mathbb{R}^4$.

29. A subspace of $\mathbb{R}^\infty$ that is isomorphic to P.

30. A proper subspace of $\mathbb{R}^\infty$ that is isomorphic to $\mathbb{R}^\infty$.

TRUE OR FALSE Exercises 31–38: Determine if the statement is true or false, and justify your answer.

31.

(a) Every linear transformation is also an isomorphism.

(b) Every finite dimensional vector space *V* is isomorphic to $\mathbb{R}^n$ for some *n*.

32.

(a) If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a one-to-one linear transformation, then *T* is an isomorphism.

(b) There exists a subspace of P^{10} that is isomorphic to $\mathbb{R}^{3\times 3}$.

33.

(a) If *V* and *W* are isomorphic, then there is a unique linear transformation $T : V \to W$ that is an isomorphism.

(b) Every three-dimensional subspace of a vector space *V* is isomorphic to P^2.

34.

(a) If $T_1 : V \to W$ and $T_2 : V \to W$ are isomorphisms, then so is T_1+T_2.

(b) If $T : V \to W$ is an isomorphism and $\{v_1, v_2, v_3\}$ is a basis for *V*, then $\{T(v_1), T(v_2), T(v_3)\}$ is a basis for *W*.

35.

(a) If $T : V \to W$ is an isomorphism, then so is $S(v)=-T(v)$.

(b) If $T : V \to W$ and $S : W \to Y$ are isomorphisms, then $R : V \to Y$ given by $R(v)=S(T(v))$ is also an isomorphism.

36.

(a) If $T : P \to P$ is a linear transformation and $\text{range}(T)=P$, then *T* is an isomorphism.

(b) If $T : P \to P$ is defined by $T(p(x))=p(2x+1)$, then *T* is an isomorphism.

37.

(a) **Calculus required** Let $C^\infty(a, b)$ denote the set of functions that have an infinite number of continuous derivatives on the interval (a, b). Then $T : C^\infty(a, b) \to C^\infty(a, b)$ given by $T(f)=f'(x)$ is an isomorphism.

(b) **Calculus required** Let $T : C[a, b] \to \mathbb{R}$ be given by

$$T(f)=\int_a^b f(x)\,dx$$

Then *T* is an isomorphism.

38.

(a) **Calculus required** The linear transformation $T : P_4 \to P_4$ with $T(p) = xp'(x)$ is an isomorphism.

(b) **Calculus required** If $T : P_1 \to P_1$ with

$$T(p) = x - \int_a^b p(x)\,dx$$

then T is an isomorphism.

39. Prove that a linear tranformation T is an isomorphism if and only if T has an inverse.

40. Prove that if the linear transformation T has an inverse, then it is unique.

41. Prove that if $T : V \to W$ is an isomorphism, then $T^{-1} : W \to V$ is also an isomorphism.

42. Prove that if a linear transformation $T : V \to W$ is either onto *or* one-to-one, and $\dim(V) = \dim(W)$ are both finite, then T is an isomorphism.

43. Suppose that $T : V \to W$ is a one-to-one linear transformation. Prove that V and $\mathrm{range}(T)$ are isomorphic.

44. Suppose that $T_1 : V \to W$ and $T_2 : W \to Y$ are isomorphisms. Prove that V and Y are isomorphic vector spaces.

9.3 The Matrix of a Linear Transformation

In Section 3.1 we defined the linear transformation $T: \mathbb{R}^m \to \mathbb{R}^n$ and showed that it has the form $T(\mathbf{x}) = A\mathbf{x}$ for an $n \times m$ matrix A. If V and W are finite-dimensional vector spaces, then we can establish a similar connection between linear transformations $T: V \to W$ and matrices. Before getting to that, we first revisit coordinate vectors to see how they are applied to vector spaces.

Coordinate Vectors

To establish a connection between linear transformations and vector spaces, we need a way to express vectors in the form of elements of $\mathbb{R}^m$. This can be done using coordinate vectors as defined below.

DEFINITION 9.15 ▶

Coordinate Vector

Let V be a vector space with basis $\mathcal{G} = \{\mathbf{g}_1, \ldots, \mathbf{g}_m\}$. For each $\mathbf{v} = c_1\mathbf{g}_1 + \cdots + c_m\mathbf{g}_m$ in V, we define the **coordinate vector of v with respect to** $\mathcal{G}$ by

$$[\mathbf{v}]_{\mathcal{G}} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$$

▶ In this section and the next, we will sometimes need to refer to more than one basis for a vector space. To avoid the appearance of favoring one basis over another (and additional cluttering subscripts), we depart from our customary use of $\mathcal{V}$ and $\mathcal{W}$ to represent bases for $\mathcal{V}$ and $\mathcal{W}$, respectively.

Regarding Definition 9.15:

- Although $[\mathbf{v}]_{\mathcal{G}}$ is expressed in terms of a vector in Euclidean space, it represents a vector in V.
- The choice of basis matters. Different bases will yield different coordinate vectors for the same vector $\mathbf{v}$.
- The order of the basis vectors matters. Our convention is to read a list of basis vectors from left to right, taking them in that order.
- Since $\mathcal{G}$ is a basis for V, the scalars $c_1, \ldots, c_m$ are unique, so that for each $\mathbf{v}$ there is exactly one coordinate vector with respect to a given basis $\mathcal{G}$.

Example 1

Two bases for $\mathbf{P}^2$ are

$$\mathcal{G} = \{x^2, x, 1\} \text{ and } \mathcal{H} = \{x^2+2x-4,\ x+1,\ x^2-x\}$$

Find the coordinate vector of $\mathbf{v} = x^2-6x+2$ with respect to $\mathcal{G}$ and with respect to $\mathcal{H}$.

Solution Starting with $\mathcal{G}$, it is not difficult to see that

$$\mathbf{v} = (1)x^2 + (-6)x + (2) \Rightarrow [\mathbf{v}]_{\mathcal{G}} = \begin{bmatrix} 1 \\ -6 \\ 2 \end{bmatrix}$$

For the basis $\mathcal{H}$, we need to find scalars a, b, and c such that

$$\mathbf{v} = x^2-6x+2 = a(x^2+2x-4) + b(x+1) + c(x^2-x)$$

Using our standard methods, it can be shown that $a=-1$, $b=-2$, and $c=2$. Hence

$$\mathbf{v} = (-1)(x^2+2x-4) + (-2)(x+1) + (2)(x^2-x) \Rightarrow [\mathbf{v}]_{\mathcal{H}} = \begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix}$$

Example 2

Find the coordinate vector of v=[3257] with respect to the basis for R2×2 given by

$$\mathcal{G}=\{[1000], [0100], [0010], [0001]\}$$

Solution The set $\mathcal{G}$ is the standard basis for R2×2 and can be especially easy to use. We see that

$$v=[3257]=3\ [1000]+2\ [0100]+5\ [0010]+7\ [0001]\Rightarrow[\ v]\mathcal{G}=[3257]$$

Example 3

Suppose that $\mathcal{G}$={sin(x), cos(x), e−x} is a basis for a subspace of C [0, 1]. Find v if

$$[\ v]\mathcal{G}=[5-32]$$

Solution Here we are given the coordinate vector v$\mathcal{G}$ and need to extract v. All we have to do is multiply the basis vectors by the scalars given in v$\mathcal{G}$. We have

$$v=5\ \sin(x)-3\ \cos(x)+2e-x$$

Transformation Matrices

Now that we have an understanding of coordinate vectors, let's consider how matrices can be used to represent linear transformations. Suppose that T : V→W is a linear transformation, and that $\mathcal{G}$={g1,… , gm} and $\mathcal{Q}$={q1,… , qn} are bases of *V* and *W*, respectively. Our goal is to find a matrix *A* such that

$$[T(v)]_{\mathcal{Q}} = A[\,v]_{\mathcal{G}}$$

where

$$[\,v]_{\mathcal{G}} = \text{coordinate vector of v with respect to } \mathcal{G} A[\,v]_{\mathcal{G}} = \text{coordinate vector of T(v) with respect to } \mathcal{Q}$$

If $A=[a_1 \ldots a_m]$ and $v=c_1 g_1+\cdots+c_m g_m$, then

$$A[\,v]_{\mathcal{G}}=A[c_1 \vdots c_m]=c_1 a_1+\cdots+c_m a_m$$

On the other hand, we also have

$$T(v)=c_1 T(g_1)+\cdots+c_m T(g_m)$$

Thus, in order for $A[\,v]_{\mathcal{G}}=[T(v)]_{\mathcal{Q}}$, we should set $a_i=[T(g_i)]_{\mathcal{Q}}$ for each $i=1,\ldots,m$. This brings us to our next definition.

DEFINITION 9.16 ▶

Matrix of a Linear Transformation

Let $T : V \to W$ be a linear transformation, $\mathcal{G}=\{g_1,\ldots,g_m\}$ a basis of *V*, and let $\mathcal{Q}=\{q_1,\ldots,q_n\}$ a basis of *W*. If $A=[a_1 \cdots a_m]$ with

$$a_i=[T(g_i)]_{\mathcal{Q}}$$

for each $i=1,\ldots,m$, then *A* is the **matrix of *T* with respect to $\mathcal{G}$ and $\mathcal{Q}$**.

Transformation Matrix

The matrix *A* in Definition 9.16 is also called a **transformation matrix**.

Example 4

Let T : P2→P1 be given by

$$T(ax2+bx+c)=(2a+c-3b)x+(c+4b+3a)$$

and let $\mathcal{G}$={x2, x, 1} be a basis for P2 and $\mathcal{Q}$={x, 1} a basis for P1. Find the matrix of T with respect to $\mathcal{G}$ and $\mathcal{Q}$, and then use it to compute T(2x2−4x+1).

Solution Finding A requires us to compute [T(gi)]$\mathcal{Q}$ for each basis vector of $\mathcal{G}$. We have

$$T(x2)=2x+3\Rightarrow[T(x2)]\mathcal{Q}=[23]T(x\ )=-3x+4\Rightarrow[T(x\ )]\mathcal{Q}=[-34]T(1)=x+1\Rightarrow[T(1)]\mathcal{Q}=[11]$$

Therefore, by Definition 9.16,

$$A=[[T(x2)]\mathcal{Q}\ [T(x)]\mathcal{Q}\ [T(1)]\mathcal{Q}]=[2-31341]$$

Now let's use A to compute T(2x2−4x+1). We have [2x2−4x+1]$\mathcal{G}$=[2−41], so that

$$[T(2x2-4x+1)]\mathcal{Q}=A[2-41]=[17-9]=[17x-9]\mathcal{Q}$$

To check our work, let's compute directly,

$$T(2x2-4x+1)=(2(2)+(1)-3(-4))x+((1)+4(-4)+3(2))=17x-9$$

In Example 4 we used the standard bases for P2 and P1. While this simplified computations, there is no reason why other bases cannot be used. Let's try Example 4 again, this time with different choices for $\mathcal{G}$ and $\mathcal{Q}$.

Example 5

Repeat Example 4, this time with bases

$$\mathcal{G}=\{x^2-2,\ 2x-1,\ x^2+4x-1\} \text{ and } \mathcal{Q}=\{2x+1,\ 5x+3\}$$

Solution As before, the columns of our transformation matrix *A* are given by $[T(g_i)]_{\mathcal{Q}}$. Starting with g_1, we have $T(x^2-2)=1$. Finding the coordinate vector with respect to $\mathcal{Q}$ requires more work than before. Here we need to find scalars *a* and *b* such that

$$1=a(2x+1)+b(5x+3)$$

Applying our usual methods yields the solution $a=-5$ and $b=2$. Hence

$$[T(x^2-2)]_{\mathcal{Q}}=\begin{bmatrix}-5\\2\end{bmatrix}$$

Applying the same procedure to the other basis vectors yields

$$T(2x-1)=-7x+7=-56(2x+1)+21(5x+3)\ \Rightarrow\ [T(2x-1)]_{\mathcal{Q}}=\begin{bmatrix}-56\\21\end{bmatrix}$$

and

$$T(x^2+4x-1)=-11x+18=-123(2x+1)+47(5x+3)$$

$$\Rightarrow [T(x^2+4x-1)]_{\mathcal{Q}}=\begin{bmatrix}-123\\47\end{bmatrix}$$

Therefore the transformation matrix is

$$B=\begin{bmatrix}-5 & -56 & -123\\ 2 & 21 & 47\end{bmatrix}$$

To test this out by computing $T(2x^2-4x+1)$, we need to determine the coordinate vector of $2x^2-4x+1$ with respect to $\mathcal{G}$. To do so, we need to find scalars *a*, *b*, and *c* such that

$$2x^2-4x+1=a(x^2-2)+b(2x-1)+c(x^2+4x-1)$$

Our usual methods produce $a=1$, $b=-4$, and $c=1$, so that $[2x^2-4x+1]_{\mathcal{G}}=\begin{bmatrix}1\\-4\\1\end{bmatrix}$. Hence

$$[T(2x^2-4x+1)]_{\mathcal{Q}}=B\begin{bmatrix}1\\-4\\1\end{bmatrix}=\begin{bmatrix}96\\-35\end{bmatrix}=[96(2x+1)-35(5x+3)]_{\mathcal{Q}}=[17x-9]_{\mathcal{Q}}$$

which agrees with our earlier work.

▶ Here we use B to denote the transformation matrix to avoid confusing this matrix with the matrix A found in Example 4. In Section 9.4 we will see that there is a relationship between two transformation matrices A and B representing the same linear transformation with respect to different bases.

Inverses

Let $T : V \rightarrow W$ be an isomorphism of finite-dimensional vector spaces. By Theorem 9.10, we know that $\dim(V)=\dim(W)=m$ for some m. Now suppose that A is the transformation matrix of T with respect to bases $\mathcal{G}$ and $\mathcal{Q}$, and that $T(v)=w$. Then A is an $m\times m$ matrix with

$$A[\,v]_{\mathcal{G}}=[\,w]_{\mathcal{Q}} \tag{1}$$

As T is an isomorphism, T must be onto and one-to-one, so that by the Unifying Theorem (Version 3 or later), A is an invertible matrix. Multiplying by A^{-1} on both sides of (1) yields

$$A^{-1}[\,w]_{\mathcal{Q}}=[\,v]_{\mathcal{G}}$$

Thus A^{-1} reverses T, so we can conclude that A^{-1} is the matrix of T^{-1} with respect to $\mathcal{Q}$ and $\mathcal{G}$. (Note that the domain and codomain reverse when switching from T to T^{-1}.)

Example 6

In Example 5 of Section 9.2, it is shown that the linear transformation $T : P^1 \rightarrow R^2$ given by

$$T(p)=\begin{bmatrix} p(0) \\ p(1) \end{bmatrix}$$

is an isomorphism. Let $\mathcal{G}=\{x, 1\}$ and $\mathcal{Q}=\{e_1, e_2\}$ be bases for P^1 and R^2, respectively. Find the matrix of T^{-1} with respect to $\mathcal{G}$ and $\mathcal{Q}$.

Solution We have

$$T(x)=[01] \Rightarrow [T(x)]_{\mathcal{Q}}=[01]$$

$$T(1)=[11] \Rightarrow [T(1)]_{\mathcal{Q}}=[11]$$

Therefore

$$A=[0111]$$

is the matrix of T with respect to $\mathcal{G}$ and $\mathcal{Q}$. By the preceding comments, the matrix of T−1 with respect to $\mathcal{Q}$ and $\mathcal{G}$ is given by

$$A−1=[−1110]$$

For a typical vector w=[cd] in R2, we have w=ce1+de2, so that

$$[T−1(w)]_{\mathcal{G}}=A−1 [cd]=[−1110] [cd]=[−c+dc]=[(d−c)x+c]_{\mathcal{G}}$$

which matches T−1 found in Example 5 of Section 9.2.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find v given the coordinate vector [v]$_{\mathcal{G}}$ with respect to the basis $\mathcal{G}$.
   (a) [v]$_{\mathcal{G}}$=[32]; $\mathcal{G}$={[−51], [3−4]}
   (b) [v]$_{\mathcal{G}}$=[21−2]; $\mathcal{G}$={2x2−3x+1, x2+5, −x+3}
2. Find the coordinate vector of v with respect to $\mathcal{G}$.
   (a) v=19x+14; $\mathcal{G}$={−3x, x+2}
   (b) v=[3025]; $\mathcal{G}$={[1011], [0101], [1−101], [0111]}
3. A is the matrix of linear transformation T : V→W with respect to bases $\mathcal{G}$ and $\mathcal{Q}$, respectively. Find T(v) for the given [v]$_{\mathcal{G}}$.

(a) $A=[2-134]$; $[v]_{\mathcal{G}}=[12]$; $\mathcal{Q}=\{x-2, 3x+7\}$
(b) $A=[10-223-3021]$; $[v]_{\mathcal{G}}=[0-22]$; $\mathcal{Q}=\{\cos(x), ex, \sin(x)\}$

4. Find the matrix *A* of the linear transformation $T : V\to W$ with respect to bases $\mathcal{G}$ and $\mathcal{Q}$, respectively.

(a) $T([ab])=(a+b)x-b$; $\mathcal{G}=\{e1, e2\}$; $\mathcal{Q}=\{-x, 3\}$
(b) $T(ax+b)=bx2+(a-b)x-a$; $\mathcal{G}=\{1, x\}$; $\mathcal{Q}=\{x2, -x, -1\}$

5. Suppose $T : V\to W$ and $S : W\to Y$ are linear transformations that have corresponding matrices *A* and *B* with respect to some bases. Show that BA is the matrix of the composition $S\circ T : V\to Y$

EXERCISES

Exercises 1–4: Find v given the coordinate vector $[v]_{\mathcal{G}}$ with respect to the basis $\mathcal{G}$.

1. $[v]_{\mathcal{G}}=[-41]$; $\mathcal{G}=\{[32], [14]\}$

2. $[v]_{\mathcal{G}}=[25]$; $\mathcal{G}=\{-3x+1, 2x-4\}$

3. $[v]_{\mathcal{G}}=[-103]$; $\mathcal{G}=\{x2-x+3, 3x2+4, -5x-2\}$

4. $[v]_{\mathcal{G}}=[1231]$; $\mathcal{G}=\{[1121], [1222], [0210], [2103]\}$

Exercises 5–12: Find the coordinate vector of v with respect to the given basis $\mathcal{G}$.

5. $v=[89]$; $\mathcal{G}=\{[20], [03]\}$

6. $v=-14x+15$; $\mathcal{G}=\{2x, 5\}$

7. $v=12x2-15x+30$; $\mathcal{G}=\{4x2, 3x, 5\}$

8. $v=[-6620-7]$; $\mathcal{G}=\{[2000], [0300], [0050], [0001]\}$

9. $v=[5-5]$; $\mathcal{G}=\{[12], [31]\}$

10. $v=-11x$; $\mathcal{G}=\{-3x+2, 2x-5\}$

11. $v=9x+1$; $\mathcal{G}=\{x2-1, 2x+1, 3x2-1\}$

12. $v=[031-1]$; $\mathcal{G}=\{[1010], [0211], [1-101], [0011]\}$

Exercises 13–18: Let *A* be the matrix of linear transformation T : V→W with respect to bases $\mathcal{G}$ and $\mathcal{Q}$, respectively. Find T(v) for the given [v]$\mathcal{G}$.

13. A=[132−1]; [v]$\mathcal{G}$=[4−3]; $\mathcal{Q}$={[11], [23]}

14. A=[4031]; [v]$\mathcal{G}$=[22]; $\mathcal{Q}$={3x−2, x+5}

15. A=[112013011]; [v]$\mathcal{G}$=[134]; $\mathcal{Q}$={x2−2x, x2+x+4, 3x+1}

16. A=[−1314011−10]; [v]$\mathcal{G}$=[20−1]; $\mathcal{Q}$= {[31−2], [240], [123]}

17. A=[0432−1−32−2−1]; [v]$\mathcal{G}$=[2−31]; $\mathcal{Q}$={sin(x), cos(x), e−x}

18. A=[0011101020011100]; [v]$\mathcal{G}$=[321−1]; $\mathcal{Q}$= {[1012], [2−102], [1421], [3124]}

Exercises 19–26: Find the matrix *A* of the linear transformation T : V→W with respect to bases $\mathcal{G}$ and $\mathcal{Q}$, respectively.

19. T ([ab])=bx−a; $\mathcal{G}$={e1, e2}; $\mathcal{Q}$={x, 1}

20. T(ax+b)=ax2+(a+b)x−b; $\mathcal{G}$={x, 1}; $\mathcal{Q}$={x2, x, 1}

21. T(f(x))=f′(x); $\mathcal{G}$={x2, x, 1}; $\mathcal{Q}$={x, 1}

22. T(f(x))=xf′(x)+f(0); $\mathcal{G}$={x2+3, x−2, x2+x}; $\mathcal{Q}$={x2, x, 1}

23. T ([ab])=[b−aa+2b]; $\mathcal{G}$={e1, e2}; $\mathcal{Q}$={[11], [12]}

24. T(ax+b)=(b−a)x−(a+b); $\mathcal{G}$={x, 1}; $\mathcal{Q}$={x+1, x+2}

25. T ([ab])=−bx+a+b; $\mathcal{G}$={[21] [32]}; $\mathcal{Q}$={5x+3, 2x+1}

26. T ([abcd])=[a+bb+cc+d]; $\mathcal{G}$={[1000], [1100], [1110], [1111]}; $\mathcal{Q}$= {[100], [110], [111]}

Exercises 27–30: Suppose that

A=[abcdef]

is the matrix of T : V→W with respect to bases $\mathcal{G}$={g1, g2, g3} and $\mathcal{Q}$= {q1, q2}, respectively. Find the matrix of *T* with respect to the given bases $\mathcal{H}$ and $\mathcal{R}$.

27.

(a) $\mathcal{H}$={g1, g2, g3}, $\mathcal{R}$={2q1, q2}

(b) $\mathcal{H}$={3g1, g2, g3}, $\mathcal{R}$={q1, q2}

28.

(a) $\mathcal{H}=\{g3, g2, g1\}$, $\mathcal{R}=\{q1, q2\}$
(b) $\mathcal{H}=\{g1, g2, g3\}$, $\mathcal{R}=\{q2, q1\}$

29.

(a) $\mathcal{H}=\{g3, g1, g2\}$, $\mathcal{R}=\{q1, q2\}$
(b) $\mathcal{H}=\{g1, g3, g2\}$, $\mathcal{R}=\{q2, q1\}$

30.

(a) $\mathcal{H}=\{g1, 5g2, g3\}$, $\mathcal{R}=\{q1, 2q2\}$
(b) $\mathcal{H}=\{g2, g1, 3g3\}$, $\mathcal{R}=\{4q2, q1\}$

31. Suppose that T : P1→P1 has matrix A=[1301] with respect to the basis $\mathcal{G}$={x+1, x−1} for the domain and $\mathcal{Q}$={1, x} for the codomain. Use the inverse of *A* to find T−1(x).

32. Suppose that T : P1→P1 has matrix A=[2713] with respect to the basis $\mathcal{G}$={2x+1, 3x−1} for the domain and $\mathcal{Q}$={1, x+3} for the codomain. Use the inverse of *A* to find T−1(x+2).

33. Suppose that T : R2→P1 has matrix A=[2132] with respect to the basis $\mathcal{G}$={[13], [21]} for the domain and $\mathcal{Q}$={x, 2x+1} for the codomain. Use the inverse of *A* to find T−1(x+1).

34. Suppose that T : R2→P1 has matrix A=[1314] with respect to the basis $\mathcal{G}$={[25], [13]} for the domain and $\mathcal{Q}$={3x, 2x−1} for the codomain. Use the inverse of *A* to find T−1(x−2).

FIND AN EXAMPLE Exercises 35–40: Find an example that meets the given specifications. Prove your claim.

35. An element v and basis $\mathcal{G}$ of P2 such that

[v]$\mathcal{G}$=[20−3]

36. An element v and basis $\mathcal{G}$ of R2×2 such that

[v]$\mathcal{G}$=[1−74−4]

37. A basis $\mathcal{G}$ of R2 such that

[75]$\mathcal{G}$=[−34]

38. A basis $\mathcal{G}$ of P1 such that

$$[4x-11]\mathcal{G}=[3-5]$$

39. Vector spaces *V* and *W*, bases for each, and a linear transformation T : V→W that has matrix

$$A=[212031]$$

with respect to the bases.

40. Vector spaces *V* and *W*, bases for each, and a linear transformation T : V→W that has matrix

$$A=[6133-25]$$

with respect to the bases.

TRUE OR FALSE Exercises 41–44: Determine if the statement is true or false, and justify your answer.

41.

(a) The matrix of any linear transformation between finite-dimensional vector spaces must be square.

(b) The matrix of a linear transformation T : V→W is unique.

42.

(a) The matrix of a linear transformation between isomorphic finite-dimensional vector spaces must be invertible.

(b) If T : V→V is the identity linear transformation T(v)=v, where *V* is finite dimensional, then the matrix of *T* must be the identity matrix.

43.

(a) If [ab] is the coordinate vector of a vector with respect to a basis $\mathcal{G}$, then [2a−3b] is the coordinate vector with respect to $\mathcal{G}$ of some other vector.

(b) If [v]$\mathcal{G}$=[ab] for $\mathcal{G}$={g1, g2}, then [v]$\tilde{\mathcal{G}}$=[ba] for $\tilde{\mathcal{G}}$={g2, g1}.

44.

(a) If 0 is the zero vector of a finite-dimensional vector space *V*, then

$$[\ 0]\mathcal{G}=[0\vdots0]$$

for every basis $\mathcal{G}$ of *V*.

(b) If $\mathcal{G}$ and $\mathcal{H}$ are two distinct bases for a finite-dimensional vector space *V*, then $[\mathbf{v}]_{\mathcal{G}}$ and $[\mathbf{v}]_{\mathcal{H}}$ cannot have the same entries for any element v of *V*.

45. Let $\mathcal{G}$ be a basis for a vector space V of dimension m. Show that a set of vectors $\{v_1, \ldots, v_k\}$ is linearly independent in V if and only if the coordinate vectors $\{[v_1]_{\mathcal{G}}, \ldots, [v_k]_{\mathcal{G}}\}$ are linearly independent in $\mathbb{R}^m$.

46. Let $\mathcal{G}$ be a basis for a vector space V, and suppose that $[v]_{\mathcal{G}} = [w]_{\mathcal{G}}$. Prove that $v = w$.

47. Let $\mathcal{G}$ be a basis for a vector space V of dimension m. Show that a linear combination of vectors $v_1, \ldots, v_k$ is equal to v in V if and only if there is a linear combination of the coordinate vectors $[v_1]_{\mathcal{G}}, \ldots, [v_k]_{\mathcal{G}}$ that is equal to the coordinate vector $[v]_{\mathcal{G}}$ in $\mathbb{R}^m$.

48. Suppose that A is the matrix of linear transformation $T : V \to W$ with respect to bases $\mathcal{G}$ and $\mathcal{Q}$, respectively.

(a) Show that v is in the kernel of T if and only if $[v]_{\mathcal{G}}$ is in the null space of A.

(b) Show that w is in the range of T if and only if $[w]_{\mathcal{Q}}$ is in the column space of A.

49. Suppose that A is the matrix of linear transformation $T : V \to V$ with respect to basis $\mathcal{G}$ for both the domain and codomain. Let $T^2(v) = T \circ T(v) = T(T(v))$ denote the composition of T with itself.

(a) Show that A^2 is the matrix of the linear transformation $T^2 : V \to V$ with respect to basis $\mathcal{G}$ for both the domain and codomain.

(b) If T^n denotes the n-fold composition of T with itself, then show that A^n is the matrix of the linear transformation $T^n : V \to V$ with respect to basis $\mathcal{G}$ for both the domain and codomain.

9.4 Similarity

In this section we continue our exploration of matrix representatives of linear transformations, now focusing on the special case T : V→V, where the same basis $\mathcal{G}$ is used for both the domain and codomain. Let’s start with an example.

Example 1

Let T : R3→R3 be given by T(x)=Ax, where

A=[10101−421−1]

and x is with respect to the standard basis. Find the matrix *B* of *T* with respect to the basis

$\mathcal{G}$={[2−41], [−130], [011]}={g1, g2, g3}

Solution Even though we are working with a linear transformation from a vector space to itself, and the same basis is being used for both the domain and codomain, we still follow the same procedure as we used in Section 9.3. For each gi in $\mathcal{G}$, we need to find the coordinate vector of T(gi) with respect to $\mathcal{G}$. Starting with g1, we have

T(g1)=T([2−41]) =A[2−41]=[3−8−1]

To determine [T(g1)]$\mathcal{G}$, we need to find c1, c2, and c3 such that

[3−8−1]=c1[2−41]+c2 [−130]+c3 [011]=[2−10−431101][c1c2c3]

Setting S=[g1 g2 g3]=[2−10−431101], we have S−1=[31−152−2−3−12], so that

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = S^{-1} \begin{bmatrix} 3 \\ -8 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & -1 \\ 5 & 2 & -2 \\ -3 & -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -8 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$$

Therefore $[T(\mathbf{g}_1)]_{\mathcal{G}} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$. Since we have S^{-1}, computing $[T(\mathbf{g}_2)]_{\mathcal{G}}$ and $[T(\mathbf{g}_3)]_{\mathcal{G}}$ is easier. They are

$$T(\mathbf{g}_2) = \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} \Rightarrow [T(\mathbf{g}_2)]_{\mathcal{G}} = S^{-1} \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

$$T(\mathbf{g}_3) = \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix} \Rightarrow [T(\mathbf{g}_3)]_{\mathcal{G}} = S^{-1} \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

Thus the matrix B of T with respect to the basis $\mathcal{G}$ is

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

Looking back at our work, we see that since $T(\mathbf{x}) = A\mathbf{x}$, computing each column $\mathbf{b}_i$ of B amounted to first computing $T(\mathbf{g}_i) = A\mathbf{g}_i$, and from this computing $S^{-1}(T(\mathbf{g}_i)) = S^{-1}A\mathbf{g}_i$. Thus we have

$$\mathbf{b}_i = S^{-1}A\mathbf{g}_i \text{ for } i = 1, 2, 3$$

But since $B = [\mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_3]$ and $S = [\mathbf{g}_1\ \mathbf{g}_2\ \mathbf{g}_3]$, it follows that

$$B = S^{-1}AS$$

This form is reminiscent of the diagonalization of matrices, and it results from an underlying change of basis that is taking place. Specifically, recall from Section 4.4 that S is the change of basis matrix from $\mathcal{G}$ to the standard basis for $\mathbb{R}^3$. Thus we can think of the product $S^{-1}AS[\mathbf{v}]_{\mathcal{G}}$ as doing the following:

(a) Multiplying S by $[\mathbf{v}]_{\mathcal{G}}$ converts $[\mathbf{v}]_{\mathcal{G}}$ from the coordinate vector with respect to $\mathcal{G}$ to the standard basis.

(b) Multiplying A by $S[\mathbf{v}]_{\mathcal{G}}$ performs the linear transformation T.

(c) Multiplying S^{-1} by $AS[\mathbf{v}]_{\mathcal{G}}$ converts from the standard basis back to the coordinate vector with respect to $\mathcal{G}$.

The same change of basis can be performed between two bases of any finite-dimensional vector space.

Change of Basis

Suppose that $\mathcal{G}=\{g_1,\ldots,g_m\}$ and $\mathcal{H}=\{h_1,\ldots,h_m\}$ are bases for a vector space V. Then for each g_i, there exists a unique set of scalars $s_{1i},\ldots,s_{mi}$ such that

$$g_i=s_{1i}h_1+\cdots+s_{mi}h_m$$

Now set

$$S=\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mm} \end{bmatrix} \tag{1}$$

Then for $v=a_1g_1+\cdots+a_mg_m$, we have

$$v=a_1g_1+\cdots+a_mg_m =a_1(s_{11}h_1+\cdots+s_{m1}h_m)+\cdots+a_m(s_{1m}h_1+\cdots+s_{mm}h_m) =(a_1s_{11}+\cdots+a_ms_{1m})h_1+\cdots+(a_1s_{m1}+\cdots+a_ms_{mm})h_m = \left[S\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}\right]_{\mathcal{H}}$$

Therefore

$$[v]_{\mathcal{G}}=\left[S\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}\right]_{\mathcal{H}}$$

Thus multiplication by the matrix S changes the coordinate vector with respect to $\mathcal{G}$ into the coordinate vector with respect to $\mathcal{H}$.

DEFINITION 9.17 ▶

Change of Basis Matrix

The matrix S in (1) is called the **change of basis matrix** from $\mathcal{G}$ to $\mathcal{H}$.

Note that the change can be reversed with the inverse matrix: S^{-1} converts a coordinate vector with respect to $\mathcal{H}$ into the equivalent coordinate vector with respect to $\mathcal{G}$.

Example 2

Find the change of basis matrix S from basis $\mathcal{G}$ to basis $\mathcal{H}$ of P1, where

$$\mathcal{G}=\{2x+5, x+3\}, \mathcal{H}=\{2x-1, x-1\}$$

Then use S to find $[\ v]\mathcal{H}$ for $[\ v]\mathcal{G}=[-13]$, and find the change of basis matrix from $\mathcal{H}$ to $\mathcal{G}$.

Solution Starting with the first vector in $\mathcal{G}$, we have

$$2x+5=s11(2x-1)+s21(x-1)$$

Regrouping and comparing coefficients yields the linear system

$$2s11+s21=2-s11-s21=5$$

which has the unique solution s11=7, s21=−12. The second vector x+3 gives rise to the equation

$$x+3=s12(2x-1)+s22(x-1)$$

Solving the equivalent linear system yields the unique solution s12=4, s22=−7. Therefore

$$S=[74-12-7]$$

and so

$$[\ v]\mathcal{H}=S\ [\ v]\mathcal{G}=[74-12-7][-13]=[5-9]$$

Thus we can conclude that

$$-(2x+5)+3(x+3)=5(2x-1)-9(x-1)$$

(Both simplify to x+4.) The change of basis matrix from $\mathcal{H}$ to $\mathcal{G}$ is given by

$$S^{-1} = \begin{bmatrix} 7 & 4 \\ -12 & -7 \end{bmatrix}$$

Note that $S^{-1} = S$, so the change of basis matrix is the same from $\mathscr{H}$ to $\mathscr{G}$ as it is from $\mathscr{G}$ to $\mathscr{H}$. This is not typical, but it is possible.

Transformation Matrices Revisited

Now that we know how change of basis matrices are defined for a general vector space *V*, we can combine this with what we learned in Section 9.3 about transformation matrices.

THEOREM 9.18

Let $T : V \to V$ be a linear transformation. Suppose that *A* and *B* are the matrices of *T* with respect to the bases $\mathscr{G}$ and $\mathscr{H}$, respectively, and let *S* be the change of basis matrix from $\mathscr{G}$ to $\mathscr{H}$. Then $A = S^{-1}BS$.

Proof For a typical vector $\mathbf{v}$ in *V*, the product $A[\mathbf{v}]_{\mathcal{G}}$ produces $[T(\mathbf{v})]_{\mathcal{G}}$, the coordinate vector of $T(\mathbf{v})$ with respect to the basis $\mathscr{G}$. On the other hand,

(a) $S[\mathbf{v}]_{\mathcal{G}}$ converts $[\mathbf{v}]_{\mathcal{G}}$ to $[\mathbf{v}]_{\mathcal{H}}$, the coordinate vector with respect to $\mathscr{H}$.

(b) $B(S[\mathbf{v}]_{\mathcal{G}})$ yields $[T(\mathbf{v})]_{\mathcal{H}}$, the coordinate vector of $T(\mathbf{v})$ with respect to the basis $\mathscr{H}$.

(c) $S^{-1}(BS[\mathbf{v}]_{\mathcal{G}})$ converts $[T(\mathbf{v})]_{\mathcal{H}}$ back to $[T(\mathbf{v})]_{\mathcal{G}}$.

Therefore $A[\mathbf{v}]_{\mathcal{G}}$ and $S^{-1}BS[\mathbf{v}]_{\mathcal{G}}$ are the same for all $\mathbf{v}$, and hence $A = S^{-1}BS$. ■

Two matrices *A* and *B* related as in Theorem 9.18 go by a special name.

DEFINITION 9.19 ▶

Similar Matrices, Similarity Transformation

A square matrix A is **similar** to matrix B if there exists an invertible matrix S such that $A=S^{-1}BS$. The change from B to A is called a **similarity transformation**.

Note that if $A=S^{-1}BS$, then $B=SAS^{-1}$. Hence setting $R=S^{-1}$ gives

$$R^{-1}AR=(S^{-1})^{-1}AS^{-1}=SAS^{-1}=B$$

so that B is also similar to A. Thus it makes sense to simply say that A and B are similar matrices.

Example 3

Let matrices

$$A=\begin{bmatrix}4&1&-3\\0&2&5\\1&-1&3\end{bmatrix},\quad B=\begin{bmatrix}4&3&-8\\-109&-16&151\\-16&-1&21\end{bmatrix},\quad S=\begin{bmatrix}1&3&-2\\2&7&2\\1&3&-1\end{bmatrix}$$

Show that A and B are similar matrices with similarity transformation matrix S.

Solution First, we note that $\det(S)=1$, so that S is an invertible matrix. To show that A and B are similar, we shall show that $SA=BS$, which saves us the trouble of computing S^{-1}. We have

$$SA=\begin{bmatrix}1&3&-2\\2&7&2\\1&3&-1\end{bmatrix}\begin{bmatrix}4&1&-3\\0&2&5\\1&-1&3\end{bmatrix}=\begin{bmatrix}2&9&6\\10&14&35\\3&8&9\end{bmatrix}$$

and

$$BS=\begin{bmatrix}4&3&-8\\-109&-16&151\\-16&-1&21\end{bmatrix}\begin{bmatrix}1&3&-2\\2&7&2\\1&3&-1\end{bmatrix}=\begin{bmatrix}2&9&6\\10&14&35\\3&8&9\end{bmatrix}$$

Since $SA=BS$ and S is invertible, then $A=S^{-1}BS$ and hence A and B are similar matrices.

THEOREM 9.20

Two matrices A and B are similar if and only if A and B are the transformation matrices of some linear transformation $T : V \to V$ with respect to bases of V.

Proof Combining Theorem 9.18 and Definition 9.19, we see that if A and B are transformation matrices for the same linear transformation T, then A and B are similar.

Now suppose that A and B are similar matrices with $A=S^{-1}BS$ and $S=[s_1 \cdots s_m]$. Then $\mathcal{G}=\{s_1,\ldots,s_m\}$ is a basis for $\mathbb{R}^m$ because S is invertible. Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be given by $T(v)=Bv$. Then B is the transformation matrix with respect to the standard basis, and as S is the change of basis matrix from $\mathcal{G}$ to the standard basis, it follows that A is the transformation matrix of T with respect to $\mathcal{G}$. Hence A and B are both transformation matrices for T. ■ ■

Note that given two matrices A and B, generally there is no simple way to determine if they are similar. One approach that is sure to work is illustrated in the next example.

Example 4

Determine if the given matrices A and B are similar.

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

Solution A and B are similar if there exists an invertible matrix S such that $A=S^{-1}BS$, or equivalently, $SA=BS$. Letting $S=$

$\begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$ and multiplying out *SA* and *BS*, we have

$$\begin{bmatrix} -2s_{11}+2s_{12} & -9s_{11}+7s_{12} \\ -2s_{21}+2s_{22} & -9s_{21}+7s_{22} \end{bmatrix} = \begin{bmatrix} 3s_{11}-s_{21} & 3s_{12}-s_{22} \\ -2s_{11}+2s_{21} & -2s_{12}+2s_{22} \end{bmatrix}$$

Setting the components equal to one another yields the homogeneous linear system

$$\begin{aligned} 5s_{11}-2s_{12}-s_{21} \quad &= 0 \\ 9s_{11}-4s_{12} \quad -s_{22} &= 0 \\ 2s_{11} \quad -4s_{21}+2s_{22} &= 0 \\ 2s_{12}-9s_{21}+5s_{22} &= 0 \end{aligned}$$

The system has the trivial solution, but since we require *S* to be invertible, we are not interested in that solution. However, there are also nontrivial solutions—for instance, $s_{11}=1$, $s_{12}=2$, $s_{21}=1$, and $s_{22}=1$. These give us

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

As $\det(S) = -1 \neq 0$, it follows that *S* is invertible. We can check that $A = S^{-1}BS$ by computing

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

and

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

Hence $SA = BS$, so that *A* and *B* are similar.

Note that in Example 4, any nonzero scalar multiple of *S* would also serve as a solution to $SA = BS$, showing us that *S* is not unique.

Two matrices that are similar have several interesting properties in common.

THEOREM 9.21

If *A* and *B* are similar matrices, then they have the same characteristic polynomial and the same eigenvalues (including multiplicities), and det(A)=det(B).

Proof If *A* and *B* are similar matrices, then there exists an invertible matrix *S* such that A=S−1BS. Hence

det(A−λI)=det(S−1BS−λI) =det(S−1BS−S−1(λI)S) =det(S−1(B−λI)S) =det(S−1)det(B−λI)det(S) =det(B−λI)

▶ Here we use the fact that det(*CD*) = det(*C*) det(*D*) for $n \times n$ matrices *C* and *D*, and if *C* is invertible then $\det(C^{-1}) = (\det(C))^{-1}$.

Thus the characteristic polynomials of *A* and *B* are the same, and therefore the eigenvalues (including multiplicities) are also the same. Setting λ=0 above shows that det(A)=det(B), completing the proof. ■■

The last part of Theorem 9.21 tells us that if det(A)≠det(B), then *A* and *B* are not similar. However, note that if det(A)=det(B), then we cannot draw any conclusion.

Example 5

Use determinants to try to determine if the pairs of matrices *A* and *B* are similar.

(a) A=[3245], B=[5121]

(b) A=[103213022], B=[230412301]

Solution For the matrices in (a), we have det(A)=7 and det(B)=4. Since the determinants differ, the two matrices cannot be similar.

For (b), both det(A)=8 and det(B)=8. Thus determinants tell us nothing about whether or not the two matrices are similar.

Although determinants are not helpful for part (b), we can apply another part of Theorem 9.21. If A and B are similar, then they have the same eigenvalues. For the matrices in part (b), it can be shown that they have different eigenvalues and so cannot be similar matrices.

COMPUTATIONAL COMMENTS

There are a number of algorithms for estimating eigenvalues that exploit the fact that similar matrices have the same eigenvalues. The popular *QR algorithm* produces a sequence of similar matrices that become successively closer to upper triangular. It starts by producing the QR factorization (see Section 8.4) $A=Q_1R_1$, where Q is an orthogonal matrix and R is upper triangular. Next, we let $A_1=R_1Q_1$, so that

$$A_1=R_1Q_1=Q_1^{-1}Q_1R_1Q_1=Q_1^{-1}AQ_1 \tag{2}$$

Thus A and A_1 are similar matrices and hence by Theorem 9.21 have the same eigenvalues. For $i>1$, we let $A_i=Q_{i+1}R_{i+1}$ be the QR factorization for A_i and then define $A_{i+1}=R_{i+1}Q_{i+1}$. By the same reasoning as in 2, A_i and A_{i+1} are similar matrices. Therefore the sequence $A, A_1, A_2, \ldots$ of matrices all have the same eigenvalues. Under certain conditions, the sequence of matrices converges to a triangular matrix that has eigenvalues along the diagonal.

Another (older) algorithm called *Jacobi's Method* is applicable to symmetric matrices A. This method resembles matrix diagonalization, starting with $A_1=A$ and setting

$$A_{i+1}=P_i^{-1}A_iP_i \text{ for } i=1, 2, \ldots$$

Note that P_i is not the same as the orthogonal matrix found when diagonalizing a symmetric matrix, and A_i is not diagonal. (How P_i and A_i are defined is beyond the scope of this discussion.) However, the sequence of matrices $A_1, A_2, \ldots$ are all similar, and they converge to a diagonal matrix with the eigenvalues of A on the diagonal.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the change of basis matrix from $\mathcal{G}$ to $\mathcal{H}$.

(a) $\mathcal{G}$={x−2, 2x+1}, $\mathcal{H}$={x, 1}

(b) $\mathcal{G}$={3x−2, x2+x, x2−x+3}, $\mathcal{H}$={1, x2, x}

2. B is the matrix of T : V→V with respect to a basis $\mathcal{H}$, and S is the change of basis matrix from a basis $\mathcal{G}$ to $\mathcal{H}$. Find the matrix A of T with respect to $\mathcal{G}$.

(a) B=[1−421], S=[10−15]

(b) B=[1010131−1−1], S=[100310121]

3. Determine if A and B are similar matrices.

(a) A=[136−1], B=[32−15]

(b) A=[10−21−242−23], B=[1−32−112−214]

EXERCISES

Exercises 1–8: Find the change of basis matrix from $\mathcal{G}$ to $\mathcal{H}$.

1. $\mathcal{G}$={2x−1, 5x+4}, $\mathcal{H}$={x, 1}

2. $\mathcal{G}$={x2−7x+5, 3x2+1, 7x−3}, $\mathcal{H}$={x2, x, 1}

3. $\mathcal{G}$={[3210], [4002], [1751], [0623]}, $\mathcal{H}$= {[1000], [0100], [0010], [0001]}

4. $\mathcal{G}$={x+3, 4x+1}, $\mathcal{H}$={1, x}

5. $\mathcal{G}$={x−2, x2+9x, x2−x−1}, $\mathcal{H}$={x, 1, x2}

6. $\mathcal{G}$={[6100], [5238], [2597], [1320]} $\mathcal{H}$= {[0100], [1000], [0001], [0011]}
7. $\mathcal{G}$={7x+4, 3x+2}, $\mathcal{H}$={2x+1, 5x+3}
8. $\mathcal{G}$={x2+2x+1, x2+1, x2−x+2} $\mathcal{H}$={x2+x+1, x+1, 1}

Exercises 9–12: *B* is the matrix of T : V→V with respect to a basis $\mathcal{H}$, and *S* is the change of basis matrix from a basis $\mathcal{G}$ to $\mathcal{H}$. Find the matrix *A* of *T* with respect to $\mathcal{G}$.

9. B=[23−41], S=[1327]
10. B=[5017], S=[4332]
11. B=[10201112−1], S=[100210131]
12. B=[311141423], S=[010101100]

Exercises 13–16: *B* is the matrix of T : V→V with respect to the basis $\mathcal{H}$. Find the matrix *A* of *T* with respect to $\mathcal{G}$.

13. B=[1211], $\mathcal{H}$={x, 1}, $\mathcal{G}$={3x+1, 2x+1}
14. B=[121011102], $\mathcal{H}$={x2, x, 1}, $\mathcal{G}$={x2+x+1, x+1, 1}
15. B=[2241], $\mathcal{H}$={x+3, −x+5}, $\mathcal{G}$={2x−1, −3x+2}
16. B=[3−102−11114], $\mathcal{H}$={1, x2+1, x−1}, $\mathcal{G}$={x−3, x2−2x+4, 1}

Exercises 17–20: Determine if *A* and *B* are similar matrices.

17. A=[1325], B=[2132]
18. A=[21−10], B=[1011]
19. A=[1−131−3−3012], B=[11210101−1]
20. A=[12131−101−2], B=[11−1−23110−2]

FIND AN EXAMPLE Exercises 21–24: Find an example that meets the given specifications. Prove your claim.

21. A vector space *V* with bases $\mathcal{G}$ and $\mathcal{H}$ related by the matrix *S* in (1) for

S=[3423]

22. A vector space V with bases $\mathcal{G}$ and $\mathcal{H}$ related by the matrix S in (1) for

$$S=[13-214426-3]$$

23. Two similar matrices A and B that are related by

$$S=[5283]$$

24. Two similar matrices A and B that are related by

$$S=[34-223-4346]$$

TRUE OR FALSE Exercises 25–30: Determine if the statement is true or false, and justify your answer.

25.

(a) If $A=S-1BS$, then A and B are similar.

(b) The matrix of a linear transformation $T : V\rightarrow V$ is unique for a fixed basis $\mathcal{G}$ of V.

26.

(a) Matrices A and B are similar if and only if they have the same dimensions.

(b) If A and B are similar matrices, then so are AT and BT.

27.

(a) If A and B have the same rank, then they are similar.

(b) Two similar matrices have the same eigenvectors.

28.

(a) If A and B are not similar and B and C are not similar, then A and C are not similar.

(b) If A, B, and C are similar, then AB and BC are similar.

29.

(a) If there exists a matrix S such that $SA=BS$, then A and B are similar matrices.

(b) For every A there exists a distinct B such that A and B are similar matrices.

30.

(a) If A and B are similar matrices, then $\text{null}(A)=\text{null}(B)$.

(b) If A and B are similar matrices, then AB and BA are similar matrices.

31. Suppose that A and B are similar matrices, related by $A=S_1^{-1}BS_1$, and that B and C are also similar matrices, related by $B=S_2^{-1}CS_2$. Find the matrix D that relates A and C by $A=D^{-1}CD$.
32. Prove that similarity of matrices is transitive: if A is similar to B and B is similar to C, then A is similar to C.
33. Suppose that A and B are both diagonalizable matrices that have the same eigenvalues, including multiplicities. Prove that A and B are similar matrices.
34. Prove that if A and B are similar matrices, then so are A^k and B^k.
35. Prove that if A and B are similar matrices, then so are A^T and B^T.
36. Suppose that A and B are similar matrices and that A is invertible. Prove that B is also invertible and that A^{-1} and B^{-1} are also similar.

Exercises 37–40: Determine if the given matrices are similar.

37. $A=\begin{bmatrix}1&-2&4\\5&1&2\\0&1&-3\end{bmatrix}$, $B=\begin{bmatrix}1&-1&1\\5&0&0\\0&1&2\end{bmatrix}$
38. $A=\begin{bmatrix}3&2&-2\\1&4&0\\-2&1&-1\end{bmatrix}$, $B=\begin{bmatrix}1&3&-1\\3&3&1\\-2&1&2\end{bmatrix}$
39. $A=\begin{bmatrix}1&0&1&3\\-1&2&4&1\\2&3&-1&0\\0&2&-2&-2\end{bmatrix}$, $B=\begin{bmatrix}1&2&1&1\\-3&-3&4&-1\\2&5&2&-1\\0&2&0&0\end{bmatrix}$
40. $A=\begin{bmatrix}2&1&-1&2\\3&0&1&0\\-1&2&4&1\\0&0&3&-1\end{bmatrix}$, $B=\begin{bmatrix}4&0&1&-3\\2&3&4&-4\\1&0&0&2\\-2&-1&0&2\end{bmatrix}$

SUPPLEMENTARY EXERCISES

1. Let $T : V \to P_2$ be a linear transformation satisfying

$$T(v_1)=x^2-3 \quad T(v_2)=-2x+7 \quad T(v_3)=3x^2+5x$$

Find $T(v_1+3v_2+v_3)$.

2. Let $T : P_2 \to R^2$ be a linear transformation satisfying

$$T(x^2+2)=[15]$$

$$T(2x+3)=[2-3]$$

Find $T(3x^2-2x)$.

Exercises 3–4: Prove that *T* is a linear transformation.

3. $T : P_2 \to R^2$ with $T(ax^2+bx+c)=[b-aa+4c]$.

4. $T : R^{2\times 2} \to P_2$ with

$$T([abcd])=bx^2-ax+4c$$

Exercises 5–6: Determine if the given function is a linear transformation. Justify your answer.

5. $T : P_1 \to R^1$ with $T(ax+b)=a-b$.

6. $T : R^{2\times 2} \to P_1$ with $T([abcd])=ax+c$.

Exercises 7–8: Give the kernel and range of the linear transformation.

7. $T : P_2 \to R$ with $T(ax^2+bx+c)=a-b+c$.

8. $T : R^2 \to P_1$ with $T([ab])=ax-b$

Exercises 9–10: Based on dimensions, determine if the vector spaces are isomorphic. If dimensions cannot be used, explain why.

9. V=R4×6 and W=R12×2

10. V=R∞ and W=C(R)

Exercises 11–12: Prove that *T* is an isomorphism.

11. T : R3→P2 with T([abc]) =cx2+ax+b

12. T : R2×2→R2×2 with T(A)=−A

Exercises 13–14: Find T−1 for the given isomorphism.

13. T : P1→R2 with T(ax+b)=[ba]

14. T : R2×2→R4 with T ([abcd])=[b−ad−c]

Exercises 15–16: Find v given the coordinate vector [v]$\mathcal{G}$ with respect to the basis $\mathcal{G}$.

15. [v]$\mathcal{G}$=[−13]; $\mathcal{G}$={[−43], [14])

16. [v]$\mathcal{G}$=[110]; $\mathcal{G}$={x2−4x+1, −2x2+5, −x+1}

Exercises 17–18: Find the coordinate vector of v with respect to $\mathcal{G}$.

17. v=8x+5; $\mathcal{G}$={2x+5, x+1}

18. v=[1−321]; $\mathcal{G}$={[−1111], [1200], [11−11], [0111]}

Exercises 19–20: Let *A* be the matrix of linear transformation T : V→W with respect to bases $\mathcal{G}$ and $\mathcal{Q}$, respectively. Find T(v) for the given [v]$\mathcal{G}$.

19. A=[13−14]; [v]$\mathcal{G}$=[−12]; $\mathcal{Q}$={3x−1, x+4}

20. A=[1102−121−2−1]; [v]$\mathcal{G}$=[31−2]; $\mathcal{Q}$={cos(x), ex, sin(x)}

Exercises 21–22: Find the matrix *A* of the linear transformation T : V→W with respect to bases $\mathcal{G}$ and Q, respectively.

21. T ([ab])=bx−(b−a); $\mathcal{G}$={e1, e2}; $\mathcal{Q}$={2x, −1}

22. T(ax+b)=ax2+(a−b)x+b; $\mathcal{G}$={x, 1}; $\mathcal{Q}$={x2, 2 ,−x}

Exercises 23–24: Find the change of basis matrix from $\mathcal{G}$ to $\mathcal{H}$.

23. $\mathcal{G}$={x+1, −x+2}, $\mathcal{H}$={2x, 1}

24. $\mathcal{G}$={3x, x2−x, x2+3}, $\mathcal{H}$={1, 3x, x2}

Exercises 25–26: *B* is the matrix of $T : V \to V$ with respect to a basis $\mathscr{H}$, and *S* is the change of basis matrix from a basis $\mathscr{G}$ to $\mathscr{H}$. Find the matrix *A* of *T* with respect to $\mathscr{G}$.

25. $B=[1241]$, $S=[1235]$

26. $B=[1-1-120313-2]$, $S=[10131-2101]$

Exercises 27–28: Determine if *A* and *B* are similar matrices.

27. $A=[124-1]$, $B=[312-5]$

28. $A=[11-21-2-21-21]$, $B=[152-142-231]$

CHAPTER 10

Inner Product Spaces

Janine Lamontagne/Getty Images

Electric cars run on batteries instead of gasoline and do not emit pollutants from their tailpipes. Drivers must recharge the batteries regularly, either at home or at charging stations similar to the one shown. Some concerns about electric cars have been their higher cost compared to gasoline powered cars, few charging stations, and concerns about how far electric cars can travel before having to recharge again. But these concerns are being addressed as electric cars become more popular and commonplace.

In Chapter 8 we introduced dot products, which provided an algebraic way to determine when vectors in Euclidean space $\mathbf{R}^n$ are orthogonal. There we also developed applications of the dot product, including projections of vectors, the Gram–Schmidt process, and orthonormal bases.

In this chapter we introduce inner products, which extend the dot product from Euclidean space to vector spaces. In Section 10.1 we define the inner product, provide a number of examples of inner products, and give some results that are generalizations of those in Chapter 8. In Section 10.2 we develop the Gram–Schmidt process in the context of an inner product space. Section 10.3 is devoted to a few applications of inner products.

In previous chapters, topics involving calculus were separated from other material. Since some of the most important examples of inner products involve calculus, examples from calculus are more fully integrated into this chapter.

10.1 Inner Products

Since the dot product proved so useful in Euclidean space $\mathbf{R}^n$, we would like to extend the dot product to a similar product in other vector spaces. The four properties of dot products given in Theorem 8.2 of Section 8.1 are really what make the dot product useful. For instance, much of the development of the Gram–Schmidt process relies on these properties. Hence it makes sense that a generalized product on a vector space should have these same properties. In fact, it makes so much sense that we adapt the properties in Theorem 8.2 as a definition.

DEFINITION 10.1 ▶

Inner Product, Inner Product Space

Let **u**, **v**, and **w** be elements of a vector space V, and let c be a scalar. An **inner product** on V is a function that takes two vectors in V as input and produces a scalar as output. An inner product function is denoted by $\langle u, v\rangle$ and must satisfy the following conditions:

(a) $\langle u, v\rangle = \langle v, u\rangle$

(b) $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$

(c) $\langle cu, v\rangle = \langle u, cv\rangle = c\langle u, v\rangle$

(d) $\langle u, u\rangle \geq 0$, and $\langle u, u\rangle = 0$ only when $u=0$

A vector space V with an inner product defined on it is called an **inner product space**.

Since this definition is guided by properties of the dot product, it follows that the dot product is an inner product on $\mathbf{R}^n$. But this is only one of many inner products on vector spaces. In fact, we can modify the usual dot product to produce a "weighted dot product" on $\mathbf{R}^n$.

Example 1

Let $t_1, \ldots, t_n$ be positive scalars, which are the "weights." Show that

$$\langle \mathbf{u}, \mathbf{v} \rangle = t_1u_1v_1 + t_2u_2v_2 + \cdots + t_nu_nv_n \tag{1}$$

is an inner product on $\mathbf{R}^n$.

▶ In Example 1,
$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$

Solution A function taking two vectors as input and producing a scalar as output is an inner product if it satisfies conditions (a)–(d) in Definition 10.1. We verify (a) and (d) here, leaving (b) and (c) as an exercise.

To show that condition (a) is met, note that

$$\langle \mathbf{u}, \mathbf{v} \rangle = t_1u_1v_1 + t_2u_2v_2 + \cdots + t_nu_nv_n = t_1v_1u_1 + t_2v_2u_2 + \cdots + t_nv_nu_n = \langle \mathbf{v}, \mathbf{u} \rangle$$

To verify (d), we start by computing

$$\langle \mathbf{u}, \mathbf{u} \rangle = t_1u_1^2 + t_2u_2^2 + \cdots + t_nu_n^2$$

Since $u_i^2 \geq 0$ for $i = 1, \ldots, n$ and the weights $t_1, \ldots, t_n$ are all positive, we have $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$. Furthermore, because the weights are positive, the only way that $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ is if $u_1 = \cdots = u_n = 0$, which implies $\mathbf{u} = \mathbf{0}$. Hence condition (d) holds and the weighted dot product is an inner product.

Example 2

Let $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix}$ be in $\mathbf{R}^3$. Compute $\langle \mathbf{u}, \mathbf{v} \rangle$ using the weighted dot product defined in Example 1 with weights $t_1 = 2$, $t_2 = 3$, and $t_3 = 1$.

Solution We have

$$\langle u, v \rangle = t_1u_1v_1 + t_2u_2v_2 + t_3u_3v_3 = (2)(1)(4)+(3)(3)(-1)+(1)(-2)(1) = -3$$

Example 3

Let $p(x)$ and $q(x)$ be polynomials in $\mathbf{P}^n$, and suppose $x_0, x_1, \ldots, x_n$ are $n + 1$ distinct real numbers. Prove that

$$\langle p, q \rangle = p(x_0)q(x_0) + p(x_1)q(x_1) + \cdots + p(x_n)q(x_n) \tag{2}$$

is an inner product on $\mathbf{P}^n$.

Solution Properties (a)–(c) of Definition 10.1 follow readily from the properties of real numbers and are left as an exercise. For (d) we have

$$\langle p, p \rangle = (p(x_0))^2 + (p(x_1))^2 + \cdots + (p(x_n))^2$$

Since each term on the right must be nonnegative, it follows that $\langle p, p \rangle \geq 0$ for any polynomial p in $\mathbf{P}^n$. Moreover, $\langle p, p \rangle = 0$ exactly when $p(x_0) = p(x_1) = \cdots = p(x_n) = 0$. But the only way a polynomial of degree n or less can have $n + 1$ distinct roots is if it is the zero polynomial. Thus property (d) is also satisfied, so that (2) defines an inner product.

Example 4

Suppose that $p(x) = x^2 - 3x + 2$ and $q(x) = 2x^2 + 4x - 1$ are in $\mathbf{P}^2$, and that $x_0 = -1$, $x_1 = 1$, and $x_2 = 4$. Compute $\langle p, q \rangle$ using the inner product defined in Example 3.

Solution We have

$$\langle p, q\rangle = p(-1)q(-1)+p(1)q(1)+p(4)q(4)=(6)(-3)+(0)(5)+(6)(47)=264$$

Note that a weighted version of this inner product (2) can also be defined (see Exercise 49).

Example 5

Let f and g be two continuous functions in $C[-1, 1]$. Show that

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\, dx \tag{3}$$

defines an inner product on $C[-1, 1]$.

Solution Properties (a)–(c) of Definition 10.1 follow readily from basic properties of the definite integral. If f is in $C[-1, 1]$, then $(f(x))^2 \geq 0$ for all x in $[-1, 1]$, so that

$$\langle f, f\rangle = \int_{-1}^{1} (f(x))^2\, dx \geq 0$$

The second part of property (d) follows from the more subtle but plausible fact from real analysis that for a continuous function $f(x)$,

$$\int_{-1}^{1} (f(x))^2\, dx > 0$$

except when $f(x) = z(x)$, the identically zero function. Therefore this is the only function for which $\langle f, f\rangle = 0$, and hence (3) gives an inner product.

▶ $z(x)=0$ for all x in $[-1, 1]$.

Example 6

Let $f(x) = x^2 + 4x$ and $g(x) = 5x^2 - 3$. Evaluate ⟨f, g⟩ using the inner product defined in Example 5.

Solution The inner product of $f(x)$ and $g(x)$ is

$$\langle f, g\rangle = \int_{-1}^{1}(x^2+4x)(5x^2-3)\,dx = \int_{-1}^{1}(5x^4+20x^3-3x^2-12x)\,dx = 0$$

Orthogonality and Norms

In Chapter 8 two vectors **u** and **v** in $\mathbf{R}^n$ were defined to be orthogonal if $u \cdot v = 0$. We extend this to inner products in a natural way.

DEFINITION 10.2 ▶

Orthogonal Vectors

Two vectors **u** and **v** in an inner product space V are **orthogonal** if and only if ⟨**u**, **v**⟩ = 0.

For instance, in Example 6 we showed that the vectors $f(x) = x^2 + 4x$ and $g(x) = 5x^2 - 3$ are orthogonal with respect to the inner product given in Example 5.

Example 7

Which pairs among $p_1(x) = x^2 - 5x + 4$, $p_2(x) = x^2 - x - 2$, and $p_3(x) = 3x^2 - x - 4$ are orthogonal with respect to the inner product on $\mathbf{P}^2$ in Example 4?

Solution We have

$$\begin{aligned}\langle p_1, p_2\rangle &= p_1(-1)p_2(-1)+p_1(1)p_2(1)+p_1(4)p_2(4)=0\\ \langle p_1, p_3\rangle &= p_1(-1)p_3(-1)+p_1(1)p_3(1)+p_1(4)p_3(4)=0\\ \langle p_2, p_3\rangle &= p_2(-1)p_3(-1)+p_2(1)p_3(1)+p_2(4)p_3(4)=404\end{aligned}$$

Hence $p_1(x)$ and $p_2(x)$ are orthogonal, $p_1(x)$ and $p_3(x)$ are orthogonal, but $p_2(x)$ and $p_3(x)$ are not.

Example 8

If $V = C[-1, 1]$, then we can show that

$$\langle f, g\rangle = \int_{-1}^{1} (x^2+1) f(x) g(x)\, dx$$

is an inner product on V, a weighted version of the inner product in Example 5 (see Exercise 50). Compute $\langle f, g\rangle$ for $f(x)$ and $g(x)$ in Example 6.

Solution Here we have

$$\langle f, g\rangle = \int_{-1}^{1} (x^2+1)(x^2+4x)(5x^2-3)\, dx = \frac{8}{35}$$

so that $f(x)$ and $g(x)$ are not orthogonal with respect to this inner product.

In Section 8.1 we defined the norm (length) of a vector in $\mathbf{R}^n$ in terms of the dot product. Here we define the norm in terms of an inner product. Just as in Euclidean space, the norm on a vector space gives us a way to define the length of each vector in the space (the norm of the vector) and to measure the distance between vectors (the norm of the difference).

DEFINITION 10.3 ▶

Norm

Let **v** be a vector in an inner product space V. Then the **norm** of **v** is given by

$$||v||=\langle v, v\rangle$$

Example 9

If A and B are matrices in $\mathbf{R}^{3\times 3}$, then it can be shown that (see Exercise 53)

$$\langle A, B\rangle=tr(AT B)$$

is an inner product. Compute ||A|| for A=[1−2−1320−111].

▶ Recall that tr(C) denotes the *trace* of C, the sum of the diagonal entries of C.

Solution We have

$$||A||2=\langle A, A\rangle=tr(AT A)=tr([113−2393−232])=22$$

Hence ||A||=22≈4.69.

Example 10

Suppose that the vector space $\mathbf{P}^1$ has inner product

$$\langle p, q\rangle=p(0)q(0)+3p(1)q(1)+2p(3)q(3)$$

Determine which of $p(x) = 3x - 2$ and $q(x) = -2x + 4$ is longer, and find the distance between the two vectors.

Solution The length of each vector is given by the norm, so that

$$\|p\|=\langle p, p\rangle=(-2)^2+3(1)^2+2(7)^2=105\approx 10.247 \quad \|q\|=\langle q, q\rangle=(4)^2+3(2)^2+2(-2)^2=36=6$$

Since $\|p\|>\|q\|$, p is longer than q. (Remember that the results might be different with another inner product.) The distance between our vectors is given by $\|p-q\|$. Since $p(x) - q(x) = 5x - 6$, we have

$$\|p-q\|=\langle p-q, p-q\rangle=(-6)^2+3(-1)^2+2(9)^2=201\approx 14.177$$

In Euclidean space we formulated the Pythagorean Theorem in terms of norms. We can do the same with inner product spaces.

THEOREM 10.4

(PYTHAGOREAN THEOREM) Let **u** and **v** be vectors in an inner product space V. Then **u** and **v** are orthogonal if and only if

$$\|\mathbf{u}\|^2+\|\mathbf{v}\|^2=\|\mathbf{u}+\mathbf{v}\|^2 \tag{4}$$

Theorem 10.4 generalizes Theorem 8.6 in Section 8.1.

Proof The proof only uses properties from the definition of an inner product, without any reference to a specific inner product. We have

$$\|u+v\|^2=\langle u+v,\ u+v\rangle \ =\langle u,\ u\rangle+\langle u,\ v\rangle+\langle v,\ u\rangle+\langle v,\ v\rangle \ =\|u\|^2+\|v\|^2+2\langle u,\ v\rangle$$
(by (b) of Definition 10.1)

Since **u** and **v** are orthogonal if and only if
$\langle u, v\rangle=0$, then **u** and **v** are orthogonal if and only if (4) is true. ■ ■

Example 11

Verify the Pythagorean Theorem for the vectors $p_1(x) = x^2 - 5x + 4$ and $p_2(x) = x^2 - x - 2$ in the inner product space given in Example 4.

Solution We saw in Example 7 that $\langle p_1, p_2\rangle=0$, so by the Pythagorean Theorem we expect that $\|p_1\|^2+\|p_2\|^2=\|p_1+p_2\|^2$. To verify this, we compute

$$\|p_1\|^2=\langle p_1,\ p_1\rangle=(p_1(-1))^2+(p_1(1))^2+(p_1(4))^2=100 \quad \|p_2\|^2=\langle p_2,\ p_2\rangle=(p_2(-1))^2+(p_2(1))^2+(p_2(4))^2=104 \quad \|p_1+p_2\|^2=\langle p_1+p_2,\ p_1+p_2\rangle = (p_1(-1)+p_2(-1))^2+(p_1(1)+p_2(1))^2+(p_1(4)+p_2(4))^2=204$$

Therefore $\|p_1\|^2+\|p_2\|^2=\|p_1+p_2\|^2$.

Projection and Inequalities

When studying dot products in Euclidean space, we developed the projection of one vector onto another and extended this to the projection of a vector onto a subspace. Here we shall generalize projections onto a vector in a vector space. We treat projections onto subspaces in the next section.

In Section 8.2, the formula for the projection of one vector onto another was defined in terms of dot products. Hence it is reasonable to generalize projections to inner product spaces by changing the dot products into inner products.

DEFINITION 10.5 ▶

Projection onto a Vector

Let **u** and **v** be vectors in an inner product space V, with **v** nonzero. Then the **projection of u onto v** is given by

$$\operatorname{proj}_{\mathbf{v}}\mathbf{u} = \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\,\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{v}\|^2}\,\mathbf{v} \tag{5}$$

Example 12

Determine $\operatorname{proj}_{\mathbf{v}}\mathbf{u}$ for the vectors **u**, **v**, and the inner product space given in Example 2.

Solution In Example 2 we showed that $\langle \mathbf{v}, \mathbf{u}\rangle = -3$. The inner product of **v** with itself is

$$\langle \mathbf{v}, \mathbf{v}\rangle = v_1^2 t_1 + v_2^2 t_2 + v_3^2 t_3 = (4)^2(2) + (-1)^2(3) + (1)^2(1) = 36$$

Therefore

$$\operatorname{proj}_{\mathbf{v}}\mathbf{u} = \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\,\mathbf{v} = -\frac{3}{36}\begin{bmatrix}4\\-1\\1\end{bmatrix} = -\frac{1}{12}\begin{bmatrix}4\\-1\\1\end{bmatrix} = \begin{bmatrix}-\frac{1}{3}\\ \frac{1}{12}\\ -\frac{1}{12}\end{bmatrix}$$

Our new definition for projection will be a disappointment if it does not have properties similar to those of the Euclidean space version of projection. Happily, everything carries over to inner product spaces with no significant changes.

THEOREM 10.6 ▶

Let **u** and **v** be vectors in an inner product space *V*, with **v** nonzero, and let *c* be a nonzero scalar. Then

(a) $\text{proj}_{\mathbf{v}}\mathbf{u}$ is in span{**v**}.

(b) $\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}$ is orthogonal to **v**.

(c) If **u** is in span{**v**}, then $\mathbf{u}=\text{proj}_{\mathbf{v}}\mathbf{u}$.

(d) $\text{proj}_{\mathbf{v}}\mathbf{u}=\text{proj}_{c\mathbf{v}}\mathbf{u}$.

▶ Theorem 10.6 generalizes Theorem 8.16 in Section 8.2.

Proof We give the proofs of parts (a) and (b) here and leave the proofs of parts (c) and (d) as an exercise. To prove (a), we note that since

$$\text{proj}_{\mathbf{v}}\mathbf{u}=\frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}$$

then $\text{proj}_{\mathbf{v}}\mathbf{u}$ is a scalar multiple of **v** and so is in span{**v**}.

For (b), applying of Definition 10.1 parts (b) and (c) gives us

$$\langle \mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u},\mathbf{v}\rangle=\langle \mathbf{u}, \mathbf{v}\rangle-\langle \text{proj}_{\mathbf{v}}\mathbf{u},\mathbf{v}\rangle$$

Substituting in the formula for projection and applying (c) of Definition 10.1, we have

$$\langle \mathbf{u},\mathbf{v}\rangle-\langle \text{proj}_{\mathbf{v}}\mathbf{u},\mathbf{v}\rangle=\langle \mathbf{u},\mathbf{v}\rangle-\left\langle \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\langle \mathbf{v}, \mathbf{u}\rangle}\mathbf{v}, \mathbf{u}\right\rangle =\langle \mathbf{u},\mathbf{v}\rangle-\frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\langle \mathbf{v}, \mathbf{v}\rangle=\langle \mathbf{u}, \mathbf{v}\rangle-\langle \mathbf{v}, \mathbf{u}\rangle=0$$

because $\langle \mathbf{u}, \mathbf{v}\rangle=\langle \mathbf{v}, \mathbf{u}\rangle$. Therefore $\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}$ is orthogonal to **v**. ■ ■

Combining (a) and (b) of Theorem 10.6 tells us that $\text{proj}_{\mathbf{v}}\mathbf{u}$ and $\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}$ are orthogonal (see Exercise 60), so that by the Pythagorean Theorem,

$$\|\mathbf{u}\|^2=\|\text{proj}_{\mathbf{v}}\mathbf{u}\|^2+\|\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}\|^2$$

Since $\|\mathbf{u}-\text{proj}_{\mathbf{v}}\mathbf{u}\|^2\geq 0$, it follows that

$$||projvu||\leq||u|| \qquad (6)$$

This inequality is useful for proving the next theorem.

THEOREM 10.7 ▶

(THE CAUCHY–SCHWARZ INEQUALITY) For all **u** and **v** in an inner product space *V*,

$$|\langle u,v\rangle|\leq||u||\ ||v|| \qquad (7)$$

▶ Theorem 10.7 generalizes Theorem 8.7 in Section 8.1.

Proof First, if either **u** = 0 or **v** = 0, then both sides of (7) are equal to 0 and we are done. So let's assume that both **u** and **v** are nonzero vectors. Then

$$||projvu||=|\ |\langle v,\ u\rangle||v||2\ v|\ |=|\langle v,\ u\rangle|||v||2\ ||v||=|\langle v,\ u\rangle|||v||$$

Combining this with (6) yields the inequality

$$|\langle v,\ u\rangle|||v||\leq||u||$$

Hence (7) holds, completing the proof. ■ ■

Example 13

Verify that the Cauchy–Schwarz inequality holds for the inner product given in Example 9 when applied to the matrices

$$A=[20-1112-2-10]\ \text{ and }\ B=[101-212011]$$

Solution We have

$$\langle A, B\rangle = tr(A^T B) = tr ([0 -1 2 -2 0 1 -5 2 3]) = 3 ||A|| = \langle A, A\rangle$$
$$= tr(A^T A) = tr ([9 3 0 3 2 2 0 2 5]) = 16 ||B|| = \langle B, B\rangle$$
$$= tr(B^T B) = tr ([5 -2 -3 -2 2 3 -3 3 6]) = 13$$

Since $3 < 16\ 13 \approx 14.422$, the Cauchy–Schwarz inequality is verified for this pair of matrices.

The Cauchy–Schwarz inequality makes it easy to prove a second important inequality.

THEOREM 10.8 ▶

(THE TRIANGLE INEQUALITY) For all **u** and **v** in an inner product space *V*,

$$||u+v|| \leq ||u|| + ||v|| \tag{8}$$

Proof We have

$$||u+v||^2 = \langle u+v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle = ||u||^2 + ||v||^2 + 2\langle u, v\rangle \leq ||u||^2 + ||v||^2 + 2||u||\,||v|| = (||u|| + ||v||)^2 \qquad \text{(By Cauchy-Schwarz inequality)}$$

Taking square roots on both sides yields (8) and completes the proof. ■

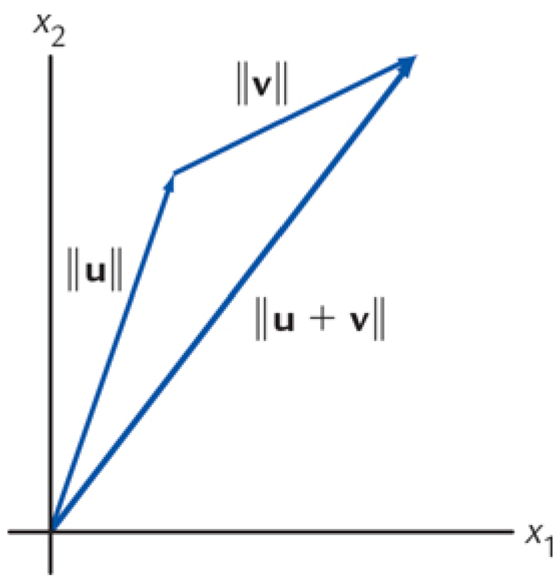


Figure 1 Graphical depiction of the triangle inequality,

$$||u+v||\le||u||+||v||$$

If we think of **u** and **v** placed tip-to-tail to form two sides of a triangle (see Figure 1), then **u** + **v** gives the third side. It is geometrically evident in $\mathbf{R}^2$ that the sum of the lengths of two sides of a triangle must be at least as great as the length of the third side. The triangle inequality tells us that the same is true in any inner product space.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Compute the indicated inner product.
   (a) ⟨u, v⟩ for **u** = [252], **v** = [130], and the inner product given in Example 2 with weights t1=2, t2=4, and t3=1.
   (b) ⟨p, q⟩ for $p(x) = 5x + 1$, $q(x) = -2x + 3$, and the inner product given in Example 3 with $x_0 = -1$, $x_1 = 1$, and $x_2 = 2$.
   (c) ⟨f, g⟩ for $f(x) = x + 2$, $g(x) = -x^2$, and the inner product given in Example 5.

(d) ⟨A, B⟩=tr(AT B) for A = [1−12−1], B = [5342].

2. Compute the norm with respect to the indicated inner product.

(a) ‖[41−2]‖ for the inner product given in Example 2.

(b) | |4x+3| | for the inner product given in Example 3 with $x_0 = -1$, $x_1 = 1$, and $x_2 = 3$.

(c) | |x2| | for the inner product given in Example 5.

(d) | |A| | for A = [2−113] and ⟨A, B⟩=tr(AT B).

3. Compute the indicated projection with respect to the given inner product.

(a) projuv for **u** = [13−2], **v** = [−120], and the inner product given in Example 2 with weights t1=2, t2=1, and t3=3.

(b) projpq for $p(x) = x + 2$, $q(x) = -x + 3$, and the inner product given in Example 3 with $x_0 = -1$, $x_1 = 0$, and $x_2 = 2$.

(c) projfg for f(x)=2(x), g(x)=−x3 and the inner product given in Example 5.

(d) projAB for A = [1−112], B = [2−10−2], and ⟨A, B⟩=tr(AT B).

4. Determine if the statement is true or false, and justify your answer. Here **u** and **v** are vectors in an inner product space *V*.

(a) If **u** and **v** are parallel, then projuv=u.

(b) ⟨−2u, 3v⟩=|−6| ⟨u, v⟩.

(c) ∫−22 f(x)g(x)ex dx is an inner product on C(−2, 2).

(d) If | |u| |=2| |v| |, then **u** = 2**v**.

EXERCISES

Exercises 1–8: Compute the indicated inner product.

1. ⟨u, v⟩ for **u** = [121], **v** = [342], and the inner product given in Example 2.

2. ⟨u, v⟩ for **u** = [252], **v** = [130], and the inner product given in Example 2 with weights t1=3, t2=1, and t3=4.

3. ⟨p, q⟩ for p(x)=3x+2, q(x)=−x+1, and the inner product given in Example 3 with x0=−1, x1=0, and x2=2.

4. ⟨p, q⟩ for $p(x)=x2+1$, $q(x)=2x-3$, and the inner product given in Example 3 with $x0=-1$, $x1=1$, $x2=2$, and $x3=5$.
5. ⟨f, g⟩ for $f(x)=x+3$, $g(x)=x2$, and the inner product given in Example 5.
6. ⟨f, g⟩ for $f(x)=x$, $g(x)=ex$, and the inner product given in Example 5.
7. ⟨A, B⟩=tr(AT B) for A = [2−134], B = [52−3−2].
8. ⟨A, B⟩=tr(AT B) for A = [7−326], B = [4205].
9. Suppose that **u** = [10−1] and **v** = [212] are orthogonal with respect to the inner product given in Example 2 with weights $t1=3$, $t2=1$, and $t3=a$. What are the possible value(s) of *a*?
10. Suppose that **u** = [322] and **v** = [5−12] are orthogonal with respect to the inner product given in Example 2 with weights $t1=1$, $t2=b$, and $t3=2$. What are the possible value(s) of *b*?
11. Suppose that $p(x)=x+2$, $q(x)=-3x+1$ are orthogonal with respect to the inner product given in Example 3 with $x0=-1$, $x1=a$, and $x2=2$. What are the possible value(s) of *a*?
12. Suppose that $p(x)=x2-3x-1$, $q(x)=x+2$ are orthogonal with respect to the inner product given in Example 3 with $x0=-2$, $x1=0$, $x2=1$, and $x3=a$. What are the possible value(s) of *a*?
13. Suppose that $f(x)=2x$ and $g(x)=x+b$. For what value(s) of *b* are *f* and *g* orthogonal with respect to the inner product in Example 5?
14. Suppose that $f(x)=x2$ and $g(x)=x+b$. For what value(s) of *b* are *f* and *g* orthogonal with respect to the inner product in Example 5?

Exercises 15–22: Compute the norm with respect to the indicated inner product.

15. ‖[1−32]‖ for the inner product given in Example 2.
16. ‖[20−5]‖ for the inner product given in Example 2 with weights $t1=1$, $t2=5$, and $t3=2$.
17. | |3x−5| | for the inner product given in Example 3 with $x0=-2$, $x1=1$, and $x2=4$.

18. $||-x2+x-4||$ for the inner product given in Example 3 with x0=0, x1=3, x2=2, and x3=6.
19. $||x3||$ for the inner product given in Example 5.
20. $||xe2x||$ for the inner product given in Example 5.
21. $||A||$ for A = [3−120] and ⟨A, B⟩=tr(AT B).
22. $||A||$ for A = [2301−3−1252] and ⟨A, B⟩=tr(AT B).

Exercises 23–30: Compute the indicated projection with respect to the given inner product.

23. projuv for **u** = [121], **v** = [342], and the inner product given in Example 2.
24. projuv for **u** = [252], **v** = [130], and the inner product given in Example 2 with weights t1=3, t2=1, and t3=4.
25. projpq for p(x)=3x+2, q(x)=−x+1, and the inner product given in Example 3 with x0=−1, x1=0, and x2=2.
26. projpq for p(x)=x2+1, q(x)=2x−3, and the inner product given in Example 3 with x0=−1, x1=1, x2=2, and x3=5.
27. projfg for f(x)=x, g(x)=x2, and the inner product given in Example 5.
28. projfg for f(x)=sin(x), g(x)=1−x2, and the inner product given in Example 5.
29. $\text{proj}_A B$ for A=[2−110],B=[230−2], and ⟨A,B⟩=tr(ATB).
30. projAB for A = [34−1−3], B = [15−21], and ⟨A, B⟩=tr(AT B).

FIND AN EXAMPLE Exercises 31–40: Find an example that meets the given specifications.

31. An orthogonal basis for $\mathbf{R}^2$ with respect to the inner product ⟨u, v⟩=3u1v1+2u2v2.
32. An orthogonal basis for $\mathbf{p}^1$ with respect to the inner product ⟨p, q⟩=∫01p(x)q(x) dx.
33. A vector **u** such that $||u||$=2 and **u** is orthogonal to [23] with respect to a weighted dot product of the form ⟨u, v⟩ =t1u1v1+t2u2v2.

34. An inner product on $\mathbf{p}^2$ such that $\|p\|=3$ for $p(x)=x^2-4x+3$.

35. A nonidentity matrix *A* such that $\langle u, v\rangle=u^TAv$ is an inner product on $\mathbf{R}^3$.

36. A nonidentity matrix *A* such that $\langle u, v\rangle=u^TAv$ is *not* an inner product on $\mathbf{R}^3$.

37. An inner product of your creation on $\mathbf{p}^2$.

38. An inner product of your own creation on $C[0, 1]$.

39. A function $\langle p, q\rangle$ that is *almost* an inner product on $\mathbf{p}^2$: It satisfies (a)–(c) of Definition 10.1, but not (d).

40. A function $\langle A, B\rangle$ that is a poor attempt at an inner product on $\mathbf{R}^{3\times3}$: It satisfies (a) of Definition 10.1, but not (b)–(d).

TRUE OR FALSE Exercises 41–46: Determine if the statement is true or false, and justify your answer. Here **u** and **v** are vectors in an inner product space *V*.

41.

(a) If $\langle u, v\rangle=3$, then $\langle 2u, -4v\rangle=-24$.

(b) $\|u+v\|^2=\|u\|^2+\|v\|^2$ for all **u** and **v** in *V*.

42.

(a) If **u** and **v** are orthogonal, and **v** and **w** are orthogonal, then **u** and **w** are orthogonal.

(b) If {**u**, **v**, **w**} is an orthogonal set, then $\|u+v+w\|^2=\|u\|^2+\|v\|^2+\|w\|^2$.

43.

(a) If **u** and **v** are orthogonal with $\|u\|=3$ and $\|v\|=4$, then $\|u+v\|=5$.

(b) If $u=cv\neq0$ for a scalar *c*, then $u=\text{proj}_u v$.

44.

(a) If {**u**, **v**} is an orthogonal set and c_1 and c_2 are scalars, then $\{c_1\mathbf{u}, c_2\mathbf{v}\}$ is also an orthogonal set.

(b) $-\|u\|\,\|v\|\leq\langle u, v\rangle$ for all **u** and **v** in *V*.

45.

(a) $\|u-v\|\leq\|u\|-\|v\|$ for all **u** and **v** in *V*.

(b) $\langle f, g\rangle=\int_{-1}^{1}xf(x)g(x)\,dx$ is an inner product on $C[-1, 1]$.

46.

(a) $\langle p, q\rangle = p(x_0)q(x_0)+p(x_1)q(x_1)$ is an inner product on P_2 when $x_0 \neq x_1$.

(b) If $T: V \to R$ is a linear transformation, then $\langle u, v\rangle = T(u) \cdot T(v)$ is an inner product.

47. Complete Example 1. Prove that the weighted dot product on $\mathbf{R}^n$ given by

$$\langle u, v\rangle = t_1u_1v_1+t_2u_2v_2+\cdots+t_nu_nv_n$$

where $t_1, t_2, \ldots, t_n$ are all positive, is an inner product

48. Complete Example 3. Show that properties (a)–(c) of Definition 10.1 are true for the inner product

$$\langle p, q\rangle = p(x_0)q(x_0)+\cdots+p(x_n)q(x_n)$$

49. A weighted version of the inner product given in Example 3 is defined as follows: For $p(x)$ and $q(x)$ in $\mathbf{P}^n$ and distinct real numbers $x_0, x_1, \ldots, x_n$, let

$$\langle p, q\rangle = t(x_0)p(x_0)q(x_0)+\cdots+t(x_n)p(x_n)q(x_n)$$

where $t(x)$ takes positive values on $x_0, \ldots, x_n$. Show that $\langle p, q\rangle$ is an inner product on $\mathbf{p}^n$.

50. Let f and g be continuous functions in $C[-1, 1]$. Show that the weighted version of (3) given by

$$\langle f, g\rangle = \int_{-1}^{1} t(x)f(x)g(x)\,dx$$

where $t(x)>0$ is continuous for all x in $[-1, 1]$, defines an inner product on $C[-1, 1]$.

51. Let f and g be continuous functions in $C[-\pi, \pi]$. Show that

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx$$

defines an inner product on $C[-\pi, \pi]$.

52. Complete the proof of Theorem 10.6, by showing that parts (c) and (d) are true.

53. Prove that $\langle A, B\rangle = \mathrm{tr}(A^T B)$ is an inner product on $R^{3\times 3}$.

54. For nonzero vectors **u** and **v**, show that there is equality in the Cauchy–Schwarz inequality exactly when $\mathbf{u}=c\mathbf{v}$ for some scalar *c.* (Hint: The key lies with the inequality (6).)

Exercises 55–64: **u**, **v**, and **w** (and their subscripted associates) are vectors in an inner product space *V.*

55. Prove that $\|c\mathbf{v}\|=|c|\,\|\mathbf{v}\|$ for every **v** and scalar *c.*

56. Prove that $\|\mathbf{v}\|\geq 0$ for every **v** in *V*, with equality holding only for $\mathbf{v}=\mathbf{0}$.

57. Prove that

$$\langle c_1\mathbf{u}_1+\cdots+c_k\mathbf{u}_k,\ \mathbf{w}\rangle=c_1\langle \mathbf{u}_1,\ \mathbf{w}\rangle+\cdots+c_k\langle \mathbf{u}_k,\ \mathbf{w}\rangle$$

58. Prove that the zero vector **0** is orthogonal to all vectors in *V.*

59. Prove that if $\mathbf{v}\neq\mathbf{0}$, then $\mathbf{w}=\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ satisfies $\|\mathbf{w}\|=1$.

60. Prove that $\operatorname{proj}_{\mathbf{v}}\mathbf{u}$ and $\mathbf{u}-\operatorname{proj}_{\mathbf{v}}\mathbf{u}$ are orthogonal.

61. For a fixed **v** in *V*, define $T_{\mathbf{v}}: V\rightarrow \mathbb{R}$ by $T_{\mathbf{v}}(\mathbf{u})=\langle \mathbf{u},\ \mathbf{v}\rangle$. Show that $T_{\mathbf{v}}$ is a linear transformation.

62. For a fixed **v** in *V*, define $T_{\mathbf{v}}: V\rightarrow V$ by $T_{\mathbf{v}}(\mathbf{u})=\operatorname{proj}_{\mathbf{v}}\mathbf{u}$. Show that $T_{\mathbf{v}}$ is a linear transformation.

63. Prove that if **u** and **v** are orthogonal, then the distance between **u** and **v** is $\|\mathbf{u}\|^2+\|\mathbf{v}\|^2$.

64. Prove that

$$\|\mathbf{u}+\mathbf{v}\|^2+\|\mathbf{u}-\mathbf{v}\|^2=2(\|\mathbf{u}\|^2+\|\mathbf{v}\|^2)$$

Exercises 65–68: If *S* is a subspace of a finite-dimensional inner product space *V*, a vector **v** is **orthogonal** to *S* if $\langle \mathbf{v},\ \mathbf{s}\rangle=0$ for every vector $\mathbf{s}$ in *S*. The set of all such vectors **v** is called the **orthogonal complement** of *S* and is denoted by $S^{\perp}$. Prove that the statement involving $S^{\perp}$ is true.

65. If *S* is a subspace, then so is $S^{\perp}$.

66. If *S* is a subspace, then $(S^{\perp})^{\perp}=S$.

67. If **s** is in *S* and $\mathbf{s}^{\perp}$ is in $S^{\perp}$, then

$$\|\mathbf{s}\pm\mathbf{s}^{\perp}\|^2=\|\mathbf{s}\|^2+\|\mathbf{s}^{\perp}\|^2.$$

68. If S is a subspace, then $S \cap S^{\perp} = \{0\}$.

10.2 The Gram–Schmidt Process Revisited

Our main goal for this section is to develop a version of the Gram–Schmidt process for an inner product space. Before we can do that, we need to carry over some concepts from Euclidean space, starting with the definition of an orthonormal set.

▶ Recall that the Gram–Schmidt process allows us to transform a basis into an orthogonal basis. See Section 8.2 for details.

DEFINITION 10.9 ▶

Orthogonal Set

The vectors {v1,… , vk} in an inner product space *V* form an **orthogonal set** if ⟨vi, vj⟩=0 for i≠j.

Example 1

Let V be the inner product space consisting of vectors in $\mathbf{R}^3$ and the weighted dot product ⟨u, v⟩=2u1v1+3u2v2+u3v3. Show that the vectors

v1=[451], v2=[−32−6],v3=[16−7−23]

form an orthogonal set.

Solution The inner products of each pair are

$$\langle v_1, v_2\rangle=(2)(4)(-3)+(3)(5)(2)+(1)(-6)=0$$
$$\langle v_1, v_3\rangle=(2)(4)(16)+(3)(5)(-7)+(1)(-23)=0$$
$$\langle v_2, v_3\rangle=(2)(-3)(16)+(3)(2)(-7)+(-6)(-23)=0$$

and therefore the set $\{v_1, v_2, v_3\}$ is orthogonal.

Example 2

Let $V=C[-\pi, \pi]$ be the inner product space of continuous functions on $[-\pi, \pi]$ with respect to the inner product

$$\langle f, g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx$$

Show that the set $\{1, \cos(x), \sin(x)\}$ is orthogonal.

Solution As in Example 1, we need to compute the inner products of the three possible pairs of vectors:

$$\langle 1, \cos(x)\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}(1)(\cos(x))\,dx=\frac{1}{\pi}(\sin(\pi)-\sin(-\pi))=0$$
$$\langle 1, \sin(x)\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}(1)(\sin(x))\,dx=-\frac{1}{\pi}(\cos(\pi)-\cos(-\pi))=0$$
$$\langle \cos(x), \sin(x)\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}(\cos(x))(\sin(x))\,dx=\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{1}{2}\sin(2x)\,dx=-\frac{1}{4\pi}(\cos(2\pi)-\cos(-2\pi))=0$$

▶ Here we use the identity $2\cos(x)\sin(x)=\sin(2x)$.

Thus our set of vectors is orthogonal.

When studying orthogonality in Euclidean space, we saw that an orthogonal set of nonzero vectors must be linearly independent. The same is true of such a set in an inner product space.

THEOREM 10.10 ▶

Orthogonal Basis

Let $\mathcal{V}$={v1,… ,vm} be an orthogonal set of nonzero vectors in an inner product space *V*. Then $\mathcal{V}$ is linearly independent.

▶ Theorem 10.10 generalizes Theorem 8.13 in Section 8.1.

The proof is left as an exercise. An interesting consequence of Theorem 10.10 is that if a given set of nonzero vectors is orthogonal with respect to *just one* inner product, then the set must be linearly independent. We also can flip this around: If the set is linearly dependent, then it cannot be orthogonal with respect to *any* inner product, no matter how cleverly selected.

Orthogonal Basis

An orthogonal set of vectors that forms a basis for an inner product space is called an **orthogonal basis**.

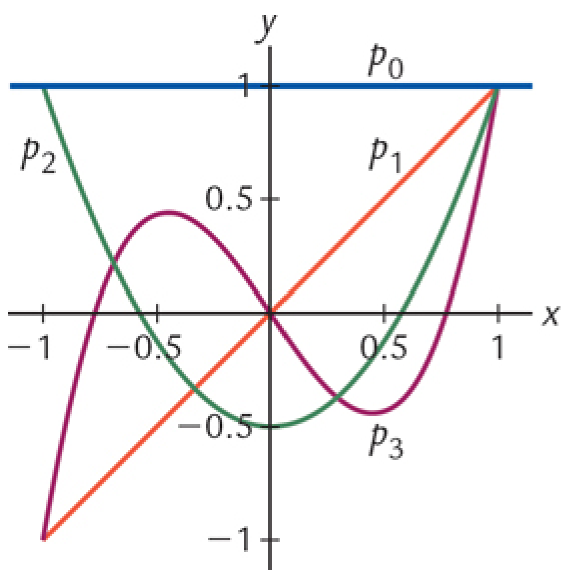


Figure 1 Graphs of the first four Legendre polynomials.

Example 3

An important class of orthogonal polynomials are the *Legendre polynomials*. There is an infinite sequence of them—the first four are

$$p_0(x)=1,\ p_1(x)=x,\ p_2(x)=\frac{1}{2}(3x^2-1),\ p_3(x)=\frac{1}{2}(5x^3-3x)$$

(See Figure 1.) Show that this set of polynomials is a basis for $\mathbf{P}^3$ by showing they are an orthogonal set with respect to the inner product

$$\langle p, q\rangle=\int_{-1}^{1} p(x)q(x)\,dx$$

Solution The inner products of each pair of polynomials are

$$\begin{aligned}
\langle p_0, p_1\rangle &= \int_{-1}^{1} x\,dx=0 & \langle p_1, p_2\rangle &= \int_{-1}^{1} \frac{1}{2}(3x^3-x)\,dx=0\\
\langle p_0, p_2\rangle &= \int_{-1}^{1} \frac{1}{2}(3x^2-1)\,dx=0 & \langle p_1, p_3\rangle &= \int_{-1}^{1} \frac{1}{2}(5x^4-3x^2)\,dx=0\\
\langle p_0, p_3\rangle &= \int_{-1}^{1} \frac{1}{2}(5x^3-3x)\,dx=0 & \langle p_2, p_3\rangle &= \int_{-1}^{1} \frac{1}{4}(15x^5-14x^3+3x)\,dx=0
\end{aligned}$$

Therefore the set $\{p_0(x), p_1(x), p_2(x), p_3(x)\}$ is orthogonal and hence by Theorem 10.10 is linearly independent. Since $\dim(\mathbf{P}^3)=4$, it follows that this set is a basis for $\mathbf{P}^3$.

An orthogonal basis is handy because the inner product can be used to easily determine how to express vectors as a linear combination of basis vectors. The next theorem shows how this is accomplished.

THEOREM 10.11

Let $\mathcal{V}=\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthogonal basis for an inner product space V. Then any vector $\mathbf{v}$ in V can be written as

$$\mathbf{v} = s_1\mathbf{v}_1 + \cdots + s_k\mathbf{v}_k$$

where $s_i=\frac{\langle \mathbf{v}_i,\mathbf{v}\rangle}{\langle \mathbf{v}_i,\mathbf{v}_i\rangle}=\frac{\langle \mathbf{v}_i,\mathbf{v}\rangle}{\|\mathbf{v}_i\|^2}$ for $i=1,\ldots,k$.

Theorem 10.11 generalizes Theorem 8.14 in Section 8.1.

The proof is left as an exercise.

Example 4

Use Theorem 10.11 to write $p(x)=10x^3+3x^2-11x+2$ as a linear combination of the Legendre polynomials in Example 3.

Solution We know that $\{p_0(x), p_1(x), p_2(x), p_3(x)\}$ is a basis for $\mathbf{P}^3$, so $p(x)$ must be a linear combination of these vectors. Since the basis is orthogonal, we can apply Theorem 10.11 to find the scalars. The squares of the norms of p_0, p_1, p_2, and p_3 are

$$\begin{aligned}
\langle p_0, p_0\rangle &= \int_{-1}^{1}(p_0(x))^2\,dx=\int_{-1}^{1}(1)^2\,dx=2\\
\langle p_1, p_1\rangle &= \int_{-1}^{1}(p_1(x))^2\,dx=\int_{-1}^{1}(x)^2\,dx=2/3\\
\langle p_2, p_2\rangle &= \int_{-1}^{1}(p_2(x))^2\,dx=\int_{-1}^{1}\left(\tfrac{1}{2}(3x^2-1)\right)^2\,dx=2/5\\
\langle p_3, p_3\rangle &= \int_{-1}^{1}(p_3(x))^2\,dx=\int_{-1}^{1}\left(\tfrac{1}{2}(5x^3-3x)\right)^2\,dx=2/7
\end{aligned}$$

It is left to the reader to verify the other required inner products, namely,

$$\langle p_0, p\rangle=6,\quad \langle p_1, p\rangle=-10/3,\quad \langle p_2, p\rangle=4/5,\quad \langle p_3, p\rangle=8/7$$

By Theorem 10.11 we have

$$\begin{aligned}
p(x)&=\frac{\langle p_0, p\rangle}{\langle p_0, p_0\rangle}p_0(x)+\frac{\langle p_1, p\rangle}{\langle p_1, p_1\rangle}p_1(x)+\frac{\langle p_2, p\rangle}{\langle p_2, p_2\rangle}p_2(x)+\frac{\langle p_3, p\rangle}{\langle p_3, p_3\rangle}p_3(x)\\
&=\frac{6}{2}p_0(x)+\frac{-10/3}{2/3}p_1(x)+\frac{4/5}{2/5}p_2(x)+\frac{8/7}{2/7}p_3(x)\\
&=3p_0(x)-5p_1(x)+2p_2(x)+4p_3(x)
\end{aligned}$$

We can check our calculations by computing

$$\begin{aligned}
3p_0(x)-5p_1(x)+2p_2(x)+4p_3(x)&=3(1)-5(x)+2\left(\tfrac{1}{2}(3x^2-1)\right)+4\left(\tfrac{1}{2}(5x^3-3x)\right)\\
&=10x^3+3x^2-11x+2=p(x)
\end{aligned}$$

Orthonormal Sets

Unit Vector, Orthonormal Basis

The formula for the scalars s_i given in Theorem 10.11 is simplified if $\|\mathbf{v}_i\| = 1$. Vectors with norm equal to 1 are called **unit vectors**. An orthogonal basis made up of unit vectors is called an orthonormal basis. When we have an **orthonormal basis**, Theorem 10.11 can be simplified to the following form.

THEOREM 10.12 ▸

Let $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthonormal basis for an inner product space V. Then any vector $\mathbf{v}$ in V can be written as

$$\mathbf{v} = \langle \mathbf{v}_1, \mathbf{v}\rangle \mathbf{v}_1 + \cdots + \langle \mathbf{v}_k, \mathbf{v}\rangle \mathbf{v}_k$$

The proof of Theorem 10.12 is left as an exercise.

Example 5

Let S be the subspace of $C[-\pi, \pi]$ with basis $\{1, \cos(x), \sin(x)\}$. Convert this to an orthonormal basis with respect to the inner product

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx$$

Then express $f(x) = \sin^2(x/2)$ (which is in S) as a linear combination of the orthonormal basis functions.

▶ Dividing a vector by its norm to create a unit vector is called **normalizing** the vector.

Solution In Example 2 we showed that the basis functions are orthogonal, so all that remains is to scale each so that they have norm 1. For any norm, if $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{w} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$, then $\|\mathbf{w}\| = 1$ (see Exercise 59 in Section 10.1). Hence we can convert an orthogonal basis to an orthonormal basis by dividing each vector by its norm. For our basis vectors we have

$$\|1\|^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} 1^2\,dx = 2 \Rightarrow g_1(x) = \frac{1}{\sqrt{2}}\,1 = \frac{1}{\sqrt{2}}$$
$$\|\cos(x)\|^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2(x)\,dx = 1 \Rightarrow g_2(x) = \frac{1}{1}\cos(x) = \cos(x)$$
$$\|\sin(x)\|^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2(x)\,dx = 1 \Rightarrow g_3(x) = \frac{1}{1}\sin(x) = \sin(x)$$

▶ The details of evaluating the integrals in this example are left to the reader.

Then $\{g_1(x), g_2(x), g_3(x)\}$ is an orthonormal basis for *S*. To express $f(x) = \sin^2(x/2)$ as a linear combination of the basis functions, we compute the inner products

$$\langle g_1, f\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{1}{\sqrt{2}}\sin^2(x/2)\,dx = \frac{1}{\sqrt{2}}$$
$$\langle g_2, f\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(x)\sin^2(x/2)\,dx = -\frac{1}{2}$$
$$\langle g_3, f\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(x)\sin^2(x/2)\,dx = 0$$

Therefore, by Theorem, 10.12

$$\sin^2(x/2) = \frac{1}{\sqrt{2}}g_1(x) - \frac{1}{2}g_2(x) + 0g_3(x) = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}\right) - \frac{1}{2}\cos(x) = \frac{1-\cos(x)}{2}$$

which agrees with the half-angle formula for sine.

Projections onto Subspaces

Theorem 10.11 provides a formula for expressing a vector **v** in a vector space *V* as a linear combination of orthogonal basis vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$. Similar to our approach in Section 8.2, we use the formula from Theorem 10.11 to serve as a guide for the formula for projecting a vector onto a subspace.

DEFINITION 10.13 ▶

Projection onto a Subspace

Let S be a subspace of an inner product space V, and suppose that S has orthogonal basis $\{v_1, \ldots, v_k\}$. Then the **projection of v onto** S is given by

$$\text{proj}_S v = \frac{\langle v_1, v\rangle}{\|v_1\|^2} v_1 + \frac{\langle v_2, v\rangle}{\|v_2\|^2} v_2 + \cdots + \frac{\langle v_k, v\rangle}{\|v_k\|^2} v_k \qquad (1)$$

As in Euclidean space, the following are true about projections onto inner product subspaces:

- If $S = \text{span}\{v_1\}$ is a one-dimensional subspace, then (1) reduces to the formula for $\text{proj}_{v_1} v$.
- The basis $\{v_1, \ldots, v_k\}$ for S must be orthogonal in order to apply the formula for $\text{proj}_S v$. If the basis is orthonormal, then (1) reduces to

$$\text{proj}_S v = \langle v_1, v\rangle v_1 + \langle v_2, v\rangle v_2 + \cdots + \langle v_k, v\rangle v_k \qquad (2)$$

- The vector $\text{proj}_S v$ does not depend on the choice of orthogonal basis for S.

THEOREM 10.14 ▶

Let S be a nonzero finite-dimensional subspace of an inner product space V, and **v** a vector in V. Then

(a) $\text{proj}_S v$ is in S.

(b) $v - \text{proj}_S v$ is orthogonal to S.

(c) If **v** is in S, then $v = \text{proj}_S v$.

(d) $\text{proj}_S v$ is independent of the choice of orthogonal basis for S.

▶ Theorem 10.14 generalizes Theorem 8.18 in Section 8.2.

The proof of Theorem 10.14 is similar to that of Theorem 8.18 in Section 8.2 and is left as an exercise.

Example 6

Let $S=\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, where

$$\mathbf{v}_1=\begin{bmatrix}4\\5\\1\end{bmatrix}, \mathbf{v}_2=\begin{bmatrix}-3\\2\\-6\end{bmatrix}, \mathbf{v}=\begin{bmatrix}3\\25\\33\end{bmatrix}$$

Find $\text{proj}_S\mathbf{v}$ using the inner product $\langle \mathbf{u}, \mathbf{v}\rangle=2u_1v_1+3u_2v_2+u_3v_3$ from Example 1.

Solution In Example 1 we showed that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal with respect to our inner product, so all that remains is to apply the formula (1). To do so, we need

$$\begin{aligned}\langle \mathbf{v}_1, \mathbf{v}\rangle&=(2)(4)(3)+(3)(5)(25)+(1)(1)(33)=432\\ \langle \mathbf{v}_2, \mathbf{v}\rangle&=(2)(-3)(3)+(3)(2)(25)+(1)(-6)(33)=-66\\ \|\mathbf{v}_1\|^2&=\langle \mathbf{v}_1, \mathbf{v}_1\rangle=(2)(4)^2+(3)(5)^2+(1)(1)^2=108\\ \|\mathbf{v}_2\|^2&=\langle \mathbf{v}_2, \mathbf{v}_2\rangle=(2)(-3)^2+(3)(2)^2+(1)(-6)^2=66\end{aligned}$$

Therefore

$$\text{proj}_S\mathbf{v}=\frac{432}{108}\mathbf{v}_1+\frac{-66}{66}\mathbf{v}_2=4\mathbf{v}_1-\mathbf{v}_2=4\begin{bmatrix}4\\5\\1\end{bmatrix}-\begin{bmatrix}-3\\2\\-6\end{bmatrix}=\begin{bmatrix}19\\18\\10\end{bmatrix}$$

Example 7

Let $S=\text{span}\{1/\sqrt{2}, \cos(x), \sin(x)\}$ be a subspace of $C[-\pi, \pi]$. Find the projection of $h(x)=x^2$ onto S with respect to the inner product

$$\langle f, g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)g(x)\,dx$$

Solution In Example 5 we showed the basis of *S* is orthonormal. Hence we can compute the projection using the simplified formula (2), which only requires the inner products

⟨1/2, h⟩=1π∫−ππ12·x2 dx=2π23⟨cos(x), h⟩
=1π∫−ππcos(x)·x2 dx=−4⟨sin(x), h⟩=1π∫−ππsin(x)·x2 dx=0

Therefore the projection is given by

projSh=2π23 (12)−4 cos(x)=π23−4 cos(x)

A graph of h(x)=x2 together with projSh is given in Figure 2. The graphs show the projection is a good approximation to the function.

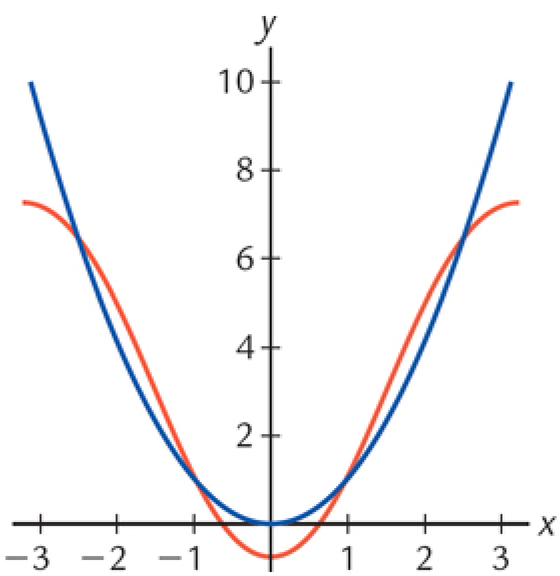


Figure 2 h(x)=x2 (blue) and projSh (red) from Example 7.

The Gram–Schmidt Process

We are now ready to recast the Gram–Schmidt process in the setting of an inner product space. The goal of the Gram–Schmidt process is the same here as in Euclidean space, to convert a basis into an orthogonal basis. (Which can then be made into an orthonormal basis by normalizing each vector.)

The basic procedure for implementing the Gram–Schmidt process in an inner product space is exactly the same as in Euclidean space. The only computational change is that we replace the dot products with inner products when computing projections.

THEOREM 10.15 ▶

(THE GRAM–SCHMIDT PROCESS) Let *S* be a subspace with basis $\{s_1, s_2, \ldots, s_k\}$. Define $v_1, v_2, \ldots, v_k$, in order, by

$$\begin{aligned} v_1 &= s_1 \\ v_2 &= s_2 - \mathrm{proj}_{v_1} s_2 \\ v_3 &= s_3 - \mathrm{proj}_{v_1} s_3 - \mathrm{proj}_{v_2} s_3 \\ v_4 &= s_4 - \mathrm{proj}_{v_1} s_4 - \mathrm{proj}_{v_2} s_4 - \mathrm{proj}_{v_3} s_4 \\ &\vdots \\ v_k &= s_k - \mathrm{proj}_{v_1} s_k - \mathrm{proj}_{v_2} s_k - \cdots - \mathrm{proj}_{v_{k-1}} s_k \end{aligned}$$

Then $\{v_1, v_2, \ldots, v_k\}$ is an orthogonal basis for *S*.

The proof is left as an exercise. After using Gram–Schmidt to find an orthogonal basis $\{v_1, \ldots, v_k\}$, we can find an orthonormal basis $\{w_1, \ldots, w_k\}$ by setting

$$w_i = \frac{1}{\|v_i\|} v_i$$

for each $i = 1, \ldots, k$.

Example 8

Starting with the vectors

$$s_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad s_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad s_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

implement the Gram–Schmidt process to find basis for $\mathbf{R}^3$ that is orthogonal with respect to the inner product $\langle u, v \rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3$ from Example 1.

Solution The first step requires no computation: $v_1=s_1$. For the second vector, we have

$$v_2=s_2-\text{proj}_{v_1}s_2=s_2-\frac{\langle v_1, s_2\rangle}{\langle v_1, v_1\rangle} v_1=[110]-\frac{5}{6}[111]=[1/6\ 1/6\ -5/6]$$

Since any multiple of an orthogonal vector is still orthogonal, we multiply v_2 by 6 to clear the fractions and make future computations easier. This gives us

$$v_2=[1\ 1\ -5]$$

Finally, for v_3 we have

$$v_3=s_3-\text{proj}_{v_1}s_3-\text{proj}_{v_2}s_3 =s_3-\frac{\langle v_1, s_3\rangle}{\langle v_1, v_1\rangle} v_1-\frac{\langle v_2, s_3\rangle}{\langle v_2, v_2\rangle} v_2 =[100]-\frac{2}{6}[111]-\frac{2}{30}[1\ 1\ -5]=[3/5\ -2/5\ 0]$$

For the sake of consistency, we again clear fractions by multiplying v_3 by 5. This leaves us with the orthogonal basis

$$\{[111] , [1\ 1\ -5] , [3\ -2\ 0]\}$$

Example 9

The monomials $\{1, x, x^2, x^3\}$ give a basis for $\mathbf{P}^3$. Apply the Gram–Schmidt process to find a basis that is orthonormal with respect to the inner product

$$\langle p, q\rangle=\int_{-1}^{1}p(x)q(x)\, dx$$

Solution We are asked for an orthonormal basis, but we start by finding an orthogonal basis, and then normalize at the end. Let $p_1(x)=1$, $p_2(x)=x$, $p_3(x)=x^2$, and $p_4(x)=x^3$. The first basis vector is $q_1(x)=p_1(x)=1$. For the second basis vector, we have

$$q_2(x)=p_2(x)-\text{proj}_{q_1} p_2=x-\frac{\langle q_1, p_2\rangle}{\langle q_1, q_1\rangle} q_1=x-\frac{0}{2/3} (1)=x$$

We have q2(x)=p2(x) because p2(x) is orthogonal to p1(x), which is why projq1p2=0. Proceeding to q3(x),

q3(x)=p3(x)−projq1p3−projq2p3 =p3(x)−⟨q1, p3⟩⟨q1, q1⟩ q1−⟨q2, p3⟩⟨q2, q2⟩ q2 =x2−2/32 (1)−02/3 (x)=x2−13=13 (3x2−1)

For the last polynomial, q4(x), we have

q4(x)=p4(x)−projq1p4−projq2p4−projq3p4 =p4(x)−⟨q1, p4⟩⟨q1, q1⟩ q1−⟨q2, p4⟩⟨q2, q2⟩ q2−⟨q3, p4⟩⟨q3, q3⟩ q3 =x3−02 (1)−2/52/3 (x)−08/45 (13 (3x2−1)) =x3−35 x=15 (5x3−3x)

The last step is to normalize each polynomial to produce an orthonormal basis. While implementing the Gram–Schmidt process, we computed

| |q1| |2=⟨q1, q1⟩=2, | |q2| |2=⟨q2, q2⟩=23, | |q3| |2=⟨q3, q3⟩=845

We also have

| |q4| |2=⟨q4, q4⟩=∫−11(15(5x3−3x))2 dx=8175

Therefore the orthonormal polynomials are

r1(x)=1||q1|| q1(x)=12r2(x)=1||q2|| q2(x)=32xr3(x)=1||q3|| q3(x)=458⋅13 (3x2−1)=58 (3x2−1)r4(x)=1||q4|| q4(x)=1758⋅15 (5x3−3x)=78 (5x3−3x)

The orthonormal basis is

{12, 32x, 58 (3x2−1), 78 (5x3−3x)}

Example 10

Use the orthonormal basis found in Example 9 to find the projection of *f* (*x*) = sin(*x*) + cos(*x*) onto P3.

Solution Since we have an orthonormal basis {r1(x), r2(x), r3(x), r4(x)}, we can apply (2) to find projp, f. The required inner products are

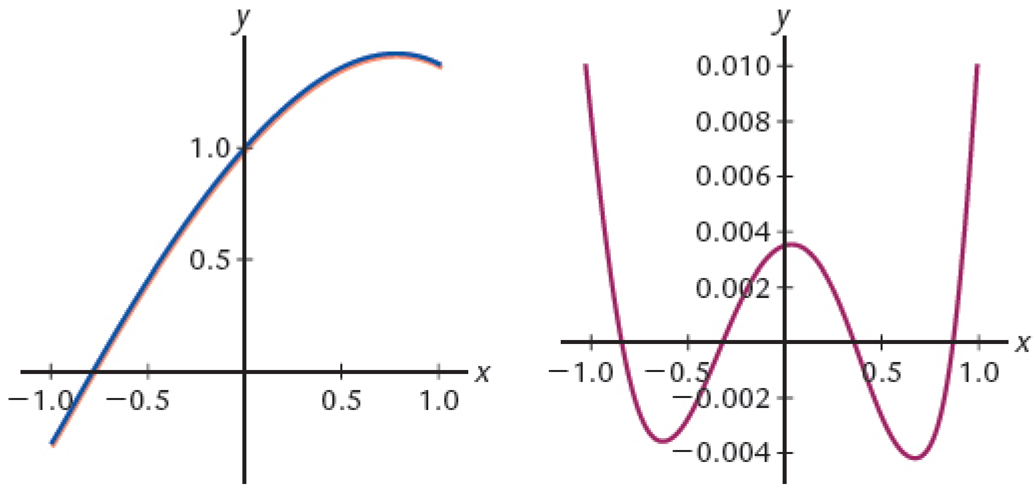
⟨r1, f⟩=∫−1112 (sin(x)+cos(x)) dx≈1.1900⟨r2, f⟩
=∫−1132 x(sin(x)+cos(x)) dx≈0.7377⟨r3, f⟩=∫−1158 (3x2−1)
(sin(x)+cos(x)) dx≈−0.1962⟨r4, f⟩=∫−1178 (5x3−3x)
(sin(x)+cos(x)) dx≈−0.0337

▶ The exact values of these integrals are somewhat complicated, so decimal approximations are reported.

Hence we have

projp, f≈1.19 12+0.737732 x−0.196258 (3x2−1)−0.033778 (5x3−3x
) ≈−0.1576x3−0.4653x2+0.9981x+0.9966

The graph of f(x)=sin(x)+cos(x) and projp, f are shown in Figure 3(a). The two graphs are virtually indistinguishable because the difference between f(x) and projP3f on [−1, 1] is quite small. The graph of the difference f(x)−projP3f in Figure 3(b) is more revealing. Among the polynomials of degree 3, the projection provides an excellent approximation to f(x)=sin(x)+cos(x).



(a) Graphs of $f(x)$ and $\mathrm{proj}_{P^3} f$ (b) Graph of $f(x) - \mathrm{proj}_{P^3} f$

Figure 3 f(x) and projP3f.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Determine the values of *a* (if any) that will make the given set of vectors orthogonal in R3 with respect to the weighted dot product with t1=1, t2=2, and t3=3. If possible, normalize the vectors to make the set orthonormal.

   {[2−24] , [2a0] , [−221]}

2. Determine the values of *a* (if any) that will make the given set of vectors orthogonal in $\mathbf{P}^2$ with respect to the inner product

   ⟨p, q⟩=p(0)q(0)+p(1)q(1)+p(2)q(2)

   If possible, normalize the vectors to make the set orthonormal.

   {x2+ax+1, −x2−2, x−3}

3. Let f(x)=x2. Find projSf for the inner product and subspace *S* in Example 7.

4. Find projSf for f(x)=e−x, where S=span{1, x} and the inner product is

⟨f, g⟩=∫02f(x)g(x) dx

5. Use the Gram–Schmidt process to convert the given set of vectors to an orthogonal basis for the span of the vectors with respect to the inner product.

(a) The set {[−110] , [011] , [10−1]} with respect to the inner product given in Example 1.

(b) The set {x2+1, −x2, x−1} with respect to the inner product given in Exercise 2.

6. Suppose *V* is an inner product space with dim(*V*) = *m*, and that *S* = {$\mathbf{v}_1$, … , $\mathbf{v}_k$} is an orthogonal subset of *V.* If *k* < *m,* prove that vectors can be added to *S* to form an orthogonal basis for *V.*

EXERCISES

1. Convert the set {v1, v2, v3} from Example 1 into an orthonormal set with respect to the inner product ⟨u, v⟩ =2u1v1+3u2v2+u3v3.

2. Verify that the set

{x+1, −9x+5, 6x2−6x+1}

is orthogonal with respect to the inner product

⟨p, q⟩=∫01p(x)q(x) dx

and then make the set orthonormal.

Exercises 3–4: Determine the values of *a* (if any) that will make the given set of vectors orthogonal in R3 with respect to the weighted dot product with $t_1=2$, $t_2=3$, and $t_3=1$. If possible, normalize the vectors to make the set orthonormal.

3. $\{[2\ -1\ 1], [2\ 1\ a], [1\ 2\ 2]\}$
4. $\{[1\ 1\ a], [-5\ 1\ 7], [3\ a\ 6]\}$

Exercises 5–6: Determine the values of *a* (if any) that will make the given set of vectors orthogonal in P_2 with respect to the inner product

$$\langle p, q\rangle=p(-1)q(-1)+p(0)q(0)+p(2)q(2)$$

If possible, normalize the vectors to make the set orthonormal.

5. $\{x^2+x, x^2+ax-2, x^2-2x\}$
6. $\{3x^2-2x-1, ax^2+x-1, 5x^2+ax-9\}$
7. Use Theorem 10.11 to express $v=[1\ 3\ 4\ 2\ 2]$ as a linear combination of the vectors given in Example 1.
8. Use Theorem 10.11 to express $f(x)=12x^2-6x-6$ as a linear combination of the vectors given in Exercise 2.
9. Let $v=[1\ 0\ -1]$. Find $\text{proj}_S v$ for the subspace *S* spanned by the vectors v_1 and v_2 and the inner product in Example 1.
10. Let $f(x)=x$. Find $\text{proj}_S f$ for the inner product and subspace *S* spanned by the functions in Exercise 2.
11. Let $f(x)=x$. Find $\text{proj}_S f$ for the inner product and subspace *S* in Example 7.
12. Find $\text{proj}_S f$ for $f(x)=e^x$, where $S=\text{span}\{1, x\}$ and the inner product is

$$\langle f, g\rangle=\int_{-1}^{1} f(x)g(x)\,dx$$

Exercises 13–18: Use the Gram–Schmidt process to convert the given set of vectors to an orthogonal basis with respect to the given inner product.

13. The set $\{[1\ -1\ 0], [2\ 0\ 1]\}$ with respect to the inner product given in Example 1.

14. The set $\left\{\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}2\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}\right\}$ with respect to the inner product given in Example 1.

15. The set $\{1, x^2\}$ with respect to the inner product given in Exercise 2.

16. The set $\{x^2+1, 4x, -3\}$ with respect to the inner product given in Exercise 2.

17. The set $\{x, 1\}$ with respect to the inner product given in Exercises 5–6.

18. The set $\{2x+1, x^2, 3\}$ with respect to the inner product given in Exercises 5–6.

FIND AN EXAMPLE Exercises 19–24: Find an example that meets the given specifications.

19. An orthogonal basis for $\mathbb{R}^2$ with respect to the inner product $\langle u, v\rangle = 3u_1v_1 + 2u_2v_2$ that includes $u_1 = \begin{bmatrix}1\\2\end{bmatrix}$.

20. An orthogonal basis for $\mathbb{R}^3$ with respect to the inner product $\langle u, v\rangle = u_1v_1 + 3u_2v_2 + 2u_3v_3$ that includes $u_1 = \begin{bmatrix}-1\\2\\1\end{bmatrix}$.

21. An orthogonal basis for P_1 with respect to the inner product $\langle p, q\rangle = \int_0^1 p(x)q(x)\,dx$ that contains $p_1(x) = 3x+1$.

22. An orthogonal basis for P_2 with respect to the inner product $\langle p, q\rangle = \int_{-1}^1 p(x)q(x)\,dx$ that contains $p_1(x) = x^2+4x-1$.

23. Three distinct functions in $C[-\pi, \pi]$ that are orthogonal, but cannot be made orthonormal, with respect to the inner product

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx$$

24. A basis for $\mathbb{R}^n$ that is orthogonal with respect to any weighted dot product.

TRUE OR FALSE Exercises 25–28: Determine if the statement is true or false, and justify your answer.

25.

(a) If $\{v_1, v_2, v_3\}$ is an orthonormal set in an inner product space V, then so is $\{c_1v_1, c_2v_2, c_3v_3\}$, where c_1, c_2, and c_3 are nonzero scalars.

(b) The set $\{1, \cos(2x), \sin(2x)\}$ is orthogonal in $C[-\pi, \pi]$ with respect to the inner product

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx$$

26.

(a) Any finite linearly independent set in an inner product space *V* can be converted to an orthonormal set by applying the Gram–Schmidt process.

(b) Every set of orthogonal vectors is linearly independent.

27.

(a) If **v** is a nonzero vector in an inner product space *V* and *S* is a nonzero finite-dimensional subspace of *V*, then $\text{proj}_S\,\mathbf{v} \neq \mathbf{0}$.

(b) If a set of vectors in an inner product space *V* is linearly dependent, then the set cannot be orthogonal.

28.

(a) If the Gram–Schmidt process is applied to a linearly dependent set, then one of the vectors produced will be the zero vector **0**.

(b) If $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is a set of nonzero vectors in $\mathbf{R}^3$, then $\mathcal{V}$ is not an orthogonal set with respect to any inner product.

29. Prove Theorem 10.10. (HINT: See the proof of Theorem 8.13.)

30. Prove Theorem 10.11. (HINT: See the proof of Theorem 8.14.)

31. Apply Theorem 10.11 to prove Theorem 10.12.

32. Prove Theorem 10.14. (HINT: See the proof of Theorem 8.18.)

Exercises 33–38: Assume **u** and **v** (and their subscripted relatives) are vectors in an inner product space *V*, and *S* is a nonzero finite-dimensional subspace of *V*.

33. Prove that $\text{proj}_S\mathbf{u} = \text{proj}_S(\text{proj}_S\mathbf{u})$.

34. Prove that $T: V \to V$ given by $T(\mathbf{u}) = \text{proj}_S\mathbf{u}$ is a linear transformation.

35. Suppose that **u** is in *S*, and that $S^\perp$ is nonzero.

(a) What is $\text{proj}_S\mathbf{u}$?

(b) What is $\text{proj}_{S^\perp}\mathbf{u}$?

36. Let $\{\mathbf{u}_1, \mathbf{u}_2\}$ be nonzero vectors, and define

$$\mathbf{v}_1 = \mathbf{u}_1, \quad \mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{\mathbf{v}_1}\mathbf{u}_2$$

Prove that $\text{span}\{\mathbf{u}_1, \mathbf{u}_2\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

37. Let $\{v_1, \ldots , v_k\}$ be an orthonormal basis of V. Prove that for any $\mathbf{v}$ in V, we have

$$\|v\|^2 = \langle v_1, v\rangle^2 + \cdots + \langle v_k, v\rangle^2$$

38. Here we prove that the Gram–Schmidt process works. Suppose that $\{\mathbf{u}_1, \ldots , \mathbf{u}_k\}$ are linearly independent vectors, and that $\{\mathbf{v}_1, \ldots , \mathbf{v}_k\}$ are as defined in the statement of the Gram–Schmidt process.

(a) Use induction to show $\{\mathbf{v}_1, \ldots , \mathbf{v}_j\}$ is an orthogonal set for $j = 1, \ldots , k$.

(b) Use induction to show $\text{span}\{u_1, \ldots , u_j\} = \text{span}\{v_1, \ldots , v_j\}$ for $j = 1, \ldots , k$.

(c) Explain why (a) and (b) imply that Gram–Schmidt yields an orthogonal basis.

10.3 Applications of Inner Products

In this section we consider a few applications of inner products. These applications exploit our ability to do the following: Given a vector **v** in an inner product space *V*, we can use the projection to find the vector s in a subspace *S* of *V* that is closest to **v**.

A vector s is "closest" to **v** when the norm of their difference $\|v-s\|$ is as small as possible. We encountered this when fitting lines to data in Euclidean space, where we found the required vector by using projections. The same approach works here, by applying this key theorem.

▶ This section is optional and can be omitted without loss of continuity.

THEOREM 10.16 ▶

Let *S* be a finite-dimensional subspace of an inner product space *V*, and suppose that **v** is in *V*. Then the closest vector in *S* to **v** is given by $\text{proj}_S v$. That is,

$$\|v-\text{proj}_S v\| \leq \|v-s\|$$

for all s in *S*, with equality holding exactly when $s=\text{proj}_S v$.

▶ Theorem 10.16 generalizes Theorem 8.32 in Section 8.5.

Proof The proof is similar to that of Theorem 8.32. If s is in *S*, then since $\text{proj}_S v$ is also in *S*, the difference $\text{proj}_S v-s$ must be in *S*. On the other hand, $v-\text{proj}_S v$ is in $S^\perp$ (by Theorem 10.14 in Section 10.2). Therefore by the Pythagorean theorem (Theorem 10.4 in Section 10.1), we have

$$\|(v-\text{proj}_S v)-(\text{proj}_S v-s)\|^2=\|v-\text{proj}_S v\|^2+\|\text{proj}_S v-s\|^2$$

As $(\mathbf{v}-\mathrm{proj}_S\mathbf{v})-(\mathrm{proj}_S\mathbf{v}-\mathbf{s})=\mathbf{v}-\mathbf{s}$, it follows that

$$\|\mathbf{v}-\mathbf{s}\|^2=\|\mathbf{v}-\mathrm{proj}_S\mathbf{v}\|^2+\|\mathrm{proj}_S\mathbf{v}-\mathbf{s}\|^2$$

Since $\|\mathrm{proj}_S\mathbf{v}-\mathbf{s}\|^2\geq 0$, we have $\|\mathbf{v}-\mathbf{s}\|\geq\|\mathbf{v}-\mathrm{proj}_S\mathbf{v}\|$. Furthermore, there is equality in this inequality exactly when $\|\mathrm{proj}_S\mathbf{v}-\mathbf{s}\|=0$. That is, when $\mathbf{s}=\mathrm{proj}_S\mathbf{v}$. ■■

Weighted Least Squares Regression

In Section 8.5 we used projection onto a subspace to fit a line to a data set of the form $(x_1, y_1),\dots, (x_n, y_n)$. There we treated each data point as being equally important, but now suppose that we view some points as more important than others. For instance, those that have the most extreme *x*-coordinates could be viewed as potential outliers. We might want to adjust our inner product so that these points have less influence on the model than those near the "center" of the data set.

We adopt the same notation as in Section 8.5. Given a line $\hat{y}=c_0+c_1x$, for each data point (x_i, y_i) we define $\hat{y}_i=c_0+c_1x_i$. The goal of ordinary **least squares regression** is to select c_0 and c_1 so that

$$(y_1-\hat{y}_1)^2+(y_2-\hat{y}_2)^2+\cdots+(y_n-\hat{y}_n)^2$$

is as small as possible. With **weighted least squares regression**, we minimize the expression

$$t_1(y_1-\hat{y}_1)^2+t_2(y_2-\hat{y}_2)^2+\cdots+t_n(y_n-\hat{y}_n)^2 \tag{1}$$

where $t_1, t_2,\dots, t_n$ are the positive **weights**. If we set

$$\mathbf{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix} \quad\text{and}\quad \hat{\mathbf{y}}=\begin{bmatrix}\hat{y}_1\\\hat{y}_2\\\vdots\\\hat{y}_n\end{bmatrix}$$

then (1) is equal to $\|\mathbf{y}-\hat{\mathbf{y}}\|^2$, where here the norm is with respect to the weighted dot product. Since $\hat{\mathbf{y}}=c_0+c_1\mathbf{x}$, if we define

$$A=\begin{bmatrix}1 & x_1\\1 & x_2\\\vdots & \vdots\\1 & x_n\end{bmatrix} \quad\text{and}\quad \mathbf{x}=\begin{bmatrix}c_0\\c_1\end{bmatrix}$$

then we need a solution to y^=Ax, where y^ is the vector in S=col(A) (the column space of *A*) that is closest to y. Thus, by Theorem 10.16, we should set

y^=projSy.

Let's consider an example that shows how this process works.

Example 1

Use least squares regression to fit a line to the data set (scatter plot shown in Figure 1)

$$(-6, 2.9), (-3, 1.5), (-2, 2), (2, 2.7), (3, 3.3), (6, 1.1)$$

Then use weighted least squares regression with weights designed to emphasize the four points in the middle of the data set.

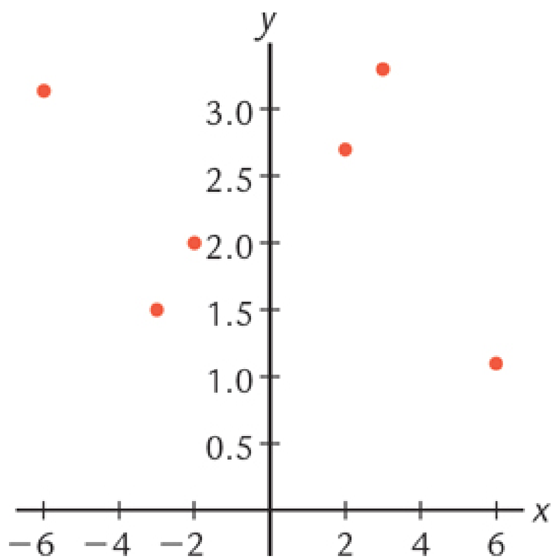


Figure 1 Scatter plot of data for Example 1.

▶ The 1's in a1 are multiplied times the constant term in the regression equation.

Solution Instead of using the formula for least squares regression developed in Section 8.5, we will use projection onto a subspace to find the equation for the regression line because that method generalizes to weighted least squares regression. Let A=[a1 a2], where

a1=[111111] and a2=[−6−3−2236]

We need to find y^=projSy, where S=span{a1, a2}. We have

a1·a2=(1)(−6)+(1)(−3)+(1)(−2)+(1)(6)+(1)(3)+(1)(2)=0

so that {a1, a2} is an orthogonal basis of *S* with respect to the dot product. Therefore we can apply the projection formula

y^=projSy=a1·ya1·a1 a1+a2·ya2·a2 a2 =13.56 a1+
−498 a2≈2.25a1−0.0408a2

Since Ax=c0a1+c1a2 and y^=2.25a1−0.0408a2, the solution to Ax=y^ is c0=2.25 and c1=−0.0408. Hence the least squares regression line is y^=2.25−0.0408x. A graph of the line together with the data is shown in Figure 2. The central four points lie roughly on a line, but the fit is poor due to the influence of the extreme points (−6, 2.9) and (6, 1.1).

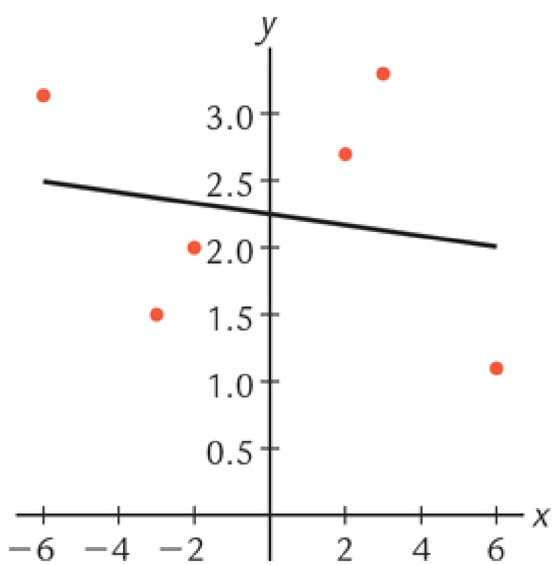


Figure 2 Scatter plot of data and y^=2.25−0.0408x.

We now repeat the analysis, but this time using a weighted dot product. We can diminish the “pull” of the extreme points with the weights

$$t=(t_1, t_2, t_3, t_4, t_5, t_6)=(1, 5, 5, 5, 5, 1)$$

This gives the four data points closest to the middle 5 times the weight of the outer two points. It is not hard to verify that the column vectors a_1 and a_2 satisfy $\langle a_1, a_2\rangle=0$ with respect to this weighted dot product. Thus we can use them for the projection function,

$$\hat{y}=\mathrm{proj}_S y=\frac{\langle a_1, y\rangle}{\langle a_1, a_1\rangle} a_1+\frac{\langle a_2, y\rangle}{\langle a_2, a_2\rangle} a_2$$

The required inner products are

$$\langle a_1, y\rangle=51.5,\langle a_2, y\rangle=23.2,\langle a_1, a_1\rangle=22,\langle a_2, a_2\rangle=202$$

Hence the projection is

$$\hat{y}=\mathrm{proj}_S y=\frac{51.5}{22} a_1+\frac{23.2}{202} a_2\approx 2.34a_1+0.115a_2$$

By the same reasoning as used previously, the equation of the weighted least squares regression line is $\hat{y}=2.34+0.115x$. The data and graph of this line are shown in Figure 3. Although the line has positive slope and is an improvement over ordinary least squares regression, it still does not fit the central data very well. Two more lines, with weights even more extreme to further diminish the effects of the extreme points, are shown in Figure 4.

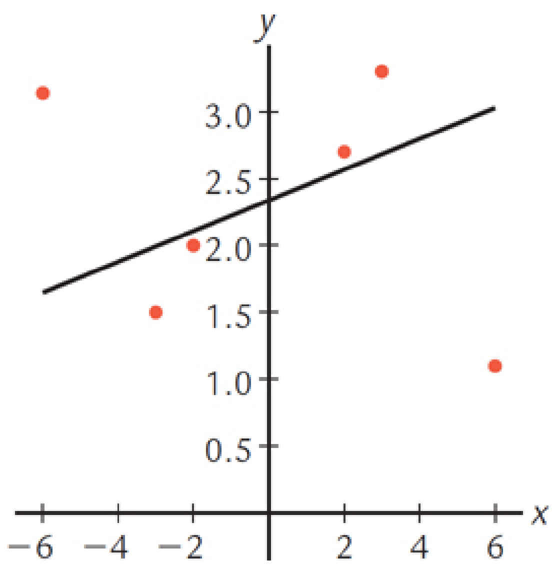


Figure 3 Scatter plot of data and y^=2.34+0.115x.

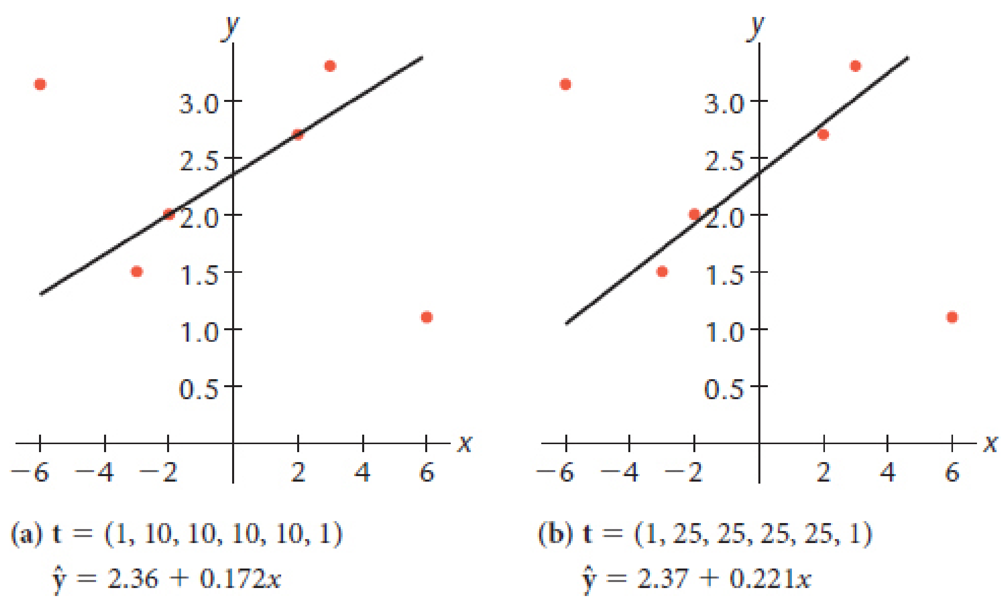


Figure 4 The graphs of weighted least squares regression lines with weights as shown.

The line in (b) of Figure 4 fits the central data fairly well, but the weights are so extreme that we are close to simply discarding the two outside points. In practice, one would typically make decisions about weights before collecting data, and it might happen that a linear equation is not appropriate for describing the data.

Fourier Approximations

We now return to the vector space $V = C[-\pi, \pi]$ together with the inner product

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx \qquad (2)$$

In Example 2 (Section 10.2), we showed that the set $\{1, \cos(x), \sin(x)\}$ forms an orthogonal set in *V*. This can be expanded to a larger orthogonal set.

THEOREM 10.17 ▶

For each integer $n \geq 1$, the set

$$\{1, \cos(x), \cos(2x), \ldots, \cos(nx), \sin(x), \sin(2x), \ldots, \sin(nx)\} \qquad (3)$$

is orthogonal in $V = C[-\pi, \pi]$ with the inner product (2).

Proof The proof involves computing a number of definite integrals to verify the orthogonality. Two are given here, and the rest are left as exercises. Starting with 1 and $\cos(kx)$, we have

$$\langle 1, \cos(kx)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(kx)\,dx = \frac{1}{k\pi}\left[\sin(kx)\right|_{-\pi}^{\pi} = 0$$

so 1 and $\cos(kx)$ are orthogonal for any $k \geq 1$.

▶ *f* is an *odd function* if $f(-x) = -f(x)$. If *f* is odd, then for any *b* we have $\int_{-b}^{b} f(x)\,dx = 0$

Since the product $\sin(jx)\cos(kx)$ is an odd function for any positive integers *j* and *k*, we have

$$\langle \sin(jx), \cos(kx)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(jx)\cos(kx)\,dx = 0$$

As noted above, the remaining integrals are left as exercises. ■

Now let F_n denote the subspace of $V = C[-\pi, \pi]$ spanned by the orthogonal basis given in Theorem 10.17. For any *f* in *V*, the best approximation in F_n to *f* is given by

$$f_n(x) = \mathrm{proj}_{F_n} f = a_0 + a_1\cos(x) + \cdots + a_n\cos(nx) + b_1\sin(x) + \cdots + b_n\sin(nx) \qquad (4)$$

_n_th-Order Fourier Approximation

The function $f_n(x)$ is called the ***n*th-order Fourier approximation** of *f*. Since the basis functions of F_n are orthogonal, from the projection formula we have

$$a_k = \frac{\langle f, \cos(kx)\rangle}{\langle \cos(kx), \cos(kx)\rangle} \quad (k \geq 1) \qquad b_k = \frac{\langle f, \sin(kx)\rangle}{\langle \sin(kx), \sin(kx)\rangle} \quad (k \geq 1)$$

For $k \geq 1$ we have $\langle \cos(kx), \cos(kx)\rangle = \langle \sin(kx), \sin(kx)\rangle = 1$ (see Exercises 35–36), so that the formulas for a_k and b_k simplify to

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\,dx \quad \text{and} \quad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx \qquad (5)$$

Since $\langle 1, 1\rangle = 2$, the constant term is

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx$$

Fourier Coefficients

The a_k's and b_k's are called the **Fourier coefficients** of *f*.

Example 2

Find the Fourier coefficients for $f(x) = x$ on $[-\pi, \pi]$.

▶ Example 2 uses $f(x) = x$ to provide an example with relatively easy calculations. In practice we would approximate more complex functions or data and use computational methods to evaluate the integrals.

Solution We start with

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} x\,dx = 0$$

because *x* is an odd function. As $x\cos(kx)$ is also an odd function, then by the same reasoning we have

$$a_k=\frac{1}{\pi}\int_{-\pi}^{\pi}x\cos(kx)\,dx=0$$

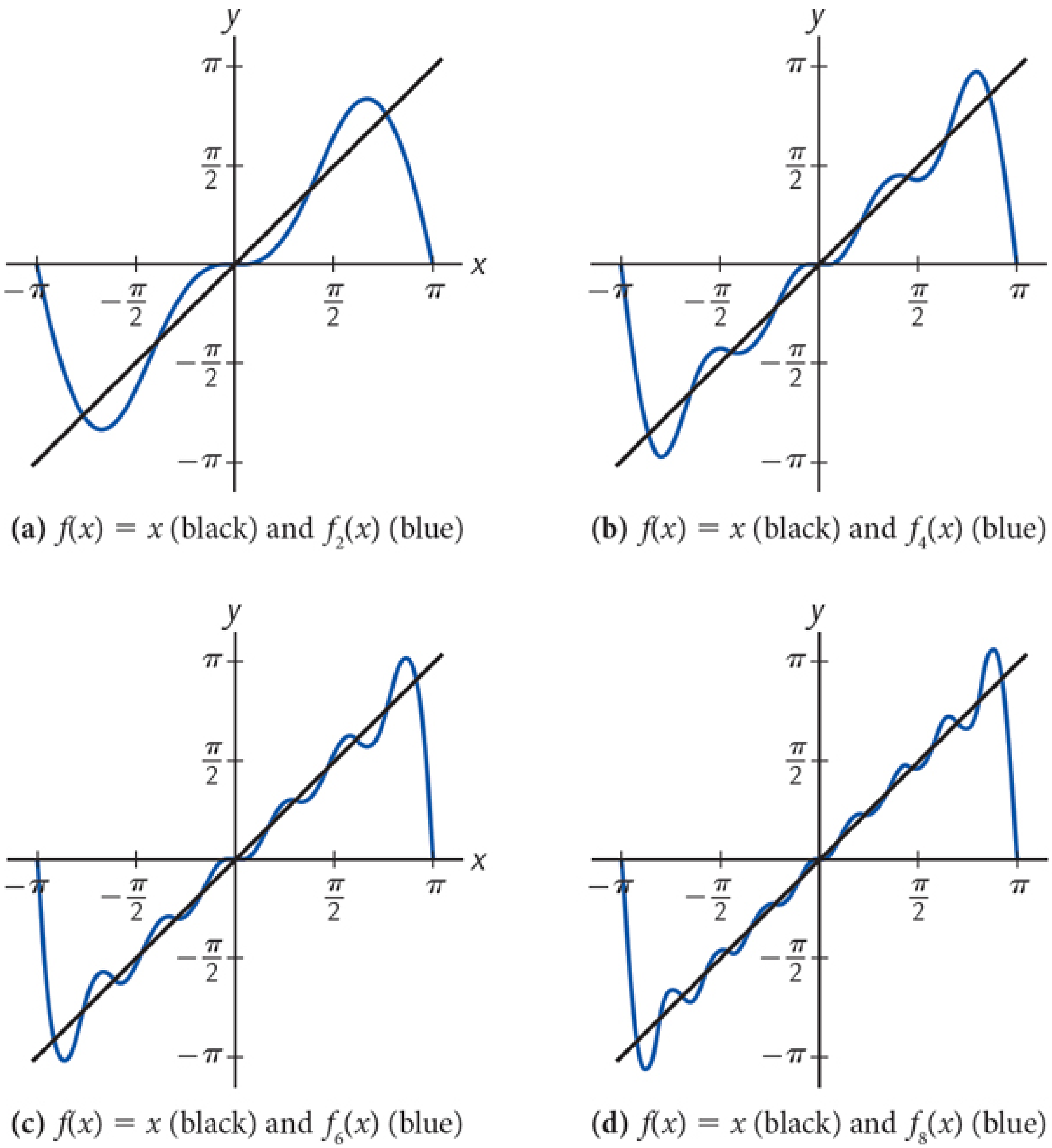
For $k\geq 1$, an application of integration by parts (the details are left as an exercise) gives us

$$b_k=\frac{1}{\pi}\int_{-\pi}^{\pi}x\sin(kx)\,dx=\frac{2}{k}(-1)^{k11}$$

Therefore our nth-order Fourier approximation to $f(x)=x$ is given by

$$f_n(x)=\sum_{k=1}^{n}\frac{2}{k}(-1)^{k+1}\sin(kx)=2\sin(x)-\sin(2x)+\frac{2}{3}\sin(3x)-\cdots+\frac{2}{n}(-1)^{n+1}\sin(nx)$$

The graphs of f_n for $n=2$, 4, 6, and 8 are given in Figure 5.



(a) $f(x) = x$ (black) and $f_2(x)$ (blue)

(b) $f(x) = x$ (black) and $f_4(x)$ (blue)

(c) $f(x) = x$ (black) and $f_6(x)$ (blue)

(d) $f(x) = x$ (black) and $f_8(x)$ (blue)

Figure 5 The graph of $f(x)=x$ and $f_n(x)$ for $n=2, 4, 6$, and 8. Note that the approximation improves with larger *n*.

If the Fourier coefficients decrease in size sufficiently quickly, then we can extend the nth-order Fourier approximation f_n to a **Fourier series**

$$a_0+\sum_{k=1}^{\infty}(a_k\cos(kx)+b_k\sin(kx))$$

Under the right conditions, the infinite series is equal to $f(x)$. The theory of Fourier series is covered in more advanced mathematics courses.

Discrete Fourier Transforms

Often in applications we want to find a Fourier approximation for a function *f* but do not have a formula for the function. Instead, we might know only values of $f(x)$ at discrete values of *x*, so we cannot directly apply the formulas given in (5) to find the Fourier coefficients.

We use a numerical integration technique to get around this problem. Suppose that for a function *g*, all we know are function values at *n* points evenly distributed in $[-\pi, \pi]$,

$$g\left(\frac{2\pi}{n}-\pi\right),\ g\left(\frac{4\pi}{n}-\pi\right),\ g\left(\frac{6\pi}{n}-\pi\right),\ \ldots,\ g\left(\frac{2n\pi}{n}-\pi\right)=g(\pi)$$

▶ Here we assume that *f* has domain $[-\pi, \pi]$. A change of variables can be used to accommodate other domains.

Then we can approximate the definite integral of *g* with the numerical integration formula

$$\frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\,dx \approx \frac{2}{n}\sum_{j=1}^{n} g\left(\frac{2j\pi}{n}-\pi\right) \qquad (6)$$

▶ (6) is just one of many numerical integration formulas.

To find a Fourier approximation for a function *f* with values known only at the discrete points

$$f\left(\frac{2\pi}{n}-\pi\right),\ f\left(\frac{4\pi}{n}-\pi\right),\ f\left(\frac{6\pi}{n}-\pi\right),\ \ldots,\ f\left(\frac{2n\pi}{n}-\pi\right)=f(\pi)$$

we use the numerical integration formula (6) to approximate a_k and b_k given by the definite integrals in (5) with the approximations

$$\begin{aligned} c_0 &= \frac{1}{n}\sum_{j=1}^{n} f\left(\frac{2j\pi}{n}-\pi\right) \\ c_k &= \frac{2}{n}\sum_{j=1}^{n} f\left(\frac{2j\pi}{n}-\pi\right)\cos\left(\frac{2jk\pi}{n}-k\pi\right) \quad (k \ge 1) \\ d_k &= \frac{2}{n}\sum_{j=1}^{n} f\left(\frac{2j\pi}{n}-\pi\right)\sin\left(\frac{2jk\pi}{n}-k\pi\right) \quad (k \ge 1) \end{aligned} \qquad (7)$$

Discrete Fourier Coefficients

The ck's and dk's are called **discrete Fourier coefficients**. In general, the larger the number of function values that are known (that is, the larger the value of *n*), the closer ck is to ak and dk is to bk. We examine this computationally in the next example.

Example 3

Let f(x)=x. Compare the values of the Fourier coefficients ak and bk of *f* with the discrete Fourier coefficients ck and dk.

Solution In Example 2 we determined that ak=0 for each k≥0, so we expect that ck should get smaller as *n* (the number of discrete function values) gets larger. Table 1 gives the values of ck for 0≤k≤10 and n=50, n=100, n=500, and n=1000. We can see that the values of ck get smaller, and hence closer to ak=0, as *n* gets larger.

▶ Other than sign, the values of ck are the same for each *k*. This is not typical and is specific to this example.

Table 1 **The Values of ak and ck for *f(x)* = *x***

k	ak	ck (*n* = 50)	ck (*n* = 100)	ck (*n* = 500)	ck (*n* = 1000)
0	0	−0.125664	−0.062832	−0.012566	−0.006283
1	0	0.125664	0.062832	0.012566	0.006283
2	0	−0.125664	−0.062832	−0.012566	−0.006283
3	0	0.125664	0.062832	0.012566	0.006283
4	0	−0.125664	−0.062832	−0.012566	−0.006283
5	0	0.125664	0.062832	0.012566	0.006283
6	0	−0.125664	−0.062832	−0.012566	−0.006283
7	0	0.125664	0.062832	0.012566	0.006283
8	0	−0.125664	−0.062832	−0.012566	−0.006283
9	0	0.125664	0.062832	0.012566	0.006283
10	0	−0.125664	−0.062832	−0.012566	−0.006283

For $f(x)=x$ we have $b_k=\frac{2}{k}(-1)^{k+1}$ for $k\ge 1$. Table 2 gives the values of b_k for $1\le k\le 10$ and $n=50$, $n=100$, $n=500$, and $n=1000$. The table values suggest that d_k is getting closer to b_k as *n* gets larger.

Table 2 The Values of a_k and d_k for $f(x) = x$

k	b_k	d_k (*n* = 50)	d_k (*n* = 100)	d_k (*n* = 500)	d_k (*n* = 1000)
1	2.00000	1.99737	1.99934	1.99997	1.99999
2	−1.00000	−0.99473	−0.99868	−0.99995	−0.99999
3	0.66667	0.65875	0.66469	0.66659	0.66665
4	−0.50000	−0.48943	−0.49737	−0.49990	−0.49997
5	0.40000	0.38675	0.39671	0.39987	0.39997
6	−0.33333	−0.31739	−0.32938	−0.33318	−0.33329
7	0.28571	0.26705	0.28109	0.28553	0.28567
8	−0.25000	−0.22858	−0.24471	−0.24979	−0.24995
9	0.22222	0.19801	0.21627	0.22199	0.22216
10	−0.20000	−0.17296	−0.19338	−0.19974	−0.19993

*n*th-Order Discrete Fourier Approximation

The ***n*th-order discrete Fourier approximation** is defined by

$$g_n(x)=c_0+c_1\cos(x)+\cdots+c_n\cos(nx)+d_1\sin(x)+\cdots+d_n\sin(nx).$$

The only difference between $f_n(x)$ and $g_n(x)$ is in how the coefficients are computed.

Example 4

Find discrete Fourier approximations for the data set shown in Figure 6.

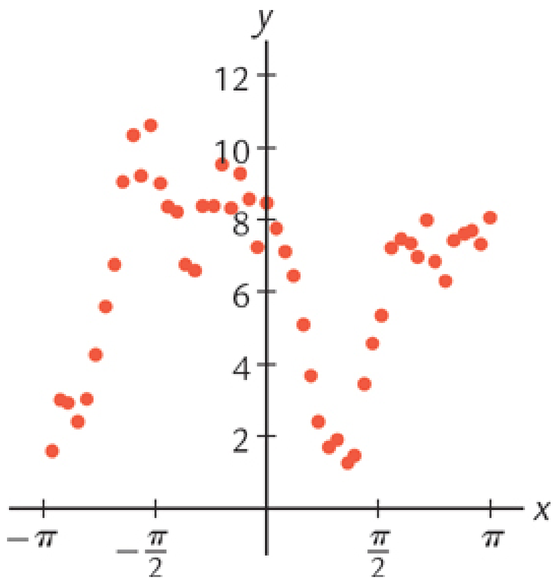
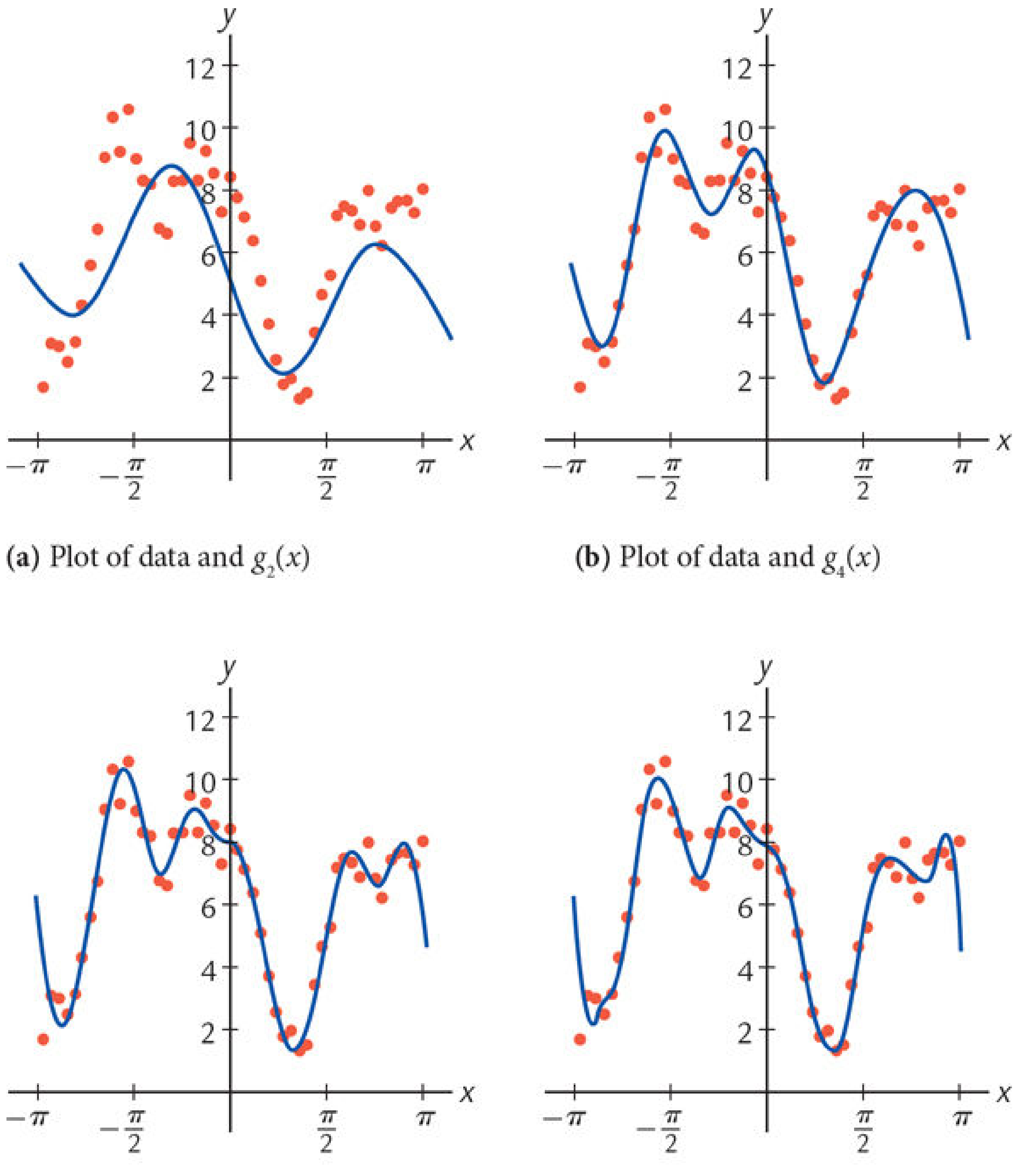


Figure 6 Data set for Example 4.

Solution There are 50 points shown in Figure 6. We have no formula for the function that generated the data, so finding the exact Fourier coefficients is out of the question. But we can use the discrete Fourier approximation and the formulas in (7) to compute approximations. The functions g2(x), g4(x), g8(x), and g12(x) are shown with the data in Figure 7.



(a) Plot of data and $g_2(x)$

(b) Plot of data and $g_4(x)$

(c) Plot of data and $g_8(x)$

(d) Plot of data and $g_{12}(x)$

Figure 7 Scatter plot of data for Example 4 and plot of discrete Fourier approximations of degree 2, 4, 8, and 12.

COMPUTATIONAL COMMENTS

Calculating coefficients for the discrete Fourier approximation can be computationally intensive. To address this, various methods known collectively as *fast Fourier transforms* (FFT) have been developed to improve computational efficiency. A popular FFT method developed by J. W. Cooley and J. W. Tukey works recursively by factoring *n*, the number of points, as $n=n_1n_2$ and then focusing on the smaller n_1 and n_2. Interestingly, there is evidence that Gauss knew about this method in the early 1800s. (There seems to be little about linear algebra that was unknown to Gauss.)

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the weighted least-squares line for the data set $\{(-2, 3), (-1, 1), (0, 0), (2, -4)\}$, with the inner two points weighted three times as much as the outer two points.
2. Find the Fourier approximation f_2 for $f(x)=1+3x$.
3. Find the Fourier approximation f_2 for $f(x)=1-|x|$.
4. Find the Fourier coefficients for $f(x)=3-\cos(3x)+5\sin(2x)$ without computing any integrals.
5. Find the discrete Fourier approximation $g_2(x)$ for $f(x)$ based on the table information.

x	0	π
$f(x)$	4	-3

EXERCISES

1. Find the weighted least-squares line for the data set {(−2, 0), (−1, 2), (1, 3), (2, 5)}, with the inner two points weighted twice as much as the outer two points.
2. Find the weighted least-squares line for the data set {(−2, −2), (−1, 0), (1, 3), (2, 6)}, with the inner two points weighted three times as much as the outer two points.

Exercises 3–4: Refer to the points shown in Figure 8.

3. Suppose that a line ℓ_1 is fitted to the points shown using ordinary least squares regression, and then a second line ℓ_2 is fitted using weighted least squares regression, with the two extreme points having half the weight of the others. How would you expect the slope of ℓ_1 to compare to that of ℓ_2? Explain your answer.

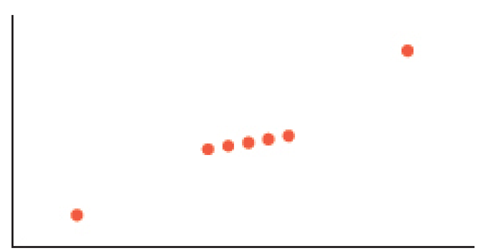


Figure 8

4. Suppose that a line ℓ_1 is fitted to the points shown using ordinary least squares regression, and then a second line ℓ_2 is fitted using weighted least squares regression, with the two extreme points having triple the weight of the others. How would you expect the slope of ℓ_1 to compare to that of ℓ_2? Explain your answer.
5. Suppose that all of the points in a weighted least squares approximation have their weights tripled. Will it change the equation of the resulting line? Explain your answer.
6. Suppose that all of the points in a weighted least squares approximation have their weights increased by a factor of 10.

Will it change the equation of the resulting line? Explain your answer.

7. Find the Fourier approximation f_2 for

$$f(x)=\begin{cases}1 & \text{if } -\frac{\pi}{2}\le x<\frac{\pi}{2}\\ 0 & \text{otherwise}\end{cases}$$

8. Find the Fourier approximation f_2 for

$$f(x)=\begin{cases}1 & \text{if } -\pi\le x<-\frac{\pi}{2} \text{ and } \frac{\pi}{2}\le x<\pi\\ 0 & \text{otherwise}\end{cases}$$

9. Find the Fourier approximation f_2 for

$$f(x)=\begin{cases}0 & \text{if } -\pi\le x<0\\ 1 & \text{if } 0\le x<\pi\end{cases}$$

10. Find the Fourier approximation f_2 for

$$f(x)=\begin{cases}1 & \text{if } -\pi\le x<0\\ 0 & \text{if } 0\le x<\pi\end{cases}$$

11. Find the Fourier approximation f_2 for $f(x)=x+1$.

12. Find the Fourier approximation f_2 for $f(x)=3-2x$.

13. Find the Fourier approximation f_2 for $f(x)=x^2$.

14. Find the Fourier approximation f_2 for $f(x)=|x|$.

Exercises 15–18: Find the Fourier coefficients for $f(x)$ without performing any integrals.

15. $f(x)=\cos(2x)-\sin(3x)$

16. $f(x)=-2+3\cos(2x)-4\sin(4x)$

17. $f(x)=1+\sin^2(4x)$

18. $f(x)=1-\cos^2(6x)$

19. Find the discrete Fourier approximation $g_1(x)$ for $f(x)$ based on the table information.

x	0	π
$f(x)$	1	2

20. Find the discrete Fourier approximation $g_2(x)$ for $f(x)$ based on the table information.

x	0	π

f(x)	−1	3

21. Find the discrete Fourier approximation g1(x) for f(x) based on the table information.

x	−π2	0	π2	π
f(x)	0	1	3	−2

22. Find the discrete Fourier approximation g2(x) for f(x) based on the table information.

x	−π2	0	π2	π
f(x)	−2	−1	0	2

FIND AN EXAMPLE Exercises 23–28: Find an example that meets the given specifications.

23. A data set such that the ordinary least squares regression line has slope zero, but the weighted least squares regression line, with triple weight on the right-most data point, has a negative slope.

24. A data set such that the ordinary least squares regression line has slope zero, but the weighted least squares regression line, with triple weight on the right-most and left-most data points, has a negative slope.

25. A function f(x) such that the Fourier coefficients are all zero except for a0 which is nonzero.

26. A function f(x) such that the Fourier coefficients b1=b2=b3=⋯ =0.

27. A function f(x) such that the Fourier coefficients a1=a2=a3=⋯=0 and a0=1.

28. A function f(x) such that the Fourier coefficients b1=b2=b3=⋯=0 and a0=−2.

TRUE OR FALSE Exercises 29–32: Determine if the statement is true or false, and justify your answer.

29.

(a) In weighted least squares regression, the weights must all be positive.

(b) In weighted least squares regression, the weights must all be greater than or equal to one.

30.

(a) Weighted least squares can only be applied to data sets where the corresponding matrix *A* has orthogonal columns.

(b) The Fourier approximation can only be applied to positive functions.

31.

(a) A Fourier approximation cannot be applied to a linear combination of sines and cosines.

(b) Weighted least squares cannot be applied to data sets with more than one independent variable.

32.

(a) If it is possible to compute Fourier coefficients for a function *f*, then it is also possible to compute discrete Fourier coefficients for *f*.

(b) In general, the higher the number of discrete function values used, the better the approximation given by the discrete Fourier approximation.

Exercises 33–36: Here we evaluate the remaining integrals required to show that the set given in Theorem 10.17 is orthogonal on $C[-\pi, \pi]$ with respect to the inner product

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\, dx$$

We also find the norms of these functions.

33. Show that $\langle 1, \sin(kx)\rangle = 0$ for all integers $k \geq 1$.

34. Show that $\|1\| = 2$.

35. Use the identity $2\sin^2(a) = 1 - \cos(2a)$ to show that $\|\sin(kx)\| = 1$ for all integers $k \geq 1$.

36. Use the identity $2\cos^2(a) = 1 + \cos(2a)$ to show that $\|\cos(kx)\| = 1$ for all integers $k \geq 1$.

37. Use integration by parts to show that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} x\cos(kx)\, dx = 0 \quad (k \geq 1)$$

38. Use integration by parts to show that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} x\sin(kx)\, dx = \frac{2}{k}(-1)^{k+1} \quad (k \geq 1)$$

39. Use the trigonometric identities

$$\cos(a-b)=\cos(a)\cos(b)+\sin(a)\sin(b) \qquad \sin(a-b)=\sin(a)\cos(b)-\sin(b)\cos(a)$$

to prove that the formulas for c_k and d_k in (7) can be simplified to

$$c_k=\frac{2}{n}\sum_{j=1}^{n}(-1)^k f\left(\frac{2j\pi}{n}-\pi\right)\cos\left(\frac{2jk\pi}{n}\right)\quad (k\geq 1) \qquad d_k=\frac{2}{n}\sum_{j=1}^{n}(-1)^k f\left(\frac{2j\pi}{n}-\pi\right)\sin\left(\frac{2jk\pi}{n}\right)\quad (k\geq 1)$$

40. Discrete values of *f* are given in the table below. Use this information to find the discrete Fourier approximation $g_3(x)$.

x	$-\frac{3\pi}{4}$	$-\frac{\pi}{2}$	$-\frac{\pi}{4}$	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π
f(x)	3.1	3.5	3.3	3.0	2.7	2.6	2.8	3.0

41. Discrete values of *f* are given in the table below. Use this information to find the discrete Fourier approximation $g_5(x)$.

x	$-\frac{3\pi}{4}$	$-\frac{\pi}{2}$	$-\frac{\pi}{4}$	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π
f(x)	2.4	2.8	3.0	3.5	2.9	2.6	2.4	2.1

42. Suppose that $f(x)=x^2+1$. Find $f_3(x)$. Then generate a sequence of 51 evenly-spaced points, starting with $x_0=-\pi$ and ending with $x_{50}=\pi$. Determine $f(x_0),\ldots, f(x_{50})$, and use these values to find $g_3(x)$.

43. Suppose that $f(x)=e^x$. Find $f_3(x)$. Then generate a sequence of 51 evenly-spaced points, starting with $x_0=-\pi$ and ending with $x_{50}=\pi$. Determine $f(x_0),\ldots, f(x_{50})$, and use these values to find $g_3(x)$.

SUPPLEMENTARY EXERCISES

Exercises 1–4: Compute the indicated inner product.

1. $\langle u, v\rangle$ for $u=[13-2]$, $v=[-143]$, and the inner product given in Section 10.1, Example 2 with weights $t_1=1$, $t_2=3$, and $t_3=2$.

2. $\langle p, q\rangle$ for $p(x)=-3x+2$, $q(x)=x+1$, and the inner product given in Section 10.1, Example 3 with $x_0=-2$, $x_1=0$, and $x_2=2$.

3. $\langle f, g\rangle$ for $f(x)=-4x+1$, $g(x)=x^2$, and the inner product given in Section 10.1, Example 5.

4. $\langle A, B\rangle=\mathrm{tr}(A^T B)$ for $A=[2011]$, $B=[-23-4-1]$.

Exercises 5–8: Compute the norm with respect to the indicated inner product.

5. $\|[32-2]\|$ for the inner product given in Section 10.1, Example 2.

6. $\|-x+2\|$ for the inner product given in Section 10.1, Example 3 with $x_0=-1$, $x_1=0$, and $x_2=2$.

7. $\|3x^2\|$ for the inner product given in Section 10.1, Example 5.

8. $\|A\|$ for $A=[0-125]$ and $\langle A, B\rangle=\mathrm{tr}(A^T B)$.

Exercises 9–12: Compute the indicated projection with respect to the inner product.

9. $\mathrm{proj}_u v$ for $u=[-10-2]$, $v=[152]$, and the inner product given in Section 10.1, Example 2 with weights $t_1=4$, $t_2=3$, and $t_3=1$.

10. $\mathrm{proj}_p q$ for $p(x)=-3x+1$, $q(x)=2x+3$, and the inner product given in Section 10.1, Example 3 with $x_0=-2$, $x_1=-1$, and $x_2=1$.

11. $\mathrm{proj}_f g$ for $f(x)=-3x+1$, $g(x)=-x^2$, and the inner product given in Section 10.1, Example 5.

12. $\mathrm{proj}_A B$ for $A=[1752]$, $B=[0-142]$, and $\langle A, B\rangle=\mathrm{tr}(A^T B)$.

Exercises 13–16: Determine if the statement is true or false, and justify your answer. Here **u** and **v** are vectors in an inner product space *V*.

13. If **u** and **v** are orthogonal, then $\text{proj}_{\mathbf{u}}\mathbf{v}=\mathbf{0}$.
14. $\langle 3\mathbf{u}, 4\mathbf{u}\rangle=12\|\mathbf{u}\|$.
15. $\int_0^2 f(x)g(x)x^3\,dx$ is an inner product on $C(0, 2)$.
16. If $\|\mathbf{u}\|=2\|\mathbf{v}\|$ and **u** is orthogonal to **v**, then $\|\mathbf{u}+\mathbf{v}\|=3\|\mathbf{u}\|$.
17. Determine the values of *a* (if any) that will make the given set of vectors orthogonal in $\mathbf{R}^3$ with respect to the weighted dot product with $t_1=1$, $t_2=5$, and $t_3=2$. If possible, normalize the vectors to make the set orthonormal.

$$\left\{\begin{bmatrix}1\\3\\-1\end{bmatrix},\ \begin{bmatrix}a\\2\\-1\end{bmatrix},\ \begin{bmatrix}1\\0\\3\end{bmatrix}\right\}$$

18. Determine the values of *a* (if any) that will make the given set of vectors orthogonal in $\mathbf{P}^2$ with respect to the inner product

$$\langle p, q\rangle=p(0)q(0)+p(1)q(1)+p(3)q(3)$$

If possible, normalize the vectors to make the set orthonormal.

$$\{x^2+x+4,\ 2ax^2-5,\ 2x+1\}$$

19. Let $f(x)=2x-1$. Find $\text{proj}_S f$ for the inner product and subspace *S* in Section 10.2, Example 7.
20. Find $\text{proj}_S f$ for $f(x)=e^{2x}$, where $S=\text{span}\{2, x\}$ and the inner product is

$$\langle f, g\rangle=\int_0^1 f(x)g(x)\,dx$$

Exercises 21–22: Use the Gram–Schmidt process to convert the given set of vectors to an orthogonal basis with respect to the inner product.

21. The set $\left\{\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}2\\1\\1\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix}\right\}$ with respect to the inner product given in Section 10.2, Example 1.
22. The set $\{-x^2+3, -x+2, 3\}$ with respect to the inner product given in Section 10.2, Exercise 2.

23. Find the weighted least-squares line for the data set $\{(-9, 2), (-2, 5), (2, 7), (3, 11)\}$, with the inner two points weighted three times as much as the outer two points.

24. Find the Fourier approximation $f2$ for $f(x)=4+x$.

25. Find the Fourier approximation $f2$ for $f(x)=1+3|x|$.

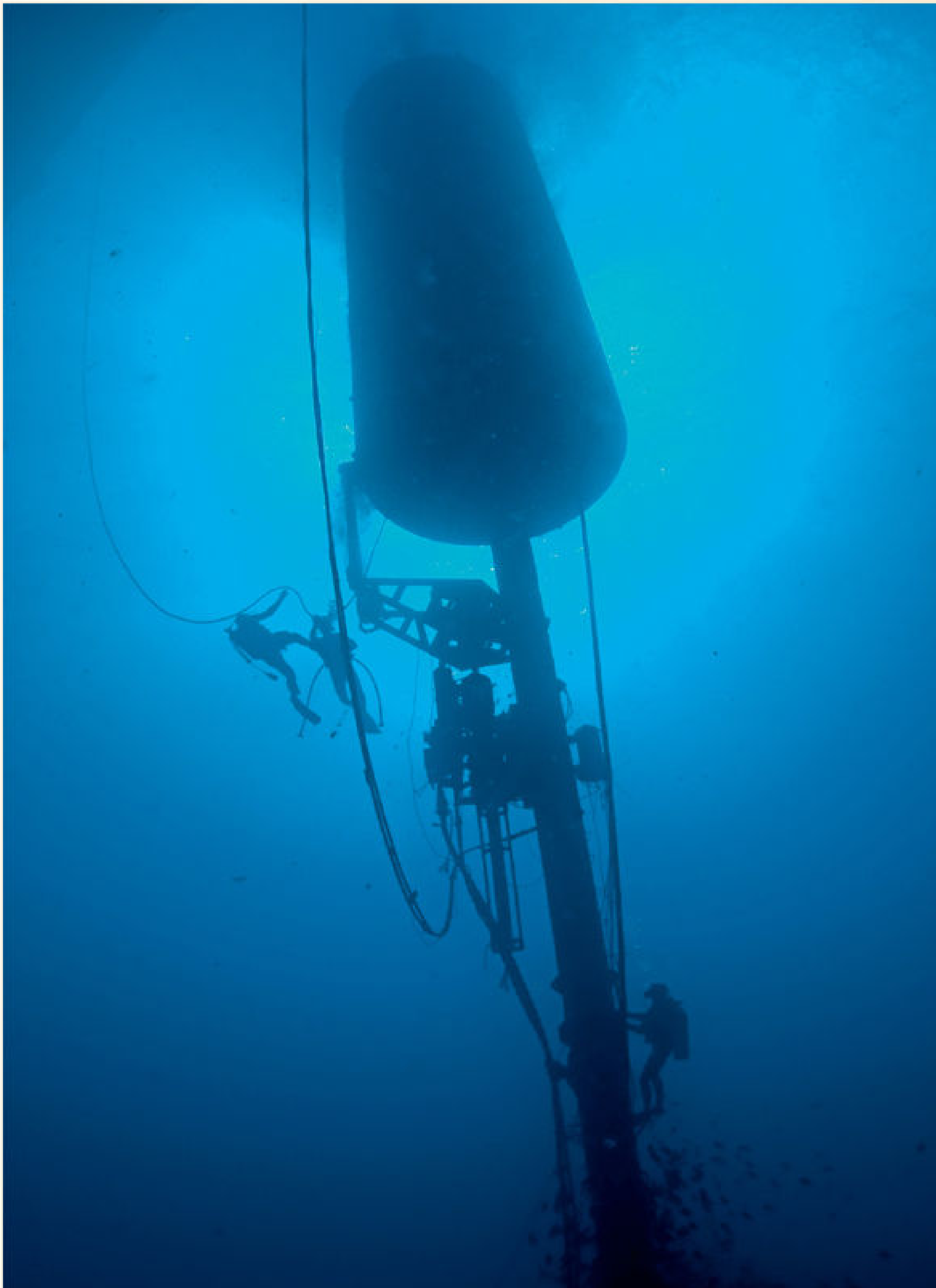
26. Find the Fourier coefficients for $f(x)=-2-4\cos(x)+2\sin(3x)$ without computing any integrals.

27. Find the discrete Fourier approximation $g2(x)$ for $f(x)$ based on the table information.

x	0	π
$f(x)$	−2	5

CHAPTER 11

Additional Topics and Applications



David Fleetham/VWPics/Dreams of the Blue Communications (Visual&Written)/Newscom

Wave power, which is still experimental and not widely used, harnesses the energy of ocean waves to generate electricity or complete other work. The photo shows divers installing a large energy buoy off Kaneohe Bay, Oahu, in Hawaii. The motion of the buoy drives an electrical generator, and the power travels through an underwater cable to the island.

11.1 Quadratic Forms

In Chapter 3 we showed that every linear transformation $T : R^m \to R^n$ has the form $T(x)=Ax$ for some $n\times m$ matrix *A*. In this section we study a function $Q : R^n \to R$ called a *quadratic form* that also can be defined in terms of matrix and vector multiplication. Such functions arise naturally in a variety of disciplines, including engineering (control theory), physics, economics, and mathematics.

▶ The sections in this chapter are optional.

DEFINITION 11.1 ▶

Quadratic Form, Matrix of the Quadratic Form

A **quadratic form** is a function $Q : R^n \to R$ that has the form

$$Q(x)=x^TAx \qquad (1)$$

where *A* is an $n\times n$ symmetric matrix called the **matrix of the quadratic form**.

Example 1

Suppose that $x=\begin{bmatrix}x_1\\x_2\end{bmatrix}$. Evaluate $Q(x)=x^TAx$ for each of the matrices

$$\text{(a)}\ \ A=\begin{bmatrix}2&0\\0&5\end{bmatrix} \qquad \text{(b)}\ \ A=\begin{bmatrix}3&1\\1&-2\end{bmatrix}$$

▶ We interpret a quadratic form x^TAx as a scalar instead of as a 1×1 matrix.

Solution (a) $x^TAx=[x_1\ \ x_2]\begin{bmatrix}2&0\\0&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=[x_1\ \ x_2]\begin{bmatrix}2x_1\\5x_2\end{bmatrix}=2x_1^2+5x_2^2.$

(b) Since A is not diagonal, this quadratic form is a bit more complicated. We have

$$\mathbf{x}^T A\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3x_1 + x_2 \\ x_1 - 2x_2 \end{bmatrix} = x_1(3x_1+x_2) + x_2(x_1-2x_2) = 3x_1^2 + 2x_1x_2 - 2x_2^2$$

Note that in both parts, the coefficients on x_1^2 and x_2^2 come directly from the diagonal entries of A. In (b), we see that the coefficient on the *cross-product* term x_1x_2 is the sum of the nondiagonal matrix entries. (This is also happening in part (a), but is not visible because the nondiagonal entries are 0.) These observations generalize so that it is not too hard to construct the matrix of a quadratic form from the equation.

Example 2

Suppose that Q is the quadratic form

$$Q(\mathbf{x}) = 3x_1^2 - 7x_3^2 - 4x_1x_2 + 10x_2x_3$$

Directly compute $Q(\mathbf{x}_0)$ for $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$. Then find the 3×3 matrix A of the quadratic form and use it to recompute $Q(\mathbf{x}_0)$ by applying the formula $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

Solution For our given $\mathbf{x}_0$ we have $x_1 = 1$, $x_2 = 3$, and $x_3 = -2$. Hence

$$Q(\mathbf{x}_0) = 3(1)^2 - 7(-2)^2 - 4(1)(3) + 10(3)(-2) = -97$$

To find the matrix A of this quadratic form, we start by noting that the terms $3x_1^2$ and $-7x_3^2$ indicate that there should be a 3 and -7 in the first and third diagonal entries, respectively. Since there is no x_2^2 term, the second diagonal entry is 0. Thus the diagonal portion of A is

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

▶ The •'s represent matrix entries that have not yet been determined.

The coefficient -4 on the cross-product term $-4x_1x_2$ should be evenly split across the (1, 2) and (2, 1) entries to ensure that *A* is symmetric. Similarly, we evenly split the coefficient 10 from $10x_2x_3$ across the (2, 3) and (3, 2) entries. Since that accounts for all of the terms of *Q*, any other entries of *A* should be zero. Therefore we end up with

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

Now we test this by computing $\mathbf{x}_0^TA\mathbf{x}_0$. We have

$$\mathbf{x}_0^TA\mathbf{x}_0=\begin{bmatrix}1&3&-2\end{bmatrix}\begin{bmatrix}3&-2&0\\-2&0&5\\0&5&-7\end{bmatrix}\begin{bmatrix}1\\3\\-2\end{bmatrix}=\begin{bmatrix}1&3&-2\end{bmatrix}\begin{bmatrix}-3\\-12\\29\end{bmatrix}=-97$$

Quadratic forms can be easier to apply when there are no cross-product terms to complicate things. Happily, we can use a change of variables to arrange for this.

THEOREM 11.2 ▶

(PRINCIPAL AXES THEOREM) If *A* is a symmetric matrix, then there exists an orthogonal matrix *P* such that the transformation $\mathbf{y}=P^T\mathbf{x}$ changes the quadratic form $\mathbf{x}^TA\mathbf{x}$ into the quadratic form $\mathbf{y}^TD\mathbf{y}$ (where *D* is diagonal) that has no cross-product terms.

▶ The "principal axes" in Theorem 11.2 are the columns of *P*, which are eigenvectors of *A*. The name will be explained shortly.

Proof Since *A* is a symmetric matrix, by the Spectral Theorem (Section 8.3) *A* can be diagonalized as $A=PDP^{-1}$, where *P* is an orthogonal matrix with eigenvectors of *A* for columns and *D* is a diagonal matrix with eigenvalues of *A* for diagonal entries. Since *P* is orthogonal, we have $P^{-1}=P^T$, so that

$$A=PDP^T \;\Rightarrow\; D=P^TAP$$

If we set $\mathbf{y}=P^{-1}\mathbf{x}=P^T\mathbf{x}$, then $\mathbf{x}=P\mathbf{y}$ so that

$$\mathbf{x}^TA\mathbf{x}=(P\mathbf{y})^TA(P\mathbf{y})=\mathbf{y}^TP^TAP\mathbf{y}=\mathbf{y}^T(P^TAP)\mathbf{y}=\mathbf{y}^TD\mathbf{y}$$

Hence the matrix of the quadratic form is diagonal with respect to the change of variables $\mathbf{y}=P^{-1}\mathbf{x}$.

Example 3

Find a change of variables to express the quadratic form with matrix

$$A=\begin{bmatrix}5 & 2\\ 2 & 8\end{bmatrix}$$

as a quadratic form with no cross-product terms.

Solution The quadratic form with matrix *A* is

$$Q(\mathbf{x})=5x_1^2+8x_2^2+4x_1x_2 \tag{2}$$

To eliminate the cross-product term, we need to find a matrix *P* that orthogonally diagonalizes *A*. Computing the roots of the characteristic polynomial reveals that the eigenvalues of this matrix are $\lambda_1=9$ and $\lambda_2=4$. The corresponding normalized eigenvectors are

$$\lambda_1=9 \;\Rightarrow\; \mathbf{p}_1=\frac{1}{\sqrt{5}}\begin{bmatrix}1\\ 2\end{bmatrix} \qquad \lambda_2=4 \;\Rightarrow\; \mathbf{p}_2=\frac{1}{\sqrt{5}}\begin{bmatrix}-2\\ 1\end{bmatrix}$$

▶ In this chapter the details of computing eigenvalues and eigenvectors are often left to the reader.

Therefore we have

$$D=[λ100λ2]=[9004] \quad \text{and} \quad P=[p1\ \ p2]=[15−252515]$$

Hence for $y=[y1y2]=PTx$,

$$Q(y)=9y12+4y22 \tag{3}$$

We can test out the two versions of the quadratic form by starting with a specific vector—say, $x0=[21]$. Evaluating (2) directly, we find that

$$Q\ ([21])=5(2)2+8(1)2+4(2)(1)=36$$

For this choice of $x0$, the corresponding $y0$ is

$$y0=PTx0=[1525−2515]\ [21]=[45−35]$$

Therefore from (3) we have

$$9y12+4y22=9(45)2+4(−35)2=9(165)+4(95)=1805=36$$

Geometry of Quadratic Forms

Principal Axes

In the Principal Axes Theorem, the columns of *P*—which are eigenvectors of *A*—are the **principal axes** of the quadratic form $xTAx$. The use of the word “axes” makes more sense when we view quadratic forms geometrically. Consider the set of vectors x in $R2$ that satisfy the equation

$$xTAx=c \tag{4}$$

where *c* is a fixed constant and *A* is an invertible 2×2 symmetric matrix. It turns out that graph of the solution set can be one of the following: an ellipse (including circles), a hyperbola, two intersecting lines, a single point, or the empty set. Here we focus on the ellipse and hyperbola.

If $A=\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ is diagonal, then (4) is equivalent to

$$ax_1^2+bx_2^2=c$$

Standard Position

When *a*, *b*, and *c* are all positive, the graph of the solution set is the ellipse in Figure 1. The graph of the solution set of a quadratic form with no cross-product terms is said to be in **standard position**. Since *A* is diagonal, the eigenvectors of *A* point in the direction of the coordinate axes, which coincide with the major and minor axes of the ellipse. This carries over to quadratic forms that have cross-product terms.

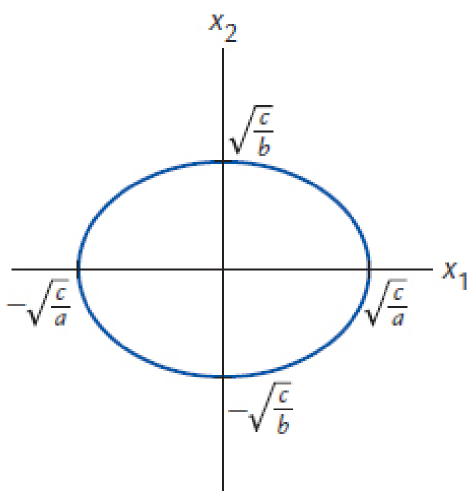


Figure 1 Graph of the solutions to $ax_1^2+bx_2^2=c$. The axis intercepts are as shown.

Example 4

Graph the set of solutions to the quadratic form

$$3x_1^2+6x_2^2+4x_1x_2=40$$

Solution The quadratic form $3x_1^2+6x_2^2+4x_1x_2$ has matrix $A=\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$. The eigenvalues and normalized eigenvectors of *A* are

$$\lambda_1=2 \;\Rightarrow\; p_1=\frac{1}{\sqrt{5}}\begin{bmatrix}2\\-1\end{bmatrix}, \qquad \lambda_2=7 \;\Rightarrow\; p_2=\frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\end{bmatrix}$$

Now let

$$D=\begin{bmatrix}\lambda_1&0\\0&\lambda_2\end{bmatrix}=\begin{bmatrix}2&0\\0&7\end{bmatrix} \quad \text{and} \quad P=\begin{bmatrix}p_1 & p_2\end{bmatrix}=\begin{bmatrix}\frac{2}{\sqrt{5}}&\frac{1}{\sqrt{5}}\\-\frac{1}{\sqrt{5}}&\frac{2}{\sqrt{5}}\end{bmatrix}$$

Then $y^TDy=40$ is equivalent to $2y_1^2+7y_2^2=40$. Since there are no cross-product terms, we can use Figure 1 as a model for the graph of $y^TDy=40$, which is given in Figure 3(a).

By the Principal Axes Theorem, $x^TAx=c$ and $y^TDy=c$ are the same equation when $y=P^Tx$, or equivalently, $x=Py$. Therefore the graph of all x that satisfy $x^TAx=c$ is the same as the graph of all y that satisfy $y^TDy=c$ after applying the transformation $x=Py$. Since *P* is an orthogonal matrix, this transformation is a rotation or reflection (see Exercise 44), so the graph of $x^TAx=c$ is a rotation or reflection of the graph of $y^TDy=c$. Furthermore, since e_1 and e_2 are parallel to the major and minor axes of the ellipse $y^TDy=c$, then $P(e_1)=p_1$ and $P(e_2)=p_2$ are parallel to the axes of $x^TAx=c$. The graph is given in Figure 3(b).

▶ Recall that $\{e_1, e_2\}$ are the standard basis for $\mathbb{R}^2$,
$e_1=\begin{bmatrix}1\\0\end{bmatrix}, \; e_2=\begin{bmatrix}0\\1\end{bmatrix}$

Summing up the solution to Example 4, we did the following:

- Found *D* and *P* for *A*.
- Graphed $y^TDy=c$, which is not difficult, because it is in standard form.
- Rotated the graph so that the axes of symmetry align with the principal axes of *A* to graph $x^TAx=c$.

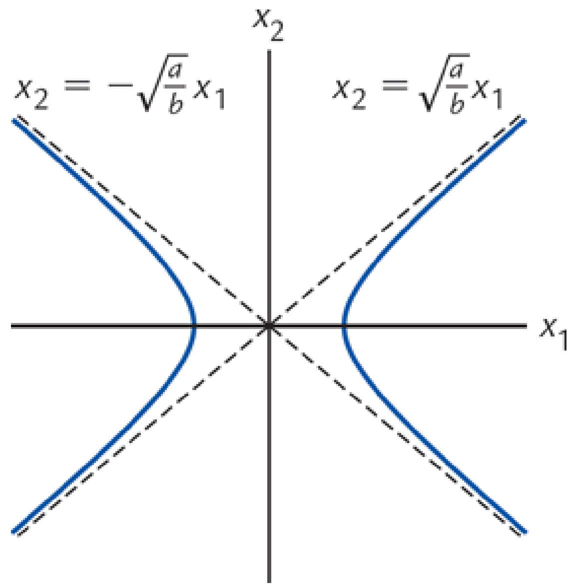
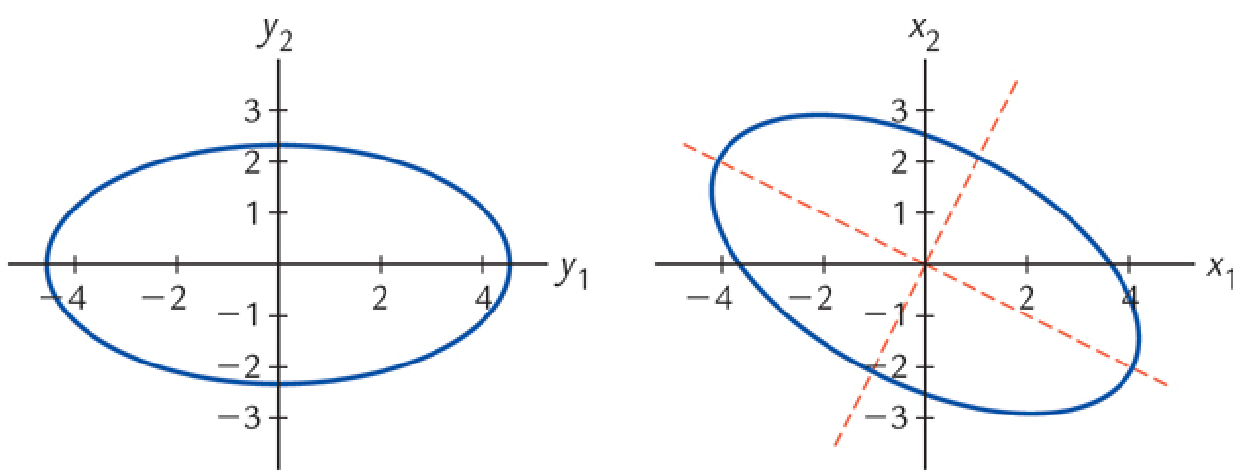


Figure 2 Graph of the solutions to ax12−bx22=c. The asymptotes (dashed) have equations shown.

Turning to hyperbolas, if *a*, *b*, and *c* are still positive, then the graph of ax12−bx22=c is a hyperbola in standard position with asymptotes x2=±abx1 (Figure 2). We can use the same approach as in Example 4 to graph hyperbolas that are not in standard position.



(a) Graph of $2y_1^2 + 7y_2^2 = 40$ **(b)** Graph of $3x_1^2 + 6x_2^2 + 4x_1x_2 = 40$

Figure 3 In (a), the graph of yTDy=40, which is in standard position. Rotating this graph to align the ellipse axes (dashed) with the eigenvectors p1 and p2 of *A* gives the graph of xTAx=40 in (b).

Example 5

Graph the set of solutions to the quadratic form

$$4x_1^2 - x_2^2 + 12x_1x_2 = 10$$

Solution Here the matrix of the quadratic form is $A = \begin{bmatrix} 4 & 6 \\ 6 & -1 \end{bmatrix}$, which has eigenvalues and normalized eigenvectors

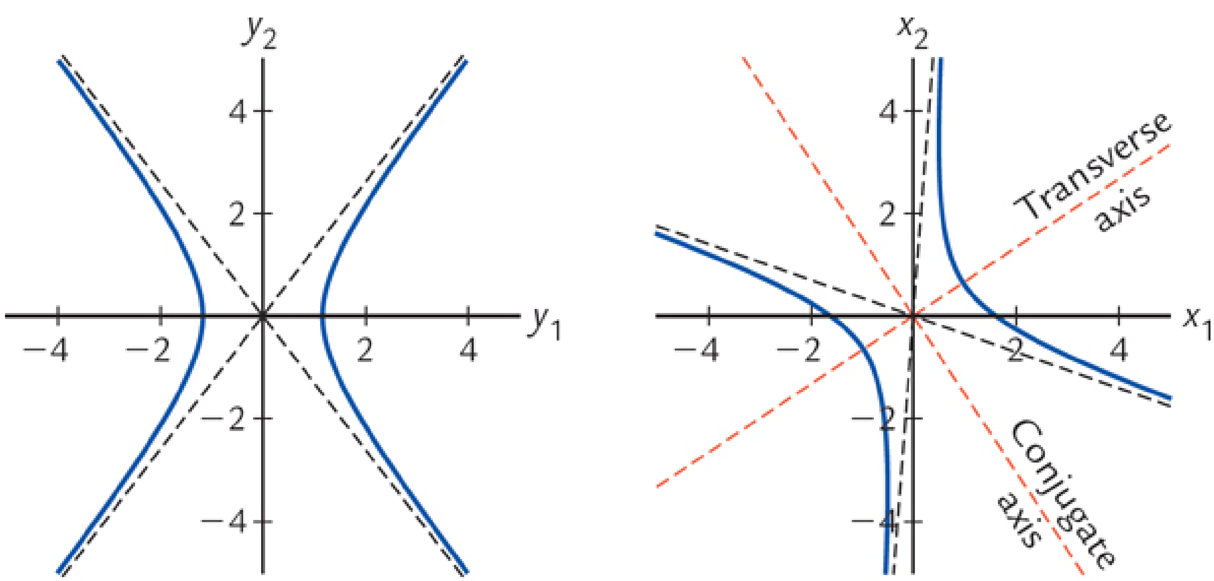
$$\lambda_1 = 8 \;\Rightarrow\; \mathbf{p}_1 = \frac{1}{\sqrt{13}}\begin{bmatrix} 3 \\ 2 \end{bmatrix}, \qquad \lambda_2 = -5 \;\Rightarrow\; \mathbf{p}_2 = \frac{1}{\sqrt{13}}\begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

Hence the matrices *D* and *P* from the Principal Axes Theorem are

$$D = \begin{bmatrix} 8 & 0 \\ 0 & -5 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} \frac{3}{\sqrt{13}} & -\frac{2}{\sqrt{13}} \\ \frac{2}{\sqrt{13}} & \frac{3}{\sqrt{13}} \end{bmatrix}$$

The equation $8y_1^2 - 5y_2^2 = 10$ is in standard form, and the graph is a hyperbola with asymptotes $y_2 = \pm\sqrt{\frac{8}{5}}\,y_1$ (Figure 4(a)). Hyperbolas also have two axes of symmetry, the transverse axis (the *x*-axis in Figure 4(a)) and the conjugate axis (the *y*-axis in Figure 4(a)).

As with ellipses, we get the graph of $\mathbf{x}^T A \mathbf{x} = 10$ by applying the transformation $\mathbf{x} = P\mathbf{y}$ to the graph of $\mathbf{y}^T D \mathbf{y} = 10$. The axes of symmetry of the graph of $\mathbf{y}^T D \mathbf{y} = 10$ are rotated to align with the eigenvectors of *A*, yielding the graph in Figure 4(b).



(a) Graph of $8y_1^2 - 5y_2^2 = 10$ **(b)** Graph of $4x_1^2 - x_2^2 + 12x_1x_2 = 10$

Figure 4 In (a), the graph of $8y_1^2-5y_2^2=10$, which is in standard position. Rotating this graph to align the axes of symmetry with the eigenvectors of *A* gives the graph of $4x_1^2-x_2^2+12x_1x_2=10$ in (b).

Types of Quadratic Forms

We can classify a quadratic form $Q(\mathbf{x})=\mathbf{x}^TA\mathbf{x}$ based on the values of $Q(\mathbf{x})$ as $\mathbf{x}$ ranges over different possibilities in $\mathbb{R}^n$.

DEFINITION 11.3 ▶

Positive Definite, Negative Definite, Indefinite, Positive Semidefinite, Negative Semidefinite

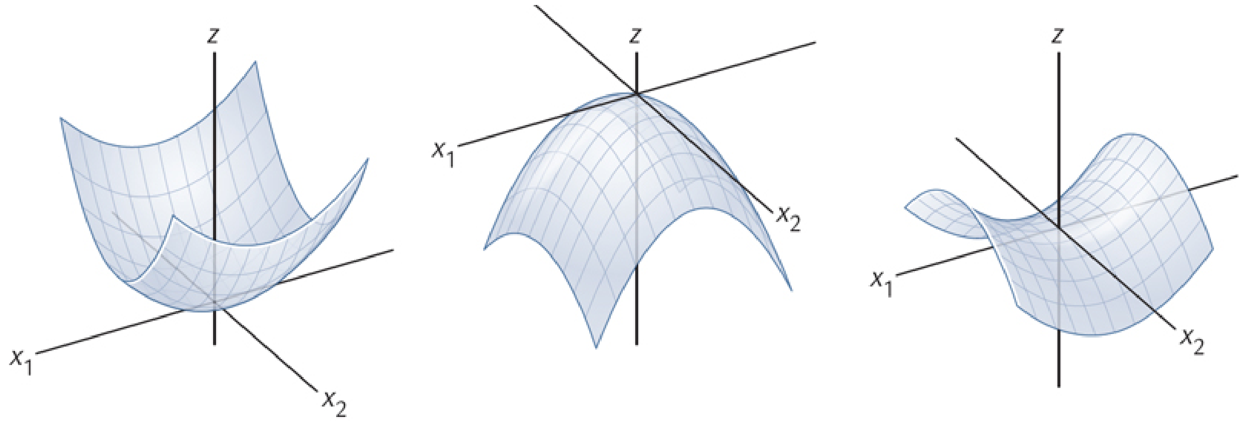
Let $Q(\mathbf{x})=\mathbf{x}^TA\mathbf{x}$ be a quadratic form.

(a) *Q* is **positive definite** if $Q(\mathbf{x})>0$ for all nonzero vectors $\mathbf{x}$ in $\mathbb{R}^n$, and *Q* is **positive semidefinite** if $Q(\mathbf{x})\geq 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

(b) *Q* is **negative definite** if $Q(\mathbf{x})<0$ for all nonzero vectors $\mathbf{x}$ in $\mathbb{R}^n$, and *Q* is **negative semidefinite** if $Q(\mathbf{x})\leq 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

(c) *Q* is **indefinite** if $Q(\mathbf{x})$ is positive for some $\mathbf{x}$'s in $\mathbb{R}^n$ and negative for others.

The graphs in Figure 5 show quadratic forms that are positive definite, negative definite, and indefinite.



(a) $z = x_1^2 + 2x_2^2$ (b) $z = -2x_1^2 - x_2^2$ (c) $z = x_1^2 - x_2^2$

Figure 5 Plots of (a) positive definite, (b) negative definite, and (c) indefinite quadratic forms.

It might seem difficult to classify a quadratic form, but it turns out that the eigenvalues of the matrix of a quadratic form tell the story.

THEOREM 11.4 ▶

Let *A* be an $n \times n$ symmetric matrix, and suppose that $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$. Then

(a) *Q* is positive definite exactly when *A* has only positive eigenvalues.

(b) *Q* is negative definite exactly when *A* has only negative eigenvalues.

(c) *Q* is indefinite exactly when *A* has positive and negative eigenvalues.

Proof Since *A* is a symmetric matrix, there exist matrices *P* and *D* such that $P^T A P = D$, where the columns of *P* are orthonormal eigenvectors of *A* and the diagonal entries of *D* are the eigenvalues $\lambda_1, \dots, \lambda_n$ of *A*. Since *P* is invertible, for a given $\mathbf{x}$ we can define $\mathbf{y} = P^{-1}\mathbf{x}$, so that $\mathbf{x} = P\mathbf{y}$. Then

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = (P\mathbf{y})^T A (P\mathbf{y}) = \mathbf{y}^T (P^T A P)\mathbf{y} = \mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

If the eigenvalues are all positive, then $Q(\mathbf{x})>0$ except when $\mathbf{y}=\mathbf{0}$, which implies $\mathbf{x}=\mathbf{0}$. Hence Q is positive definite. On the other hand, suppose that A has a nonpositive eigenvalue—say, $\lambda_1 \le 0$. If $\mathbf{y}$ has $y_1=1$ and the other components are 0, then for the corresponding $\mathbf{x} \ne \mathbf{0}$ we have

$$Q(\mathbf{x})=\lambda_1 \le 0$$

so that Q is not positive definite. This proves part (a). The other parts are similar and left as an exercise.

Example 6

Determine if $Q(\mathbf{x})=x_2^2+2x_1x_2+4x_1x_3+2x_2x_3$ is positive definite, negative definite, or indefinite.

Solution The matrix of this quadratic form is

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

The eigenvalues of A are $\lambda_1=-2$, $\lambda_2=0$, and $\lambda_3=3$. Thus, by Theorem 11.4, Q is indefinite.

Applying Theorem 11.4 requires knowing the eigenvalues of A, which can sometimes be hard to find. In the next section we will see how to use determinants to accomplish the same thing.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Evaluate $Q(\mathbf{x})$ for the given $\mathbf{x}_0$.

(a) Q(x)=3x12−4x22+14x1x2; x0=[−23]

(b) Q(x)=2x12+5x22−3x32+10x1x3−18x2x3; x0=[1−12]

2. Find a formula for the quadratic form with the given matrix *A*.

(a) A=[−3−4−47]

(b) A=[2−15−13252−4]

3. Find a matrix *A* such that Q(x)=xTAx.

(a) Q(x)=2x12+5x22−3x32+12x1x2

(b) Q(x)=−4x12+6x22−6x1x2−10x1x3+4x2x3

4. Determine if the quadratic form Q(x)=xTAx is positive definite, negative definite, indefinite, or none of these.

(a) A=[422−3]

(b) A=[100001010]

EXERCISES

Exercises 1–4: Evaluate Q(x) at x0.

1. Q(x)=x12−5x22+6x1x2; x0=[31]

2. Q(x)=x12+4x22−3x32+6x1x2; x0=[102]

3. Q(x)=3x12+x22−x32+6x1x3; x0=[031]

4. Q(x)=−2x12+7x22−8x1x2−4x1x3+2x2x3; x0=[031]

Exercises 5–12: Find a formula for the quadratic form with the given matrix *A*.

5. A=[4001]

6. A=[−3007]

7. A=[1332]

8. A=[5−2−20]

9. A=[10003000−2]

10. A=[21−3100−305]

11. A=[2000010000200003]

12. A=[5010020010030034]

Exercises 13–18: Find a matrix *A* such that Q(x)=xTAx.

13. Q(x)=x12−5x22+6x1x2

14. Q(x)=x12+4x22−3x32+6x1x2

15. Q(x)=3x12+x22−x32+6x1x3

16. Q(x)=−2x12+7x22−8x1x2−4x1x3+2x2x3

17. Q(x)=5x12−x22+3x32+6x1x3−12x2x3

18. Q(x)=x22+x32+2x1x2−4x1x3−8x2x3

Exercises 19–26: Determine if the quadratic form Q(x)=xTAx is positive definite, negative definite, indefinite, or none of these.

19. A=[1221]

20. A=[5111]

21. A=[222−1]

22. A=[588−1]

23. A=[010100001]

24. A=[001010101]

25. A=[0010010010000001]

26. A=[1100110000110011]

FIND AN EXAMPLE Exercises 27–34: Find an example that meets the given specifications.

27. A quadratic form Q(x) and a constant *c* such that Q(x)=c has no solutions.

28. A quadratic form Q(x) and a constant *c* such that Q(x)=c has exactly one solution.

29. A quadratic form Q(x) and a constant *c* such that the graph of Q(x)=c is two intersecting lines.

30. A quadratic form Q : R→R.

31. A quadratic form Q(x) that is also a linear transformation.

32. A quadratic form Q : R4→R that is indefinite.

33. A quadratic form Q : R3→R that is positive semidefinite but not positive definite.

34. A quadratic form $Q : R^2 \to R$ that is negative semidefinite but not negative definite.

TRUE OR FALSE Exercises 35–38: Determine if the statement is true or false, and justify your answer.

35.

(a) The matrix *A* of a quadratic form *Q* must be symmetric.

(b) If *Q* is a quadratic form, then $Q(x)=Q(-x)$.

36.

(a) The product of two quadratic forms is another quadratic form.

(b) If *A* is the matrix of a quadratic form *Q*, then A^{-1} is the matrix of $1/Q$.

37.

(a) If *Q* is a quadratic form, then

$$Q(x_1+x_2)=Q(x_1)+Q(x_2)$$

(b) If *A* is a diagonal matrix and the matrix of a quadratic form *Q*, then *Q* is positive definite.

38.

(a) If $Q_1(x)$ and $Q_2(x)$ are quadratic forms, then so is $Q_1(x)+Q_2(x)$.

(b) If *A* is the matrix of the quadratic form $Q(x)$, then cA is the matrix of the quadratic form $cQ(x)$.

39. Prove that a quadratic form $Q : R \to R$ cannot be indefinite.

40. Prove that if $Q(x)=x^TAx$ and $Q_T(x)=x^TA^Tx$, then $Q(x)=Q_T(x)$.

41. Show that $Q(x)=\|x\|^2$ is a quadratic form and give the matrix of the quadratic form. Then show that *Q* is positive definite.

42. Show that $Q(x)=\|Ax\|^2$ is a quadratic form and give the matrix of the quadratic form. Then show that *Q* is positive definite if and only if $null(A)=\{0\}$.

43. Prove that $Q(0)=0$ for every quadratic form *Q*.

44. In this exercise we show that if *P* is an orthogonal 2×2 matrix, then the transformation $Py=x$ is a rotation or reflection.

(a) Prove that $\|Py\|=\|y\|$ for y in R^2.

(b) Prove that if x and y are in R^2, then the angle between x and y is the same as the angle between Px and Py. HINT: Recall the formula

$$\cos(\theta)=\frac{x\cdot y}{\|x\|\,\|y\|}$$

(c) Combine (a) and (b) to explain why $Py=x$ is a rotation or reflection.

11.2 Positive Definite Matrices

Definition 11.3 in Section 11.1 gives, for a quadratic form, the meaning of positive definite, positive semidefinite, and so on. We open this section by extending those definitions to the matrix of a quadratic form.

DEFINITION 11.5 ▶

Positive Definite

A symmetric $n \times n$ matrix A is **positive definite** if the corresponding quadratic form $Q(x)=x^TAx$ is positive definite. Analogous definitions apply for **negative definite** and **indefinite**.

By Theorem 11.4 in Section 11.1, we know that one way to determine if a symmetric matrix A is positive definite is to examine the eigenvalues of A. If all the eigenvalues are positive, then A is positive definite. The shortcoming of this approach is that it can be difficult to find the eigenvalues, so it would be useful to have another method available. The next theorem gives us a start in this direction.

THEOREM 11.6 ▶

If A is a symmetric positive definite matrix, then A is nonsingular and $\det(A)>0$.

Proof Since A is positive definite, by Theorem 11.4 in Section 11.1, all of the eigenvalues of A are positive and hence nonzero. Thus, by The Unifying Theorem, Version 8 (Section 6.1), A is nonsingular. Next, recall that

$$\det(A)=\lambda_1\lambda_2\cdots\lambda_n$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A (see Exercise 66 in Section 6.1). Since the eigenvalues are positive, so is their product, and therefore $\det(A)>0$. ■

Example 1

Show that Theorem 11.6 holds for the matrix

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

Solution The eigenvalues of A are

$$\lambda_1=3-\sqrt{2}\approx 1.586, \quad \lambda_2=3, \quad \lambda_3=3+\sqrt{2}\approx 4.414$$

Since the eigenvalues are all positive, by Theorem 11.4 in Section 11.1 the associated quadratic form, and hence the matrix A, is positive definite. We also have $\det(A)=21>0$. This implies A is nonsingular and shows that Theorem 11.6 is true for this matrix.

It would be nice if the converse of Theorem 11.6 was true, so that $\det(A)>0$ would be enough to guarantee that A is positive definite, but that is not true. (See the example in the margin.) However, let's not abandon determinants yet. If

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

▶ Suppose that
$A=\begin{bmatrix}-2 & 0\\ 0 & -1\end{bmatrix}$

Then $\det(A)=2$, but *A* has negative eigenvalues −2 and −1 and so is not positive definite.

Leading Principal Submatrix

then the **leading principal submatrices** of *A* are given by

$$A_1=\begin{bmatrix}a_{11}\end{bmatrix},\quad A_2=\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix},\quad A_3=\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

and so on through $A_n=A$.

THEOREM 11.7 ▶

A symmetric positive definite matrix *A* has leading principal submatrices $A_1, A_2, \ldots, A_n$ that are also positive definite.

Proof Let $1\le m\le n$, and suppose that $x_m\ne 0$ is in $\mathbb{R}^m$. If we set

$$x=\begin{bmatrix}x_m\\ 0\\ \vdots\\ 0\end{bmatrix} \text{ in } \mathbb{R}^n$$

then we have

$$x_m^T A_m x_m = x^T A x > 0$$

because *A* is positive definite. Since this works for any nonzero x_m, it follows that A_m is positive definite. ■ ■

Example 2

Show that the leading principal submatrices of the positive definite matrix *A* in Example 1 are also positive definite.

Solution Since $A_1=[\,3]$ has associated quadratic form $Q(x)=3x^2$ that is positive definite, then A_1 is positive definite. The matrix

$$A_2=\begin{bmatrix}3 & -1\\ -1 & 3\end{bmatrix}$$

has positive eigenvalues $\lambda_1=2$ and $\lambda_2=4$. Hence A_2 is also positive definite by Theorem 11.4 in Section 11.1. We have already shown that $A_3=A$ is positive definite, so all leading principal submatrices of A are positive definite.

Combining Theorem 11.6 and Theorem 11.7 shows that if A is positive definite, then the leading principal submatrices satisfy $\det(A_1)>0$, $\det(A_2)>0\ldots$, $\det(A_n)>0$. Interestingly, the converse is also true.

THEOREM 11.8 ▶

A symmetric matrix A is positive definite if and only if the leading principal submatrices satisfy

$$\det(A_1)>0,\ \det(A_2)>0,\ \ldots,\ \det(A_n)>0 \tag{1}$$

The proof of Theorem 11.8 is given at the end of the section.

Example 3

Determine if the matrix

$$A=\begin{bmatrix}4 & 8 & -8\\ 8 & 25 & 11\\ -8 & 11 & 98\end{bmatrix}$$

is positive definite.

Solution We have

$$\det(A_1)=4, \quad \det(A_2)=|48825|=36, \quad \det(A_3)=|48-882511-81198|=36$$

Since all three determinants are positive, A is positive definite.

LU Factorization Revisited

In Section 3.4 we discussed LU factorization, which involves expressing a matrix A as the product $A=LU$, where L is lower triangular and U is upper triangular. There we showed that A has an LU factorization if A can be reduced to echelon form without the use of row interchanges. Previously, the only way to tell if this was true was to try it, but here we give a class of matrices that are guaranteed to have LU factorizations.

THEOREM 11.9 ▶

Suppose that the leading principal submatrices of a symmetric matrix A satisfy (1). Then A can be reduced to row echelon form without using row interchanges, and the pivot elements will all be positive.

Proof We proceed by induction on n, where A is an $n\times n$ symmetric positive definite matrix. First, if $n=1$, then $A=[a_{11}]$ is automatically in row echelon form, so no row interchanges are required. Moreover, we have

$$a_{11}=\det(A)>0$$

so that the sole pivot element is positive.

Now suppose that the theorem holds for $(n-1)\times(n-1)$ symmetric matrices, and let A be an $n\times n$ symmetric matrix that satisfies (1). We can partition *A* as

$$A = \left[\begin{array}{ccc|c} & & & a_{1n} \\ & A_{n-1} & & \vdots \\ & & & a_{(n-1)n} \\ \hline a_{n1} & \cdots & a_{n(n-1)} & a_{nn} \end{array}\right]$$

where the leading principal submatrix A_{n-1} is symmetric and satisfies the induction hypothesis. When reducing *A* to row echelon form, the values of all but the last pivot are dictated entirely by elements of A_{n-1}. Thus, by the induction hypothesis, we can reduce *A* to the form

$$A^* = \begin{bmatrix} a^*_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a^*_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & a^*_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$$

where the pivots $a^*_{11}, \ldots, a^*_{(n-1)(n-1)}$ are all positive and no row interchanges are required. Because there are no row interchanges, the determinant is unchanged, so that

$$\det(A) = \det(A^*) = \det(A_{n-1}) \cdot a_{nn} \ \Rightarrow \ a_{nn} = \frac{\det(A)}{\det(A_{n-1})}$$

Since both $\det(A) > 0$ and $\det(A_{n-1}) > 0$, we have $a_{nn} > 0$, which completes the proof. ■ ■

The matrix in Example 3 satisfies the hypotheses of Theorem 11.9, so it must have an LU factorization.

Example 4

Find an LU factorization for the matrix in Example 3,

$$A = \begin{bmatrix} 4 & 8 & -8 \\ 8 & 25 & 11 \\ -8 & 11 & 98 \end{bmatrix}$$

Solution Since Section 3.4 contains several examples showing how to find the LU factorization, some details are omitted here. Recall that we obtain *U* by reducing *A* to echelon form, and build up *L* one column at a time as we transform *A*.

Step 1a: Take the first column of *A*, divide each entry by the pivot 4, and use the resulting values to form the first column of *L*.

A=[48−882511−81198] ⇒ L=[1••2••−2••]

Step 1b: Perform row operations as usual to introduce zeros down the first column of *A*.

A=[48−882511−81198] ∼ [48−8092702782]=A1

Step 2a: Take the second column of A1, starting down from the pivot entry 9, and divide each entry by the pivot. Use the resulting values to form the lower portion of the second column of *L*.

A1=[48−8092702782] ⇒ L=[1••21•−23•]

Step 2b: Perform row operations as usual to introduce zeros down the second column of A1.

A1=[48−8092702782] ∼ [48−80927001]=A2

Step 3: Set *U* equal to A2, and finish filling in *L*.

L=[100210−231] and U=[48−80927001]

Standard matrix multiplication can be used to verify that A=LU.

In Section 3.4 we extended the LU factorization to A=LDU, where *L* is as before, *D* is diagonal, and *U* is an upper triangular matrix with 1's along the diagonal. Although the LU factorization is not unique, the LDU factorization is unique (see Exercise 35).

To find the LDU factorization, we start by finding the LU factorization. Once that is done, we write *U* as the product of a diagonal matrix and an upper triangular matrix,

$$U=\begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nn} \end{bmatrix} = \begin{bmatrix} u_{11} & 0 & 0 & \cdots & 0 \\ 0 & u_{22} & 0 & \cdots & 0 \\ 0 & 0 & u_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nn} \end{bmatrix} \begin{bmatrix} 1 & \frac{u_{12}}{u_{11}} & \frac{u_{13}}{u_{11}} & \cdots & \frac{u_{1n}}{u_{11}} \\ 0 & 1 & \frac{u_{23}}{u_{22}} & \cdots & \frac{u_{2n}}{u_{22}} \\ 0 & 0 & 1 & \cdots & \frac{u_{3n}}{u_{33}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

with the left matrix being D and the right the new U. For instance, taking U from the factorization in Example 4, we have

$$\begin{bmatrix} 4 & 8 & -8 \\ 0 & 9 & 27 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} = DU$$

The matrix U looks familiar—it is L^T. This is not a coincidence.

THEOREM 11.10 ▸

A symmetric matrix A that satisfies (1) can be uniquely factored as $A=LDL^T$, where L is lower triangular with 1's on the diagonal, and D is a diagonal matrix with all positive diagonal entries.

Proof The matrix A can be expressed uniquely as $A=LDU$ (see Exercise 35), and since A is symmetric, we have

$$LDU=A=A^T=(LDU)^T=U^TD^TL^T=U^TDL^T$$

Because the factorization is unique, it follows that $U=L^T$. That the diagonal entries of D are positive follows from Theorem 11.9. ■ ■

Since the diagonal entries of D are all positive in the factorization given in Theorem 11.10, we can define the matrix

$$D^{1/2}=\begin{bmatrix} u_{11} & 0 & 0 & \cdots & 0 \\ 0 & u_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nn} \end{bmatrix}$$

If we set $L_c=LD^{1/2}$, then we have

$$A=LDL^T=LD^{1/2}D^{1/2}L^T=(LD^{1/2})(LD^{1/2})^T=L_cL_c^T$$

Cholesky Decomposition

The factorization A=LcLcT is called the **Cholesky decomposition** of *A*. By Theorem 11.10, every symmetric matrix *A* satisfying (1) has a Cholesky decomposition.

Example 5

Find the Cholesky decomposition for the matrix *A* in Example 4.

▶ Roughly speaking, the Cholesky decomposition can be thought of as the square root of a matrix. When it exists, it can be used to find the solution to a linear system with significantly more efficiency than LU decomposition.

Solution We have

Lc=LD1/2=[100210−231] [400090001] =
[100210−231] [200030001]=[200430−491]

We can verify the decomposition by computing

LcLcT=[200430−491] [24−4039001]=[48−882511−81198]=A

Proof of Theorem 11.8

We are now in the position to prove Theorem 11.8. Recall the statement of the theorem.

THEOREM 11.8 ▶

A symmetric matrix *A* is positive definite if and only if the leading principal submatrices satisfy

det(A1)>0, det(A2)>0, …, det(An)>0 (1)

Proof As noted earlier, combining Theorem 11.6 and Theorem 11.7 shows that if A is positive definite, then the leading principal submatrices satisfy $\det(A_1)>0$, $\det(A_2)>0$, … , $\det(A_n)>0$. This completes one direction of the proof.

To complete the second direction of the proof, suppose that A is a symmetric matrix and that the leading principal submatrices satisfy $\det(A_1)>0$, $\det(A_2)>0$, … , $\det(A_n)>0$. Then A has a Cholesky decomposition $A=L_cL_c^T$. Moreover, since $\det(A)>0$, it follows that A is nonsingular and hence L_c^T is also nonsingular. Thus if $x\neq 0$, then $L_c^Tx\neq 0$. Therefore

$$x^TAx=x^TL_cL_c^Tx=(L_c^Tx)^T(L_c^Tx)=\|L_c^Tx\|^2>0$$

so that A is positive definite. ■ ■

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the principal submatrices of A.
   (a) A=[455−2]
   (b) A=[−12−4201−413]
2. Determine if the given matrix is positive definite.
   (a) A=[1225]
   (b) A=[11−1143−1311]

EXERCISES

Exercises 1–6: Find the principal submatrices of the given matrix.

1. A=[3557]

2. A=[1−6−62]
3. A=[14−3402−325]
4. A=[3020−4121−1]
5. A=[210−11341040−2−11−21]
6. A=[1−230−242−332310−310]

Exercises 7–12: Determine if the given matrix is positive definite.

7. A=[211−2]
8. A=[1335]
9. A=[12−1251−1111]
10. A=[1−32−310−72−76]
11. A=[1121121021631036]
12. A=[1−221−25−1−32−114−11−3−13]

Exercises 13–16: Show that the given matrix is positive definite, and then find the LU factorization.

13. A=[1−2−25]
14. A=[13310]
15. A=[1−22−25−52−56]
16. A=[11−1110−4−1−411]

Exercises 17–20: Show that the given matrix is positive definite, and then find the LDU factorization.

17. A=[1−2−28]
18. A=[4−4−45]
19. A=[1323108289]
20. A=[111151115]

Exercises 21–24: Show that the given matrix is positive definite, and then find the Cholesky decomposition.

21. A=[1225]
22. A=[42210]

23. A=[1232563610]

24. A=[1414257176]

FIND AN EXAMPLE Exercises 25–30: Find an example that meets the given specifications.

25. A 2×2 matrix *A* such that det(A1)<0 but det(A2)>0.

26. A 2×2 matrix *A* such that det(A1)>0 but det(A2)<0.

27. A 3×3 matrix *A* such that det(A1)<0 and det(A3)<0, but det(A2)>0.

28. A 3×3 matrix *A* such that det(A1)>0, but det(A2)<0, and det(A3)<0.

29. A 3×3 matrix *A* such that det(A)>0 but *A* is not positive definite.

30. A 3×3 matrix *A* such that det(A)<0 but *A* is not negative definite.

TRUE OR FALSE Exercises 31–34: Determine if the statement is true or false, and justify your answer.

31.

(a) A symmetric matrix *A* is positive definite if and only if det(A)>0.

(b) If LcLcT is the Cholesky decomposition of *A*, then LcTLc is the Cholesky decomposition of AT.

32.

(a) For any matrix *A*, the product ATA is positive definite.

(b) If *A* is a square upper triangular matrix with positive diagonal entries, then *A* is positive definite.

33.

(a) If *A* and *B* are n×n positive definite matrices, then so is AB.

(b) If *A* and *B* are n×n positive definite matrices, then so is A+B.

34.

(a) If *A* is a positive definite matrix, then so is A−1.

(b) If the determinants of the leading principal submatrices of a symmetric matrix *A* are all negative, then *A* is negative definite.

35. Let A=L1D1U1 and A=L2D2U2 be two LDU factorizations of *A*. Prove that L1=L2, D1=D2, and U1=U2, which shows that the LDU factorization is unique.

11.3 Constrained Optimization

In this section we consider the following problem: Find the maximum and/or minimum value of a quadratic form $Q(x)$, where x ranges over a set of vectors x that satisfy some constraint.

Our problem is easiest to solve when the quadratic form has no cross-product terms, so let's start by looking at an example of that case.

Example 1

Find the maximum and minimum values of $Q(x)=4x_1^2-3x_2^2+7x_3^2$, subject to the constraint $||x||=1$.

Solution First, note that if $\|x\|=1$, then $\|x\|^2=1$ and therefore

$$x_1^2+x_2^2+x_3^2=1$$

Each of x_1^2, x_2^2, and x_3^2 is nonnegative, so that

$$Q(x)=4x_1^2-3x_2^2+7x_3^2\le 7x_1^2+7x_2^2+7x_3^2=7(x_1^2+x_2^2+x_3^2)=7$$

Thus $Q(x)\le 7$. Moreover, if $x_1=x_2=0$ and $x_3=1$, then $\|x\|=1$ and $Q(x)=7$. Therefore, subject to the constraint $\|x\|=1$, the maximum value is $Q(x)=7$.

A similar argument can be used to show that subject to the same constraint, the minimum value is equal to the minimum coefficient, so that the minimum value is $Q(x)=-3$.

Example 1 illustrates a more general fact that is described in the next theorem.

THEOREM 11.11 ▶

Suppose that $Q : R^n \to R$ is a quadratic form

$$Q(x)=q_1x_1^2+q_2x_2^2+\cdots+q_nx_n^2$$

that has no cross-product terms. Let q_i and q_j be the maximum and minimum, respectively, of the coefficients $q_1, q_2, \ldots, q_n$. Then subject to the constraint $\|x\|=1$, we have

(a) The maximum value of $Q(x)$ is q_i, attained at $x=e_i$.

(b) The minimum value of $Q(x)$ is q_j, attained at $x=e_j$.

▶ Recall that
$e_k=[0 \vdots 1 \vdots 0]$ ← entry *k*

The proof is similar to the solution to Example 1 and is left as an exercise.

Example 2

Find the maximum and minimum values of

$$Q(x)=2x_1^2-4x_2^2+5x_3^2-x_4^2$$

subject to the constraint $\|x\|=1$.

Solution By Theorem 11.11, the maximum value of $Q(x)$ is 5 and the minimum value is −4. The maximum and minimum values are attained at, respectively,

$$x_3=\begin{bmatrix}0\\0\\1\\0\end{bmatrix} \quad \text{and} \quad x_2=\begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

Theorem 11.11 tells us what to do with a quadratic form that is free of cross-product terms, so now we consider general quadratic forms $Q : \mathbb{R}^n \to \mathbb{R}$.

THEOREM 11.12 ▶

Let $Q(x)=x^TAx$, where *A* is a symmetric $n\times n$ matrix. Suppose that *A* has eigenvalues $\lambda_1\le\lambda_2\le\ldots\le\lambda_n$, and let $u_1,\ldots,u_n$ be the associated normalized eigenvectors. Then subject to the constraint $\|x\|=1$, we have

(a) The maximum value of $Q(x)$ is λ_n, attained at $x=u_n$.

(b) The minimum value of $Q(x)$ is λ_1, attained at $x=u_1$.

Proof Recall from Section 11.1 that for such a matrix *A*, there exists an orthogonal matrix *P* (with columns that are eigenvectors of *A*) and a diagonal matrix *D* (diagonal entries are the eigenvalues of *A*) such that if $x=Py$, then

$$x^TAx=y^TDy$$

Since *D* is diagonal, the quadratic form y^TDy has no cross-product terms, so that Theorem 11.11 applies. Moreover, if $\|y\|=1$, then

$$\|x\|=\|Py\|=1$$

(See Exercise 60, Section 8.3.) Therefore the maximum value of x^TAx subject to the constraint $\|x\|=1$ is equal to the maximum value of y^TDy subject to the constraint $\|y\|=1$. (The same holds if "maximum" is replaced with "minimum.")

Returning to Theorem 11.11, the matrix of the quadratic form is diagonal with diagonal entries $q_1,\ldots,q_n$. For our matrix *D*, the diagonal entries are the eigenvalues of *A*, so that Theorem 11.11 applies to the set $\{\lambda_1,\ldots,\lambda_n\}$, which yields the claimed maximum and minimum values. Observing that $Pe_k=u_k$ completes the proof. ■

Example 3

Find the maximum and minimum values of

$$Q(\mathbf{x})=x_2^2+2x_1x_2+4x_1x_3+2x_2x_3$$

subject to the constraint $\|\mathbf{x}\|=1$.

Solution The matrix of this quadratic form is

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

Using our usual methods, we find that the eigenvalues of *A* are $\lambda_1=-2$, $\lambda_2=0$, and $\lambda_3=3$. Thus, by Theorem 11.12, the maximum value of $Q(\mathbf{x})$ is 3 and the minimum value is -2. The maximum and minimum values are attained at the eigenvectors

$$\text{Maximum attained at } \mathbf{u}_3=\begin{bmatrix}\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix} \quad \text{Minimum attained at } \mathbf{u}_1=\begin{bmatrix}-\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix}$$

Varying the Constraint

In the next two examples, we modify the constraint requirements. In both cases, the problem can be solved by changing variables and then using earlier methods.

Example 4

Find the maximum and minimum values of

$$Q(\mathbf{x})=5x_2^2+8x_2^2+4x_1x_2$$

subject to the constraint $\|x\|=c$, where $c>0$ is a positive constant.

Solution Let's start by solving the problem with the constraint $\|x\|=1$. The matrix of the quadratic form is

$$A=\begin{bmatrix}5 & 2\\ 2 & 8\end{bmatrix}$$

which has eigenvalues $\lambda_1=4$ and $\lambda_2=9$. Hence, subject to $\|x\|=1$, the maximum value of $Q(x)$ is 9 and the minimum value is 4. To finish, note that $\|x\|=1$ if and only if $\|cx\|=c$ and $Q(cx)=c^2Q(x)$ (see Exercise 37). Therefore, subject to the constraint $\|x\|=c$, the maximum value of $Q(x)$ is $9c^2$ and the minimum value is $4c^2$.

Example 5

Find the maximum and minimum values of

$$Q(x)=2x_2^2+5x_2^2+4x_1x_2$$

subject to the constraint $4x_1^2+25x_2^2=100$.

Solution The first step is to use a change of variables to adjust the constraint equation so that it once again describes unit vectors. Let

$$w_1=\frac{x_1}{5} \quad \text{and} \quad w_2=\frac{x_2}{2}$$

so that the constraint equation becomes

$$4(5w_1)^2+25(2w_2)^2=100 \;\Rightarrow\; w_1^2+w_2^2=1$$

and hence we have the constraint $\|w\|=1$. With this variable change, the quadratic form becomes

$$Q(x)=Q\left(\begin{bmatrix}x_1\\ x_2\end{bmatrix}\right)=Q\left(\begin{bmatrix}5w_1\\ 2w_2\end{bmatrix}\right)$$

$$=2(5w1)2+5(2w2)2+4(5w1)(2w2)$$

$$=50w12+20w22+40w1w2$$

The matrix of the quadratic form 50w12+20w22+40w1w2 is

$$A=[50202020]$$

which has eigenvalues λ1=10 and λ2=60. Hence the maximum value of Q(x) is 60 and the minimum value is 10. To determine where the maximum and minimum values are attained, we note that the normalized eigenvectors of *A* are

$$λ1=10 \Rightarrow w1=[1525], \qquad λ2=60 \Rightarrow w2=[2515]$$

Thus the maximum of 50w12+20w22+40w1w2 subject to ‖w‖=1 is attained at

$$[2515]=w2=[w1w2]=[x1/5x2/2]$$

so that the maximum of Q(x)=2x22+5x22+4x1x2 subject to 4x12+25x22=100 is attained at

$$x2=[5w12w2]=[255]$$

By a similar argument, the minimum is attained at x1=[−525].

Adding Orthogonality to the Constraint

In some applications, it is handy to be able to constrain x so that both ‖x‖=1 and x is orthogonal to un, the eigenvector associated with the largest eigenvalue λn.

Example 6

Find the maximum value of $Q(x)=4x_1^2-3x_2^2+7x_3^2$, subject to the constraints $\|x\|=1$ and $x\cdot u_3=0$, where $u_3=(0, 0, 1)$.

Solution This is the quadratic form from Example 1, where we asked for the maximum subject only to the constraint $\|x\|=1$. Since we have added another constraint, the maximum can be no larger (and possibly is smaller) than the maximum found before.

Due to the extra constraint, we cannot apply Theorem 11.11, but we can apply the same line of reasoning as used in Example 1. The constraint $x\cdot u_3=0$ implies that the third component $x_3=0$, so that the two constraints together impose the requirement $x_1^2+x_2^2=1$. Therefore we have

$$Q(x)=4x_1^2-3x_2^2+7x_3^2=4x_1^2-3x_2^2\le 4x_1^2+4x_2^2=4(x_1^2+x_2^2)=4$$

This puts an upper bound of 4 on the maximum. Moreover, if $x=(1, 0, 0)$, then x satisfies the constraints and $Q(x)=4$. Thus the maximum value is 4, attained when $x=(1, 0, 0)$.

In Example 6 the vector $u_3=(0, 0, 1)$ is the eigenvector associated with the largest eigenvalue $\lambda_3=7$ of *A*, the matrix of the quadratic form. Note that the maximum 4 is equal to the *second-largest* eigenvalue of *A* and is attained when x is an associated eigenvector. This is not a coincidence.

THEOREM 11.13 ▶

Let $Q(x)=x^TAx$, where *A* is a symmetric $n\times n$ matrix with eigenvalues $\lambda_1\le\lambda_2\le\cdots\le\lambda_n$ and associated orthonormal eigenvectors $u_1, u_2,\dots, u_n$. Then subject to the constraints $\|x\|=1$ and $x\cdot u_n=0$,

(a) The maximum value of $Q(x)$ is λ_{n-1}, attained at $x=u_{n-1}$.

(b) The minimum value of $Q(x)$ is λ_1, attained at $x=u_1$.

▶ Theorem 11.13 assumes that $n>1$.

Proof Since the eigenvectors $u_1, u_2, \ldots, u_n$ are orthonormal, they form a basis for $\mathbb{R}^n$. Moreover, for a given x if $c_i = x \cdot u_i$ for $i=1, \ldots, n$, then

$$x = c_1u_1 + c_2u_2 + \cdots + c_nu_n$$

Since one constraint is that $x \cdot u_n = 0$, it follows that $c_n = 0$, so that

$$x = c_1u_1 + \cdots + c_{n-1}u_{n-1}$$

Because $u_1, \ldots, u_{n-1}$ are orthonormal, we have

$$\|x\|^2 = c_1^2 + \cdots + c_{n-1}^2$$

(See Exercise 46 of Section 8.2.) Since $\|x\| = 1$, we have $c_2^2 + \cdots + c_n^2 = 1$. Thus

$$\begin{aligned} x^TAx &= x^TA(c_1u_1 + \cdots + c_{n-1}u_{n-1}) \\ &= (c_1u_1 + \cdots + c_{n-1}u_{n-1}) \cdot (c_1\lambda_1u_1 + \cdots + c_{n-1}\lambda_{n-1}u_{n-1}) \\ &= c_1^2\lambda_1 + \cdots + c_{n-1}^2\lambda_{n-1} \\ &\le c_1^2\lambda_{n-1} + \cdots + c_{n-1}^2\lambda_{n-1} \\ &= \lambda_{n-1}(c_1^2 + \cdots + c_{n-1}^2) = \lambda_{n-1} \end{aligned}$$

This shows that $Q(x) \le \lambda_{n-1}$. Furthermore, if $x = u_{n-1}$, then $c_{n-1} = 1$ and $c_1 = \cdots = c_{n-2} = 0$, and hence

$$Q(u_{n-1}) = \lambda_{n-1}$$

Therefore the maximum value of $Q(x)$ is λ_{n-1}, attained when $x = u_{n-1}$. The minimum value is not changed by the constraint $x \cdot u_n = 0$, so (b) follows from Theorem 11.12. ■ ■

Example 7

Find the maximum and minimum values of

$$Q(\mathbf{x}) = x_2^2 + 2x_1x_2 + 4x_1x_3 + 2x_2x_3$$

subject to the constraints $\|\mathbf{x}\| = 1$ and $\mathbf{x} \cdot \mathbf{u}_3 = 0$, where $\mathbf{u}_3 = (1, 1, 1)$.

Solution This is the quadratic form in Example 3, which has matrix

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

There we showed that the eigenvalues of A are $\lambda_1 = -2$, $\lambda_2 = 0$, and $\lambda_3 = 3$. Also note that $A\mathbf{u}_3 = 3\mathbf{u}_3$, so that $\mathbf{u}_3$ is an eigenvector associated with the largest eigenvalue $\lambda_3 = 3$. Therefore by Theorem 11.13, subject to our constraints, the maximum value of $Q(\mathbf{x})$ is $\lambda_2 = 0$ and the minimum value is $\lambda_1 = -2$.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find the maximum and minimum values of the quadratic form $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\| = 1$.
   (a) $Q(\mathbf{x}) = 5x_1^2 - 2x_2^2$
   (b) $Q(\mathbf{x}) = 2x_1^2 - x_2^2 - 4x_3^2$
   (c) $Q(\mathbf{x}) = x_1^2 + 2x_2^2 + 6x_1x_2$
   (d) $Q(\mathbf{x}) = x_1^2 + 3x_2^2 + 2x_3^2 + 6x_1x_3$
2. Find the maximum and minimum values of the quadratic form $Q(\mathbf{x})$ subject to the given constraint.
   (a) $Q(\mathbf{x}) = x_1^2 + 9x_2^2 + 6x_1x_2$; $\|\mathbf{x}\| = 2$
   (b) $Q(\mathbf{x}) = x_2^2 + 2x_1x_3$; $\|\mathbf{x}\| = 10$
   (c) $Q(\mathbf{x}) = x_1^2 + x_2^2 + 8x_1x_2$; $4x_1^2 + 9x_2^2 = 100$
   (d) $Q(\mathbf{x}) = 4x_1^2 + 9x_2^2 + 4x_1x_2$; $x_1^2 + 9x_2^2 = 9$

EXERCISES

Exercises 1–6: Find the maximum and minimum values of the quadratic form $Q(x)$ subject to the constraint $\|x\|=1$.

1. $Q(x)=2x_1^2-3x_2^2$
2. $Q(x)=6x_1^2+5x_2^2$
3. $Q(x)=3x_1^2-3x_2^2-5x_3^2$
4. $Q(x)=-x_1^2+4x_2^2+8x_3^2$
5. $Q(x)=x_1^2-x_2^2-4x_3^2+2x_4^2$
6. $Q(x)=3x_1^2-4x_2^2-2x_3^2+x_4^2$

Exercises 7–12: Find the maximum and minimum values of the quadratic form $Q(x)$ subject to the constraint $\|x\|=1$.

7. $Q(x)=4x_1^2+x_2^2+4x_1x_2$
8. $Q(x)=3x_1^2+3x_2^2+8x_1x_2$
9. $Q(x)=x_2^2+2x_1x_3$
10. $Q(x)=x_1^2+4x_2^2+x_3^2+4x_2x_3$
11. $Q(x)=x_1^2+2x_1x_2+2x_1x_3+2x_2x_3+x_2^2+x_3^2$
12. $Q(x)=x_1^2+4x_1x_2+6x_1x_3+6x_2x_3+x_2^2$ (HINT: The eigenvalues are $\lambda=-3,\ -1,\ 6$.)

Exercises 13–16: Find the maximum and minimum values of the quadratic form $Q(x)$ subject to the given constraint.

13. $Q(x)=x_1^2+4x_2^2+4x_1x_2$; $\|x\|=2$
14. $Q(x)=3x_1^2+3x_2^2+8x_1x_2$; $\|x\|=0.5$
15. $Q(x)=x_1^2+2x_2x_3$; $\|x\|=10$
16. $Q(x)=4x_1^2+x_2^2+x_3^2+4x_1x_3$; $\|x\|=5$

Exercises 17–20: Find the maximum and minimum values of the quadratic form $Q(x)$ subject to the given constraint.

17. $Q(x)=4x_1^2+x_2^2+4x_1x_2$; $4x_1^2+25x_2^2=100$
18. $Q(x)=2x^2_1+2x_2^2+10x_1x_2$; $9x_1^2+16x_2^2=144$

19. $Q(\mathbf{x})=4x_1^2+4x_2^2+6x_1x_2$; $9x_1^2+x_2^2=9$

20. $Q(\mathbf{x})=3x_1^2+3x_2^2+8x_1x_2$; $x_1^2+4x_2^2=4$

Exercises 21–24: Find the maximum and minimum values of the quadratic form $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\|=1$ and $\mathbf{x}\cdot\mathbf{u}_3=0$.

21. $Q(\mathbf{x})=4x_1^2+x_2^2+3x_3^2$; $\mathbf{u}_3=(1, 0, 0)$

22. $Q(\mathbf{x})=-2x_1^2+x_2^2-5x_3^2$; $\mathbf{u}_3=(0, 1, 0)$

23. $Q(\mathbf{x})=x_1^2+4x_2^2+x_3^2+4x_2x_3$; $\mathbf{u}_3=(0, 2, 1)$

24. $Q(\mathbf{x})=x_1^2+4x_1x_2+6x_1x_3+6x_2x_3+x_2^2$; $\mathbf{u}_3=(1, 1, 1)$ (HINT: Eigenvalues are $\lambda=-3,-1, 6$.)

FIND AN EXAMPLE Exercises 25–30: Find an example that meets the given specifications.

25. A quadratic form $Q:\mathbb{R}^2\to\mathbb{R}$ that has maximum value 5, subject to the constraint that $\|\mathbf{x}\|=1$.

26. A quadratic form $Q:\mathbb{R}^3\to\mathbb{R}$ that has minimum value 4, subject to the constraint that $\|\mathbf{x}\|=1$.

27. A quadratic form $Q:\mathbb{R}^2\to\mathbb{R}$ that has maximum value 6 and minimum value 1, subject to the constraint that $\|\mathbf{x}\|=1$.

28. A quadratic form $Q:\mathbb{R}^3\to\mathbb{R}$ that has maximum value -2 and minimum value -7, subject to the constraint that $\|\mathbf{x}\|=1$.

29. A quadratic form $Q:\mathbb{R}^2\to\mathbb{R}$ that has maximum value 4 and minimum value -1, subject to the constraint that $\|\mathbf{x}\|=3$.

30. A quadratic form $Q:\mathbb{R}^3\to\mathbb{R}$ that has maximum value 8 and minimum value -3, subject to the constraint that $\|\mathbf{x}\|=4$.

TRUE OR FALSE Exercises 31–34: Determine if the statement is true or false, and justify your answer.

31.

(a) A quadratic form $Q:\mathbb{R}^n\to\mathbb{R}$ has a maximum and minimum when constrained to $\|\mathbf{x}\|=1$.

(b) If M is the maximum value of a quadratic form $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\|=1$, then cM is the maximum value of $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\|=c$.

32.

(a) A quadratic form cannot have the same maximum and minimum values unless the matrix of the quadratic form is 1×1.

(b) If a quadratic form is positive definite, then it cannot have an unconstrained maximum value.

33.

(a) If *M* and *m* are, respectively, the maximum and minimum values of a quadratic form $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\|=1$, then $m<M$.

(b) If $\mathbf{x}_1$ maximizes a quadratic form $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\|=1$, then so does $-\mathbf{x}_1$.

34.

(a) The minimum value of a quadratic form $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\|=2$ must be less than the minimum value of $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\|=1$.

(b) It is possible for a quadratic form to attain its constrained maximum at more than one point.

35. Prove Theorem 11.11: Suppose that $Q : R^n\to R$ is a quadratic form

$$Q(\mathbf{x})=q_1x_1^2+q_2x_2^2+\cdots+q_nx_n^2$$

that has no cross-product terms. Let q_i and q_j be the maximum and minimum, respectively, of the coefficients $q_1, q_2,\ldots, q_n$. Then subject to the constraint $\|\mathbf{x}\|=1$, we have

(a) Maximum value of $Q(\mathbf{x})=q_i$, attained at $\mathbf{x}=\mathbf{e}_i$.

(b) Minimum value of $Q(\mathbf{x})=q_j$, attained at $\mathbf{x}=\mathbf{e}_j$.

36. Prove the following generalized version of Theorem 11.12: Let *A* be a symmetric $n\times n$ matrix, and let $Q : R^n\to R$ be the quadratic form

$$Q(\mathbf{x})=\mathbf{x}^TA\mathbf{x}$$

If *A* has eigenvalues $\lambda_1\le\ldots\le\lambda_n$, then subject to the constraint $\|\mathbf{x}\|=c$ $(c>0)$, the maximum value of $Q(\mathbf{x})$ is $c^2\lambda_n$ and the minimum value is $c^2\lambda_1$.

37. Prove that if $Q(\mathbf{x})$ is a quadratic form and *c* is a constant, then $Q(c\mathbf{x})=c^2Q(\mathbf{x})$.

38. Prove the following extension of Theorem 11.13. Let *A* be a symmetric $n\times n$ matrix with eigenvalues $\lambda_1\le\lambda_2\le\ldots\le\lambda_n$ and associated orthonormal eigenvectors $u_1, u_2, \ldots, u_n$. Then subject to the constraints

$$\|x\|=1,\ \ x\cdot u_n=0,\ \ x\cdot u_{n-1}=0, \ldots,\ \ x\cdot u_{n-j+1}=0$$

prove that the maximum value of the quadratic form $Q(x)=x^TAx$ is λ_{n-j}, with the maximum attained at u_{n-j}.

11.4 Complex Vector Spaces

In this section we define *complex vector spaces*, which are a fairly straightforward extension of the real vector spaces defined in Definition 7.1 in Section 7.1. In fact, the only change when moving to complex vector spaces is that we allow the scalars to be complex numbers.

▶ Here we assume a basic understanding of the properties of complex numbers. A review of some of these properties is given in Section 6.3.

DEFINITION 11.14 ▶

Complex Vector Space, Vector

A **complex vector space** consists of a set V of **vectors** together with operations of addition and scalar multiplication on the vectors that satisfy each of the following:

(1) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in V, then so is $\mathbf{v}_1+\mathbf{v}_2$. Hence V is closed under addition.

(2) If c is a complex scalar and $\mathbf{v}$ is in V, then so is $c\mathbf{v}$. Hence V is closed under scalar multiplication.

(3) There exists a **zero vector** $\mathbf{0}$ in V such that $\mathbf{0}+\mathbf{v}=\mathbf{v}$ for all $\mathbf{v}$ in V.

(4) For each $\mathbf{v}$ in V, there exists an **additive inverse** (or **opposite**) vector $-\mathbf{v}$ in V such that $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$ for all $\mathbf{v}$ in V.

(5) For all $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ in V and complex scalars c_1 and c_2, we have

   (a) $\mathbf{v}_1+\mathbf{v}_2=\mathbf{v}_2+\mathbf{v}_1$

   (b) $(c_1+c_2)\mathbf{v}_1=c_1\mathbf{v}_1+c_2\mathbf{v}_1$

   (c) $(\mathbf{v}_1+\mathbf{v}_2)+\mathbf{v}_3=\mathbf{v}_1+(\mathbf{v}_2+\mathbf{v}_3)$

   (d) $(c_1c_2)\mathbf{v}_1=c_1(c_2\mathbf{v}_1)$

(e) $c_1(v_1+v_2)=c_1v_1+c_1v_2$
(f) $1\cdot v_1=v_1$

We have seen that all finite-dimensional real vector spaces are isomorphic to $\mathbb{R}^n$ for some *n*, so these are arguably the most important of the real vector spaces. If we allow complex numbers in place of real numbers, then we get $\mathbb{C}^n$, the set of vectors of the form

$$v=\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

where $v_1, \ldots, v_n$ are complex numbers. With addition and scalar multiplication defined as it is on $\mathbb{R}^n$, the set $\mathbb{C}^n$ is a vector space.

Example 1

Compute v_1+v_2 and iv_1 for

$$v_1=\begin{bmatrix} i+3 \\ -2i \end{bmatrix}, \quad v_2=\begin{bmatrix} 2-3i \\ 5i+1 \end{bmatrix}$$

Solution Adding real and imaginary parts in each component, we have

$$v_1+v_2=\begin{bmatrix} (i+3)+(2-3i) \\ -2i+(5i+1) \end{bmatrix}=\begin{bmatrix} 5-2i \\ 1+3i \end{bmatrix}$$

Using the identity $i^2=-1$ allows us to simplify:

$$iv_1=\begin{bmatrix} i(i+3) \\ i(-2i) \end{bmatrix}=\begin{bmatrix} i^2+3i \\ -2i^2 \end{bmatrix}=\begin{bmatrix} -1+3i \\ 2 \end{bmatrix}$$

Example 2

Show that the set $\mathbb{C}^{2\times2}$ of 2×2 matrices with complex entries forms a complex vector space when addition and scalar multiplication are defined in the usual manner for matrices.

Solution We show that two of the requirements of Definition 11.14 hold and leave the rest as an exercise. First, if

$$A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } B=\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

are both in $C^{2\times 2}$, then

$$A+B=\begin{bmatrix} (a_{11}+b_{11}) & (a_{12}+b_{12}) \\ (a_{21}+b_{21}) & (a_{22}+b_{22}) \end{bmatrix}$$

is also in $C^{2\times 2}$. Thus the set is closed under addition, so that (1) of Definition 11.14 is satisfied. Second, the zero matrix

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

from $R^{2\times 2}$ also serves as the zero matrix for $C^{2\times 2}$, so that (3) is satisfied. The remaining properties are left as an exercise.

Example 3

Suppose that : : $R\to R$ and $g : R\to R$, and define

$$h(x)=f(x)+ig(x)$$

Then $h : R\to C$ is a complex-valued function of a real variable. Show that the set of all such functions forms a complex vector space under the usual definition for addition and scalar multiplication of functions.

Solution As with Example 2, here we verify some conditions of Definition 11.14 and leave the rest to the exercises. First, suppose that $h(x)=f(x)+ig(x)$ is in our set of functions, and let $c=a+ib$ be a complex scalar. Then

$$ch(x)=(a+ib)(f(x)+ig(x)) =(af(x)-bg(x))+i(ag(x)+bf(x))$$

Since both $af-bg$ and $ag+bf$ are real functions of real variables, it follows that $ch(x)$ is a complex-valued function of a real variable. Hence our set is closed under scalar multiplication, as required by (2) of Definition 11.14.

Next, for $h(x)=f(x)+ig(x)$, if we let $-h(x)=-f(x)-ig(x)$, then $-h(x)$ is also in our set, and

$$h(x)+(-h(x))=(f(x)+ig(x))+(-f(x)-ig(x))=0+0i=0$$

Thus each vector in our set has an additive inverse in the set, which shows that (4) of Definition 11.14 is also true. Verification of the remaining conditions is left as an exercise.

Concepts such as linear combination, linear independence, span, basis, and subspace carry over to complex vector spaces in a natural and essentially unchanged manner. For instance, to show that a set of vectors $\{v_1,\dots,v_k\}$ is linearly independent in a complex vector space, we need to show that the only solution among complex scalars $c_1,\dots,c_k$ to

$$c_1v_1+\cdots+c_kv_k=0$$

is the trivial $c_1=\cdots=c_k=0$.

Complex Inner Product Spaces

In Chapter 10 we developed the notion of an inner product on a real vector space and considered some of the properties of an inner product. Here we provide an account of the analog for complex vector spaces, starting with the definition of an inner product. (Note the similarities between this and Definition 10.1.) Complex inner product spaces have applications in a number of fields, including physics and engineering.

DEFINITION 11.15 ▶

Inner Product, Complex Inner Product Space, Unitary Space

Let u, v, and w be elements of a complex vector space *V*, and let *c* be a complex scalar. An **inner product** on *V* is a function denoted by $\langle u, v\rangle$ that takes any two vectors in *V* as input and produces a scalar as output. An inner product on a complex vector space satisfies the conditions:

(a) $\langle u, v\rangle=\overline{\langle v, u\rangle}$

(b) $\langle u+v, w\rangle=\langle u, w\rangle+\langle v, w\rangle$

(c) $\langle cu, v\rangle=c\langle u, v\rangle$

(d) $\langle u, u\rangle$ is a nonnegative real number, and $\langle u, u\rangle=0$ only when $u=0$

A complex vector space *V* with an inner product defined on it is called a **complex inner product space** or a **unitary space**.

▶ Note that unlike in real inner product spaces, typically $\langle u, v\rangle\neq\langle v, u\rangle$

Complex Dot Product

A complex inner product space of particular importance is the vector space $\mathbb{C}^n$ together with the **complex dot product**, defined by

$$u\cdot v=u_1\bar{v}_1+u_2\bar{v}_2+\cdots+u_n\bar{v}_n$$

For example, if

$$u=\begin{bmatrix}1+i\\-2i\\4\end{bmatrix} \quad\text{and}\quad v=\begin{bmatrix}5i\\2-i\\1+3i\end{bmatrix}$$

then

$$\mathbf{u}\cdot\mathbf{v}=(1+i)(\overline{5i})+(-2i)(\overline{2-i})+(4)(\overline{1+3i})=(1+i)(-5i)+(-2i)(2+i)+(4)(1-3i)=11-21i$$

Note that

$$\mathbf{u}\cdot\mathbf{u}=u_1\overline{u}_1+u_2\overline{u}_2+\cdots+u_n\overline{u}_n=|u_1|^2+|u_2|^2+\cdots+|u_n|^2\geq 0$$

with equality holding only if $\mathbf{u}=\mathbf{0}$. Thus (d) of Definition 11.15 holds for the complex dot product. Proving that the other conditions also hold is left as an exercise (see Exercise 44).

The following theorem gives three properties of inner products that follow from Definition 11.15. Of these, the first two are the same as for real inner products, but the third is different because of the complex conjugation in (a) of the definition.

THEOREM 11.16

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be elements of a complex vector space *V*, and let *c* be a complex scalar. Then an inner product defined on *V* satisfies each of the following:

(a) $\langle \mathbf{0}, \mathbf{u}\rangle=\langle \mathbf{u}, \mathbf{0}\rangle=0$

(b) $\langle \mathbf{u}, \mathbf{v}+\mathbf{w}\rangle=\langle \mathbf{u}, \mathbf{v}\rangle+\langle \mathbf{u}, \mathbf{w}\rangle$

(c) $\langle \mathbf{u}, c\mathbf{v}\rangle=\overline{c}\langle \mathbf{u}, \mathbf{v}\rangle$

The proof is left as an exercise (see Exercise 45).

Example 4

Show that the set $\mathbb{C}^{2\times 2}$ of 2×2 matrices with complex entries (as in Example 2) is a complex inner product space using the following for an inner product: For each

$$A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{bmatrix} \quad\text{and}\quad B=\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22}\end{bmatrix}$$

in $C^{2\times 2}$, we set

$$\langle A, B\rangle = a_{11}\bar{b}_{11} + a_{12}\bar{b}_{12} + a_{21}\bar{b}_{21} + a_{22}\bar{b}_{22}$$

Solution We already know from Example 2 that $C^{2\times 2}$ is a complex vector space. Thus all that remains is to show that our proposed inner product is an inner product.

To proceed, we could work through the four conditions of Definition 11.15. However, looking at our function closely, we can see that it is really just a thinly disguised version of the complex dot product on C^4. The only difference (essentially a cosmetic one) is that the entries of our matrices are arranged in two rows of two each, instead of a single column as in C^4. Since the complex dot product is an inner product, so is the function defined here.

Given a vector u in a complex inner product space V, the **norm** (or **length**) of u is defined just as it is in a real inner product space, by

$$\|u\| = \sqrt{\langle u, u\rangle}$$

Once we have the norm, we define the **distance** from u to v by

$$\text{Distance from u to v} = \|v - u\|$$

The norm has the properties given in the next theorem.

THEOREM 11.17

Let u and v be elements of a complex inner product space V, and suppose that c is a complex scalar. Then

(a) $\|u\| \geq 0$ (with equality only if $u = 0$)

(b) $\|cu\| = |c|\ \|u\|$

(c) $|\langle u, v\rangle| \leq \|u\|\|v\|$ (This is the **Cauchy–Schwarz inequality**)

(d) $\|u+v\| \le \|u\|+\|v\|$ (This is the **Triangle inequality**)

The proof is similar to the real case and is left as exercises.

Example 5

Let $u=[-1+i \; -2i]$ and $v=[2 \; 1-i]$ be in C^2. Compute $\langle u, v\rangle$, $\|u\|^2$, $\|v\|^2$, and $\|u+v\|^2$ using the complex dot product for the inner product.

Solution We have

$$\langle u, v\rangle=(-1+i)(2)+(-2i)(1+i)=(-2+2i)+(2-2i)=0 \quad \|u\|^2=(-1+i)(-1-i)+(-2i)(2i)=2+4=6 \quad \|v\|^2=(2)(2)+(1-i)(1+i)=4+2=6 \quad \|u+v\|^2=(1+i)(1-i)+(1-3i)(1+3i)=2+10=12$$

Regarding complex inner products:

- Note that we have $\langle u, v\rangle=0$. The definition of *orthogonal* is the same for complex inner product spaces as it is for the real counterparts:

  Vectors u and v are **orthogonal** if and only if $\langle u, v\rangle=0$.

- Thus our two vectors are orthogonal with respect to the complex dot product.
- In Example 5 we have $\langle u, v\rangle=0$, $\|u\|^2=6$, $\|v\|^2=6$, and $\|u+v\|^2=12$. In general, for a complex inner product, if $\langle u, v\rangle=0$ then $\|u\|^2+\|v\|^2=\|u+v\|^2$. Interestingly, the converse is not true. (See Exercise 31 and Exercise 46.)
- The definitions of **orthonormal**, **orthogonal set**, **orthonormal set**, and other related concepts are analogous to those for real vector spaces.
- The Gram–Schmidt process also works when applied to a complex inner product space.

▶ There is nothing special about using 0 and 1 for the limits of integration. As long as $a<b$, we also get an inner product using
$\int_a^b h_1(x)\overline{h_2(x)}\,dx$

Suppose that $h_1(x)$ and $h_2(x)$ are from the set of complex-valued functions of a real variable, described in Example 3. Let's add one additional condition—namely, that both functions are continuous. Then a complex inner product is given by

$$\langle h_1, h_2\rangle=\int_0^1 h_1(x)\overline{h_2(x)}\,dx \tag{1}$$

Note that

$$\langle h_1, h_1\rangle=\int_0^1 h_1(x)\overline{h_1(x)}\,dx=\int_0^1 |h_1(x)|^2\,dx$$

Thus $\langle h_1, h_1\rangle\geq 0$, with equality holding only if $h_1(x)=0$, the zero function. (This is where the continuity requirement is used.) The other conditions required of an inner product can be readily verified.

Example 6

Let $h_1(x)=x+i$ and $h_2(x)=1-3xi$, and let S be the subspace spanned by these two functions. Use the Gram–Schmidt process to find an orthogonal basis for S.

Solution If we let $j_1(x)$ and $j_2(x)$ denote the orthogonal basis vectors, then by Gram–Schmidt we have

$$j_1(x)=h_1(x)\qquad j_2(x)=h_2(x)-\frac{\langle h_2, j_1\rangle}{\langle j_1, j_1\rangle}\,j_1(x)$$

We start by setting $j_1(x)=x+i$. To find $j_2(x)$, we compute

$$\langle h_2, j_1\rangle=\int_0^1(1-3xi)(x-i)dx=\int_0^1(-2x-i(1+3x^2))dx=-1-2i \qquad \langle j_1, j_1\rangle=\int_0^1|(x+i)|^2\,dx=\int_0^1(x^2+1)dx=\frac{4}{3}$$

Therefore

$$j_2(x)=(1-3xi)-\frac{3}{4}(-1+2i)(x+i)=\frac{1}{4}\left((3x-2)+i(3-6x)\right)$$

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Let

$$u=[1-4i-1+i2+3i],\ \ v=[1+i2-i1+3i],\ \ w=[2+i2-5i2-3i]$$

(a) Find $2u+w$ and $-3v+u-2w$.

(b) Determine if $(i+1,\ -2+i,\ 3-4i)$ is in span$\{u,\ v,\ w\}$.

(c) Find $\langle u,\ w\rangle$ using the complex dot product.

(d) Compute $\|2iv\|$ using the complex dot product.

(e) Normalize v and w using the complex dot product.

(f) Normalize $v-u$ and $2u+w$ using the complex dot product.

EXERCISES

Exercises 1–10: Let

$$u=[2+3i\ -13+4i],\ \ v=[23-i1+5i],\ \ w=[2+i2-i4-3i]$$

1. Find each linear combination:

(a) $u-v$

(b) $w+3v$

(c) $2u+2iw-5v$

2. Find each linear combination:

(a) $2u+w$

(b) $w-iu$

(c) $2iw-3v+2iu$

3. Determine if there exists a constant c such that $u+iv=cw$.

4. Determine if there exists a nontrivial linear combination such that c1u+c2v+c3w=0.
5. Is (−5+2i, −3, −5+i) in the span of {u, v, w}?
6. Is (−2+i, −9, −1+16i) in the span of {u, v, w}?
7. Compute each using the complex dot product.
   (a) ⟨u, v⟩
   (b) ⟨iv, −2w⟩
   (c) ‖w‖
8. Compute each using the complex dot product.
   (a) ⟨2w, 3iu⟩
   (b) ⟨2iv, 5v⟩
   (c) ‖w+ iu‖
9. Normalize **u** and **v** using the complex dot product.
10. Normalize **u** + **v** and **v** − **w** using the complex dot product.

Exercises 11–20, Let

A=[2+i31+i2+3i], B=[−i42+2i1+4i], C=[03+i−4i1+i]

11. Find each linear combination:
    (a) A−iC
    (b) 2B−A−4iC
12. Find each linear combination:
    (a) C−(1+i)A
    (b) iA−(1−i)B−3C
13. Determine if there exists a constant *c* such that A−cB=iC.
14. Determine if there exists a nontrivial linear combination such that c1A+c2B+c3C=[0000].
15. Is [3+2i−3−7i4+8i5+2i] in the span of {A, B, C}?
16. Is [4−3i6+5i2+5i13−7i] in the span of {A, B, C}?
17. Compute each using the inner product given in Example 4.
    (a) ⟨A, C⟩
    (b) ⟨iB, −2A⟩
    (c) ‖B‖

18. Compute each using the inner product given in Example 4.

(a) $\langle C, B\rangle$

(b) $\langle 3C, (1-i)A\rangle$

(c) $\|A+iC-B\|$

19. Normalize *A* and *C* using the inner product given in Example 4.

20. Normalize $A-C$ and $A+B-C$ using the inner product given in Example 4.

Exercises 21–30: Let $h_1(x)=1+ix$, $h_2(x)=i-x$, and $h_3=3-(1+i)x$, and (when appropriate) use the inner product given in (1).

21. Find each linear combination.

(a) $h_1(x)+(4-i)h_2(x)$

(b) $ih_1(x)-h_2(x)+3h_3(x)$

22. Find each linear combination.

(a) $ih_3(x)-(2-i)h_1(x)$

(b) $h_3(x)+2ih_2(x)-4h_1(x)$

23. Determine if there exists a constant *c* such that $h_3(x)+h_2(x)=ch_1(x)$.

24. Determine if there exists a nontrivial linear combination such that $c_1h_1(x)+c_2h_2(x)+c_3h_3(x)=0$.

25. Determine if $(2+i)+(3-2i)x$ is in

$$\text{span}\{h_1(x), h_2(x), h_3(x)\}$$

26. Determine if $(2-2i)-(2-3i)x$ is in

$$\text{span}\{h_1(x), h_2(x), h_3(x)\}$$

27. Compute each of the following.

(a) $\langle h_1, h_3\rangle$

(b) $\langle ih_2, -2h_3\rangle$

(c) $\|h_1\|$

28. Compute each of the following.

(a) $\langle h_3, h_2\rangle$

(b) $\langle 3h_2, ih_1\rangle$

(c) $\|h_2+3ih_3\|$

29. Normalize $\mathbf{h}_1$ and $\mathbf{h}_2$.

30. Normalize $\mathbf{h}_2 - i\mathbf{h}_3$ and $\mathbf{h}_2 + 3\mathbf{h}_3 - (1+i)\mathbf{h}_1$.

FIND AN EXAMPLE Exercises 31–36: Find an example that meets the given specifications.

31. A complex vector space *V* and vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u}+\mathbf{v}\|^2$ but $\langle \mathbf{u}, \mathbf{v} \rangle \neq 0$.

32. A complex vector space *V* not given in this section.

33. An inner product on $\mathbb{C}^n$ other than one given in this section.

34. An inner product on $\mathbf{P}_c^n$, the polynomials of degree *n* or less with complex coefficients.

35. A complex vector space *V* and a nonempty subset *S* that is closed under addition and multiplication by real scalars.

36. A pair of vectors that are linearly independent when the scalars are restricted to the reals but linearly dependent when the scalars are complex.

TRUE OR FALSE Exercises 37–40: Determine if the statement is true or false, and justify your answer.

37.

(a) A complex vector space *V* must be closed under addition and scalar multiplication of the vectors.

(b) A complex vector space can have only one inner product defined on it.

38.

(a) The subset of $\mathbb{C}^n$ consisting of vectors with real entries is a subspace of $\mathbb{C}^n$.

(b) The norm of a vector in $\mathbb{C}^n$ that has all real entries is the same as the norm of that vector in $\mathbb{R}^n$.

39.

(a) The norm of any nonzero vector in a complex inner product space must be a positive real number.

(b) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in a complex vector space *V*, then so is $i\mathbf{v}_1 - (1+i)\mathbf{v}_2$.

40.

(a) If $\mathbf{u}$ and $\mathbf{v}$ are in a complex inner product space *V*, then $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

(b) If S_1 and S_2 are subspaces of a complex vector space *V*, then the intersection $S_1 \cap S_2$ is also a subspace of *V*.

41. Finish Example 2 by showing that the remaining unverified conditions for a complex vector space hold.

42. Finish Example 3 by showing that the remaining unverified conditions for a complex vector space hold.

43. Prove that $\mathbb{R}^{2\times 2}$ with standard operations is not a complex vector space, by finding a condition of Definition 11.14 that is not met.

44. Prove that the complex dot product is an inner product on $\mathbb{C}^n$. (Condition (d) has already been verified, so you need only show that (a)–(c) are true.)

45. Prove Theorem 11.16: Let u, v, and w be elements of a complex vector space *V*, and let *c* be a complex scalar. Then an inner product defined on *V* satisfies each of the following:

(a) $\langle 0, u\rangle = \langle u, 0\rangle = 0$

(b) $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$

(c) $\langle u, cv\rangle = \bar{c}\langle u, v\rangle$

46. Prove that if $\langle u, v\rangle = 0$, then $\|u\|^2 + \|v\|^2 = \|u+v\|^2$.

Exercises 47–50: Here we prove each part of Theorem 11.17. Assume that u and v are elements of a complex vector space *V*, and suppose that *c* is a complex scalar.

47. Prove that $\|u\| \geq 0$, with equality only if $u = 0$.

48. Prove that $\|cu\| = |c|\ \|u\|$.

49. Prove that $|\langle u, v\rangle| \leq \|u\| \cdot \|v\|$. (This is the Cauchy–Schwarz inequality.)

50. Prove that $\|u+v\| \leq \|u\| + \|v\|$. (This is the Triangle inequality.)

11.5 Hermitian Matrices

In Section 8.3 we encountered the Spectral Theorem, which says that a matrix A with real entries is orthogonally diagonalizable exactly when A is symmetric. In this section we consider what happens if we allow A to have complex entries. Since the reals are a subset of the complex numbers, a reasonable guess is that the Spectral Theorem just generalizes, so that a matrix A with complex entries is orthogonally diagonalizable if and only if A is symmetric. However, this is not quite true. The central goal of this section is to find the correct analog of the Spectral Theorem for complex matrices, but before we can do that, we need to develop a few new ideas.

▶ See Section 6.3 for a review of the basic properties of complex numbers.

Unitary Matrices

If Complex Conjugate, Conjugate Transpose

A is a complex matrix, then the **complex conjugate** of A is denoted by $\bar{A}$ and is found by taking the complex conjugate of each entry of A. The **conjugate transpose** of A is denoted by A^* and defined

$$A^* = \overline{A^T}$$

▶ Recall that if $c = a + ib$, then the complex conjugate of c is $\bar{c} = a - ib$.

Example 1

Find A^* for

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

Solution We have

$$A=\begin{bmatrix}2+i & 3\\ -5i & 1+4i\\ 6 & 1-2i\end{bmatrix} \Rightarrow \overline{A}=\begin{bmatrix}2-i & 3\\ 5i & 1-4i\\ 6 & 1+2i\end{bmatrix} \Rightarrow A^{*}=\overline{A}^{T}=\begin{bmatrix}2-i & 5i & 6\\ 3 & 1-4i & 1+2i\end{bmatrix}$$

Note that the order of conjugation and transposition makes no difference in A^*. For the matrix in Example 1, we could just as well have computed

$$A^{T}=\begin{bmatrix}2+i & -5i & 6\\ 3 & 1+4i & 1-2i\end{bmatrix} \Rightarrow A^{*}=\overline{A^{T}}=\begin{bmatrix}2-i & 5i & 6\\ 3 & 1-4i & 1+2i\end{bmatrix}$$

A general proof that $\overline{A}^{T}=\overline{A^{T}}$ is left as Exercise 28.

Conjugate transposes have properties similar to those of transposes of real matrices. These are summarized in the theorem below, with the proof of each part left as exercises.

THEOREM 11.18

Suppose that A and B are matrices with complex entries and that c is a complex scalar. Then

(a) $(A^*)^* = A$

(b) $(A + B)^* = A^* + B^*$

(c) $(AB)^* = B^*A^*$

(d) $(cA)^* = \overline{c}A^*$

Unitary Matrix

Recall that a square matrix with real entries is *orthogonal* if the matrix columns form an orthonormal set with respect to the usual dot product. Equivalently, a real matrix A is orthogonal if and only if $A^{-1}=A^{T}$. The counterpart of orthogonal matrices for a matrix A with complex entries is called a Unitary Matrix **unitary matrix**, which requires that

$$A−1=A*$$

It can be shown that a square matrix *A* is unitary if and only if the columns of *A* are orthonormal with respect to the complex dot product (see Exercise 33).

Example 2

Show that

$$A=[12−1+i2121+i2]$$

is a unitary matrix.

Solution One way to solve this is to compute *AA**. If the product is the identity matrix, then we know that $A*=A−1$ and can conclude that *A* is unitary. Instead, we shall show directly that the columns are orthonormal with respect to the complex dot product. Setting $A=[a1\ \ a2]$, we have

$$a1\ .\ a2=(12)\ (−1−i2)+(12)\ (1−i2)=0‖a1‖=a1\ .\ a1=(12)2+(12)2=1‖a2‖=a2·a2=(−1+i2)(−1−i2)+(1+i2)(1−i2)=1$$

Hence the columns are orthonormal and therefore *A* is unitary.

Diagonalizing Matrices

Unitarily Diagonalizable

We say that a complex matrix *A* is **unitarily diagonalizable** if there exist a diagonal matrix *D* and a unitary matrix *P* such that

$$A=PDP−1=PDP*$$

As with real matrices, the diagonal entries of *D* are the eigenvalues of *A*, and the columns of *P* are the corresponding eigenvectors (see

Theorem 6.9 in Section 6.2). The question is, when will a complex matrix be diagonalizable?

▶ "Unitarily diagonalizable" gets tiresome to say over and over, so sometimes we will just say "diagonalizable" with the understanding that P must be unitary.

Let's recall what happens for real matrices. A real matrix A is orthogonally diagonalizable if and only if $A = A^T$—that is, when A is symmetric. The analog of symmetric for complex matrices is

$$A = A^* \qquad (1)$$

Hermitian

A matrix A satisfying (1) is called **Hermitian**. From the definition, we see that a Hermitian matrix is unchanged by taking its conjugate transpose. For example, the matrix A below is Hermitian, because

$$A=[32-i2+i4] \Rightarrow \bar{A}=[32+i2-i4] \Rightarrow A^*=[32-i2+i4]=A$$

▶ **Charles Hermite** (1822–1901) was a French mathematician who made contributions to a variety of areas of mathematics, among them linear algebra.

Note that any Hermitian matrix must have real diagonal entries. (Why?) Unfortunately, the Hermitian matrices still are not exactly the set that we seek.

While all Hermitian matrices are unitarily diagonalizable, it turns out that there are some complex matrices that are unitarily diagonalizable but not Hermitian. We need to expand the set of Hermitian matrices to the larger set of **normal** matrices, which are those complex matrices A such that

$$A^*A = AA^*$$

Note that all unitary matrices are normal, because

$$A^*A=A^{-1}A=I=AA^{-1}=AA^*$$

Similarly, Hermitian matrices are also normal, because

$$A^*A = AA = AA^*$$

There are normal matrices that are not Hermitian (or unitary, for that matter).

Example 3

Show that

$$A = \begin{bmatrix} i & -i \\ i & i \end{bmatrix}$$

is normal but is not Hermitian or unitary.

Solution We have

$$A^*A = \begin{bmatrix} -i & -i \\ i & -i \end{bmatrix} \begin{bmatrix} i & -i \\ i & i \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$AA^* = \begin{bmatrix} i & -i \\ i & i \end{bmatrix} \begin{bmatrix} -i & -i \\ i & -i \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Hence *A* is normal. On the other hand, since $A^* \neq A$ our matrix is not Hermitian, and as $A^*A \neq I_2$, it follows that $A^* \neq A^{-1}$ and thus *A* is not unitary.

The following is a complex version of the Spectral Theorem (Section 8.3), given without proof.

THEOREM 11.19 ▶

A complex matrix *A* is unitarily diagonalizable if and only if *A* is normal.

Now we have an easy way to determine if a complex matrix is diagonalizable, by checking if it is normal. If *A* is an $n \times n$

diagonalizable matrix, then we find the diagonalization of *A* using the same procedure as with real matrices.

1. Find the eigenvalues and eigenvectors of *A*.
2. For each distinct eigenvalue, apply Gram–Schmidt as needed to find an orthonormal basis for the associated eigenspace. As with real symmetric matrices, eigenvectors associated with distinct eigenvalues of a normal matrix are orthogonal. Once we have orthonormal bases for each eigenspace, they can be combined to form an orthonormal basis for $\mathbb{C}^n$.
3. Define *D* to be the diagonal matrix with the eigenvalues of *A* along the diagonal, and define *P* to be the unitary matrix with the corresponding eigenvectors for columns.

Applying this procedure by hand to a large, complicated matrix is difficult. But it is manageable if the matrix is not too complicated.

Example 4

Diagonalize the matrix $A=\begin{bmatrix} i & -i \\ i & i \end{bmatrix}$.

Solution We have already seen that *A* is normal and so must be diagonalizable. Our first step is to find the eigenvalues, which are the roots of the characteristic polynomial

$$\det(A-\lambda I_2)=(i-\lambda)^2+i^2=(i-\lambda)^2-1$$

Setting this equal to 0 and solving for λ yields two eigenvalues, $\lambda_1=1+i$ and $\lambda_2=-1+i$. To find the eigenvectors associated with λ_1, we need to find the solutions to the homogeneous system with coefficient matrix $A-\lambda_1 I_2=A-(1+i)I_2$. The augmented matrix is

$$\left[\begin{array}{cc|c} i-(1+i) & -i & 0 \\ i & i-(1+i) & 0 \end{array}\right]=\left[\begin{array}{cc|c} -1 & -i & 0 \\ i & -1 & 0 \end{array}\right] \xrightarrow{iR_1+R_2 \to R_2} \sim \left[\begin{array}{cc|c} -1 & -i & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Back substitution and normalization produces the eigenvector

$$\mathbf{p}_1=\begin{bmatrix} -\frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Following a similar procedure, we find that a normalized eigenvector associated with $\lambda_2=-1+i$ is

$$p_2=[i212]$$

Since each eigenspace has dimension 1, we are spared the work of applying the Gram–Schmidt process to find orthogonal eigenvectors. Thus all that remains is to define *P* and *D*, which are

$$P=[-i2i21212] \quad \text{and} \quad D=[1+i00-1+i]$$

We can check our work by computing

$$PDP^*=[-i2i21212]\,[1+i00-1+i]\,[i212-i212]=[i-iii]=A$$

In Section 8.3 it is noted that a real symmetric matrix must have real eigenvalues. (This follows from the Spectral Theorem.) On the other hand, the preceding example shows that a normal matrix can have complex eigenvalues. In between these two sets are the Hermitian matrices, which also happen to have real eigenvalues.

THEOREM 11.20 ▸

If *A* is a Hermitian matrix, then *A* has real eigenvalues.

Proof Suppose that λ is an eigenvalue of *A* with associated eigenvector u. Then $Au=\lambda u$, and hence

$$u^*Au=u^*(\lambda u)=\lambda(u^*u)=\lambda\|u\|^2$$

We know that $\|u\|^2$ is a real number. Also, because *A* is Hermitian we have

$$(u^*Au)^*=u^*A^*(u^*)^*=u^*Au$$

This shows that u^*Au is also Hermitian and therefore has real diagonal entries. But u^*Au has only one entry (it is the complex dot product of u and Au), so this implies that u^*Au is real. Since $u^*Au=\lambda\|u\|^2$, we may conclude that λ is also real. ■■

Example 5

Show that

$$A=[52i-2i2]$$

is Hermitian and has real eigenvalues.

Solution We have

$$\bar{A}=[5-2i2i2] \quad \Rightarrow \quad A^*=\bar{A}^T=[52i-2i2]=A$$

so A is Hermitian. The characteristic polynomial of A is

$$\det(A-\lambda I_2)=(5-\lambda)(2-\lambda)-(-2i)(2i)=\lambda^2-7\lambda+6=(\lambda-6)(\lambda-1)$$

Hence the eigenvalues are $\lambda_1=1$ and $\lambda_2=6$, which are both real.

PRACTICE PROBLEMS

Do the practice problems and then check your answers against the full solutions in the back of the book.

1. Find A^* for the given A.

(a) $A=[-i4+5i-3+7i1-6i]$

(b) $A=[1+i2+3i3-i3-i-5+i5+i7+i2+i4+4i]$

2. Determine if A is Hermitian.

(a) $A=[1+2i5i-5i2-i]$

(b) A=[2i1−i−i−301+i04]

EXERCISES

Exercises 1–6: Find A∗ for the given *A*.

1. A=[1+i3i2−i1+4i]

2. A=[−7i3−2i1+5i8]

3. A=[3+i5i1−i1−4i−86+i2+2i0−7i]

4. A=[5−i2+7i4i5i315−i6i13]

5. A=[1−2i34i2i5−6i1+i36i73+2i−4i1−i3−2i11]

6. A=[41−2i121+3i11i6−5i6i3i−7i1−i3i4−5i2+i1−2i4−7i0]

Exercises 7–12: Determine if *A* is Hermitian.

7. A=[1+i3i3i2]

8. A=[43−2i3+2i3]

9. A=[35i1−i−5i−501+i07]

10. A=[5−i2+7ii36i2+7i6i4]

11. A=[1−2i34i2i5−6i1+i36i73+2i−4i1−i3−2i11]

12. A=[01−2i−122−i1+2i56i3i12−6i24−5i2+i−3i4+5i8]

Exercises 13–18: Determine if *A* is normal.

13. A=[12−5i2+5i3]

14. A=[33−2i1+i−4]

15. A=[−i−ii−i]

16. A=[2iii3i]

17. A=[−1−i1−ii5−2i1+i2i0]

18. A=[23−ii3−i1−2i−i3i−1]

FIND AN EXAMPLE Exercises 19–22: Find an example that meets the given specifications.

19. A 3×3 matrix *A* that is symmetric but not Hermitian.

20. A 3×3 matrix *A* that is normal but not Hermitian.

21. A 3×3 unitary matrix that is not in $\mathbb{R}^{3\times 3}$.

22. A 2×2 matrix that has eigenvalues $\lambda_1=2+i$ and $\lambda_2=2-i$.

TRUE OR FALSE Exercises 23–26: determine if the statement is true or false, and justify your answer.

23.

(a) A matrix *A* is unitarily diagonalizable if and only if *A* is normal.

(b) If *A* is normal, then *A* is symmetric.

24.

(a) If *A* has complex entries and $A=A^T$, then *A* is Hermitian.

(b) If *A* has complex entries, then $\det(A)$ cannot be real.

25.

(a) If *A* and *B* are $n\times n$ Hermitian matrices, then so is $A+B$.

(b) If *A* and *B* are $n\times n$ complex matrices, then A^*B is Hermitian.

26.

(a) A unitary matrix has orthonormal columns.

(b) A Hermitian matrix must have some real entries.

27. Prove that if *A* has real entries, then $A^*=A^T$.

28. Prove that if *A* has complex entries, then $(\overline{A})^T=\overline{(A^T)}$. This shows that the order of conjugation and transposition in A^* does not matter.

Exercises 29–32: Assume that *A* and *B* are matrices with complex entries and that *c* is a complex scalar.

29. Prove that $(A^*)^*=A$.

30. Prove that $(A+B)^*=A^*+B^*$.

31. Prove that $(AB)^*=B^*A^*$.

32. Prove that $(cA)^*=\overline{c}A^*$.

33. Show that a square matrix *A* is unitary if and only if the columns of *A* are orthonormal with respect to the complex dot product.

34. Prove that any Hermitian matrix must have real diagonal entries.

35. Show that if *A* is upper (or lower) triangular and normal, then *A* must be a diagonal matrix.

36. Suppose that $A=PDP^{-1}$, where *D* is diagonal and *P* is unitary. Show that the diagonal entries of *D* are the eigenvalues of *A*, and the columns of *P* are the corresponding eigenvectors.

SUPPLEMENTARY EXERCISES

Exercises 1–2: Evaluate Q(x) for the given x0.

1. Q(x)=3x12−4x22+14x1x2; x0=[−23]

2. Q(x)=2x12+5x22−3x32+10x1x3−18x2x3; x0=[1−12]

Exercises 3–4: Find a formula for the quadratic form with the given matrix *A*.

3. A=[−3−4−47]

4. A=[2−15−13252−4]

Exercises 5–6: Find a matrix *A* such that Q(x)=xTAx.

5. Q(x)=2x12+5x22−3x32+12x1x2

6. Q(x)=−4x12+6x22−6x1x2−10x1x3+4x2x3

Exercises 7–8: Determine if the quadratic form Q(x)=xTAx is positive definite, negative definite, indefinite, or none of these.

7. A=[422−3]

8. A=[100001010]

Exercises 9–10: Find the principal submatrices of *A*.

9. A=[455−2]

10. A=[−12−4201−413]

Exercises 11–12: Determine if the given matrix is positive definite.

11. A=[1225]

12. A=[11−1143−1311]

Exercises 13–16: Find the maximum and minimum values of the quadratic form Q(x) subject to the constraint ‖x‖=1.

13. Q(x)=5x12−2x22

14. Q(x)=2x12−x22−4x32

15. Q(x)=x12+2x22+6x1x2

16. Q(x)=x12+3x22+2x32+6x1x3

Exercises 17–20: Find the maximum and minimum values of the quadratic form Q(x) subject to the given constraint.

17. Q(x)=x12+9x22+6x1x2; ‖x‖=2

18. Q(x)=x22+2x1x3; ‖x‖=10

19. Q(x)=x12+x22+8x1x2; 4x12+25x22=100

20. Q(x)=4x12+9x22+4x1x2; x12+9x22=9

Exercises 21–26: Let

u=[2−4i−1+2i3i], v=[8+i2−2i1], w=[3+5i4−7i5−3i]

21. Find u+3w and 2v+u−4w.

22. Determine if (5−i, −1+i, 3i) is in span{u, v, w}.

23. Find ⟨v, w⟩ using the complex dot product.

24. Compute ‖u‖ using the complex dot product.

25. Normalize u and w using the complex dot product.

26. Normalize u−w and 2u+v using the complex dot product.

Exercises 27–28: Find A∗ for the given *A*.

27. A=[2−i2+i8+3i6−i]

28. A=[2+i2+i1−i3−2i−5+ii7+i7+i2+2i]

Exercises 29–30: Determine if *A* is Hermitian.

29. A=[1+ii−i3−2i]

30. A=[2−i1−2ii−32−i1+2i2+i5]

GLOSSARY

Below is a glossary of definitions and other terms presented in this book. In some cases, due to the complicated nature of a definition or term, it is only described here in general terms. Visit the section referenced for more details.

Additive identity matrix

The $n \times m$ consisting of all zeros satisfies $A + 0_{nm} = A$ for all $n \times m$ matrices A. (Sect. 3.2)

Adjoint matrix

The adjoint of an $n \times n$ matrix A is given by

$$\mathrm{adj}(A) = C^T = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

where C is the cofactor matrix of A. (Sect. 5.3)

Argument

The argument of a nonzero complex number z, denoted by $\arg(z)$, is the angle θ (in radians) in the counterclockwise direction from the positive x-axis to the ray that points from the origin to z. Note that the argument is not unique, because we can always add or subtract multiples of 2π. (Sect. 6.3)

Associated homogeneous linear system

A linear system of the form $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} \neq \mathbf{0}$, has associated homogeneous linear system $A\mathbf{x} = \mathbf{0}$. (Sect. 2.3)

Augmented matrix

A matrix that contains all of the coefficients of a linear system, including the constant terms. (Sect. 1.2)

Back substitution

A method of solution applicable to a system of linear equations in echelon form. Implemented by substituting known values back into remaining equations. (Sect. 1.1)

Basis

A set $\mathcal{B}$={u1,⋯,um} is a basis for a subspace S if $\mathcal{B}$ spans S and $\mathcal{B}$ is linear independent. (Sect. 4.2; see also Sect. 7.3)

Block diagonal matrix

A partitioned matrix with nondiagonal blocks that are zero matrices. (Sect. 3.3)

Change of basis matrix

A square matrix used to express a vector in $\mathbf{R}^n$ given in terms of one basis into a vector given in terms of a different basis. (Sect. 4.4; see also Sect. 9.4)

Characteristic equation

Let A be an $n \times n$ matrix. Then the characteristic equation is given by $\det(A - \lambda I_n) = 0$, where I_n is the identity matrix. The solutions to the characteristic equation are the eigenvalues of A. (Sect. 6.1)

Characteristic polynomial

Let A be an $n \times n$ matrix. Then the characteristic polynomial is given by $\det(A - \lambda I_n)$, where I_n is the identity matrix. (Sect. 6.1)

Cholesky decomposition

The factorization of A=LcLcT. (Sect. 11.2)

Closed under addition

S is closed under addition if $\mathbf{u}$ and $\mathbf{v}$ in S implies $\mathbf{u} + \mathbf{v}$ is also in S. (Sect. 4.1)

Closed under scalar multiplication

S is closed under scalar multiplication if r is a real number and $\mathbf{u}$ in S implies $r\mathbf{u}$ is also in S. (Sect. 4.1)

Codomain

A set containing all possible outputs for a function. (Note that this contains the range, which is equal to the set of possible outputs for a function.) (Sect. 3.1; see also Sect. 9.1)

Cofactor

Given a matrix A, the cofactor of a_{ij} is equal to

Cij=(-1)i+jdet(Mij)

where M_{ij} is the $(n-1) \times (n-1)$ matrix that we get from A after deleting the row and column containing a_{ij}. Put another way, C_{ij} is equal to $(-1)^{i+j}$ times the minor of a_{ij}. (Sect. 5.1)

Cofactor expansions

Let A be the $n \times n$ matrix $[a_{ij}]$ and let C_{ij} denote the cofactor of a_{ij}. Then the cofactor expansions are given by

(a) $\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$ (Expand across row i)

(b) $\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$ (Expand down column j)

(Sect. 5.1)

Cofactor matrix

For an $n \times n$ matrix A, the cofactor matrix is given by

$$C = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix}$$

where C_{ij} is the cofactor of a_{ij}. (Sect. 5.3)

Column space

Let A be an $n \times m$ matrix. The column space of A is the subspace of $\mathbf{R}^n$ spanned by the column vectors of A, and is denoted by col(A). (Sect. 4.3)

Column vector

A vector in Euclidean space expressed in the form of a column matrix. Also used to indicate the vectors formed from the columns of a matrix A. (Sect. 2.1)

Complex conjugate

The complex conjugate of $z = a + ib$ is given by $\overline{z} = a - ib$. (Sect. 6.3)

Complex dot product

The complex dot product is defined on $\mathbf{C}^n$ by

$$\mathbf{u} \cdot \mathbf{v} = u_1\overline{v}_1 + u_2\overline{v}_2 + \cdots + u_n\overline{v}_n$$

(Sect. 11.4)

Complex inner product space

A complex vector space V with an inner product defined on it. (Also called a unitary space.) (Sect. 11.4)

Complex vector space

A complex vector space consists of a nonempty set V of vectors together with operations of addition and scalar multiplication on the vectors that satisfy each of the following:

(1) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in V, then so is $\mathbf{v}_1 + \mathbf{v}_2$. Hence V is closed under addition.

(2) If c is a complex scalar and $\mathbf{v}$ is in V, then so is $c\mathbf{v}$. Hence V is closed under scalar multiplication.

(3) There exists a zero vector $\mathbf{0}$ in V such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for all in V.

(4) For each $\mathbf{v}$ in V there exists an additive inverse (or opposite) vector $-\mathbf{v}$ in V such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v}$ in V.

(5) For all $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ in V and complex scalars c_1 and c_2, we have

  (a) $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$

  (b) $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$

  (c) $c_1(\mathbf{v}_1 + \mathbf{v}_2) = c_1\mathbf{v}_1 + c_1\mathbf{v}_2$

  (d) $(c_1 + c_2)\mathbf{v}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_1$

  (e) $(c_1c_2)\mathbf{v}_1 = c_1(c_2\mathbf{v}_1)$

  (f) $1 \cdot \mathbf{v}_1 = \mathbf{v}_1$

(Sect. 11.4)

Component

A single entry in a vector in Euclidean space. (Sect. 2.1)

Conjugate transpose

The conjugate transpose of a complex matrix A is denoted by A^* and defined

$$A^* = \overline{A}^T$$

(Sect. 11.5)

Consistent linear system

A linear system that has at least one solution. (Sect. 1.1)

Converge

An iterative process is said to converge if the outcome of a sequence of steps approaches a specific value. (Sect. 1.4; see also Sect. 3.5 and Sect. 6.5)

Coordinate vector

Suppose that $\mathcal{B}=\{\mathbf{u}_1,\dots,\mathbf{u}_n\}$ forms a basis for $\mathbf{R}^n$. If $\mathbf{y} = y_1\mathbf{u}_1 + \cdots + y_n\mathbf{u}_n$, then

$$[\mathbf{y}]_{\beta}=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

is the coordinate vector of $\mathbf{y}$ with respect to $\mathcal{B}$. (Sect. 4.4; see also Sect. 9.3)

Cramer’s Rule

Let $A=[\mathbf{a}_1 \cdots \mathbf{a}_n]$ be an invertible $n \times n$ matrix. Then the components of the unique solution x to $A\mathbf{x} = \mathbf{b}$ are given by

$$x_i=\frac{\det(A_i)}{\det(A)} \quad \text{for } i=1,2,\dots,n$$

where

$$A_i=[\mathbf{a}_1 \cdots \mathbf{a}_{i-1}\ \mathbf{b}\ \mathbf{a}_{i+1} \cdots \mathbf{a}_n]$$

(A_i is just A with column i replaced by $\mathbf{b}$.) (Sect. 5.3)

Determinant

If $A = [a_{11}]$ is a 1×1 matrix, then the determinant of A is given by $\det(A) = a_{11}$.

If

$$A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then the determinant is given by

$$\det(A)=a_{11}a_{22}-a_{12}a_{21}$$

For the $n \times n$ matrix

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

the determinant of A is defined recursively by

$$\det(A)=a_{11}C_{11}+a_{12}C_{12}+\cdots+a_{1n}C_{1n}$$

where $C_{11}, \dots, C_{1n}$ are the cofactors of $a_{11}, \dots, a_{1n}$, respectively. (Sect. 5.1)

Diagonal matrix

A diagonal matrix has the form

$$A=\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

(Sect. 3.2)

Diagonalizable matrix

An $n \times n$ matrix A is diagonalizable if there exist $n \times n$ matrices D and P, with D diagonal and P invertible, such that

$$A=PDP^{-1}$$

(Sect. 6.2)

Diagonally dominant

A linear system with the same number of equations and variables is diagonally dominant if the diagonal coefficients (a_{11}, a_{22},…) are each larger in absolute value than the sum of the absolute values of the other terms in the same row. (Sect. 1.4)

Dimension (subspace)

Let S be a nonzero subspace. Then the dimension of S is the number of vectors in any basis of S. (Sect. 4.2; see also Sect. 7.3)

Distance between vectors

For two vectors **u** and **v**, the distance between **u** and **v** is given by $\|\mathbf{u} - \mathbf{v}\|$. (Sect. 8.1; see also Sect. 10.1)

Diverge

An iterative process is said to diverge if the outcome of a sequence of steps fails to approach a specific value. (Sect. 1.4; see also Sect. 3.5 and Sect. 6.5)

Domain

The set of possible inputs for a function. (Sect. 3.1; see also Sect. 9.1)

Dominant eigenvalue

Suppose that a square matrix A has eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ such that

$$|\lambda_1|>|\lambda_2|\geq|\lambda_3|\geq\cdots\geq|\lambda_n|$$

In this case λ_1 is the dominant eigenvalue of A. (Sect. 6.5)

Dot product

Suppose

$$\mathbf{u}=\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \text{ and } \mathbf{v}=\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

are both in $\mathbf{R}^n$. Then the dot product of **u** and **v** is given by

$$\mathbf{u}\cdot\mathbf{v}=u_1v_1+\cdots+u_nv_n$$

The dot product can also be expressed $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v}$. (Sect. 8.1)

Doubly stochastic matrix

A square matrix A that has nonnegative entries, and has rows and columns that each add to 1. (Sect. 3.5)

Echelon form (linear system)

A linear system satisfying the following conditions: Every variable is the leading variable of *at most* one equation; the system is organized in a descending “stair step” pattern so that the index of the leading variables increases from the top to bottom; and every equation has a leading variable. Such a system is called an echelon system. (Sect. 1.1)

Echelon form (matrix)

Also called row echelon form, a matrix is in echelon form if every leading term is in a column to the left of the leading term of the row below it, and any zero rows are at the bottom of the matrix. (Sect. 1.2)

Eigenspace

Let A be a square matrix with eigenvalue λ. The subspace of all eigenvectors associated with λ, together with the zero vector, is the eigenspace of λ. Each distinct eigenvalue of A has its own associated eigenspace. (Sect. 6.1)

Eigenvalue

Let A be an $n \times n$ matrix. Suppose that λ is a scalar and $\mathbf{u} \neq \mathbf{0}$ is a vector satisfying

$$A\mathbf{u}=\lambda\mathbf{u}$$

The scalar λ is called an eigenvalue of A. (Sect. 6.1)

Eigenvector

Let A be an $n \times n$ matrix. Suppose that λ is a scalar and $\mathbf{u} \neq \mathbf{0}$ is a vector satisfying

Au=λu Then **u** is called an eigenvector of *A*. (Sect. 6.1)

Elementary (equation) operations

Three operations that can be performed on a linear system that do not change the set of solutions, so they yield an equivalent system. They are (1) interchanging two equations, (2) multiplying an equation by a nonzero constant, and (3) adding a multiple of one equation to another. (Sect. 1.2)

Elementary matrix

A square matrix E such that the product EA induces an elementary row operation on A. (Sect. 3.2; see also Sect. 3.4)

Elementary row operations

Three row operations that can be performed on an augmented matrix that do not change the set of solutions to the corresponding linear system. They are (1) interchanging two rows, (2) multiplying a row by a nonzero constant, and (3) adding a multiple of one row to another. (Sect. 1.2)

Equivalent matrices

Two matrices are equivalent if one can be transformed into the other through a sequence of elementary row operations. If the matrices in question are augmented matrices, then the corresponding linear systems have the same set of solutions. (Sect. 1.2)

Equivalent systems

Two linear systems are equivalent if they have the same set of solutions. (Sect. 1.2)

Euclidean space

The set of all vectors in $\mathbf{R}^n$ together with the "standard" definitions for vector arithmetic. (Sect. 2.1)

Fourier approximation

The nth order Fourier approximation for a function f in $C[-\boldsymbol{\pi}, \boldsymbol{\pi}]$ is given by

$$f_n(x)=a_0+a_1\cos(x)+\cdots+a_n\cos(nx) \ +b_1\sin(x)+\cdots+b_n\sin(nx)$$

where the coefficients are given by

$$a_0=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)\,dx \quad a_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos(kx)\,dx\quad(k\geq 1)\quad b_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin(kx)\,dx\quad(k\geq 1)$$

The a_k's and b_k's are called the Fourier coefficients of f. (Sect. 10.3)

Free parameter

An unspecified numerical quantity that can be equal to any real number. (Sect. 1.1)

Free variable

Any variable in a linear system in echelon form that is not a leading variable. (Sect. 1.1)

Full pivoting

An extension of partial pivoting where both row and column interchanges are performed to reduce round-off error when implementing elimination methods. This method is not covered in this text, but it is described in more advanced texts on numerical linear algebra. (Sect. 1.4)

Gauss–Jordan elimination

An algorithm that extends Gaussian elimination, applying row operations in a manner that will transform a matrix to reduced echelon form. (Sect. 1.2)

Gauss–Seidel iteration

A variant of Jacobi iteration, Gauss–Seidel iteration is an iterative method for approximating the solutions to a linear system that has the same number of equations as variables. (Sect. 1.4)

Gaussian elimination

An algorithm for applying row operations in a manner that will transform a matrix to echelon form. (Sect. 1.2)

General solution

A description of the set of all solutions to a linear equation or linear system. (Sect. 1.1)

Hermitian matrix

A complex matrix A is Hermitian if

$$A = A^{\star}$$

(Sect. 11.5)

Homogeneous equation

A linear equation is homogeneous if it has the form

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = 0$$

Such equations always have the trivial solution, so are consistent. (Sect. 1.2)

Homogeneous system

A linear system is homogeneous if it has the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= 0 \\ \vdots \qquad \vdots \qquad \vdots \qquad & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

Such systems always have the trivial solution, so are consistent. This system can also be expressed by $A\mathbf{x} = \mathbf{0}$. (Sect. 1.2; see also Sect. 2.3)

Hyperplane

The set of all solutions to a linear equation in four or more variables. (Sect. 1.1)

Idempotent matrix

A square matrix A is idempotent if $A^2 = A$. (Sect. 3.2)

Identity matrix

The $m \times m$ identity matrix is given by

$$I_m = [\mathbf{e}_1 \; \mathbf{e}_2 \; \cdots \; \mathbf{e}_m] = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

(Sect. 3.2)

Image

The output of a function from a particular input. (Sect. 3.1; see also Sect. 9.1)

Imaginary part

If *a* and *b* are real numbers, then a typical complex number has the form

$$z=a+ib$$

where *i* satisfies $i^2 = -1$. Here *b* is the imaginary part of *z*, denoted by Im(*z*). (Sect. 6.3)

Inconsistent linear system

A linear system that has no solutions. (Sect. 1.1)

Indefinite matrix

A symmetric matrix *A* is indefinite if *A* is the matrix of an indefinite quadratic form. (Sect. 11.2)

Indefinite quadratic form

Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form. Then *Q* is indefinite if $Q(\mathbf{x})$ is positive for some $\mathbf{x}$'s in $\mathbf{R}^n$ and negative for others. (Sect. 11.1)

Initial state vector

A vector with nonnegative entries that add to 1. This vector typically represents the initial probability distribution for a Markov chain. (Sect. 3.5)

Inner product (complex)

Let **u**, **v**, and **w** be elements of a complex vector space *V*, and let *c* be a complex scalar. An inner product on *V* is a function denoted by $\langle \mathbf{u}, \mathbf{v}\rangle$ that takes any two vectors in *V* as input and produces a scalar as output. An inner product on a complex vector space satisfies the following conditions:

(a) $\langle u,v\rangle=\overline{\langle v,u \rangle}$

(b) $\langle u,+v,w \rangle=\langle u,w \rangle+\langle v,w \rangle$

(c) $\langle cu,v \rangle=c\langle u,v \rangle$

(d) $\langle u,u \rangle\ge 0$, and $\langle u,u \rangle=0$ only when $\mathbf{u} = \mathbf{0}$

(Sect. 11.4)

Inner product (real)

Let **u**, **v**, and **w** be elements of a vector space V, and let c be a scalar. An inner product on V is a function that takes two vectors in V as input and produces a scalar as output. An inner product function is denoted by $\langle u,v \rangle$, and satisfies the following conditions:

(a) $\langle u,v \rangle=\langle v,u \rangle$

(b) $\langle u+v,w \rangle=\langle u,w \rangle+\langle v,w \rangle$

(c) $\langle cu,v \rangle=\langle u,cv \rangle=c\langle u,v \rangle$

(d) $\langle u,u \rangle\geq 0$, $\langle u,u \rangle=0$ only when $\mathbf{u} = \mathbf{0}$

(Sect. 10.1)

Inner product space

A vector space V with an inner product defined on it. (Sect. 10.1)

Inverse linear transformation

A linear transformation $T : \mathbf{R}^m \to \mathbf{R}^n$ is invertible if T is one-to-one and onto. When T is invertible, the inverse function $T^{-1}: \mathbf{R}^n \to \mathbf{R}^m$ is defined by

$$T^{-1}(y)=x \text{ if and only if } T(x)=y$$

(Sect. 3.3; see also Sect. 9.2)

Invertible matrix

An $n \times n$ matrix A is invertible if there exists an $n \times n$ matrix B such that $AB = I_n$. The matrix B is called the inverse of A and is denoted A^{-1}. (Sect. 3.3)

Isomorphic vector spaces

V and W are isomorphic vector spaces if there exists an isomorphism $T : V \to W$. (Sect. 9.2)

Isomorphism

A linear transformation $T : V \to W$ is an isomorphism if T is both one-to-one and onto. If such an isomorphism exists, then we say that V and W are isomorphic vector spaces. (Sect. 9.2)

Jacobi iteration

An iterative method for approximating the solutions to a linear system that has the same number of equations as variables. (Sect. 1.4)

Kernel

Given a linear transformation T, the set of all vectors $\mathbf{v}$ such that $T(\mathbf{v}) = \mathbf{0}$ is the kernel of T (denoted ker(T)) and is a subspace of the domain of T. (Sect. 4.1; see also Sect. 9.1)

LDU factorization

A variant of LU factorization, with $A = LDU$, where U is unit upper triangular, D is diagonal, and L is lower triangular with 1’s on the diagonal. (Sect. 3.4)

Leading principal submatrix

Let

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

Then the leading principal submatrices of A are given by

$$A_1=\begin{bmatrix} a_{11} \end{bmatrix}, A_2=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} A_3=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

and so on through $A_n = A$. (Sect. 11.2)

Leading term

In a matrix, the leading term for a row is the leftmost nonzero entry in that row. A row of zeros has no leading term. (Sect. 1.2)

Leading variable

The leftmost variable in a linear equation that has a nonzero coefficient. In a linear system in echelon form, the leading variables are the leftmost variables in each equation. (Sect. 1.1)

Least squares solution

If A is an $n \times m$ matrix and $\mathbf{y}$ is in $\mathbf{R}^n$, then a least squares solution to $A\mathbf{x} = \mathbf{y}$ is a vector $\hat{x}$ in $\mathbf{R}^m$ such that

$$\|A\hat{x}-y\| \le \|Ax-y\|$$

for all in $\mathbf{R}^m$. (Sect. 8.5)

Linear combination

If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are vectors and $c_1, c_2, \ldots, c_m$ are scalars, then

$$c_1u_1+c_2u_2+\cdots+c_mu_m$$

is a linear combination of the vectors. Note that it is possible for scalars to be negative or equal to zero. (Sect. 2.1; see also Sect. 7.2)

Linear equation

An equation of the form

$$a_1x_1+a_2x_2+x_3a_3+\cdots+a_nx_n=b$$

wherw $a_1, a_2, \ldots, a_n$ and b are constants and $x_1, x_2, \ldots, x_n$ are variables or unknowns. (Sect. 1.1)

Linear transformation

A function $T : \mathbf{R}^m \to \mathbf{R}^n$ is a linear transformation if for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbf{R}^m$ and all scalars r we have

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

(b) $T(r\mathbf{u}) = rT(\mathbf{u})$

(Sect. 3.1; see also Sect. 9.1)

Linearly dependent

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors. If the vector equation

$$x_1\mathbf{u}_1+x_2\mathbf{u}_2+\cdots+x_m\mathbf{u}_m=\mathbf{0}$$

has nontrivial solutions, then the set is linearly dependent. (Sect. 2.3; see also Sect. 7.2)

Linearly independent

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors. If the only solution to the vector equation

$$x_1\mathbf{u}_1+x_2\mathbf{u}_2+\cdots+x_m\mathbf{u}_m=\mathbf{0}$$

is the trivial solution $x_1 = x_2 = \cdots = x_m = 0$, then the set is linearly independent. (Sect. 2.3; see also Sect. 7.2)

Lower triangular matrix

An $n \times n$ matrix A is lower triangular if the terms above the diagonal are all zero,

$$A=\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

(Sect. 3.2)

LU factorization

If $A = LU$, where U is upper triangular and L is lower triangular with 1's on the diagonal, then the product is called an LU factorization of A. (Sect. 3.4)

Markov chain

A sequence of state vectors $\mathbf{x}_0$, $\mathbf{x}_1$, ⋯ generated recursively by $\mathbf{x}_{i+1} = A\mathbf{x}_i$, where A is a transition matrix. (Sect. 3.5)

Matrix

A rectangular table of numbers, upon which various algebraic operations are defined and can be performed. The plural of matrix is matrices. (Sect. 1.2)

Matrix addition

The component-wise rule for adding one matrix to another of identical dimension. (Sect. 3.2)

Matrix dimensions

The number of rows and columns for a matrix. Generally displayed as $n \times m$, where n is the number of rows and m the number of columns. (Sect. 3.1)

Matrix multiplication

The rules for multiplying two matrices to produce a new matrix. If A is $n \times m$ and B is $r \times s$, then the product AB is defined when = r. If this is true, then $n \times s$ are the dimensions of AB. (Sect. 3.2)

Matrix of a linear transformation

Let $T : V \to W$ be a linear transformation, $\mathcal{G} = \{g_1, \ldots, g_m\}$ a basis of V, and $\mathcal{Q} = \{q_1, \ldots, q_n\}$ a basis of W. If $A = [\mathbf{a}_1 \cdots \mathbf{a}_m]$ with

$$\text{ai} = [\text{T(gi)}]_{\mathcal{Q}}$$

for each $i = 1, \ldots, m$, then A is the matrix of T with respect to $\mathcal{G}$ and $\mathcal{Q}$. (Sect. 9.3)

Matrix of a quadratic form

An $n \times n$ matrix A used to define the quadratic form

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

(Sect. 11.1)

Matrix powers

If A is an $n \times n$ matrix, then

$$A^k = \underbrace{A \cdot A \cdots A}_{k \text{ terms}}$$

(Sect. 3.2)

Matrix-vector multiplication

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ be vectors in $\mathbf{R}^n$. If

$$A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_m] \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

then $A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m$. (Sect. 2.2)

Minor

If A is an $n \times n$ matrix, let M_{ij} denote the $(n - 1) \times (n - 1)$ matrix that we get from A after deleting the row and column containing a_{ij}. The determinant $\det(M_{ij})$ is the minor of a_{ij}. (Sect. 5.1)

Modulus

The modulus of a complex number $z = a + ib$ is given by

$$|z| = \sqrt{a^2 + b^2}$$

(Sect. 6.3)

Multiplicity of a root

Given a polynomial $P(x)$, a root α of $P(x) = 0$ has multiplicity r if $P(x) = (x - \alpha)^r Q(x)$ with $Q(\alpha) \neq 0$. (Sect. 6.1)

Negative definite matrix

A symmetric matrix A is negative definite if A is the matrix of a negative definite quadratic form. (Sect. 11.2)

Negative definite quadratic form

Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form. Then Q is negative definite if $Q(\mathbf{x}) < 0$ for all nonzero vectors $\mathbf{x}$ in $\mathbf{R}^n$, and Q is negative semidefinite if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ in $\mathbf{R}^n$. (Sect. 11.1)

Nonhomogeneous linear system

A linear system of the form $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b} \neq \mathbf{0}$. (Sect. 2.3)

Nonsingular matrix

A square matrix that has an inverse. (Sect. 3.3)

Nontrivial solution

A solution to a homogeneous linear equation (or homogeneous linear system) where some of the variables are nonzero. Although all homogeneous linear equations (systems) have the trivial solution, not all have nontrivial solutions. (Sect. 1.2)

Norm of a vector

Let $\mathbf{x}$ be a vector in $\mathbf{R}^n$. Then the norm (or length) of $\mathbf{x}$ is given by

$$\| x\| = x \cdot x$$

If $\mathbf{x}$ is in an inner product space, then the norm is given by

$$\| x\| = \langle x, x \rangle$$

(Sect. 8.1; see also Sect. 10.1)

Normal equations

Given a linear system $A\mathbf{x} = \mathbf{y}$, the normal equations for this system are

$$A^TAx = A^Ty$$

The set of solutions to the normal equations is the same as the set of least squares solutions to $A\mathbf{x} = \mathbf{y}$. If A has linearly independent columns, then there is a unique least squares solution given by

$$\hat{x} = (A^TA)^{-1}A^Ty$$

(Sect. 8.5)

Normal matrix

A complex matrix A is normal if

$$A^{\star}A = AA^{\star}$$

All unitary matrices are normal, but the reverse is not true. (Sect. 11.5)

Null space

If A is an $n \times m$ matrix, then the set of solutions to $A\mathbf{x} = \mathbf{0}$ is called the null space of A and is denoted by null(A). It is a subspace of $\mathbf{R}^m$. (Sect. 4.1)

Nullity

If A is an $n \times m$ matrix, then the nullity of A (denoted nullity(A)) is the dimension of null(A). (Sect. 4.2)

One-to-one

A linear transformation $T : \mathbf{R}^m \to \mathbf{R}^n$ is one-to-one if for every vector $\mathbf{w}$ in $\mathbf{R}^n$ there exists *at most* one vector $\mathbf{u}$ in $\mathbf{R}^m$ such that $T(\mathbf{u}) = \mathbf{w}$. Alternate definition: A linear transformation T is one-to-one if $T(\mathbf{u}) = T(\mathbf{v})$ implies that $\mathbf{u} = \mathbf{v}$. (Sect. 3.1; see also Sect. 9.1)

Onto

A linear transformation $T : \mathbf{R}^m \to \mathbf{R}^n$ is onto if for every vector $\mathbf{w}$ in $\mathbf{R}^n$ there exists *at least* one vector $\mathbf{u}$ in $\mathbf{R}^m$ such that $T(\mathbf{u}) = \mathbf{w}$. (Sect. 3.1; see also Sect. 9.1)

Orthogonal basis

A basis is orthogonal if it is an orthogonal set. (Sect. 8.1; see also Sect. 10.2)

Orthogonal complement

Let S be a subspace. A vector $\mathbf{u}$ is orthogonal to S if $\mathbf{u}$ is orthogonal to every vector $\mathbf{s}$ in S. The set of all such vectors $\mathbf{u}$ is called the orthogonal complement of S and is denoted by $S^\perp$. (Sect. 8.1; see also Sect. 10.1)

Orthogonal matrix

A square matrix is orthogonal if the columns form an orthonormal set. (Sect. 8.3)

Orthogonal set

A set of vectors is orthogonal if each pair of distinct vectors is orthogonal to each other. (Sect. 8.1; see also Sect. 10.2)

Orthogonal vectors

Vectors **u** and **v** in $\mathbf{R}^n$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$. If **u** and **v** are in an inner product space, then they are orthogonal if $\langle \mathbf{u},\mathbf{v} \rangle = 0$. (Sect. 8.1; see also Sect. 10.1)

Orthogonally diagonalizable matrix

A square matrix A is orthogonally diagonalizable if there exists an orthogonal matrix P and a diagonal matrix D such that $A = PDP^{-1}$. (Sect. 8.3)

Orthonormal basis

A basis is orthonormal if it forms an orthonormal set. (Sect. 8.2; see also Sect. 10.2)

Orthonormal set

A set of vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is orthonormal if the set is orthogonal and $||\mathbf{w}_j|| = 1$ for each of $j = 1, 2, \ldots k$. (Sect. 8.2; see also Sect. 10.2)

Parallelogram Rule

A geometric interpretation of vector addition that involves viewing vectors as two of the four sides of a parallelogram. (Sect. 2.1)

Partial pivoting

An additional step in Gaussian elimination (or Gauss–Jordan elimination) where a row interchange is performed to move the largest term (in absolute value) for a column into the pivot position. This is done to reduce roundoff error. (Sect. 1.4)

Particular solution

Any specific solution to a linear system $A\mathbf{x} = \mathbf{b}$. (Sect. 2.3)

Partitioned matrix

A matrix that has been divided into smaller submatrices. (Sect. 3.2)

Pivot columns

The columns containing pivot positions for a matrix in echelon form. (Sect. 1.2)

Pivot positions

For a matrix in echelon form, the pivot positions are the locations of the leading terms. (Sect. 1.2)

Pivot rows

The rows containing pivot positions for a matrix in echelon form. (Sect. 1.2)

Positive definite matrix

A symmetric matrix A is positive definite if A is the matrix of a positive definite quadratic form. (Sect. 11.2)

Positive definite quadratic form

Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form. Then Q is positive definite if $Q(\mathbf{x}) > 0$ for all nonzero vectors $\mathbf{x}$ in $\mathbf{R}^n$, and Q is positive semidefinite if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ in $\mathbf{R}^n$. (Sect. 11.1)

Power method

An iterative method for approximating an eigenvector and corresponding eigenvalue for a square matrix. (Sect. 6.5)

Principal axes

The normalized orthogonal eigenvectors of a symmetric matrix A used to define a quadratic form. (Sect. 11.1)

Probability vector

A vector with nonnegative entries that add to 1. These vectors are encountered in the context of Markov chains. (Sect. 3.5)

Projection onto subspaces

Let S be a subspace with orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$. Then the projection of $\mathbf{u}$ onto S is given by

$$\mathrm{proj}_S \mathbf{u} = \frac{\mathbf{v}_1 \cdot \mathbf{u}}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\mathbf{v}_2 \cdot \mathbf{u}}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\mathbf{v}_k \cdot \mathbf{u}}{\|\mathbf{v}_k\|^2}\mathbf{v}_k$$

For an inner product space, the dot products are replaced by inner products in the definition. (Sect. 8.2; see also Sect. 10.2)

Projection onto vectors

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbf{R}^n$ with $\mathbf{v}$ nonzero. Then the projection of $\mathbf{u}$ onto $\mathbf{v}$ is given by

$$\mathrm{proj}_\mathbf{v} \mathbf{u} = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{v}\|^2}\mathbf{v}$$

If **u** and **v** are in an inner product space, then the projection of **u** onto **v** is given by

$$\mathrm{proj}_{\mathbf{v}}\mathbf{u} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\|^2}\mathbf{v}$$

(Sect. 8.2; see also Sect. 10.1)

QR factorization

The QR factorization of an $n \times m$ matrix A is given by $A = QR$, where Q is $n \times m$ with orthonormal columns and R is $m \times m$, upper triangular, and has positive diagonal terms. (Sect. 8.3)

Quadratic form

A quadratic form is a function $Q : \mathbf{R}^n \to \mathbf{R}$ that has the form

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

where A is an $n \times n$ symmetric matrix called the matrix of the quadratic form. (Sect. 11.1)

R^n

The set of all vectors with n real numbers for components. (Sect. 2.1)

Range

The set of outputs for a function. (Sect. 3.1; see also Sect. 9.1)

Rank of a matrix

The dimension of the row space of a matrix A, or the dimension of the column space of A, which is the same. (Sect. 4.3)

Rank–Nullity theorem

Given an $n \times m$ matrix A, the Rank–Nullity theorem states that

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = m$$

(Sect. 4.3)

Real part

If a and b are real numbers, then a typical complex number has the form

$$z = a + ib$$

where i satisfies $i^2 = -1$. Here a is called the real part of z, denoted by Re(z). (Sect. 6.3)

Reduced echelon form

Also called row reduced echelon form, a matrix in this form is in echelon form and each pivot column consists entirely of zeros except for in the pivot position, which contains a 1. (Sect. 1.2)

Regular matrix

A stochastic matrix A is regular if for some integer $k \geq 1$ the matrix A^k has all strictly positive entries. (Sect. 3.5)

Row space

Let A be an $n \times m$ matrix. The row space of A is the subspace of $\mathbf{R}^m$ spanned by the row vectors of A and is denoted by row(A). (Sect. 4.3)

Row vector

A vector in Euclidean space expressed in the form of a horizontal n-tuple. This term is also used to indicate a vector formed from a row of a matrix A. (Sect. 2.1; see also Sect. 4.3)

Scalar

A real number when viewed as a multiple of a vector. (Sect. 2.1)

Scalar multiplication (matrices)

The multiplication of each term of a matrix by a real number. (Sect. 3.2)

Scalar multiplication (vectors)

The multiplication of each component of a vector by a real number. (Sect. 2.1)

Similar matrices

A square matrix A is similar to matrix B if there exists an invertible matrix S such that $A = S^{-1}BS$. The change from B to A is called a similarity transformation. (Sect. 9.4)

Singular matrix

A square matrix that does not have an inverse. (Sect. 3.3)

Singular value decomposition

Suppose that A is an $n \times m$ matrix. If $n \geq m$, then the singular value decomposition is the factorization of A as the product $A = U\Sigma V^T$, where

- U is an $n \times n$ orthogonal matrix.
- Σ is an $n \times m$ matrix of the form ∑=[D0(n-m)m], where D is a diagonal matrix with

  D=[σ10⋯00σ2⋯0⋮⋮⋱⋮00⋯σm]

  and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$ are the singular values of A. The singular values are given by σi=λi, where λ_i is an eigenvalue of A^TA.
- V is an $m \times m$ orthogonal matrix.

If $n < m$, then ∑=[D0n(m-n)] with everything else the same. (Sect. 8.4)

Skew symmetric matrix

A square matrix A is skew symmetric if $A^T = -A$. (Sect. 3.2)

Solution, linear equation

An n-tuple $(s_1, \ldots, s_n)$ that satisfies a linear equation. (Sect. 1.1)

Solution, linear system

An n-tuple $(s_1, \ldots, s_n)$ that satisfies a linear system. (Sect. 1.1)

Solution set

The set of all solutions to a linear equation or a linear system. (Sect. 1.1)

Span

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set of vectors. The span of this set is denoted span $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ and is defined to be the set of all linear combinations

x1u1+x2u2+⋯+xmum

where $x_1, x_2, \ldots, x_m$ can be any real numbers. (Sect. 2.2; see also Sect. 7.2)

Square matrix

A matrix with the same number of rows and columns. (Sect. 3.1)

Standard basis (R^n)

Given by the vectors

$$\mathbf{e}_1=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix},\ \mathbf{e}_2=\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix},\ \cdots,\ \mathbf{e}_n=\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$$

(Sect. 4.2)

State vector
A probability vector that is part of the sequence of vectors in a Markov chain. (Sect. 3.5)

Steady-state vector
A state vector $\mathbf{x}$ that satisfies $A\mathbf{x} = \mathbf{x}$, where A is a transition matrix for a Markov chain. (Sect. 3.5)

Stochastic matrix
A square matrix A that has nonnegative entries and columns that each add to 1. (Sect. 3.5)

Subspace
A subset S of vectors is a subspace if S satisfies the following three conditions: (1) S contains $\mathbf{0}$, the zero vector, (2) if $\mathbf{u}$ and $\mathbf{v}$ are in S, then $\mathbf{u} + \mathbf{v}$ is also in S, and (3) if r is a real number and $\mathbf{u}$ is in S, then $r\mathbf{u}$ is also in S. (Sect. 4.1; see also Sect. 7.1)

Subspace spanned
If $S = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$, then S is the subspace spanned (or subspace generated) by $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$. (Sect. 4.1)

Symmetric matrix
A matrix A is symmetric if $A^T = A$. (Sect. 3.2)

System of linear equations
A collection of linear equations of the form

$$\begin{array}{ccccccccccc} a_{11}x_1 &+& a_{12}x_2 &+& a_{13}x_3 &+& \cdots &+& a_{1n}x_n &=& b_1 \\ a_{21}x_1 &+& a_{22}x_2 &+& a_{23}x_3 &+& \cdots &+& a_{2n}x_n &=& b_2 \\ a_{31}x_1 &+& a_{32}x_2 &+& a_{33}x_3 &+& \cdots &+& a_{3n}x_n &=& b_3 \\ \vdots & & \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 &+& a_{m2}x_2 &+& a_{m3}x_3 &+& \cdots &+& a_{mn}x_n &=& b_m \end{array}$$

(Sect. 1.1)

System of linear first-order differential equations
The general form for such a system is

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3 + \cdots a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + a_{23}y_3 + \cdots a_{2n}y_n \\ y_3' &= a_{31}y_1 + a_{32}y_2 + a_{33}y_3 + \cdots a_{3n}y_n \\ &\vdots \qquad \vdots \qquad \vdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + a_{n3}y_3 + \cdots a_{nn}y_n \end{aligned}$$

Here we assume that $y_1 = y_1(t), \ldots, y_n = y_n(t)$ are each differentiable functions. The system is linear because the functions are linearly related, and it is first-order because only the first derivative appears. (Sect. 6.4)

Theorem

A mathematical statement that has been rigorously proved to be true. (Sect. 1.1)

Tip-to-Tail rule

A geometric interpretation of vector addition that involves translating one vector so that its initial point (the tail) is situated at the end point (the tip) of the other. (Sect. 2.1)

Transformation matrix

Another term for the matrix of a linear transformation. (Sect. 9.3)

Transition matrix

A stochastic matrix A used to proceed from one state vector to the next in a Markov chain via the relationship $\mathbf{x}_{i+1} = A\mathbf{x}_i$. (Sect. 3.5)

Transpose

The transpose of a matrix A is denoted by A^T and results from interchanging the rows and columns of A. (Sect. 3.2)

Triangular form

A linear system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\ \ddots \quad \vdots \quad &\vdots \\ a_{nn}x_n &= b_n \end{aligned}$$

where $a_{11}, a_{22}, \cdots, a_{nn}$ are all nonzero. Also called a triangular system. (Sect. 1.1)

Trivial solution

The solution to a homogeneous linear equation (or homogeneous linear system) where all variables are set equal to zero. (Sect. 1.2)

Trivial subspaces

$S = \{\mathbf{0}\}$ and $S = \mathbf{R}^n$ are the trivial subspaces of $\mathbf{R}^n$. For a vector space V, the trivial subspaces are $S = \{\mathbf{0}\}$ and $S = V$. (Sect. 4.1; see also Sect. 7.1)

Unit vector

A vector with a norm (length) equal to 1. (Sect. 10.2)

Unitarily diagonalizable

A complex matrix A is unitarily diagonalizable if there exist a diagonal matrix D and a unitary matrix P such that

$$A = PDP^{-1} = PDP^{*}$$

As with real matrices, the diagonal entries of D are the eigenvalues of A, and the columns of P are the corresponding eigenvectors. (Sect. 11.5)

Unitary matrix

The counterpart of orthogonal matrices for a matrix A with complex entries is called a unitary matrix, which requires that

$$A^{-1} = A^{*}$$

(Sect. 11.5)

Unitary space

A complex vector space V with an inner product defined on it. (Also called a complex inner product space.) (Sect. 11.4)

Upper triangular matrix

An $n \times n$ matrix A is upper triangular if it has the form

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

That is, A is upper triangular if the entries below the diagonal are all zero. (Sect. 3.2)

VecMobile

An imaginary vehicle created to illustrate the notion of span. (Sect. 2.2)

Vector

In Euclidean space, an ordered list of real numbers usually presented in a vertical column. In general, a vector can be any number of different mathematical objects, including matrices and continuous functions. (Sect. 2.1; see also Sect. 7.1)

Vector arithmetic

The “standard” definition of equality, addition, and scalar multiplication as it is applied to vectors in Euclidean space. (Sect. 2.1)

Vector form

A specific way to describe the general solution to a linear system using a linear combination of vectors. (Sect. 2.1)

Vector space

A vector space consists of a nonempty set V of vectors together with operations of addition and scalar multiplication on the vectors that satisfy each of the following:

(1) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in V, then so is $\mathbf{v}_1 + \mathbf{v}_2$. Hence V is closed under addition.

(2) If c is a real scalar and $\mathbf{v}$ is in V, then so is $c\mathbf{v}$. Hence V is closed under scalar multiplication.

(3) There exists a zero vector $\mathbf{0}$ in V such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for all v in V.

(4) For each $\mathbf{v}$ in V there exists an additive inverse (or opposite) vector $-\mathbf{v}$ in V such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v}$ in V.

(5) For all $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ in V and real scalars c_1 and c_2, we have

(a) $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$

(b) $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$

(c) $c_1(\mathbf{v}_1 + \mathbf{v}_2) = c_1\mathbf{v}_1 + c_1\mathbf{v}_2$

(d) $(c_1 + c_2)\mathbf{v}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_1$

(e) $(c_1c_2)\mathbf{v}_1 = c_1(c_2\mathbf{v}_1)$

(f) $1 \cdot \mathbf{v}_1 = \mathbf{v}_1$

(Sect. 7.1)

Zero column

A matrix column consisting entirely of zeros. (Sect. 1.2)

Zero row

A matrix row consisting entirely of zeros. (Sect. 1.2)

Zero subspace

The zero subspace $S = \{\mathbf{0}\}$ is the subspace consisting solely of the zero vector $\mathbf{0}$. It is the only subspace of $\mathbf{R}^n$ that does not have a basis. (Sect. 4.2)

Zero vector

In $\mathbf{R}^n$, a vector with zeros for each component. In a vector space, a vector that is the counterpart to 0 in the real numbers. (Sect. 2.1)

SOLUTIONS TO PRACTICE PROBLEMS

Chapter 1

Section 1.1

1.

(a) $-2x_1+8x_2=5 \Rightarrow x_2=58+14x_1$. Substitute into the second equation to obtain $3x_1-12(58+14x_1)=4 \Rightarrow -152 \neq 4$. Thus no solution exists.

(b) $x_1-2x_2=3 \Rightarrow x_2=12x_1-32$. Substitute into the second equation to obtain $-3x_1+6(12x_1-32)=-9 \Rightarrow -9=-9$, which is true for all x_1. Therefore we may set x_1 as a free variable, $x_1 = s_1$ and then $x_2=12s_1-32$.

2.

(a) The fourth equation, $0 = -2$, does not hold true, so no solutions exist.

(b) x_2 and x_4 are free variables, so let $x_2 = s_1$ and $x_4 = s_2$. From the third equation, $x_5 = 4$. Substitute into the second equation to obtain

$$x_3-2s_2+4=2x_3=2s_2-2$$

Now substitute into the first equation to obtain

$$x_1-s_1-2(2s_2-2)+s_2-2(4)=1x_1=5+s_1+3s_2$$

3.

(a) False, by Property (c) of triangular systems.

(b) True. It will have 5 pivot variables, so it must have 3 free variables.

(c) False. For example

$$x+y=12x+2y=2x-y=1$$

has exactly one solution.

(d) False. In the system

$$x_1=1x_2=2$$

there are no free variables or free parameters.

4.

(a) There are 4 leading variables.

(b) There are 5 free variables.

(c) There are 5 free parameters.

(d) There are infinitely many solutions.

5. Let x be the number of floor seats, and y the number of balcony seats. We have $x + y = 280$, because the theater capacity is 280. And we have $22x + 14y = 5320$, because the sales total is \$5320. From the first equation, $y = 280 - x$. Substitute into the second equation to obtain

$$22x+14(280-x)=5320\Rightarrow 8x+3920=5320\Rightarrow 8x=1400\Rightarrow x=175$$

So there are 175 floor seats, and $y = 280 - 175 = 105$ balcony seats.

6. Let x be the number of nickels, y the number of dimes, and the z number of quarters. Because the quarters are worth \$2.75, we have $25z = 275$, so $z = 11$. The dimes and quarters are worth \$3.65, so we have

$$10y+25z=365\Rightarrow 10y+275=365\Rightarrow 10y=90\Rightarrow y=9$$

There are 31 coins, so $x + y + z = 31$, and because $z = 11$ and $y = 9$, we have $x = 11$.

Section 1.2

1.

(a) [1-2122-355-132-5]-2R1+R2→R2~ [1-2120131-132-5] R1+R3→R3~ [1-2120131013-3] -R2+R3→R3~ [1-2120131000-4]

(b) [13-112126-1536-1-336-4413-2-141]-2R1+R2→R2R1+R3→R3-R1+R4→R4~[13-11210013-140027-2500-1-220]-2R2+R3→R3R2+R4→R4~ [13-11210013-1400010-3000114]-R3+R4→R4~ [13-11210013-1400010-3000017]

2.

(a) [1-232-35-14-5]-2R1+R2→R2R1+R3→R3~[1-2301-102-2] -2R2+R3→R3~ [1-2301-1000] 2R2+R1→R1~ [10101-1000]

(b) [1421-1-4-102864]R1+R2→R2-2R1+R3→R3~ [142100110022] -2R2+R3→R3~ [142100110000] -2R2+R1→R1~ [140-100110000]

3.

(a) [-12-3-1-13-1-32-210-2]-R1+R2→R22R1+R3→R3~[-12-31012-2024-4] -2R2+R3→R3~ [-12-3-1012-20000]

Free variable, $x_3 = s$. Row 2 ⇒ $x_2 + 2s = -2$ ⇒ $x_2 = -2 - 2s$. Row 1 ⇒ $-x_1 + 2(-2 - 2s) - 3s = -1$ ⇒ $x_1 = -3 - 7s$.

(b) [1-1-21-201-1-1-1-10-111-360] -R1+R2→R2R1+R3→R3~[1-1-21-20001-21000-1-240] R2+R3→R3~ [1-1-21-20001-210000-450]
Free variables, $x_5 = s_1$ and $x_2 = s_2$.
Row 3⇒-4x4+5s1=0⇒x4=54s1.Row 2⇒x3-2(54s1)+s1=0⇒x3=32s1.Row 1⇒x1-(s2)-2(32s1)+(54s1)-2(s1)=0⇒x1=154s1+s2.

4.

(a) [12141121]-R1+R2→R2~[12140-11-3] 2R2+R1→R1~[103-20-11-3] -R2→R2~[103-201-13]
Free variable $x_3 = s$. Row 2 ⇒ $x_2 - (s) = 3$ ⇒ $x_2 = 3 + s$. Row 1 ⇒ $x_1 + 3(s) = -2$ ⇒ $x_1 = -2 - 3s$.

(b) [1-13-12-14-1-13-64]-2R1+R2→R2R1+R3→R3~[1-13-101-2102-33] -2R2+R3→R3~ [1-13-101-210011] 2R3+R2→R2-3R3+R1→R1~[1-10-401030011] R2+R1→R1~ [100-101030011]
From Row 1, $x_1 = -1$, from Row 2, $x_2 = 3$ and from Row 3, $x_3 = 1$.

5.

(a) False. Every matrix can be transformed to reduced row echelon form.

(b) True. Suppose the matrix is $m \times n$, with $m > n$. No column can have more than one pivot if the matrix is in echelon form, so the number of rows with a pivot is, at most, n. Since there are rows, and $m > n$, the matrix must have a zero row.

(c) True. The reverse operation of Ri↔Rj is Rj↔Ri. The reverse operation of $cR_i \to R_i$ is 1cRi→Ri. And the reverse operation of $cR_i + R_j \to R_j$ is $-cR_i + R_j \to R_j$.

(d) True. If there exists a solution, then there will be infinitely many solutions because any free variable can take infinitely many values in **R**.

6.

(a) 4

(b) 3

(c) 6

(d) 4

Section 1.3

1. At equilibrium, we have

x3=x1+x2+403x2=x1+x3+703x1=x2+x3+303

Rearranging, we have

x1+x2-3x3=-40x1-3x2+x3=-70-3x1+x2+x3=-30

Row-reduce the augmented matrix, and obtain [11-3-401-31-70-311-30]-R1+R2→R23R1+R3→R3~[11-3-400-44-3004-8-150] R2+R3→R3~ [11-3-400-44-3000-4-180]
So Row 3⇒x3=-180-
4=45. Row 2⇒-4x2+4(45)=-30⇒x2=1052. Row 1⇒x1+
(1052)-3(45)=-40⇒x3=852.

2. As in Example 3, we determine output to satisfy consumer and between-industry demand, and obtain the equations

a=50+0.2bb=80+0.35b

We may substitute the second equation into the first to obtain

a=50+0.2(80+0.35a) =0.07a+660.93a=66a≈71.0

Then substitute into equation 2 to obtain $b \approx 80 + 0.35(71.0) = 104.9$.

3. We consider $x_1CO_2 + x_2H_2O \rightarrow x_3C_6H_{12}O_6 + x_4O_2$, which implies

x1-6x3=02x1+x2-6x3-2x4=02x2-12x3=0

Row-reduce the augmented matrix [10-60021-6-2002-1200]-2R1+R2→R2~ [10-600016-2002-1200] -2R2+R3→R3~ [10-600016-2000-2440]

We set $x_4 = s_1$ as a free variable. From row 3,
-24x3+4s1=0⇒x3=16s1. From row 2,
x2+6(16s1)-2(s1)=0⇒x2=s. From row 1, x1-6(16s1)=0⇒x1=s1.
We set $s_1 = 6$ to obtain $x_1 = 6$, $x_2 = 6$, $x_3 = 1$, and the balanced equation

6CO2+6H2O→C6H12O6+6O2

4. Assuming $p = ad^b$, so that $\ln(p) = \ln(a) + b\ln(d)$, and letting $a_1 = \ln(a)$, we obtain the following equations using the data for Earth and Neptune

a1+b ln(149.6)=ln(365.2)a1+b ln(4495.1)=ln(59800)

The solution to this system is $a_1 = -1.6029$ and $b = 1.4983$. Thus, a=ea1=e-1.6029=0.2013. Therefore, $p = (0.2013)d^{1.4983}$.

5. Multiply both sides of the equation by $(2x + 1)(x - 1)$ to obtain $(x + 5) = A(x - 1) + B(2x + 1) = (A + 2B)x + (-A + B)$. Equate coefficients of and the constant terms to obtain

A+2B=1-A+B=5

The solution to this system is $A = -3$ and $B = 2$.

6.

(a) False. See Example 1.

(b) True. Any positive integer multiple of a solution will also balance the equation.

(c) False. For example, no parabola or the form $y = ax^2 + bx + c$ will pass through the three points (0, 0), (0, 1), (0, 2).

(d) False. For example, $f(x) = 5e^x$ and $f(x) = 5e^{-2x}$ are both of the form $f(x) = ae^x + be^{-2x}$ and $f(0) = 5$.

Section 1.4

1.

(a) Using Gaussian elimination with 3 significant digits of accuracy:

[15625249-78-11] -49R1+R2→R2~ [1562520-27600-2560]

Row 2⇒x2=-2560-27600=9.28×10-2. Row 1 ⇒ $x_1 + 562(9.28 \times 10^{-2}) = 52 \Rightarrow x_1 = -0.154$.

Using partial pivoting:

[15625249-78-11]R1↔R2~[49-78-11156252] (-1/49)R1+R2→R2~[49-78-11056452.2]

Row 2⇒x2=52.2564=9.26×10-2. Row 1 ⇒ $49x_1 - 78(9.26 \times 10^{-2}) = -11 \Rightarrow x_1 = -7.71 \times 10^{-2}$.

(b) Using Gaussian elimination with 3 significant digits of accuracy:

[2-859815-37913567-398411](3/2)R1+R2→R2(-67/2)R1+R3→R3~ [2-8598150-5.01810.027.50229.0-19900.0-492.0](229.0/5)R2+R3→R3~ [2-8598150-5.01810.027.500.0063000.0768.0]

Row 3⇒x3=768.063000.0=1.22×10-2. Row 2 ⇒ $-5.0x_2 + (1810.0)(1.22 \times 10^{-2}) = 27.5 \Rightarrow x_2 = -1.08$. Row 1 ⇒ $2x_1 - 8(-1.08) + 598(1.22 \times 10^{-2}) = 15 \Rightarrow x_1 = -0.468$.

Using partial pivoting.

[2-859815-37913567-398411]R1↔R3~ [67-398411-3791352-859815] (3/67)R1+R2→R2(-2/67)R1+R3→R2~[67-39841105.25917.05.490-6.84595.014.7] R2↔R3~ [67-3984110-6.84595.014.705.25917.05.49] (5.25/6.84)R2+R3→R3~ [67-3984110-6.84595.014.700.001370.016.8]

Row 3⇒x3=16.81370.0=1.23×10-2. Row 2 ⇒ $-6.84x_2 + (595.0)(1.23 \times 10^{-2}) = 14.7 \Rightarrow x_2 = -08$. Row 1 ⇒ $67x_1 - 39(-1.08) + 84(1.23 \times 10^{-2}) = 11 \Rightarrow x_1 = -0.480$.

2.

n	x_1	x_2
0	0	0
1	−2.25	0.385
2	−2.15	0.731
3	−2.07	0.716

Exact solution: $x_1 = -2.07$, $x_2 = 0.704$.

(a)

n	x_1	x_2	x_3
0	0	0	0
1	0.913	−3	1.38
2	0.935	−2.49	2.73
3	0.655	−1.89	2.54

Exact solution: $x_1 = 0.689$, $x_2 = -2.05$, $x_3 = 2.32$.

3.

(a) Gauss–Seidel iteration of given linear system:

n	x_1	x_2
0	0	0
1	−2.25	0.731
2	−2.07	0.703
3	−2.07	0.704

Exact solution: $x_1 = -2.07$, $x_2 = 0.704$.

(b) Gauss–Seidel iteration of given linear system:

n	x_1	x_2	x_3
0	0	0	0
1	0.913	−3.10	2.77
2	0.702	−1.85	2.24
3	0.684	−2.07	2.32

Exact solution: $x_1 = 0.689$, $x_2 = -2.05$, $x_3 = 2.32$.

Chapter 2

Section 2.1

1. u-w=[-434]-[50-2]=[-4-53-04-(-2)]=[-936]v+3w=[-162]+3[50-2]=[-1+3(5)6+3(0)2+3(-2)]=[146-4]-2w+u+3v=-2[50-2]+[-43-4]+3[-162] =[-2(5)+(-4)+3(-1)-2(0)+3+3(6)-2(-2)+4+3(2)]=[-172114]

2.

   (a) -x1+4x2=37x1+6x2=102x1-6x2=5

   (b) 3x1 -x3=44x1-2x2+2x3=7 -5x2+9x3=112x1+6x2+5x3=-6

3.

   (a) x1[1-54]+x2[170]+x3[-26-8]=[3120]

   (b) x1[404]+x2[-3212]+x3[-156]+x4[5-20]=[0610]

4.

   (a) [x1x2x3]=[570]+s1[3-21]

   (b) [x1x2x3x4]=[10170]+s1[2011]+s2[13100]

5.

   (a) x1a1+x2a2=b⇔x1[1-5]+x2[36]=[59]⇔[x1+3x2-5x1+6x2]=[59]⇔

   the augmented matrix [135-569] has a solution:

   [135-569]5R1+R2→R2~[13502134]

   From row 2, 21x2=34⇒x2=3421. From row 1, x1+3(341)=5⇒x1=17. Thus, **b** is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$, with b=17a1+3421a2.

   (b) x1a1+x2a2=b⇔x1[1-38]+x2[-23-3]=[75-4]⇔[x1-2x2-3x1+3x28x1-3x2]=[75-4]⇔

   the augmented matrix [1-27-3358-3-4] yields a solution.

   [1-27-3358-3-4]3R1+R2→R2-8R1+R3→R3~[1-270-326013-60] (133)R2+R3→R3~ [1-270-326001583]

   From the third equation, we have 0=1583, and thus the system does not have a solution. Thus, **b** is *not* a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

6.

(a) False. Addition of vectors is associative and commutative.

(b) True. The scalars may be any real number.

(c) True. The solutions to a linear system with variables $x_1, \dots, x_n$ can be expressed as a vector **x**, which is the sum of a fixed vector with n components and a linear combination of k vectors with n components, where k is the number of free variables.

(d) False. The Parallelogram Rule gives a geometric interpretation of vector addition.

Section 2.2

1.

(a) 0u1+0u2=0[2-3]+0[41]=[00],   1u1+0u2=1[2-3]+0[41]=[2-3],0u1+1u2=0[2-3]+1[41]=[41]

(b) 0u1+0u2=0[614]+0[-23-3]=[000], 1u1+0u2=1[614]+0[-23-3]=[614], 0u1+1u2=0[614]+1[-23-3]=[-23-3]

2. Set x1u1+x2u2=b⇒x1[12-2]+x2[043]=[-125]⇒[x12x1+4x2-2x1+3x2]=[-125]. From the first equation, $x_1 = -1$. Then the second equation is $2(-1) + 4x_2 = 2 \Rightarrow x_2 = 1$. The third equation is now $-2(-1) + 3(1) = 5 \Rightarrow 5 = 5$. So **b** is in the span of $\{\mathbf{u}_1, \mathbf{u}_2\}$, with $(-1)\mathbf{u}_1 + (1)\mathbf{u}_2 = \mathbf{b}$.

3.

(a) A=[7-2-2-1743-1-2], x=[x1x2x3], b=[6111]

(b) A=[4-3-1531260], x=[x1x2x3x4], b=[010]

4.

(a) Row-reduce to echelon form:

[23-1-2](1/2)R1+R2→R2~[230-12]

There is not a row of zeros, so every choice of **b** is in the span of the columns of the given matrix and, therefore, the columns of the matrix span $\mathbf{R}^2$.

(b) Row-reduce to echelon form:

[411-3](-1/4)R1+R2→R2~[410-134]

Since there is not a row of zeros, every choice of **b** is in the span of the columns of the given matrix, and therefore the columns of the matrix span $\mathbf{R}^2$.

5.

(a) Row-reduce to echelon form:

[13-1-1-23025]R1+R2→R2~ [13-1012025] -2R2+R3→R3~ [13-1012001]

There is not a row of zeros, so every choice of **b** is in the span of the columns of the given matrix and, therefore, the columns of the matrix span $\mathbf{R}^3$.

(b) Row-reduce to echelon form:

[2061-21-141](-1/2)R1+R2→R2(1/2)R1+R3→R3~[2060-2-2044] 2R2+R3→R3~ [2060-2-2000]

Because there is a row of zeros, there exists a vector **b** that is not in the span of the columns of the matrix and, therefore, the columns of the matrix do not span $\mathbf{R}^3$.

6.

(a) False. If the vectors span $\mathbf{R}^3$, then vectors have three components, and cannot span $\mathbf{R}^2$.

(b) True. Every vector **b** in $\mathbf{R}^2$ can be written as

b=x1u1+x2u2 =x12(2u1)+x23(3u2)

which shows that $\{2\mathbf{u}_1, 3\mathbf{u}_2\}$ spans $\mathbf{R}^2$.

(c) True. Every vector **b** in $\mathbf{R}^3$ can be written as $\mathbf{b} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + x_3\mathbf{u}_3$. So $A\mathbf{x} = \mathbf{b}$ has the solution

x=[x1x2x3].

(d) True. Every vector **b** in $\mathbf{R}^2$ can be written as $\mathbf{b} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + 0\mathbf{u}_3$, so $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ spans $\mathbf{R}^2$.

Section 2.3

1.

(a) Consider $x_1\mathbf{u}_1 + x_2\mathbf{u}_2 = \mathbf{0}$, and solve using the corresponding augmented matrix:

[240-310](3/2)R1+R2→R2~[240070]

The only solution is the trivial solution, so the vectors are linearly independent.

(b) Consider $x_1\mathbf{u}_1 + x_2\mathbf{u}_2 = \mathbf{0}$, and solve using the corresponding augmented matrix:

[6-201304-30](-1/6)R1+R2→R2(-2/3)R1+R3→R3~[6-20010300-530] (1/2)R2+R3→R3~ [6-2001030000]

The only solution is the trivial solution, so the vectors are linearly independent.

2.

(a) We solve the homogeneous system of equations using the corresponding augmented matrix:

[1503-40]-3R1+R2→R2~[1500-190]

The only solution is the trivial solution, so the columns of the matrix are linearly independent.

(b) We solve the homogeneous system of equations using the corresponding augmented matrix:

[10302-240-3720]-2R1+R2→R23R1+R3→R3~[10300-2-2007110] (7/2)R2+R3→R2~ [10300-2-200040]

There is only the trivial solution; the columns of the matrix are linearly independent.

3.

(a) We solve the homogeneous equation using the corresponding augmented matrix:

[14202840]-2R2+R3→R3~[14200000]

Because there exist nontrivial solutions, the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

(b) We solve the homogeneous equation using the corresponding augmented matrix:

[10-110-1-1010-22100]R1+R2→R22R1+R3→R3~[10-1100-1-12002-120] 2R2+R3→R3~ [10-1100-1-12000-360]

Because there exist nontrivial solutions, the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

4.

(a) False, because {[100], [010]} is linearly independent in $\mathbf{R}^3$ but does not span $\mathbf{R}^3$.

(b) True, by the Unifying Theorem.

(c) True. Because $\mathbf{u}_1 - 4\mathbf{u}_2 = 4\mathbf{u}_2 - 4\mathbf{u}_2 = \mathbf{0}$, $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly dependent.

(d) False. Suppose A=[1100], then the columns of *A* are linearly dependent, and Ax=[01] has no solutions.

Chapter 3

Section 3.1

1.

(a) T([2-1])=[3(2)+2(-1) -(2)+ (-1)-4(2)-3(-1)]=[4-3-5]

(b) A=[32-11-4-3]

(c) Because $n = 3 > m = 2$, by Theorem 3.7 T is not onto. To determine if T is one-to-one, we row-reduce the corresponding augmented matrix:

[320-110-4-30](1/3)R1+R2→R2(4/3)R1+R3→R3~[32005300-130] (1/5)R2+R3→R3~ [3200530000]

Because $T(\mathbf{x}) = A\mathbf{x} = \mathbf{0}$ has only the trivial solution, by Theorem 3.5 T is one-to-one.

2.

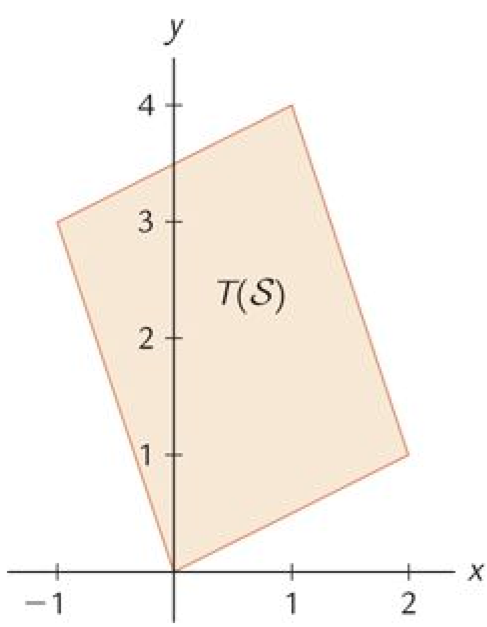
(a) T(u1+u2)=T(u1)+T(u2)=[23]+[-41]=[-24]

(b) T(3u1)=3T(u1)=3[23]=[69]

(c) T(2u1-u2)=2T(u1)-T(u2)=2[23]-[-41]=[85]

3. Because T([00])=[0+2(0)0-2]=[0-2]≠[00], T is not a linear transformation.

4.



5.

(a) False. T(4[01])=T([04])=[0-2(4)3(0)+4]=[-82], but 4T([01])=4[0-2(1)3(0)+1]=[-84]≠T(4[01]).

(b) False. For example, A=[100001000010] is onto.

(c) True. By definition, a linear transformation is a function.

(d) False. For example, T([x])=[x2x] satisfies $T(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$, but T is not onto.

Section 3.2

1.

(a) A+B=[253-4]+[314-5]=[567-9] *AC* is not defined.

(b) B-3I2=[314-5]-3[1001]=[014-8] *DB* is not defined.

(c) CB=[21540-1] [314-5]=[10-331-15-45]; A2=[253-4] [253-4]=[19-10-631]

(d) CT-D=[21540-1]T-[2-2-3031] =[25014-1]-[2-2-3031]=[07311-2];DC=[2-2-3031] [21540-1]=[-6-31511]

2. Set $A^2 = A$,

[a2-11]2=[a2-11][a2-22a+2-a-1-1]=[a2-11]

Because row 2, column 2 requires $-1 = 1$, we conclude that there are no solutions.

3. We have T1(x)=[1-2-13]x, and T2(x)=[-131-1]x.

(a) T1(T2(x))=[1-2-13]([-131-1] [x1x2]) =([1-2-13] [-131-1])[x1x2] =[-354-6] [x1x2], so A=[-354-6]

(b) T2(T2(x))=[-131-1]([-131-1] [x1x2]) =([-131-1] [-131-1])[x1x2] =[4-6-24] [x1x2], so A=[4-6-24]

4.

(a) $-2R_1 + R_2 \to R_2$

(b) $4R_2 \to R_2$

(c) $3R_1 + R_2 \to R_2$

(d) R1↔R3

5.

(a) True, because $A^2 = AA$ is defined only if $n = m$.

(b) True, as the transpose operation distributes over addition and scalar multiplication.

(c) False. For example, AB=[1101] [1011]=[2111] is not diagonal.

(d) False. For example, E1E2=[2001] [1002]=[2002] is not an elementary matrix.

Section 3.3

1. [21115]-1=12(5)-11(1)[5-11-12] =-1[5-11-12]=[-5111-2]

2. The linear system is equivalent to $A\mathbf{x} = \mathbf{b}$, with A=[21115], and b=[52]. Thus, x=A-1b=[21115]-1[52]=[-5111-2] [52]=[-31]. Therefore, $x_1 = -3$ and $x_2 = 1$.

3. T([x1x2])=T(x)=Ax, where A=[3253]. We determine A^{-1}:

[32105301](-5/3)R1+R2→R2~[32100-13-531] 6R2+R1→R1~[30-960-13-531] (1/3)R1→R1-3R2→R2~[10-32015-3],

so A-1=[3253]-1=[-325-3]. Consequently, T-1([x1x2])=T-1(x)=A-1x=([-3x1+2x25x1-3x2]).

4. [1-21100210010-20-1001]-2R1+R2→R22R1+R3→R3~[1-2110005-2-2100-41201](4/5)R2+R3→R3~ [1-2110005-2-21000-3525451]5R3→R3~ [1-2110005-2-21000-3245] (-2/3)R3+R2→R2(1/3)R3+R1→R1~[1-20534353050-103-53-10300-3245](2/5)R2+R1→R1~ [100132313050-103-53-10300-3245](1/5)R2→R2(-1/3)R3→R3~[100132313010-23-13-23001-23-43-53],

so [1-21210-20-1]-1=[132313-23-13-23-23-43-53]

5.

(a) True. For if $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$.

(b) True, because $(AB)^{-1} = B^{-1}A^{-1}$.

(c) True, because A can be row-reduced to the identity matrix.

(d) False. For example, if A=[0000] is singular, but Ax=[10] has no solutions.

Section 3.4

1.

(a) Solve Ly=b, [1031] y=[13], using back substitution, to obtain y=[10]. Now solve Ux=y, [-2501] x=[10], using back substitution, to obtain x=[-120].

(b) Solve *L***y** = **b**, [100310-211]y=[-121], using back substitution, to obtain y=[-15-6]. Now solve Ux=y, [31-102-2002] x=[-156], using back substitution, to obtain x=[-76-12-3].

2.

(a) [21-1440-216]-2R1+R2→R2R1+R3→R3~[21-1022025]⇒L=[100210-1•1][21-1022025]-R2+R3→R3~ [21-1022003]⇒L=[100210-111]

Thus L=[100210-111] and U=[21-1022003].

(b) [213-1010223242122]-R1+R3→R3-R1+R4→R4~ [213-1010202-1500-13]⇒L=[100001001•101••1]

[213-1010202-1500-13] -2R2+R3→R3~ [213-1010200-1100-13]⇒L=[10000100121010•1]

[213-1010200-1100-13] -R3+R4→R4~ [213-1010200-110002]⇒L=[1000010012101011]

Thus L=[1000010012101011] and U=[213-1010200-110002].

(c) [210-2223-1-4078]-R1+R2→R22R1+R3→R3~[210-201310274]⇒L=[100110-2•1]

[210-201310274]-2R2+R3→R3~ [210-201310012]⇒L=[100110-221]

Thus L=[100110-221] and U=[210-201310012].

(d) [212-2-2-1427209]R1+R2→R2-2R1+R3→R3-R1+R4→R4~[2120-110030-17]⇒L=[1000-11002•101••1]

[2120-110030-17] -R2+R4→R4~ [2120-11003006]⇒L=[1000-1100201011•1]

[2120-11003006] -2R3+R4→R4~ [2120-11003000]⇒L=[1000-110020101121]

Thus,L=[1000-110020101121] and U=[2120-11003000].

3.

(a) L=[100210-111] . We divide the rows of U=[21-1022003] by the diagonal entries to obtain D=[200020003] and U=[112-12011001] .

(b) L=[1000010012101011] . We divide the rows of U=[213-1010200-110002] by the diagonal entries to obtain D=[2000010000-100002] and U=[11232-120102001-10001] .
(c) L=[100110-221]. We divide the rows of U=[210-201310012] by the diagonal entries to obtain D=[200010001] and U=[1120-101310012].
(d) L=[1000-110020101121]. We divide the rows of U=[2120-11003000] by the diagonal entries to obtain D=[20000-10000300001] and U=[112101-1001000].

Section 3.5

1.

(a) Stochastic, because each entry is nonnegative and all column sums are one.
(b) Not stochastic, because the third column does not sum to one.

2.

(a) From column 1, we would need a = 0.7, but then row 1 has sum 1.4. Therefore, A cannot be a doubly stochastic matrix.
(b) Setting column and row sums equal to one, we first obtain a = 0.4 (from row 1), b = 0.3 (from column 1), d = 0.4 (from column 2), and we obtain c = 0.3 from row 3. We check that the resulting matrix [0.40.20.40.30.40.30.30.40.3] is doubly stochastic.

3. x1=Ax0=[0.40.70.60.3] [0.50.5]=[0.550.45];x2=Ax1= [0.40.70.60.3] [0.550.45]=[0.5350.465]

4.

(a) Solve (A − I)**x** = **0** by row-reducing the augmented matrix.

[-0.750.500.75-0.50]R1+R2→R2~[-0.750.50000] and we obtain x=s[2/31]. Setting the column sum of **x** equal to one, we need s=35, and so x=35[2/31]=[2/53/5].

(b) Solve (A − I)**x** = **0** by row-reducing the augmented matrix.

[-0.80.50.500.4-0.5000.40-0.50] (1/2)R1+R2→R2(1/2)R1+R3→R3~[-0.80.50.500-0.250.25000.25-0.250] R2+R3→R3~ [-0.80.50.500-0.250.2500000] and we obtain x=s[5/411]. Setting the column sum of equal to one, we need s=413, and so x=413[5411]=[513413413].

5.

(a) False. Entries may be 0.
(b) True, by Theorem 3.29(c).
(c) True, by Theorem 3.31(a).

(d) False. For example, if A=[1100], then AAT=[1100][1100]T=[2000] is not doubly stochastic.

Chapter 4

Section 4.1

1.

(a) Not a subspace. The vector 0=[000] is not in the set.

(b) Let S be the set of vectors of the form [0a-b-2a]. Letting $a = 0$ and $b = 0$, we see that $\mathbf{0} \in S$. Suppose **u** and **v** are in S, then u=[0a1-b1-2a1] and v=[0a2-b2-2a2]. Thus u+v=[0a1-b1-2a1]+[0a2-b2-2a2]=[0(a1+a2)-(b1+b2)-2(a1+a2)]∈S. Let $r \in \mathbf{R}$ and $\mathbf{u} \in S$, then ru=r[0a1-b1-2a1]=[0(ra1)-(rb1)-2(ra1)]∈S, and we conclude that S is a subspace. (*Alternatively, we have* S=span{[01-2], [0-10]}, *and thus S is a subspace.*).

2.

(a) We row-reduce

[13-230-4]~[13-20-92]

Thus, $A\mathbf{x} = \mathbf{0}$ has solutions of the form

x=s[43291]
and, therefore, null(A)=span{[43291]}.

(b) We row-reduce

[214-11513]~[214-1092-172]

Thus, $A\mathbf{x} = \mathbf{0}$ has solutions of the form x=s[-1992910]+t[89-7901], and, therefore, null(A)=span{[-1992910], [89-7901]}.

3.

(a) Ab=[1339] [3-1]=[00]=0, so $\mathbf{b} \in \ker(T)$. We row-reduce to determine a solution of $A\mathbf{x} = \mathbf{c}$,

[132395]~[13200-1]

So $A\mathbf{x} = \mathbf{c}$ has no solution, and $\mathbf{c} \notin \text{range}(T)$.

(b) Ab=[1-236-24] [3-2]=[7-3-14]≠0, so $\mathbf{b} \notin \ker(T)$. We row-reduce to determine a solution of $A\mathbf{x} = \mathbf{c}$,

[1-22361-24-1]~[1-22012-5003]

So $A\mathbf{x} = \mathbf{c}$ has no solution, and $\mathbf{c} \notin \text{range}(T)$.

4.

(a) False. ker(T) is a subset of the domain space.

(b) False. ker(T) is a subspace of $\mathbf{R}^3$.

(c) False. The trivial subspace {**0**} contains a single vector.

(d) False. If S is a subspace, then $\mathbf{0} \in S$, so $\mathbf{0} \notin S^C$, and S^C cannot be a subspace.

Section 4.2

1.

(a) Row-reduce the matrix with the given vectors as rows,

[2-76-14]~[2-707]

The basis for S is given by the non-zero row vectors, {[2-7], [07]}, and dim(S) = 2.

(b) Row-reduce the matrix with the given vectors as rows,

[1-13210-31-4]~[1-1303-6001]

The basis for S is given by the non-zero row vectors, {[1-13], [03-6], [001]}, and dim(S) = 3.

2.

(a) Row-reduce the matrix with the given vectors as columns,

[26-7-14]~[2607]

The basis for S is given by columns 1 and 2 of the original matrix corresponding to the pivot column of the row-reduced matrix. Thus, our basis for S is {[2-7], [6-14]}, and dim(S) = 2.

(b) Row-reduce the matrix with the given vectors as columns,

[12-3-11130-4]~[12-303-2001]

The basis for S is given by columns 1, 2, and 3 of the original matrix corresponding to the pivot columns of the row-reduced matrix. Thus, our basis for S is {[1-13], [210], [-31-4]}, and dim(S) = 3.

3.

(a) The second vector is -12 times the first, so we eliminate the second vector and obtain the basis {[2-6]}. The dimension is 1.

(b) The second vector is −2 times the first, and is eliminated as a dependent vector. Because the remaining vectors are not multiples of one another, they are linearly independent, and the basis is {[1-23], [3-68]}. The dimension is 2.

4.

(a) One extension is {[100], [010], [12-1]}, as these three vectors are linearly independent.

(b) One extension is {[100], [120], [-1-22]}, as these three vectors are linearly independent.

5.

(a) We row-reduce

[13-230-4]~[13-20-92]

Thus, $A\mathbf{x} = \mathbf{0}$ has solutions of the form

x=s[43291]

Therefore, the null space has basis {[43291]}, and the dimension is 1.

(b) We row-reduce

[214-11513]~[214-1092-172].

Thus, $A\mathbf{x} = \mathbf{0}$ has solutions of the form

x=s1[-1992910]+s2[89-7901]

Therefore, the null space has basis {[-1992910], [89-7901]}, and the dimension is 2.

6.

(a) False. If $S = \{\mathbf{0}\}$, then $\dim(S) = 0$.

(b) False. If the subspace is non-trivial, then it has a basis, and the dimension is the unique number of vectors in a basis. If $S = \{\mathbf{0}\}$, then $\dim(S) = 0$.

(c) True. With a basis of two vectors, the span of a basis forms a plane.

(d) False. If $A = [0\ 0]$, then $n = 1 < 2 = m$, but $\text{nullity}(A) = 2 \nleq 1$.

Section 4.3

1.

(a) We reduce A to echelon form:

[14-6-2-812] ~ [14-6000]

A basis for the column space, which is determined from the pivot column 1, is {[1-2]}. A basis for the row space is determined from the nonzero rows of the echelon form, {[14-6]}. We solve $A\mathbf{x} = \mathbf{0}$, to obtain x=s[-410]+s2[601], and so our nullspace basis is {[-410], [601]}. We have $\text{rank}(A) = 1$, $\text{nullity}(A) = 2$, and $\text{rank}(A) + \text{nullity}(A) = 1 + 2 = 3 = m$.

(b) We reduce A to echelon form:

A basis for the column space, determined from the pivot columns 1, 2, and 3, is {[2-10], [143], [022]}. A basis for the row space is determined from the nonzero rows of the echelon form, {[2103], [092252], [0023-83]}. We solve $A\mathbf{x} = \mathbf{0}$, to obtain x=s[-13-7341], and so our nullspace basis is {[-13-7341]}. We have rank(A) = 3, nullity(A) = 1, and rank(A) + nullity(A) = 3 + 1 = 4 = m.

2.

(a) The dimension of the row space is 7 − 2 = 5, the number of nonzero rows in the echelon form.

(b) The dimension of the column space is also 5.

(c) The dimension of the null space is 12 − 5 = 7, so nullity(A) = 7.

(d) The dimension of the row space is 5, so rank(A) = 5.

3. nullity(A) = m − rank(A) = 8 − 5 = 3.

4.

(a) False. A=[10] has rank(A) = 1 = nullity(A).

(b) False. A = [0] has rank(A) = 0.

(c) False. A=[10] has nullity(A) = 0.

(d) True. A has, at most, n pivots, so rank(A) ≤ n.

Section 4.4

1.

(a) x=U[x]B=[32-76] [-14]=[531]

(b) x=U[x]B=[11-22036-54] [30-1]=[5314]

2.

(a) [x]B=U-1x=[1225]-1[1-3] =[5-2-21] [1-3]=[11-5]

(b) [x]B=U-1x=[1211-10-100]-1[21-1] =[00-10-1-1123] [21-1]=[101]

3.

(a) [4311]-1[1325]=[-5-12717]

(b) [11-2212010]-1[121010-10-2]=[1153-10-2-12-12-23]

4.

(a) True. Because $A = W^{-1}V$, where V has columns given by the basis vectors of $\mathcal{B}1$ and W has columns given by the basis vectors of $\mathcal{B}2$, it follows that A is invertible, with $A^{-1} = V^{-1}W$.

(b) True. Let $A = W^{-1}V$, where V has columns given by the basis vectors of $\mathcal{B}1$ and W has columns given by the basis vectors of $\mathcal{B}2$. Then $A^{-1} = V^{-1}W$, which is the change of basis matrix from $\mathcal{B}2$ to $\mathcal{B}1$.

(c) True. If $A = W^{-1}V$, where V has columns given by the basis vectors of $\mathcal{B}1$ and W has columns given by the basis vectors of $\mathcal{B}2$, it follows that $W = V$, so $A = W^{-1}V = V^{-1}V = I$.

(d) True, by Theorem 4.28.

Chapter 5

Section 5.1

1.

(a) C23=(-1)2+3|M23|=-|135-1| =-((1)(-1)-(3)(5))=16.C31= (-1)3+1|M31|=|3-20-4|=3(-4)-(-2)0=-12.

(b) C31=(-1)2+3|M23|=-|202313223| =-(2det[1323]-(0)det[3323]+2det[3122]) =-(2(-3)+2(4))=-2.C31=(-1)3+1|M31|=|0-12112213| =(0)det[1213]-(-1)det[1223]+2det[1121] =0-1+2(-1)=-3.

2.

(a) Using the first row,

det(A)=a11C11+a12C12+a13C13=(3)C11+(1)C12+(0)C13 =(3)(-1)1+1|M11|+(1)(-1)1+2|M12| =3det[2-213]+(-1)det[2-243] =3(8)+(-1)(14)=10.

Because det(*A*) ≠ 0, *A* is invertible, and *T* is also invertible.

(b) Using the last column, det(A)=a34C34+a44C44=(2)(-1)3+4|M34|+(2)(-1)4+4|M44|= (-2)|521312401|+2|521312-134|=(-2)(11)+2(-28)=-78. Because det(*A*) ≠ 0, *A* is invertible, and *T* is also invertible.

3.

(a) det([3a13])=3(3)-a(1)=0⇒a=9. *A* is not invertible if *a* = 9.

(b) det([a120-a31-22])=-2a2+8a+3=0⇒a=2±1222. *A* is not invertible if a=2±1222.

4.

(a) det(A)=det([2-516])=17.det(AT)=det([21-56])=17.

(b) det(A)=det([12-5-3142-12])=29.det(AT)=det([1-3221-1-542])=29.

5.

(a) det(A-λI2)=det([231-4]-λ[1001])=det([2-λ31-4-λ])=λ2+2λ-11=0⇒λ=-1+23, or λ=-1-23

(b) det(A-λI2)=det([1411]-λ[1001])=det([1-λ411-λ])=λ2-2λ-3=0⇒λ=-1 or λ=3

6.

(a) True. Use the zero row to compute the determinant, then det(*A*) = 0, and *A* is not invertible.

(b) False. For example, if A=[0000], then det(A) = 0, but T is not one-to-one.

(c) False. det(3A) = 3^4det(A) = 81det(A). (The statement is only true if det(A) = 0.)

(d) False. For example, if A=[1001] and B=[0110], then $A \sim B$ by interchanges rows, but det(A) = 1 and det(B) = −1.

Section 5.2

1.

(a) Because A is in echelon form, det(A) = 3(−2)(−5) = 30.

(b) Row-reduce to echelon form:

[000-2001-102-172137]R1↔R4R2↔R3~ [213702-17001-1000-2]

Thus, det(A) = $(-1)^2$(2)(2)(1)(−2) = −8.

2.

(a) det(A^4) = $(\det(A))^4$ = $(3)^4$ = 81

(b) det($-2A^2$) = $(-2)^3(\det(A))^2$ = $-8(3)^2$ = −72

(c) det(AB^2) = det(A)$(\det(B))^2$ = $3(-1)^2$ = 3

(d) det($3A^2B$) = $3^3(\det(A))^2$det(B) = $27(3)^2(-1)$ = −243

3.

(a) det(A) = $(-1)^1$ det(B)=-|270-3|=-2(-3)=6≠0, so A is invertible.

(b) det(A)=(-1)1det(B)=-|2-140-23005| =-(2)(-2)(5)=20≠0, so A is invertible.

4.

(a) False. Either an elementary row operation changes the sign of the determinant, leaves the determinant unchanged, or multiplies the determinant by a non-zero scalar.

(b) False. For example, A=[1001] satisfies det(A) = det(A^2).

(c) False. For example,

1=det ([100010001])=det ([100010001]-[00000000-1]),
but det ([100010001])-det ([00000000-1])=0-0=0

(d) False. For example, if A=[1001] and B=[0000], then A is invertible, B is not, and AB is not invertible.

Section 5.3

1.

(a) Let A=[5-3-12] and b=[7-4]. Then A1=[7-3-42], and A2=[57-1-4]. By Cramer’s Rule, x1=det(A1)det(A)=27 and x2=det(A2)det(A)=-137.

(b) Let A=[21-21-35-1-42] and b=[11-6]. Then A1=[11-21-35-6-42], A2=[21-2115-1-62], and A3=[2111-31-1-4-6]. By Cramer’s Rule, x1=det(A1)det(A)=2635, x2=det(A2)det(A)=6735, and x3=det(A3)det(A)=4235=65.

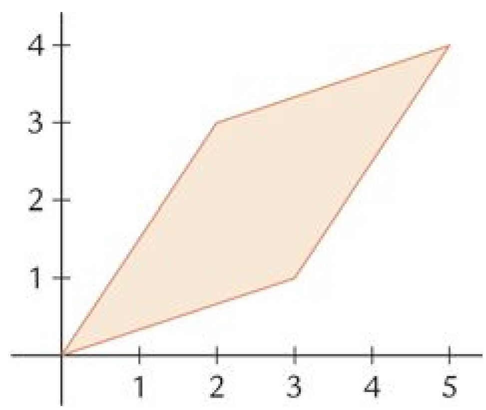
2.

(a) adj(A)=adj([133-4])=[-4-3-31];A-1=1det(A)adj(A)=1-13[-4-3-31]=[413313313-113]

(b) adj(A)=adj([214150211])=[53-20-1-64-909];A-1=1det(A)adj(A)=1-27[53-20-1-64-909]=[-527-19202712729-427130-13]

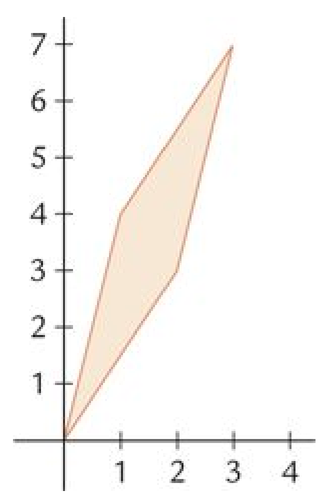
3.

(a)



area=|det([3213])|=7

(b)



area=|det([2134])|=5

4.

(a) True. Otherwise, determinants cannot be used.

(b) False. adj([0000])=[0000].

(c) True. adj(3[abcd])=adj([3a3b3c3d])=[3d-3b-3c3a]=3[d-b-ca]=3adj(abcd).

(d) False. [2002] ~ [1001], but adj([2002])=[2002], and adj([1001])=[1001].

Chapter 6

Section 6.1

1. Ax1=[-2-42-212425] [-201]=[66-3]≠λx1 for any λ, so $\mathbf{x}_1$ is not an eigenvector.

   Ax2=[-2-42-212425] [-231]=[-693]=3x2, so $\mathbf{x}_2$ is an eigenvector.

   Ax3=[-2-42-212425] [121]=[-8213]≠λx3, for any λ, so $\mathbf{x}_3$ is not an eigenvector.
2. We row-reduce to obtain the null space of A-3I2=[-1-4-1-4] ~ [1400]. Solving, we obtain x=s[-41]. A basis for the $\lambda = 3$ eigenspace is {[-41]}.
3. Characteristic polynomial:

   det(A-λI2)=det([3122]-λ[1001]) =det([3-λ122-λ])=λ2-5λ+4.

   Eigenvalues: $\lambda^2 - 5\lambda + 4 = (\lambda - 1)(\lambda - 4) = 0 \Rightarrow \lambda = 1$ and $\lambda = 4$.

   Eigenspace of $\lambda = 1$:

   A-(1)I2=[2121] ~ [2100],

   so a basis for this eigenspace is {[1-2]}.

   Eigenspace of $\lambda = 4$:

   A-4I2=[-112-2] ~ [1-100],

   so a basis for this eigenspace is {[11]}.
4. Characteristic polynomial:

   det(A-λI3)=det([100200-211]-λ[100010001])=det([1-λ002-λ0-211-λ])=-λ3+2λ2-λ.

Eigenvalues: $-\lambda^3 + 2\lambda^2 - \lambda = -\lambda(\lambda - 1)^2 = 0 \Rightarrow \lambda = 0$ and $\lambda = 1$.

Eigenspace of $\lambda = 0$:

A-(0)I3=[100200-211] ~ [100011000],

so a basis for this eigenspace is {[01-1]}.

Eigenspace of $\lambda = 1$:

A-(1)I3=[0002-10-110] ~ [1-120000000],

so a basis for this eigenspace is {[120],[001]}.

5. For example, A=[-2005].

6.

(a) False. The zero vector is not an eigenvector.

(b) True, because $A(\mathbf{u}_1 + \mathbf{u}_2) = A\mathbf{u}_1 + A\mathbf{u}_2 = \lambda\mathbf{u}_1 + \lambda\mathbf{u}_2 = \lambda(\mathbf{u}_1 + \mathbf{u}_2)$.

(c) False. For example, A=[1000] has 0 as an eigenvalue, but the column space of A is span {[10]}≠R2.

(d) False. [2002] ~ [1001] but these matrices have different eigenvalues.

Section 6.2

1.

(a) A4=PD4P-1=[1225][100-1]4[1225]-1 =[1225] [1400(-1)4] [5-2-21]= [1001]

(b) A4=PD4P-1=[1211-1002-1] [-100010002]4[1211-1002-1]-1 =[1211-1002-1] [(-1)4000(1)400024] [15451515-151525-25-35] =[7-6-9010-6610]

2.

(a) We may obtain A as

[1112] [200-1][1112]-1=[5-36-4]

(b) We may obtain A as

[-1110-11110] [-200020001][-1110-11110]-1=[132373-1343-1343-43-23]

3.

(a) det([-3-2106]-λ[1001])=λ2-3λ+2=0⇒λ1=1, and $\lambda_2 = 2$.

Because [-3-2106]-(1)[1001]=[-4-2105], a basis for the eigenspace of $\lambda_1 = 1$ is {[1-2]}. Because [-3-2106]-(2)[1001]=[-5-2104], a basis for the eigenspace of $\lambda_2 = 2$ is {[2-5]}.

We therefore have [-3-2106]=[12-2-5] [1002] [12-2-5]-1.

(b) det([-2-82020-1-21]-λ[100010001])=-λ3+λ2+2λ =-λ(λ+1)(λ-2)=0⇒λ1=0,λ2=-1, and $\lambda_3 = 2$.

Because [-2-82020-1-21]-(0)[100010001]=[-2-82020-1-21] ~ [10-1010000],

a basis for the eigenspace of $\lambda_1 = 0$ is {[101]}. Because [-2-82020-1-21]-(-1)[100010001]=[-1-82030-1-22] ~ [10-2010000],

a basis for the eigenspace of $\lambda_2 = -1$ is {[201]}. Because [-2-82020-1-21]-2[100010001]=[-4-82000-1-2-1] ~ [120001000],

a basis for the eigenspace of $\lambda_2 = 2$ is {[2-10]}. We therefore have [-2-82020-1-21]=[12200-1110] [0000-10002] [12200-1110]-1

4. For example, A=[121122127] has eigenvalues 0, 2, and 8.

5. One of the eigenvalues must be a repeated eigenvalue, of multiplicity 2. Its eigenspace must have dimension 2, because the matrix is diagonalizable. The other 3 eigenvalues have eigenspaces of dimension 1.

Section 6.3

1.

(a) $z_1 - z_2 = (2 + i) - (5 - 3i) = (2 - 5) + (1 - (-3))i = -3 + 4i$

(b) $4z_1 + 3z_2 = 4(2 + i) + 3(5 - 3i) = (4(2) + 3(5)) + (4(1) + 3(-3))i = 23 - 5i$

(c) z2z1=5-3i2+i=(5-3i)(2-i)(2+i)(2-i)=7-11i5=75-115i

(d) [z1]=|2+i|=22+11=5

(e) z2¯=5-3i¯=5+3i

(f) $\mathrm{Re}(z_1z_2) = \mathrm{Re}((2 + i)(5 - 3i)) = \mathrm{Re}(13 - i) = 13$

2.

(a) Characteristic polynomial: det([2-332]-λ[1001])=([2-λ-332-λ])=λ2-4λ+13.

Eigenvalues: $\lambda^2 - 4\lambda + 13 = 0 \Rightarrow \lambda_1 = 2 + 3i$, and $\lambda_2 = 2 - 3i$.

Eigenspace of $\lambda_1 = 2 + 3i$: A-(2+3i)I2=[-3i-33-3i] ~ [1-i00], so a basis for this eigenspace is {[i1]}.

A basis for the eigenspace of λ2=2-3i=λ1¯ is {[-i1]}.

(b) Characteristic polynomial: det([12-54]-λ[1001])=det([1-λ2-54-λ])=λ2-5λ+14.

Eigenvalues: $\lambda^2 - 5\lambda + 14 = 0 \Rightarrow$ λ1=52+312i, and λ2=52-312i.

Eigenspace of λ1=52+312i: A-(52+312i)I2=[-32-312i2-532-312i] ~ [1-310+3110i00], so a basis for this eigenspace is {[310-3110i1]}. A basis for the eigenspace of λ2=52-312i=λ1¯ is {[310+3110i1]}.

3.

(a) An eigenvalue is $\lambda = 4 + i$. So the rotation is by $\tan^{-1}(1/4) \approx 0.245$ radians, and the dilation is by 42+12=17.

(b) An eigenvalue is $\lambda = 2 + 5i$. So the rotation is by $\tan^{-1}(5/2) \approx 1.190$ radians, and the dilation is by 22+52=29.

4.

(a) Solve det([1-131]-λ[1001])=det([1-λ-131-λ])=λ2-2λ+4=0⇒λ=1±3i. The rotation–dilation matrix is B=[1-331].

(b) Solve det([25-51]-λ[1001])=det([2-λ5-51-λ])=λ2-3λ+27=0⇒λ=32±3112i. The rotation–dilation matrix is B=[32-3112311232].

Section 6.4

1.

(a) y=c1e4t[1-2]+c2e-t[52] $\Rightarrow y_1 = c_1e^{4t} + 5c_2e^{-t}$ and $y_2 = -2c_1e^{4t} + 2c_2e^{-t}$.

(b) y=c1e-2t[1-21]+c2et[-101]+c3et[2-30] $\Rightarrow y_1 = c_1e^{-2t} - c_2e^t + 2c_3e^t$, $y_2 = -2c_1e^{-2t} - 3c_3e^t$, and $y_3 = c_1e^{-2t} + c_2e^t$.

(c) y=c1(cos(t)[31]-sin(t)[21])+c2(sin(t)[31]+cos(t)[21]) $\Rightarrow y_1 = (3c_1 + 2c_2)\cos t + (-2c_1 + 3c_2)\sin t$ and $y_2 = (c_1 + c_2)\cos t + (-c_1 + c_2)\sin t$.

2.

(a) To solve y′=Ay=[1-241]y, we determine the eigenvalues and eigenvectors of A, from the characteristic polynomial, $\det(A - \lambda I_2) = \lambda^2 - 2\lambda + 9 = 0 \Rightarrow$ λ=1±22i. λ=1+22i⇒u=[22i1], λ=1-22i⇒u=[-22i1] Thus, y1=(-22c1sin(22t)+22c2cos(22t))et, and y2=(c1cos(22t)+c2sin(22t))et.

(b) To solve y′=Ay=[12-1-3-1152-2] y, we determine the eigenvalues and eigenvectors of A, from the characteristic polynomial, $\det(A - \lambda I_3) = \lambda^3 + 2\lambda^2 + 8\lambda + 1 = 0$.
λ=-0.128⇒u=[0.2060.3420.917], λ≈-0.936+2.624i⇒u=[0.493-0.224-0.365i0.507+0.563i], λ≈-0.936-2.624i⇒u=[0.493-0.224+0.365i0.507-0.563i]
Thus,

y1=(0.206)c1e-0.128t+e-0.936t((0.493)c2cos(2.624t)+(0.493)c3sin(2.624t)) y2=(0.342)c1e-0.128t+c2((-0.224)cos(2.624t)+(0.365)sin(2.624t))e-0.936t+ c3((-0.224)sin(2.624t)-(0.365)cos(2.624t))e-0.936ty3=(0.917)c1e-0.128t+c2((0.507)cos(2.624t)-(0.563)sin(2.624t))e-0.936t+ c3((0.507)sin(2.624t)+(0.563)cos(2.624t))e-0.936t

3.

(a) To solve y′=Ay=[12-1-1]y, we determine the eigenvalues and eigenvectors of A, from the characteristic polynomial, $\det(A - \lambda I_2) = \lambda^2 + 1$.

λ=i⇒u=[-1-i1] , λ=-i⇒u=[-1+i1]

Thus,

y=c1(cos(t)[-11]-sin(t)[-10])+c2et(sin(t)[-11]+cos(t)[-10]).

Thus, $y_1 = c_1(-\cos t + \sin t) + c_2(-\cos t - \sin t)$ and $y_2 = c_1 \cos t + c_2 \sin t$.

Setting $y_1(0) = -2$ and $y_2(0) = 1$, we obtain

-c1-c2=-2c1=1

Thus, $c_1 = 1$ and $c_2 = 1$. Therefore, $y_1 = -2 \cos t$ and $y_2 = \cos t + \sin t$.

(b) To solve y′=Ay=[1-3-12123-21]y, we determine the eigenvalues and eigenvectors of A, from the characteristic polynomial, $\det(A - \lambda I_2) = -\lambda(\lambda^2 - 3\lambda + 16)$. λ=0⇒u=[-57-471], λ=32+552i⇒u=[12+5510i12-5510i1], λ=32-12i55⇒u=[12-5510i12+5510i1] Thus, y1=-57c1+c2(12cos(552t)-5510sin(5510t))e32t+c3(12sin(5510t)+5510cos(552t))e32ty2=-47c1+c2(12cos(552t)+5510sin(5510t))e32t+c3(12sin(5510t)-5510cos(552t))e32ty3=c1+c2cos(552t)e32t+c3sin(5510t)e32t Setting $y_1(0) = -2$, $y_2(0) = 1$, and $y_3 = 3$, we obtain

-57c1+12c2+5510c3=-2-47c1+12c2-5510c3=1c1+c2=3

Thus, c1=74, c2=54 and c3=-554. Therefore,

y1=-54-34cos(552t)e32t-554(sin5510t)e32ty2=-1+2cos(552t)e32ty3=74+54cos(552t)e32t-554sin(5510t)e32t

Section 6.5

1.

(a) x1=Ax0=[3-112] [10]=[31],x2=Ax1=[3-112] [31]=[85]

(b) x1=Ax0=[2031140-1-7] [010]=[01-1],x2=Ax1=[2031140-1-7] [01-1]=[-3-36]

2.

(a) Ax0=[11-21] [01]=[11], s1=1,x1=1s1Ax0=11[11]=[11]Ax1=[11-21] [11]=[1-1], s2=2,x2=1s2Ax1=12[2-1]=[1-12]=[1.00-0.50]

(b) Ax0=[1-1-2203-121] [001]=[-231], s1=3,x1=1s1Ax0=13[-231]=[-0.671.000.33]Ax1=[1-1-2203-121] [-0.671.000.33]=[-2.33-0.353.00], s2=3.00,x2=1s2Ax1=13.00[-2.33-0.353.00]=[-0.78-0.121.00]

Chapter 7

Section 7.1

1. Property (1): If q_1 and q_2 are polynomials with real coefficients and degree no greater than 2, then $q_1 + q_2$ is a polynomial with real coefficients and degree no greater than 2. Thus, V is closed under addition.
Property (2): If c is a real scalar and q is a polynomial with real coefficients and degree no greater than 2, then cq is a polynomial with real coefficients and degree no greater than 2. Thus, V is closed under scalar multiplication.
Property (3): Let $z(x) = 0$, then z is in V, and $(z + q)(x) = z(x) + q(x) = 0 + q(x) = q(x)$. Therefore, $z + q = q$ for every q in V. Thus z is the zero vector in V.
Property (4): Given q_1 in V, let $q_2(x) = -q_1(x)$. Then q_2 is in V, and $(q_1 + q_2)(x) = q_1(x) + q_2(x) = q_1(x) - q_1(x) = 0 = z(x)$. Thus $q_1 + q_2 = z$, where q_2 is an additive inverse of q_1.
Property (5): Let q_1, q_2, and q_3 belong to V, and c_1 and c_2 scalars. Then,

(a) $(q_1 + q_2)(x) = q_1(x) + q_2(x) = (q_2 + q_1)(x)$; thus, $q_1 + q_2 = q_2 + q_1$.

(b) $((q_1 + q_2) + q_3)(x) = (q_1 + q_2)(x) + q_3(x) = (q_1(x) + q_2(x)) + q_3(x) = q_1(x) + (q_2(x) + q_3(x)) = q_1(x) + (q_2 + q_3)(x) = (q_1 + (q_2 + q_3))(x)$; thus, $(q_1 + q_2) + q_3 = q_1 + (q_2 + q_3)$.

(c) $(c_1(q_1 + q_2))(x) = c_1((q_1 + q_2)(x)) = c_1(q_1(x) + q_2(x)) = c_1(q_1(x)) + c_1(q_2(x)) = (c_1q_1)(x) + (c_1q_2)(x) = (c_1q_1 + c_1q_2)(x)$; thus, $c_1(q_1 + q_2) = c_1q_1 + c_1q_2$.

(d) $((c_1 + c_2)q_1)(x) = (c_1 + c_2)(q_1(x)) = c_1(q_1(x)) + c_2(q_1(x)) = (c_1q_1)(x) + (c_2q_1)(x) = (c_1q_1 + c_2q_1)(x)$; thus, $(c_1 + c_2)q_1 = c_1q_1 + c_2q_1$.

(e) $((c_1c_2)q_1)(x) = (c_1c_2)(q_1(x)) = c_1(c_2(q_1(x))) = c_1((c_2q_1)(x)) = (c_1(c_2q_1))(x)$; thus, $(c_1c_2)q_1 = c_1(c_2q_1)$.

(f) $(1 \cdot q_1)(x) = (1)(q_1(x)) = q_1(x)$; thus, $1 \cdot q_1 = q_1$.

2. Let S be the set of 2×2 diagonal matrices. S contains the zero vector, [0000]. Suppose U and V belong to S. Then U+V= [u1100u22]+[v1100v22]=[u11+v1100u22+v22] is also diagonal and, therefore, in S. Also, cU=[cu1100cu22] is diagonal and thus in S. Consequently, S is a subspace of $\mathbf{R}^{2\times 2}$.

3. Let S be the set of functions in $C(\mathbf{R})$ such that $f(0) = 0$. S contains the zero vector, because $z(x) = 0$ satisfies $z(0) = 0$, so z is in S. Suppose f and g are in S; then $f(0) = 0$ and $g(0) = 0$. Thus, $(f + g)(0) = f(0) + g(0) = 0 + 0 = 0$, and $f + g$ is in S. Also, $(cf)(0) = cf(0) = c(0) = 0$; thus cf is in S. Therefore S is a subspace of $C(\mathbf{R})$. Next, let Q be the set of functions in $C(\mathbf{R})$ such that $f(0) = 1$. Then Q is not a subspace, because the zero vector, $z(x) = 0$, does not belong to Q, because $z(0) = 0 \neq 1$.

Section 7.2

1. $\{x^2 - 2, 2x + 1, 2x^2 + x - 3\}$ is linearly independent in $\mathbf{P}^2$ if the equation $c_1(x^2 - 2) + c_2(2x + 1) + c_3(2x^2 + x - 3) = 0$ has only the trivial solution. Expanding, we obtain $(c_1 + 2c_3)x^2 + (2c_2 + c_3)x + (-2c_1 + c_2 - 3c_3) = 0$. Equate coefficients and solve the system

c1+2c3=02c2+c3=0-2c1+c2-3c32=0

to obtain $c_1 = 0$, $c_2 = 0$, and $c_3 = 0$. Thus, $\{x^2 - 2, 2x + 1, 2x^2 + x - 3\}$ is linearly independent in $\mathbf{P}^2$.

To determine if $\{x^2 - 2, 2x + 1, 2x^2 + x - 3\}$ spans $\mathbf{P}^2$, let $p(x) = a_2x^2 + a_1x + a_0$ be in $\mathbf{P}^2$, and consider the equation

c1(x2-2)+c2(2x+1)+c3(2x2+x-3)=a2x2+a1x+a0

We equate coefficients as above, and obtain the equations

c1+2c3=a22c2+c3=a1-2c1+c2-3c3=a0

The coefficient matrix [102021-21-3] row reduces to the identity matrix, so the system of equations has a solution for any choice of coefficients a_0, a_1, a_2. Thus, $\{x^2 - 2, 2x + 1, 2x^2 + x - 3\}$ spans $\mathbf{P}^2$.

2. {[120-1],[01-20],[1-123],[201-1]} is linearly independent in $\mathbf{R}^{2\times2}$ if the equation c1[120-1]+c2[01-20]+c3[1-123]+c4[201-1]=[0000] has only the trivial solution. Expanding, we obtain [c1+c3+2c42c1+c2-c3-2c2+2c3+c4-c1+3c3-c4]=[0000].

Equate entries and solve the system

c1+c3+2c4=0,2c1+c2-c3=0-2c2+2c3+c4=0,-c1+3c3-c4=0

The coefficient matrix [101221-100-221-103-1] row reduces to the identity, so $c_1 = 0$, $c_2 = 0$, $c_3 = 0$, $c_4 = 0$. Thus, {[120-1],[01-23],[1-123],[201-1]} is linearly independent in $\mathbf{R}^{2\times2}$.

To determine whether the matrices span $\mathbf{R}^{2\times2}$, let A=[a1a2a3a4]∈R2×2, and consider the equation

c1[120-1]+c2[01-2-0]+c3[1-123]+c4[201-1]=[a1a2a3a4]

We obtain

c1+c3+2c4=a1,2c1+c2-c3=a2-2c2+2c3+c4=a3,-c1+3c3-c4=a4.

Again, the coefficient matrix reduces to the identity, so this system has a solution for any choice of A=[a1a2a3a4]∈R2×2.

Thus, {[120-1],[01-20],[1-123],[201-1]} spans $\mathbf{R}^{2\times2}$.

Section 7.3

1.

(a) Because $\mathcal{V}$ has 3 vectors and $\dim(\mathbf{P}^2) = 3$, we need only show that $\mathcal{V}$ is linearly independent. Consider $c_1(x^2 + x - 1) + c_2(-2x + 3) + c_3(x^2 - 5) = 0 \Rightarrow (c_1 + c_3)x^2 + (c_1 - 2c_2)x + (-c_1 + 3c_2 - 5c_3) = 0$. We solve

c1+c3=0c1-2c2=0-c1+3c2-5c3=0

and obtain $c_1 = c_2 = c_3 = 0$. Thus, $\mathcal{V}$ is linearly independent. Therefore, by Theorem 7.17, $\mathcal{V}$ is a basis.

(b) Because $\mathcal{V}$ has 4 vectors and dim($\mathbf{R}^{2\times 2}$) = 4, we need only determine if $\mathcal{V}$ is linearly independent. Consider c1[10-11]+c2[0123]+c3[1-110]+c4[-120-1]=[0000]⇒[c1+c3-c4c2-c3+2c4-c1+2c2+c3c1+3c2-4c4]=[0000] We solve

c1+c3-c4=0,c2-c3+2c4=0-c1+2c2+c3=0,c1+3c2-4c4=0

and obtain $c_1 = c_2 = c_3 = c_4 = 0$. Thus, $\mathcal{V}$ is linearly independent. Therefore, by Theorem 7.17, $\mathcal{V}$ is a basis.

2. dim(S) = 3, and a basis for S is

{[1000],[0010],[0001]}

3. We extend $\{x - 2, x^2 + 1\}$ to $\{x - 2, x^2 + 1, 1\}$ to obtain a basis for $\mathbf{P}^2$.

Chapter 8

Section 8.1

1.

(a) u1·u3=[2-103]·[110-3] =2(1)+(-1)(1)+0(0)+3(-3)=-8-u2·(2u1)=(-[1-142])·(2[2-103])=[-11-4-2]·[4-206]=(-1)(4)+(1)(-2)+(-4)(0)+(2)(6)=-18

(b) ‖u2-(-2u3)‖=‖[1-142]-(-2)[110-3]‖=‖[314-4]‖ =(3)2+(1)2+(4)2+(-4)2=42

2. u1·u2=[2a-1]·[120]=2a+2, u1·u3=[2a-1]·[b-13]=2b-a-3, and u2·u3=[120]·[b-13]=b-2. Thus, we obtain the system

2a+2=02b-a-3=0b-2=0

This system has no solution. Thus, there are no values *a* and *b* such that {$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} forms an orthogonal set.

3. Because $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal, by the Pythagorean Theorem ‖2u1+3u2‖=‖2u1‖2+‖3u2‖2=(2‖u1‖)2+(3‖u2‖)2=(2(4))2+(3(1))2=73.

4. We have u1·u2=[31-1]·[125]=0, so $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal. By the Pythagorean Theorem, we check

‖u1+u2‖2=‖u1‖2+‖u2‖2‖[31-1]+[125]‖2=‖[31-1]‖2+‖[125]‖2‖[434]‖2=11+3041=41

For the Cauchy–Schwarz inequality, we need

|u1·u2|≤‖u1‖‖u2‖0≤1130=330

And for the Triangle inequality, we need

|u1+u2|≤‖u1‖+‖u2‖41≤11+30,

which holds if and only if

41≤11+330+30=41+330,

which is true.

5. We determine the null space of [10-2] and obtain span {[010],[201]}. Thus, a basis for $S^{\perp}$ is {[010],[201]}.

Section 8.2

1. proju1u2=u1·u2 ‖u1‖2u1=([2-10]·[-113]/‖[2-10]‖2)[2-10]=(-3/5)[2-10]=[-65350]
2. projSu3=([2-10]·[-120]/‖[2-10]‖2)[2-10]+([12-2]·[-120]/‖[12-2]‖2)[12-2] =(-4/5)[2-10]+(3/9)[12-2]=[-19152215-23]
3. v1=s1=[1203]·v2=s2-projv1s2=[302-1]-v1·s2‖v1‖2v1 =[302-1]([1203]·[302-1]/‖[1203]‖2)[1203] =[302-1]-(0/14)[1203]=[302-1]

   Thus, an orthogonal basis for *S* is {[1203],[302-1]}.
4. v1=s1=[20-1]·v2=s2-projv1s2=[11-1]-v1·s2‖v1‖2v1 =[11-1]-([20-1]·[11-1]/‖[20-1]‖2)[20-1] =[11-1]-[650-35]=[-151-25]

   Thus, an orthogonal basis for *S* is {[20-1],[-151-25]}. Dividing by the norm of each vector gives the orthonormal basis {15[20-1],130[-15-2]}

Section 8.3

1. The matrix is not orthogonal, because

   [14-1212-121214140-14]T[14-1212-121214140-14]=[38-38-116-3812-18-116-1838]≠I3

2. We determine the corresponding orthonormal eigenvectors. Because det(λ[1001]-[3221])=λ2-4λ-1, we obtain λ1=2+5⇒u1=[1+510+25210+25] and λ2=2-5⇒u2=[1-510-25210-25]. Thus, P=[1+510+251-510-25210+25210-25] and D=[λ100λ2]=[2+5002-5].
3. ATA=[310221]T[310221]=[13556], det(λI2-ATA) =det(λ[1001]-[13556])=λ2-19λ+53=0 ⇒λ1=12149+192≥0 and λ2=192-12149≥0
4. We apply to Gram–Schmidt process to obtain an orthonormal set

q1=[2131321313], and q2=[-3131321313]. Thus, Q=[21313-313133131321313], and R=QTA[21313-313133131321313]T[2-332]=[130013].

Section 8.4

1. det(λ[1001]-[31-1102]T[31-1102])=λ2-16λ+56=0⇒λ1=8+22, and λ2=8-22⇒σ1=8+22≈3.291 and σ2=8-22≈2.274.
2. Let $B = A^T$, and determine the eigenvalues and normalized eigenvectors of BTB=[2-1-16], λ1=4+5⇒v1=[2-510-45110-45], λ2=4-5⇒v2=[2+510+45110+45]. Thus V=[2-510-452+510+45110-45110+45]. The singular values are σ1=λ1=4+5 and σ2=λ2=4-5. Thus, Σ=[4+5004-500]. Also, u1=1σ1Bv1=14+5[110-1-12] [2-510-45110-45]=[32-10210-35-2210-3510210-35], and u2=1σ2Bv2=14-5×[110-1-12] [5+245+10145+10]=[10+32210+35-2210+35-10210+35]. To obtain $\mathbf{u}_3$, we determine null(BT)=null([10-11-12])=span{[131]}, so u3=[111113111111111]. Thus, U=[32-10210-3510+32210+3511111-2210-35-2210+353111110210-35-10210+3511111]. The singular value decomposition for A is $V\Sigma^T U^T$.
3. $b = \sigma_1 \cdot \epsilon \cdot n = 22 \cdot 10^{-7} \cdot 4 = 8.8 \cdot 10^{-6}$. Because $\sigma_3 = 10^{-5} > b$, but $\sigma_4 = 10^{-8} < b$, it follows that the numerical rank of A is 3.

Section 8.5

1.
   (a) An orthonormal basis for S is given by v=[1-4]/‖[1-4]‖=[11717-41717] The vector in S closest to **y** is projSy=(v·y)v=([11717-41717]·[21])[11717-41717]=[-217817]

   (b) The spanning vectors for S are already orthogonal, so we only need to normalize the vectors. An orthonormal basis for S is given by v1=[12-1]/‖[12-1]‖=[166136-166] and v2=[-31-1]/‖[-31-1]‖=[-3111111111-11111]. The vector in S closest to **y** is projSy=(v1·y)v1+(v2·y)v2=([166136-166]·[10-1])[166136-166]+([-3111111111-11111]·[10-1])×[-3111111111-11111]=[29331633-533]

2. A=[3-2211-3], x=[x1x2], y=[1-23]. We calculate ATA=[3-2211-3]T[3-2211-3]=[14-7-714] and ATy=[3-2211-3]T[1-23]=[2-13]. The normal equations, $A^TA\mathbf{x} = A^T\mathbf{y}$, are thus given by [14-7-714] [x1x2]=[2-13], or

14x1-7x2=2-7x1+14x2=-13

3. We consider a line of the form $y = c_1 + c_2x$. We evaluate at the given points

(-1, 1.7)⇒c1-c2=1.7(0, 1.2)⇒c1=1.2(1, 0.5)⇒c1+c2=0.5

This is equivalent to $A\mathbf{c} = \mathbf{y}$, where A=[1-11011], c=[c1c2] and y=[1.71.20.5]. Because A has linearly independent columns, the least squares solution of $A\mathbf{c} = \mathbf{y}$ is c=(ATA)-1ATy= [3002]-1[111-101] [1.71.20.2]≈[1.133-0.6] Thus the equation of the line that best fits the data is $y \approx 1.133 - 0.6x$.

Chapter 9

Section 9.1

1. $T(\mathbf{v}_1 - 4\mathbf{v}_2 + 2\mathbf{v}_3) = T(\mathbf{v}_1) - 4T(\mathbf{v}_2) + 2T(\mathbf{v}_3) = (3x - 2) - 4(-5x) + 2(x + 7) = 25x + 12$.
2. Write $-x + 3 = c_1(x - 1) + c_2(-2x + 3) \Rightarrow c_1 = 3$ and $c_2 = 2$, so $-x + 3 = 3(x - 1) + 2(-2x + 3)$. Thus T(-x+3)=T(3(x-1)+2(-2x+3))=3T(x-1)+2T(-2x+3)=3[-12]+2[41]=[58]
3.
   (a) Let $a_1x^2 + b_1x + c_1$ and $a_2x^2 + b_2x + c_2$ belong to $\mathbf{P}^2$, and let s and t be scalars. Then

   T(s(a1x2+b1x+c1)+t(a2x2+b2x+c2))=T((sa1+ta2)x2+(sb1+tb2)x+(sc1+tc2))=[(sa1+ta2)+(sb1+tb2)(sb1+tb2)-2(sc1+tc2)]=s[a1+b1b1-2c1]+t[a2+b2b2-2c2]=sT(a1x2+b1x+c1)+tT(a2x2+b2x+c2)

   Thus, by Theorem 9.2, T is a linear transformation.

   (b) Let [a1b1c1d1] and [a2b2c2d2] belong to $\mathbf{R}^{2\times 2}$, and let s and t be scalars. Then

   T(s[a1b1c1d1]+t[a2b2c2d2])=T([sa1+ta2sb1+tb2sc1+tc2sd1+td2])=(sa1+ta2)x2-(sd1+td2)x+5(sb1+tb2)=s(a1x2-d1x+5b1)+t(a2x2-d2x+5b2)=sT([a1b1c1d1])+tT([a2b2c2d2]). Thus, by Theorem 9.2, T is a linear transformation.
4.
   (a) Because $T(2x) = (2)^3 = 8$, but $2(T(x)) = 2(1)^3 = 2 \neq 8$, T is not a linear transformation.

   (b) Let A and B belong to $\mathbf{R}^{2\times 3}$, and let and t be scalars. Then

   T(sA+tB)=(sA+tB)T =(sA)T+(tB)T =sAT+tBT =s(T(A))+t(T(B))

   Thus, by Theorem 9.2, T is a linear transformation.
5.
   (a) ker(T) is the set of all polynomials $p(x) = ax + b$ such that T(ax+b)=[6a-4b-9a+6b]=[00]. Thus, a=23b, and ker(T)={p(x):p(x)=23bx+b}=span{23x+1}.
   The range of T consists of all matrices of the form [6a-4b-9a+6b]=(3a-2b)[2-3]; thus, range(T)=span{[2-3]}.

(b) $\ker(T)$ is the set of all polynomials $p(x) = ax^2 + bx + c$ such that $T(ax^2 + bx + c) = a - 3b + c = 0$. Therefore, $a = 3b - c$, and thus $\ker(T) = \{p(x) : p(x) = (3b - c)x^2 + bx + c\} = \text{span}\{3x^2 + x, -x^2 + 1\}$. The range of T consists of all polynomials $p(x) = a - 3b + c$; thus, $\text{range}(T) = \text{span}\{1\}$.

(c) $\ker(T)$ is the set of all $\mathbf{v}$ such that $T(\mathbf{v}) = \mathbf{v} = \mathbf{0}$; thus, $\ker(T) = \{\mathbf{0}_V\}$. The range of T is the vector space V, because every $\mathbf{v} \in V$ is obtained from $T(\mathbf{v}) = \mathbf{v}$.

(d) $\ker(T)$ is the set of all $\mathbf{v}$ such that $T(\mathbf{v}) = \mathbf{0}_W$; thus, $\ker(T) = V$, because every $\mathbf{v} \in V$ satisfies $T(\mathbf{v}) = \mathbf{0}_W$. The range of T is the set $\{\mathbf{0}_W\}$, because $T(\mathbf{v}) = \mathbf{0}_W$ for all $\mathbf{v} \in V$.

Section 9.2

1.

(a) $\dim(V) = \dim(\mathbf{R}^{4\times5}) = 4(5) = 20$, and $\dim(W) = \dim(\mathbf{P}^{19}) = 20$. Because $\dim(V) = \dim(W)$, the vector spaces are isomorphic.

(b) Both V and W are infinite dimensional, so dimensions cannot be used to determine if V and W are isomorphic.

2.

(a) First check that T is a linear transformation. Let [a1b1] and [a2b2] be two vectors in $\mathbf{R}^2$, and let and t be scalars. Then

T=(s[a1b1]+t[a2b2])=T([sa1+ta2sb1+tb2])=(sb1+tb2)x+(sa1+ta2)-(sb1+tb2)=s(b1x+a1-b1)+t(b2x+a2-b2)=sT([a1b1])+tT([a2b2])

Thus, T is a linear transformation. Suppose T(ab)=bx+a-b=0x+0, the zero vector in $\mathbf{P}^2$. Then $b = a = 0$, and [ab]=[00], the zero vector in $\mathbf{R}^2$. Thus ker(T)={0R2} and, therefore, T is one-to-one. By Theorem 9.3, $\dim(\text{range}(T)) = \dim(\mathbf{R}^2) - \dim(\ker(T)) = 2 - 0 = 2$. Thus, because $\dim(\mathbf{P}^1) = 2$, $\text{range}(T) = \mathbf{P}^1$, and T is onto. Therefore, T is an isomorphism.

(b) First check that T is a linear transformation. Let A and B belong to $\mathbf{R}^{2\times3}$, and let s and t be scalars. Then

T(sA+tB)=(sA+tB)T =(sA)T+(tB)T =sAT+tBT =s(T(A))+t(T(B))

Thus, by Theorem 9.2, T is a linear transformation. Suppose $T(A) = A^T = 0$, the zero matrix in $\mathbf{R}^{3\times2}$. Then $A = 0T$, the zero matrix in $\mathbf{R}^{2\times3}$. Thus, ker(T)={0R2×3} and, therefore, T is one-to-one. By Theorem 9.3, $\dim(\text{range}(T)) = \dim(\mathbf{R}^{2\times3}) - \dim(\ker(T)) = 6 - 0 = 6$. Thus, because $\dim(\mathbf{R}^{3\times2}) = 6$, $\text{range}(T) = \mathbf{R}^{3\times2}$ and T is onto. Therefore, T is an isomorphism.

3.

(a) T^{-1} : $\mathbf{R}^3 \rightarrow \mathbf{P}^2$. Suppose T(ax2+bx+c)=[bcd]=[def]. Then $b = d$, $c = e$, and $a = f$. Thus, T-1([def])=fx2+dx+e.

(b) T^{-1} : $\mathbf{P}^3 \rightarrow \mathbf{R}^{2\times2}$. Suppose T([uvwz])=vx3+wx2+zx-u=ax3+bx2+cx+d. Then $v = a$, $w = b$, $z = c$, and $u = -d$. Thus, T-1(ax3+bx2+cx+d)=[-dabc], because T-1(T([uvwz]))=T-1(ax3+bx2+cx+d)=T-1(vx3+wx2+zx-u)=[uvwz].

Section 9.3

1.

(a) v=(3)[-51]+(2)[3-4]=[-9-5]

(b) $\mathbf{v} = (2)(2x^2 - 3x + 1) + (1)(x^2 + 5) + (-2)(-x + 3) = 5x^2 - 4x + 1$

2.

(a) Set [v]$\mathcal{G}$=[ab]. Then $19x + 14 = a(-3x) + b(x + 2) = (-3a + b)x + 2b$. So $2b = 14 \Rightarrow b = 7$, and $-3a + (7) = 19 \Rightarrow a = -4$. Thus [v]$\mathcal{G}$=[-47].

(b) Set [v]$\mathcal{G}$=[abcd]. Then [3025]=a[1011]+b[0101]+c[1-101]+d[0111]=[a+cb-c+da+da+b+c+d]. We solve the system

a+c=3b-c+d=0a+d=2a+b+c+d=5

and obtain $a = 1$, $b = 1$, $c = 2$, and $d = 1$. Thus, [v]$\mathcal{G}$=[1121].

3.

(a) [T(v)]$\mathcal{Q}$=A[v]$\mathcal{G}$=[2-134] [12]=[011]=[(0)(x-2)+(11)(3x+7)]$\mathcal{Q}$, so $T(\mathbf{v}) = 33x + 77$.

(b) [T(v)]$\mathcal{Q}$=A[v]$\mathcal{G}$=[10-223-3021] [0-22]=[-4-12-2]=[(-4)cos(x)+(-12)ex+(-2)sin(x)]$\mathcal{Q}$, so $T(\mathbf{v}) = -4\cos x - 2\sin x - 12e^x$.

4.

(a) T(e1)=T([10])=(1+0)x-0=x⇒[T(e1)]$\mathcal{Q}$=[-10], and T(e2)=T([01])=(0+1)x-1=x-1⇒[T(e2)]$\mathcal{Q}$=[-1-13]. Therefore, A=[-1-10-13].

(b) T(1)=T(0x+1)=1x2+(0-1)x-0=x2-x⇒[T(x)]$\mathcal{Q}$=[1-120], and T(x)=T(1x+0)=0x2+(1-0)x-1⇒[T(1)]$\mathcal{Q}$=[0121]. Therefore, A=[10-121201].

5. Let $\mathcal{B}$ be a basis for V, $\mathcal{C}$ a basis for W, and $\mathcal{D}$ a basis for Y such that T is represented by A with these bases, and S is represented by B with these bases. Let $\mathbf{v} \in V$. Then

[(S∘T)(v)]$\mathcal{D}$=[S(T(v))]$\mathcal{D}$ =B[T(v)]$\mathcal{C}$ =B[A(v)$\mathcal{B}$] =(BA)[v]$\mathcal{B}$

So BA is the matrix representation of $S \circ T$ with respect to these bases.

Section 9.4

1.

(a) Set $x - 2 = s_{11}(x) + s_{21}(1)$, to obtain $s_{11} = 1$ and $s_{21} = -2$. Set $2x + 1 = s_{21}(x) + s_{22}(1)$, to obtain $s_{21} = 2$ and $s_{22} = 1$. Thus, S=[12-21].

(b) Set $3x - 2 = s_{11}(1) + s_{21}(x^2) + s_{31}(x)$, to obtain $s_{11} = -2$, $s_{21} = 0$, and $s_{31} = 3$. Set $x^2 + x = s_{12}(1) + s_{22}(x^2) + s_{32}(x)$, to obtain $s_{12} = 0$, $s_{22} = 1$, and $s_{32} = 1$. Set $x^2 - x + 3 = s_{13}(1) + s_{23}(x^2) + s_{23}(x)$, to obtain $s_{13} = 3$, $s_{23} = 1$, and $s_{33} = -1$. Thus, S=[-20301131-1].

2.

(a) A=S-1BS=[10-15]-1[1-421] [10-15]=[5-2065-3]

(b) A=S-1BS=[100310121]-1[1010131-1-1] [100310121]=[221010-5-7-2]

3.

(a) det(A)=det([136-1])=-19, and det(B)=det([32-15])=17. Because det(A) ≠ det(B), A and B are not similar matrices.

(b) det(A)=det([10-21-242-23])=-2, and det(B)=det([1-32-112-214])=4. Because det(A) ≠ det(B), A and B are not similar matrices.

Chapter 10

Section 10.1

1.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = t_1u_1v_1 + t_2u_2v_2 + t_3u_3v_3 = (2)(2)(1) + (4)(5)(3) + (1)(2)(0) = 64$

(b) $\langle p, q \rangle = p(-1)q(-1) + p(1)q(1) + p(2)q(2) = (5(-1) + 1)(-2(-1) + 3) + (5(1) + 1)(-2(1) + 3) + (5(2) + 1)(-2(2) + 3) = -20 + 6 + -11 = -25$

(c) ⟨f,g⟩=∫-11(x+2)(-x2)dx=∫-11(-x3-2x2)dx=-43

(d) ⟨A,B⟩=tr(ATB)=tr([1-12-1]T[5342])=tr([137-9-5])=8

2.

(a) Let u=[41-2]. Then ‖u‖=⟨u,u⟩=(2)(4)(4)+(3)(1)(1)+(1)(-2)(-2)=39.

(b) Let $p(x) = 4x + 3$. Then ‖p‖=⟨p,p⟩=p(-1)p-1+p(1)p(1)+p(3)p(3)=(-1)2+(7)2+(15)2=511.

(c) Let $f(x) = x^2$. Then ‖f‖=⟨f,f⟩=∫-11(x2)(x2)dx=25=1510

(d) ‖A‖=⟨A,A⟩=tr(ATA)=(tr([2-113]T[2-113]))1/2=(tr([51110]))1/2=15

3.

(a) projuv=⟨u,v⟩⟨u,u⟩u =(2)(1)(-1)+(1)(3)(2)+(3)(-2)(0)(2)(1)(1)+(1)(3)(3)+(3)(-2)(-2)[13-2] =423[13-2]=[4231223-823]

(b) projpq=⟨p,q⟩⟨p,q⟩
p =p(-1)q(-1)+p(0)q(0)+p(2)q(2)p(-1)p(-1)+p(0)p(0)+p(2)p(2)(x+2) =(1)(4)+(2)(3)+(4)(1)(1)2+(2)2+(4)2(x+2) =23(x+2)=23x+43

(c) projfg=⟨f,g⟩⟨f,f⟩f=∫-11(2x)(-x3)dx∫-11(2x)(2x)dx(2x)=-4583(2x)=-35x

(d) projAB=⟨A,B⟩⟨A,A⟩A=tr(ATB)tr(ATA)A= =tr([1-112]T[2-10-2])tr([1-112]T[1-112])[1-112] =tr([2-3-2-3])tr([2115])[1-112] =-17[1-112]=[-1717-17-27]

4.

(a) False. For example, let v=[20] and u=[10]. Then **u** and **v** are parallel, but projuv=[20]≠u.

(b) False. $\langle -2\mathbf{u}, 3\mathbf{v} \rangle = -6\langle \mathbf{u}, \mathbf{v} \rangle$, not $|-6|\langle \mathbf{u}, \mathbf{v} \rangle = 6\langle \mathbf{u}, \mathbf{v} \rangle$.

(c) True, because the properties of an inner product are all fulfilled.

(d) False. If $\mathbf{u} = -2\mathbf{v}$, then $||\mathbf{u}|| = 2||\mathbf{v}||$, but $\mathbf{u} \neq 2\mathbf{v}$ unless $\mathbf{v} = \mathbf{0}$.

Section 10.2

1. We have $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = (1)(2)(2) + (2)(-2)(a) + (3)(4)(0) = 4 - 4a$, which is 0 when $a = 1$. We have $\langle \mathbf{v}_2, \mathbf{v}_3 \rangle = (1)(2)(-2) + (2)(a)(2) + (3)(0)(1) = -4 + 4a$, which also is 0 when $a = 1$. Finally, we have $\langle \mathbf{v}_1, \mathbf{v}_3 \rangle = (1)(2)(-2) + (2)(-2)(2) + (3)(4)(1) = 0$. Therefore the set is orthogonal when $a = 1$. The norms of these vectors are ‖v1‖=60, ‖v2‖=6 (when $a = 1$), and ‖v3‖=15. Therefore the orthonormal set is {160v1, 16v2,115v3}.

2. We check $\langle p_2, p_3 \rangle = p_2(0)p_3(0) + p_2(1)p_3(1) + p_2(2)p_3(2) = (-2)(-3) + (-3)(-2) + (-6)(-1) = 18 \neq 0$, so no value of a will make the set of vectors orthogonal.

3. Because {1/2,cos(x),sin(x)} is orthonormal, projSf=⟨f1,f⟩f1+⟨f2,f⟩f2+⟨f3,f⟩f3=(1π∫-ππ(1/2)(x2)dx)(1/2)+(1π∫-ππ(cos(x))(x2)dx)(cos(x))+ (1π∫-ππ(sin(x))(x2)dx)(sin(x))=(132π2)(1/2)+(-4)(cos(x))+(0)(sin(x))=13π2-4cosx.

4. We apply Gram–Schmidt to orthogonalize $\{1, x\}$. Let $f_1(x) = 1$, and $f_2(x) = x$. Set $g_1(x) = f_1(x) = 1$. Let g2(x)=f2(x)-(projg1f2)(x)=f2(x)-⟨g1,f2⟩‖g1‖2g1(x)=x-∫02(1)(x)dx∫02(1)(1)dx(1)=x-22(1)=x-1. Thus, projSf=⟨g1,f⟩‖g1‖2g1+⟨g2,f⟩‖g2‖2g2=∫02(1)(e-x)dx∫02(1)(1)dx(1)+∫02(x-1)(e-x)dx∫02(x-1)(x-1)dx(x-1)=1-e-22(1)+-2e-223(x-1)=12-12e-2-3e-2(x-1).

5.

(a) Let s1=[-110], s2=[011], and s3=[10-1]. Set v1=s1=[-110], and S_1 = span{$\mathbf{s}_1$}.

Let v2=s2-projS1s2=s2-⟨v1,s2⟩‖v1‖2v1 =[011]-2(-1)(0)+3(1)(1)+1(0)(1)2(-1)2+3(1)2+1(0)2[-110] =[011]-35[-110]=[35251]

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for S_2 = span{$\mathbf{s}_1$, $\mathbf{s}_2$}.

Let v3=s3-projS2s3=s3-⟨v1,s3⟩‖v1‖2v1-⟨v2,s3⟩‖v2‖2v2 =[011]-2(-1)(1)+3(1)(0)+1(0)(-1)2(-1)2+3(1)2+1(0)2[-110]-2(35)(1)+3(25)(0)+1(1)(-1)2(35)2+3(25)2+1(1)2[35251] =[10-1]--25[-110]-15115[35251]=[611411-1211]

Then {$\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$} is an orthogonal basis for span{$\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$}.

(b) Let $f_1(x) = x^2 + 1$, $f_2(x) = -x^2$, and $f_3(x) = x - 1$. Set $g_1(x) = f_1(x) = x^2 + 1$, and S_1 = span{g_1}. Let g2(x)=f2(x)-(projS1f2)(x)=f2(x)-⟨g1,f2⟩‖g1‖2g1(x)= (-x2)-∫01(x2+1)(-x2)dx∫01(x2+1)(x2+1)dx(x2+1)=-x2- -8152815(x2+1)=-57x2+27, and S_2 = span{g_1, g_2}. Let g3(x)=f3(x)-(projS2f3)(x)=(x-1)-⟨g1,f3⟩‖g1‖2g1(x)-⟨g2,f3⟩‖g2‖2g2(x)= (x-1)-∫01(x2+1)(x-1)dx∫01(x2+1)(x2+1)dx(x2+1)-∫01(-57x2+27)(x-1)dx∫01(-57x2+27)(-57x2+27)dx (-57x2+27)=(x-1)--7122815(x2+1)- -112121(-57x2+27)=-1516x2+x-316. We conclude that {x2+1,-57x2+27,-1516x2+x-316} is an orthogonal basis for S = span{f_1, f_2, f_3}.

6. By Theorem 7.14, we can extend the vectors to form an linearly independent set {$\mathbf{v}_1$, … , $\mathbf{v}_k$, $\mathbf{s}_{k+1}$, … , $\mathbf{s}_m$}. By the Gram–Schmidt process, retaining the first k vectors $\mathbf{v}_1$, … , $\mathbf{v}_k$, we form an orthogonal set {$\mathbf{v}_1$, … , $\mathbf{v}_k$, $\mathbf{v}_{k+1}$, … , $\mathbf{v}_m$} An orthogonal set of vectors is linearly independent and, therefore, this is an orthogonal basis for V.

Section 10.3

1. We need to find projspan{a1,a2}y, where a1=[1111] and a2=[-2-102], so we first find an othogonal basis for {$\mathbf{a}_1$, $\mathbf{a}_2$} using Gram–Schimdt. Let $\mathbf{v}_1 = \mathbf{a}_1$, and v2=a2-projv1a2=[-2-102]-⟨v1,a2⟩⟨v1,v1⟩v1=[-2-102]-1(1)(-2)+3(1)(-1)+3(1)(0)+1(1)(2)1(1)(1)+3(1)(1)+3(1)(1)+1(1)(1)[1111]=[-138-5838198] We evaluate ⟨v1,y⟩⟨v1,v1⟩=1(1)(3)+3(1)(1)+3(1)(0)+1(1)(-4)1(12)+3(12)+3(12)+1(12)=14 and ⟨v2,y⟩⟨v2,v2⟩=1(-138)(3)+3(-58)(1)+3(38)(0)+1(198)(-4)1(-138)2+3(-58)2+3(382)+1(198)2=-13079, and obtain the weighted least squares line y=14-13079x.

2. Evaluate

a0=12π(∫-ππ(1+3x)dx)=1,a1=1π(∫-ππ(1+3x)cos(x)dx)=0,a2=1π(∫-ππ(1+3x)cos(2x)dx)=0,b1=1π(∫-ππ(1+3x)sin(x)dx)=6,b2=1π(∫-ππ(1+3x)sin(2x)dx)=-3.

Thus, $f_2(x) = 1 + 6\sin(x) - 3\sin(2x)$.

3. Evaluate

a0=12π(∫-ππ(1-|x|)dx)=1-12π,a1=1π(∫-ππ(1-|x|)cos(x)dx)=4π,a2=1π(∫-ππ(1-|x|)cos(2x)dx)=0,b1=1π(∫-ππ(1-|x|)sin(x)dx)0,b2=1π(∫-ππ(1-|x|)sin(2x)dx)=0.

Thus, f2(x)=1-12π+4πcos(x).

4. Because 1, $\cos(3x)$, and $\sin(2x)$ are basis functions, we have $a_0 = 3$, $a_3 = -1$, and $b_2 = 5$, with all other Fourier coefficients 0.

5. Evaluate

c0=12(f(0)+f(π))=12(4+(-3))=12,c1=22(f(0)cos(0)+f(π)cos(π))=22(4cos(0)+(-3)cos(π))=7,d1=22(f(0)sin(0)+f(π)sin(π))=22(4sin(0)+(-3)sin(π))=0.

Thus, g1(x)=12+7cos(x).

Chapter 11

Section 11.1

1.

(a) Q(x0)=Q([-23])=3(-2)2-4(3)2+14(-2)(3)=-108

(b) Q(x0)=Q([1-12])=2(1)2+5(-1)2-3(2)2+10(1)(2)-18(-1)(2)=51

2.

(a) Q(x)=Q([x1x2])=-3x12+7x22-8x1x2

(b) Q(x)=Q([x1x2x3])=2x12+3x22-4x32-2x1x2+10x1x3+4x2x3

3.

(a) A=[26065000-3]

(b) A=[-4-3-5-362-520]

4.

(a) det(λI2-A)=det(λ[1001]-[422-3])=λ2-λ-16=0⇒λ1=12-1265<0 and λ2=1265+12>0. Because *A* has both a positove and negative eigenvalue, by Theorem 11.4(c), *A* is indefinite.

(b) det(λI3-A)=det(λ[100010001]-[100001010])= λ3-λ2-λ+1=(λ+1)(λ-1)2=0⇒λ1=-1<0 and λ_2 = 1 > 0. Because *A* has both a positive and negative eigenvalue, by Theorem 11.4(c), *A* is indefinite.

Section 11.2

1.

(a) A_1 = [4], A2=[455-2]

(b) A_1 = [−1], A2=[-1220], A3=[-12-4201-413]

2.

(a) Apply Theorem 11.8, because *A* is symmetric. det(A_1) = det([1]) = 1 > 0, and det(A2)=det([1215])=1>0, so *A* is positive definite.

(b) Apply Theorem 11.8, because *A* is symmetric.

det(A_1) = det([1]) = 1 > 0, and
det(A2)=det([1114])=3>0,
det(A3)=det([11-1143-1311])=14>0 so *A* is positive definite.

Section 11.3

1.

(a) max{q_1, q_2} = max{5, −2} = 5 and min{q_1, q_2} = min{5, −2} = −2. So, by Theorem 11.11, the maximum and minimum values of $Q(\mathbf{x})$ subject to $||\mathbf{x}||$ = 1 are 5 and −2, respectively.

(b) max{q_1, q_2, q_3} = max{2, −1, −4} = 2, and min{q_1, q_2, q_3} = min{2,−1,−4} = −4. So, by Theorem 11.11, the maximum and minimum values of $Q(\mathbf{x})$ subject to $||\mathbf{x}||$ = 1 are 2 and −4, respectively.

(c) Q(x)=xT[1332]x, and det(λI2-A)=det(λ[1001]-[1332])=λ2-3λ-7=0

⇒λ1=32-1237<λ2=32+1237. By Theorem 11.12, subject to $||\mathbf{x}||$ = 1, the maximum value of $Q(\mathbf{x})$ is 32+1237, and the minimum value of $Q(\mathbf{x})$ is 32-1237.

(d) Q(x)=xT[103030302]x, and det(λI3-A)=det(λ[100010001]-[103030302])=λ3-6λ2+2λ+21=(λ-3)(λ2-3λ-7)=0 ⇒λ1=32-1237<λ2=3<λ3=32+1237. By Theorem 11.12, subject to $||\mathbf{x}||$ = 1, the maximum value of $Q(\mathbf{x})$ is 32+1237, and the minimum value of $Q(\mathbf{x})$ is 32-1237.

2.

(a) Q(x)=xT[1339]x, and det(λI2-A)=det(λ[1001]-[1339])=λ2=10λ=λ(λ-10)=0⇒λ1=0<λ2=10. Thus, subject to $||\mathbf{x}||$ = c = 2, the maximum value of $Q(\mathbf{x})$ is $c^2(10) = (2)^2(10) = 40$, and the minimum value of $Q(\mathbf{x})$ is $c^2(0) = (2)^2 0 = 0$.

(b) Q(x)=xT[001010100]x, and det(λI3-A)=det(λ[100010001]-[001010100])=λ3-λ2-λ+1

=(λ+1)(λ-1)2=0⇒-1<λ2=λ3=1. Thus, subject to $||\mathbf{x}||$ = c = 10, the maximum value of $Q(\mathbf{x})$ is $c^2(1) = (10)^2(1) = 100$, and the minimum value of $Q(\mathbf{x})$ is $c^2(-1) = (10)^2(-1) = -100$.

(c) Let $w_1 = x_1/5$ and $w_2 = 3x_2/10$. So $x_1 = 5w_1$ and x2=103w2. Then Q(x)=(5w1)2+(103w2)2+8(5w1)(103w2)=25w12+1009w22+4003w1w2=wT[25200320031009]w. Therefore, det(λ[1001]-[25200320031009])=λ2-3259λ-125003=0⇒λ1=32518-25182329<λ2=32518+25182329. Thus, subject to 4x12+9x22=100⇔‖w‖=1, the maximum value of $Q(\mathbf{x})$ is 32518+25182329, and the minimum value of $Q(\mathbf{x})$ is 32518-25182329.

(d) Let $w_1 = x_1/3$ and $w_2 = x_2$. So, $x_1 = 3w_1$ and $x_2 = w_2$. Then Q(x)=4(3w1)2+9(w2)2+4(3w1)(w2)=36w12+9w22+12w1w2=wT[36669]w. Therefore, det(λ[1001]-

[36669])=λ2-45λ+288=0⇒λ1=452-3297<λ2=452+3297. Thus, subject to x12+9x22=9⇔‖w‖=1, the maximum value of $Q(\mathbf{x})$ is 452+3297, and the minimum value of $Q(\mathbf{x})$ is 452-3297.

Section 11.4

1.

(a) $2\mathbf{u} + \mathbf{w} = 2(1- 4i, -1+i, 2+3i)+(2+i, 2-5i, 2-3i) = (4 - 7i, -3i, 6 + 3i)$ and $-3\mathbf{v} + \mathbf{u} - 2\mathbf{w} = -3(1 + i, 2 - i, 1 + 3i) + (1 - 4i, -1 + i, 2 + 3i)-2(2 + i, 2 - 5i, 2 - 3i) = (-6 - 9i, -11 + 14i, -5)$

(b) We apply Gaussian elimination [1-4i1+i2+i1+i-1+i2-i2-5i-2+i2+3i1+3i2-3i3-4i]×(--1+i1-4i)R1+R2→R2(-2+3i1-4i)R1+R3→R3[1-4i1+i2+i1+i03617-917i4117-7417i-3217+2517i03817+5017i6517-6317i7217-6917i]×(-3817+5017i3617-917i)R2+R3→R3[1-4i1+i2+i1+i03617-917i4117-7417i-3217+2517i00-419-419i709-199i] and conclude that $(1 + i, -2 + i, 3 - 4i) \in \text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$.

(c) ⟨u,w⟩=u1w1¯+u2w2¯+u3w3¯=(1-4i)(2+i)¯+(-1+i)(2-5i)¯+(2+3i)(2-3i)¯ =(1-4i)(2-i)+(-1+i)(2+5i)+(2+3i)(2+3i)=(-2-9i)+(-7-3i)+(-5+12i)=-14

(d) ‖2iv‖=|2i| ‖v‖=2|v1|2+|v2|2+|v3|2=2(12+11)(22+(-1)2)+(12+32)=217

(e) v‖v‖=1(12+12)+(22+(-1)2)+(12+32)[1+i2-i1+3i]=[(134+134i)17(117-134i)17(134+334i)17]w‖w‖=1(22+12)+(22+(-5)2)+(22+(-3)2)[2+i2-5i2-3i]=[(247+147i)17(247-547i)17(247-347i)17]

(f) v-u‖v-u‖=1‖v-u‖([1+i2-i1+3i]-[1-4i-1+i2+3i]) =1‖v-u‖[5i3-2i-1] =1(02+52)+(32+(-2)2)+((-1)2+02)[5i3-2i-1] =139[5i3-2i-1] = [539i39(113-239i)39-13939]2u+w‖2u+w‖ =1‖2u+w‖(2[1-4i-1+i2+3i]+[2+i2-5i2-3i]) =1‖2u+w‖[4-7i-3i6+3i] =1(42+(-7)2)+(02+(-3)2)+(62+32)[4-7i-3i6+3i] =1119[4-7i-3i6+3i]=[(4119-117i)119-3119i119(6119+3119i)119]

Section 11.5

1.

(a) A*=[-i4+5i-3+7i1-6i]*=[-i4+5i-3+7i1-6i]T¯ =[-i-3+7i4+5i1-6i]¯=[-i-3-7i4-5i1+6i]

(b) A*=[1+i2+3i3-i3-i-5+i5+i7+i2+i4+4i]* =[1+i2+3i3-i3-i-5+i5+i7+i2+i4+4i]T =[1+i3-i7+i2+3i-5+i2+i3-i5+i4+4i]¯ =[1-i3+i7-i2-3i-5-i2-i3+i5-i4+-i]

2.

(a) A*=[1+2i5i-5i2-i]*=[1-2i5i-5i2+i]≠A, so A is not Hermitian.

(b) A*=[2i1-i-i-301+i04]*=[2i1-i-i-301+i04]=A, so A is Hermitian.

ANSWERS TO SELECTED EXERCISES

Chapter 1

Section 1.1

1. Only (−3, −3) lies on line.

3. Only (−2, 5) lies on both lines.

5. None satisfies the linear system.

7. Only (b), (c), and (d) are solutions to the linear system.

9. $x_1 = 3$, $x_2 = -1$.

11. x1=s1,x2=12+52s1

13. x1=-841,x2=-541

15. Echelon form; x_1, x_2 leading variables, no free variables.

17. Echelon form; x_1, x_3 leading variables, x_2 a free variable.

19. Not in echelon form, x_2 the leading variable in two equations.

21. Echelon form; x_1, x_3 leading variables, and x_2, x_4 free variables.

23. x1=-195,x2=5

25. x1=-23,+43s1,x2=s1

27. x1=10-12s1,x2=-2+12s1,x3=s1,x4=5

29. x1=56+12s1,+13s2,x2=s1,x3=43+13s2,x4=s2

31.

- (a) Reverse order of equations; x1=1315,x2=−45.
- (b) Interchange equations 1 and 3; $x_1 = 9$, $x_2 = -17$, $x_3 = 1$.

33. $x_1 = -3 + 5s_1$, $x_2 = 2 - 2s_1, x_3 = s_1$, $x_4 = 0$

35. $x_1 = 3 + 2s_1$, $x_2 = 1 - s_1, x_3 = s_1$

37.

- (a) k≠-152
- (b) k=23

39. 9 variables

41. 7 leading variables

43. For example,

x1=0x2=0x3=0

45. For example,

x1+x2 =0x1+x2-x3=0   x3=0x1+x2+x3=0

47. On Monday, I bought 3 apples and 4 oranges and spent \$0.55. On Tuesday I bought 6 oranges and spent \$0.60. How much does each apple and orange cost?
Answer: Apples cost 5 cents each and oranges cost 10 cents each.

49. For example,

x1-x2 =-33x1 -x3=4

51.

(a) False
(b) False

53.

(a) True
(b) False

55.

(a) True
(b) True

57. 298 adults and 87 children

59. a1=145anda2=115

61. 196.875 liters of 18% solution, 103.125 liters of 50% solution

63. \$33,333 in safe bond, \$66,667 in risky bond

65. 36.923 gallons hot water, 23.077 gallons cold water

67. a=59andb=-1609

69. The published values from the U.S. Mint are q = 0.955 in and n = 0.835 in.

71. $x_1 = 12$, $x_2 = 5$

73. $x_1 = 0.625 + 14.25s_1$, $x_2 = -1.75 + 2.25s_1$, $x_3 = s_1$

75. $x_1 = -8.417 - 4.125s_1$, $x_2 = 4.333 - 2s_1$, $x_3 = 1.5 + 1.75s_1$, $x_4 = s_1$

77. $x_1 = 30.985 + 14.625s_1$, $x_2 = 1.966 - 2.125s_1$, $x_3 = 8.227 - 0.5s_1$, $x_4 = s_1$, $x_5 = 0.182$

Section 1.2

1. 4x1+2x2-x3=2-x1 +5x3=7

3. 12x2-3x3-9x4=17-12x1+5x2-3x3+11x4=06x1+8x2+2x3+10x4=-817x1 +13x4=-1

5. Echelon form.

7. Not in echelon form.

9. Echelon form.

11. $-2R_1 \rightarrow R_1$

13. $-2R_2 + R_3 \rightarrow R_3$

15. R1↔R2,[-14337-250-3]

17. 2R1↔R1,[06-24-1-9415072]

19. [211-4-13];x1=-2,x2=5

21. [-25-1041-23-17-1734-16];x1=-12,x2=-10,x3=-3

23. [22-18-1-10-33317];x1=3-s1,x2=s1,x3=-2

25. [26-9-40-3-119-1014-210];x1=35s1,x2=-8s1,x3=2s1,x4=s1

27. [-2-50131];x1=-5,x2=2

29. [2102-1-1-11];x1=3+s1,x2=-4-2s1,x3=s1

31. $x_1 = 1 + s_1$, $x_2 = -1 - 2s_1$, $x_3 = 2 - s_1$, $x_4 = s_1$

33.

(a) (15)R1→R1

(b) (-12)R3→R3

35.

(a) $5R_2 + R_6 \rightarrow R_6$

(b) $3R_1 + R_3 \to R_3$

37. An example: [111110111100111]

39. An example: [1004010300120001]

41. An example:

x1 =0 x2 =0 x3+x4=0

43.

(a) True

(b) True

45.

(a) False

(b) False

47.

(a) -R2→R2R3+R2→R2

(b) 3R1→R1R4+R1→R1

49.

(a) -4R6→R6R3+R6→R6

(b) Not a combination of row operations.

51. Exactly one solution.

53. HINT: Show that the system must have at least one free variable.

55. HINT: Show that the system must have at least one free variable.

57. $f(x) = 2x^2 - 3x + 5$

59. E(x)=-110x2+495x+132

61. x1=-157181,x2=20181,x3=-58181

63. x1=79-s1,x2=-239-s1,x3=-2227+s1,x4=s1

65. No solutions.

67. x1=46579s1,x2=-745579s1,x3=-2264579s1,x4=-655386s1,x5=s1

Section 1.3

1. Minimum = 20 vehicles

3. Minimum = 25 vehicles

5. $x_1 = 62$, $x_2 = 44$

7. $x_1 = 55$, $x_2 = 65$, $x_3 = 50$

9. $a = 76.596$, $b = 55.319$

11. $a = 55.774$, $b = 63.688$, $c = 81.104$

13. $2H_2 + O_2 \rightarrow 2H_2O$

15. $4Fe + 3O_2 \rightarrow 2Fe_2O_3$

17. $C_3H_8 + 5O_2 \rightarrow 3CO_2 + 4H_2O$

19. $4KO_2 + 2CO_2 \rightarrow 2K_2CO_3 + 3O_2$

21. $p = (0.19847)d^{1.5011}$

23. $p = (0.20120)d^{1.49835}$

25. $d = 0.045s^2$

27. $A = 1$, $B = -1$

29. $A = -1$, $B = -1$, $C = 1$

31. $x = 4$

33. $z = 5$

35. $a = -1$, $b = 3$, $c = 2$

37. f(x)=-14x3-16x2+6512x-3

39. $f(x) = -2e^x + 3e^{2x} + e^{-3x}$

Section 1.4

1. $x_1 = 1$, $x_2 = 2$

3. x1=7949,x2=2249,x3=12449

5. No partial pivot: $x_1 = -0.219$, $x_2 = 0.0425$
With partial pivot: $x_1 = -0.180$, $x_2 = 0.0424$

7. No partial pivot: $x_1 = -0.407$, $x_2 = -0.757$, $x_3 = 0.0124$
With partial pivot: $x_1 = -0.392$, $x_2 = -0.755$, $x_3 = 0.0124$

9.

n	x_1	x_2
0	0	0
1	−1.2	0.2
2	−1.12	0.56
3	−0.976	0.536

Exact solution: $x_1 = -1$, $x_2 = 0.5$

11.

n	x_1	x_2	x_3
0	0	0	0
1	−1.3	2.3	2.6
2	−2.295	3.34	1.42
3	−2.156	3.185	0.805

Exact solution: $x_1 = -2$, $x_2 = 3$, $x_3 = 1$

13.

n	x_1	x_2
0	0	0
1	−1.2	0.56
2	−0.976	0.4928
3	−1.0029	0.5009

Exact solution: $x_1 = -1$, $x_2 = 0.5$

15.

n	x_1	x_2	x_3
0	0	0	0
1	−1.3	2.56	1.316
2	−2.013	3.0974	0.9584
3	−2.0042	2.9884	1.0038

Exact solution: $x_1 = -2$, $x_2 = 3$, $x_3 = 1$

17. Not diagonally dominant. Not possible to reorder to obtain diagonal dominance.

19. Not diagonally dominant. Not possible to reorder to obtain diagonal dominance.

21. Jacobi iteration of given linear system:

n	x_1	x_2
0	0	0
1	−1	−1
2	−3	−3
3	−7	−7
4	−15	−15

Diagonally dominant system:

2x1-x2=1x1-2x2=-1

Jacobi iteration of diagonally dominant system:

n	x_1	x_2
0	0	0
1	0.5	0.5
2	0.75	0.75
3	0.875	0.875
4	0.9375	0.9375

23. Jacobi iteration of given linear system:

n	x_1	x_2	x_3
0	0	0	0
1	−1	8	−0.3333
2	16.67	12.33	27
3	−111.3	−21.33	29.67
4	−192	624	2.778

Diagonally dominant system:

5x1+x2-2x3=82x1-10x2+3x3=-1x1-2x2+5x3=-1

Jacobi iteration of diagonally dominant system:

n	x_1	x_2	x_3
0	0	0	0
1	1.6	0.1	−0.2
2	1.5	0.36	−0.48
3	1.336	0.256	−0.356
4	1.406	0.2604	−0.3648

25. Gauss–Seidel iteration of given linear system:

n	x_1	x_2
0	0	0
1	−1	−3
2	−7	−15
3	−31	−63
4	−127	−255

Diagonally dominant system:

2x1-x2=1x1-2x2=-1

Gauss–Seidel iteration of diagonally dominant system:

n	x_1	x_2
0	0	0
1	0.5	0.75
2	0.875	0.9375
3	0.9688	0.9844
4	0.9922	0.9961

27. Gauss–Seidel iteration of given linear system:

n	x_1	x_2	x_3
0	0	0	0
1	−1	13	43.67
2	−193.3	1062	3669
3	-1.622×10^4	8.844×10^4	3.056×10^5

n	x_1	x_2	x_3
4	-1.351×10^6	7.367×10^6	2.546×10^7

Diagonally dominant system:

5x1+x2-2x3=82x1-10x2+3x3=-1x1-2x2+5x3=-1

Gauss–Seidel iteration of diagonally dominant system:

n	x_1	x_2	x_3
0	0	0	0
1	1.6	0.42	−0.352
2	1.375	0.2694	−0.3673
3	1.399	0.2697	−0.3720
4	1.397	0.2679	−0.3723

29. $x_1 = -3$, $x_2 = 18$

31. $x_1 = 27$, $x_2 = 52$

Supplementary Exercises

1. $x_1 = -22$, $x_2 = -9$

3. $x_1 = 1 + 4s_1$, $x_2 = s_1$

5. No solutions.

7. $x_1 = -10$, $x_2 = 11$, $x_3 = 3$

9. $x_1 = 2 - 5s_1 + s_2$, $x_2 = s_1$, $x_3 = s_2$

11. $x_1 = 18 + 7s_1$, $x_2 = -5$, $x_3 = s_1$, $x_4 = -5$

13. $x_1 = 12 - 7s_1$, $x_2 = -11 + 4s_1$, $x_3 = s_1$, $x_4 = -2$, $x_5 = 4$

15. 2x1-4x2+3x3=1-3x1+5x2+11x3=0

17. 4x1+2x2+5x3=17x1-2x2 =13x1+x2+2x3=-4

19. [1-2130-23-10004]

21. [-32-2020032-10002-1]

23. [1-30001000]

25. [1010001-20300013]

27. $x_1 = 2 + 2s_1$, $x_2 = 1 + s_1$, $x_3 = s_1$

29. $x_1 = 11 + 8s_1$, $x_2 = 1 + 3s_1$, $x_3 = s_1$, $x_4 = -3$

31. $x_1 = 90$, $x_2 = 60$

33. $a = 70.59$, $b = 41.18$

35. $C_6H_{12}O_6 \rightarrow 2C_2H_5OH + 2CO_2$

37. No partial pivot: $x_1 = -0.200$, $x_2 = -0.0422$
With partial pivot: $x_1 = -0.170$, $x_2 = -0.0421$

39. Jacobi iteration of given linear system:

n	x_1	x_2
0	0	0
1	−0.8	0.6
2	−0.62	0.92
3	−0.524	0.848

Exact: x1=-3156≈-0.5536, x2=2328≈0.8214

41. Gauss–Seidel iteration of given linear system:

n	x_1	x_2
0	0	0
1	−0.8	0.92
2	−0.524	0.8096
3	−0.5571	0.8228

Exact: x1=-3156≈-0.5536, x2=2328≈0.8214

Chapter 2

Section 2.1

1. u-v=[7-3-5],6w=[12-42-6]

3. w+3v=[-10-414],2w-7v=[32-21-37]

5. -u+v+w=[-5-44],2u-v+3w=[16-26-8]

7. 3x1-x2=82x1+5x2=13

9. -6x1+5x2=45x1-3x2+2x3=16

11. x1[2-1]+x2[8-3]+x3[-45]=[-104]

13. x1[1-2-3]+x2[-12-3]+x3[-3610]+x4[-120]=[-1-15]

15. [x1x2]=[-40]+s1[31]

17. [x1x2x3]=[7-30]+s1[-201]

19. [x1x2x3x4]=[40-90]+s1[6031]+s2[-5100]

21. 1u+0v= u = [3−2], 0u+1v=v=[−1−4], 1u+1v=[3−2]+[−1−4]=[2−6]

23. 1u+0v+0w=u=[-40-3],0u+1v+0w=v=[-2-15],0u+0v+1w=w=[9611]

25. $a = 2$, $b = 7$

27. $a = 3$, $b = 5$, and $c = 7$

29. $a = -12$, $b = 0$, $c = 4$, and $d = 5$

31. $\mathbf{b} = 3\mathbf{a}_1 + 2\mathbf{a}_2$

33. **b** is not linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$.

35. b=-1711a1-211a2+a3

37. 76 pounds of nitrogen, 31 pounds of phosphoric acid, and 14 pounds of potash

39. Two bags of Vigoro and three bags of Parker’s.

41. Three bags of Vigoro and two bags of Parker’s.

43. No solution possible.

45. No solution possible.

47. Two cans of Red Bull and one can of Jolt Cola.

49. Two cans of Red Bull and two cans of Jolt Cola.

51. Three servings of Lucky Charms and _ve servings of Raisin Bran.

53. Two servings of Lucky Charms and three servings of Raisin Bran.

55.

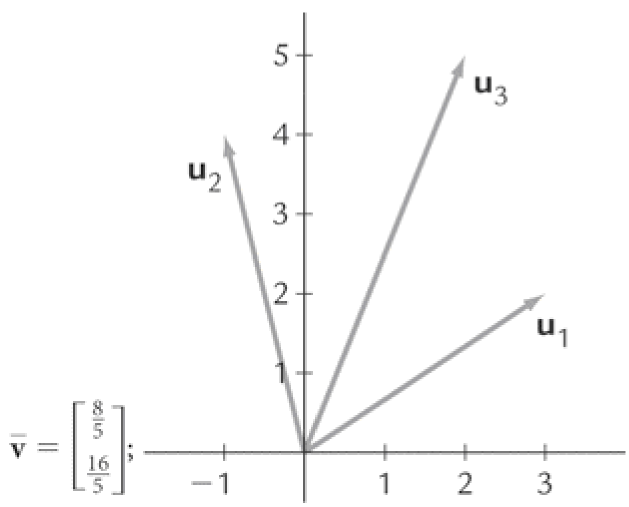
(a) a=[20008000],b=[300010,000]

(b) 8b=(8)[300010,000]=[24,00080,000]
The company produces 24,000 computer monitors and 80,000 flat panel televisions at facility B in 8 weeks.

(c) 30,000 computer monitors and 108,000 flat panel televisions

(d) 9 weeks of production at facility A and 2 weeks of production at facility B

57.



59. 6kg of $\mathbf{u}_1$, 3kg of $\mathbf{u}_2$, and 2kg of $\mathbf{u}_3$

61. For example, $\mathbf{u} = (0, 0, -1)$ and $\mathbf{v} = (3, 2, 0)$.

63. For example, $\mathbf{u} = (1, 0, 0)$, $\mathbf{v} = (1, 0, 0)$, and $\mathbf{w} = (-2, 0, 0)$.

65. For example, $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (2, 0)$.

67. For example, $\mathbf{a}_1 = (1, 2, 3)$, $\mathbf{a}_1 = (2, 4, 6)$, and $\mathbf{a}_1 = (3, 6, 9)$.

69. $x_1 = 3$, $x_2 = -2$

71.

(a) True

(b) False

73.

(a) True

(b) False

75.

(a) True

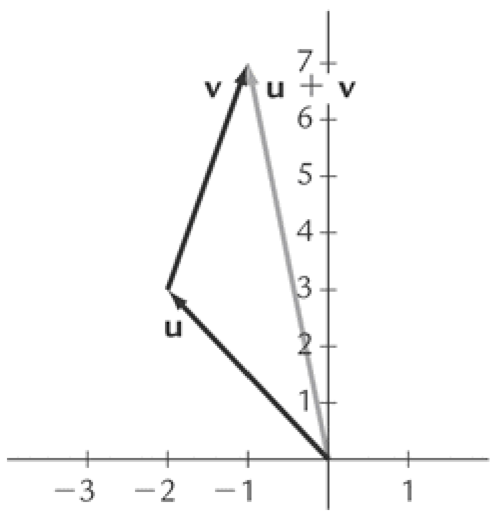
(b) True

77.

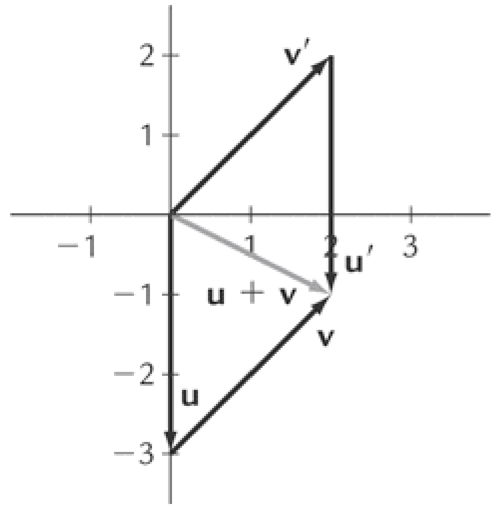
(a) False

(b) False

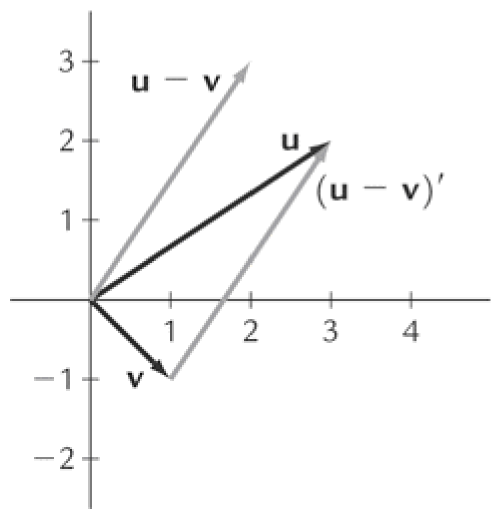
81.



83.



85.



87. $x_1 = 4$, $x_2 = -6.5$, and $x_3 = 1$

Section 2.2

1. 0u1+0u2=[00],1u1+0u2=[26],0u1+1u2=[915]

3. 0u1+0u2=[000],1u1+0u2=[25-3],0u1+1u2=[104]

5. 0u1+0u2+0u3=[000],1u1+0u2+0u3=[200],0u1+1u2+0u3=[416]

7. **b** is not in the span of $\mathbf{a}_1$.

9. **b** is not in the span of $\mathbf{a}_1$.

11. **b** is not in the span of $\mathbf{a}_1$ and $\mathbf{a}_2$.

13. A=[28-4-1-35],x=[x1x2x3],b=[-104]

15. A=[1-1-3-1-2262-3-3100],x=[x1x2x3x4],b=[-1-15]

17. x1[51]+x2[7-5]+x3[-2-4]=[92]

19. x1[403]+x2[-2-58]+x3[-372]+x4[53-1]=[1262]

21. Columns do not span $\mathbf{R}^2$.

23. Columns span $\mathbf{R}^2$.

25. Columns span $\mathbf{R}^3$.

27. Columns do not span $\mathbf{R}^3$.

29. For every choice of **b** there is a solution of $A\mathbf{x} = \mathbf{b}$.

31. Since A has more rows than columns, by Theorem 2.9 there is a choice of **b** for which there is no solution to $A\mathbf{x} = \mathbf{b}$.

33. The echelon form of A has a row of zeros, so there is a choice of **b** for which there is no solution to $A\mathbf{x} = \mathbf{b}$.

35. Example: b=[01]

37. Example: b=[001]

39. [01]

41. [01]

43. [001]

45. [001]

47. $h \neq 3$

49. $h \neq 4$

51. Example: $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (0, 1, 0)$, $\mathbf{u}_3 = (0, 0, 1)$, $\mathbf{u}_4 = (1, 1, 1)$

53. Example: $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (2, 0, 0)$, $\mathbf{u}_3 = (3, 0, 0)$, $\mathbf{u}_4 = (4, 0, 0)$

55. Example: $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (0, 1, 0)$

57. Example: $\mathbf{u}_1 = (1, -1, 0)$, $\mathbf{u}_2 = (1, 0, -1)$

59.

(a) True

(b) False

61.

(a) False

(b) True

63.

(a) False

(b) True

65.

(a) False

(b) True

67. (c) and (d) can possibly span $\mathbf{R}^3$.

69. HINT: Show that span{$\mathbf{u}$} ⊆ span{$\mathbf{cu}$} and that span{$c\mathbf{u}$} ⊆ span{$\mathbf{u}$}.

71. HINT: Let $S_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a subset of S_2 and show that every linear combination $c_1\mathbf{u}_1 + \ldots c_k\mathbf{u}_k$ is in span $\{S_2\}$.

73. PARTIAL HINT: Start with a linear combination $\mathbf{b} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3$. Show how to reorganize to write $\mathbf{b}$ as a linear combination of the set $\{\mathbf{u}_1 + \mathbf{u}_2, \mathbf{u}_1 + \mathbf{u}_3, \mathbf{u}_2 + \mathbf{u}_3\}$.

75. HINT: Generalize the argument given in Example 5.

77. True

79. False

Section 2.3

1. Linearly independent.

3. Linearly independent.

5. Linearly independent.

7. Linearly dependent.

9. Linearly independent.

11. Linearly independent.

13. System has only a trivial solution.

15. System has only a trivial solution.

17. System has only a trivial solution.

19. Linearly dependent.

21. Linearly dependent.

23. Linearly dependent.

25. Vectors are linearly independent; none in span of the others.

27. Vectors are linearly independent; none in span of the others.

29. System has a unique solution for all **b**.

31. System does not have a unique solution for all **b**.

33. System does not have a unique solution for all **b**.

35. System does not have a unique solution for all **b**.

37. $\mathbf{u} = (1, 0, 0, 0)$, $\mathbf{v} = (0, 1, 0, 0)$, $\mathbf{w} = (1, 1, 0, 0)$

39. $\mathbf{u} = (1, 0)$, $\mathbf{v} = (2, 0)$, $\mathbf{w} = (3, 0)$

41. $\mathbf{u} = (1, 0, 0)$, $\mathbf{v} = (0, 1, 0)$, $\mathbf{w} = (1, 1, 0)$

43.

(a) False

(b) False

45.

(a) False

(b) True

47.

(a) False

(b) False

49.

(a) False

(b) True

51.

(a) False

(b) True

53. (a), (b), and (c) can be linearly independent; (d) cannot.

55. HINT: Start by assuming that $\{c_1\mathbf{u}_1, c_2\mathbf{u}_2, c_3\mathbf{u}_3\}$ is linearly dependent, so the equation $x_1(c_1\mathbf{u}_1) + x_2(c_2\mathbf{u}_2) + x_3(c_3\mathbf{u}_3) = \mathbf{0}$

has a nontrivial solution. Show this implies that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is also linearly dependent, a contradiction.

57. HINT: Start by assuming that $\{\mathbf{u}_1 + \mathbf{u}_2, \mathbf{u}_1 + \mathbf{u}_3, \mathbf{u}_2 + \mathbf{u}_3\}$ is linearly dependent, so the equation $x_1(\mathbf{u}_1 + \mathbf{u}_2) + x_2(\mathbf{u}_1 + \mathbf{u}_3) + x_3(\mathbf{u}_2 + \mathbf{u}_3) = \mathbf{0}$ has a nontrivial solution. Show this implies that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is also linearly dependent, a contradiction.

59. HINT: Write the initial set of vectors as a nontrivial linear combination equal to **0**, and then show that this linear combination can be extended to the new larger set of vectors.

61. HINT: If $\mathbf{u} = c\mathbf{v}$, then $\mathbf{u} - c\mathbf{v} = \mathbf{0}$.

63. HINT: Modify the proof of part (a) of eorem 2.18.

65. Linearly independent.

67. Linearly independent.

69. Linearly dependent.

71. Unique solution for all **b**.

73. Does not have a unique solution for all **b**.

Supplementary Exercises

1. u+v=[-113],3w=[3-1521]

3. 2w+3v=[-4217],2u-5w=[-319-31]

5. 2u+v+3w=[3-1726],u-3v+2w=[9-2513]

7. x1-2x2=1-3x1+4x2=-52x1+x2=7

9. u+v=[-113],u-v=[3-71],v-u=[-37-1]

11. Yes, $\mathbf{w} = 3\mathbf{u} + \mathbf{v}$

13. **{u, v, w}** is not linearly independent.

15. x1[41]+x2[13-7]+x3[-14]=[-712]

17. x=[-100]+s1[231]

19. x=[30-10]+s1[-5081]+s2[-1100]

21. a=132,b=-52

23. **b** is not a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$.

25. A=[23-816-14-2],x=[x1x2x3x4],b=[59].

27. $\mathbf{b}$ is not in span$\{\mathbf{a}_1, \mathbf{a}_2\}$.

29. span$\{\mathbf{a}_1\} \neq \mathbf{R}^2$

31. span$\{\mathbf{a}_1, \mathbf{a}_2\} = \mathbf{R}^2$

33. span$\{\mathbf{a}_1\} \neq \mathbf{R}^3$

35. span$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} = \mathbf{R}^3$

37. $\{\mathbf{a}_1, \mathbf{a}_2\}$ is linearly independent.

39. $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is linearly independent.

41. $\{\mathbf{a}_1, \mathbf{a}_2\}$ is linearly dependent.

43. $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is linearly dependent.

Chapter 3

Section 3.1

1. T(u1)=[-102],T(u2)=[-4-33]

3. T(u1)=[-69],T(u2)=[1611]

5. **y** is in the range of *T*.

7. **y** is in the range of *T*.

9. T(-2u1+3u2)=[-134]

11. T(-u1+4u2-3u3)=[11-19]

13. Linear transformation, with A=[31-24].

15. Not a linear transformation.

17. Linear transformation, with A=[-401650].

19. Linear transformation, with A=[0sinπ4ln20].

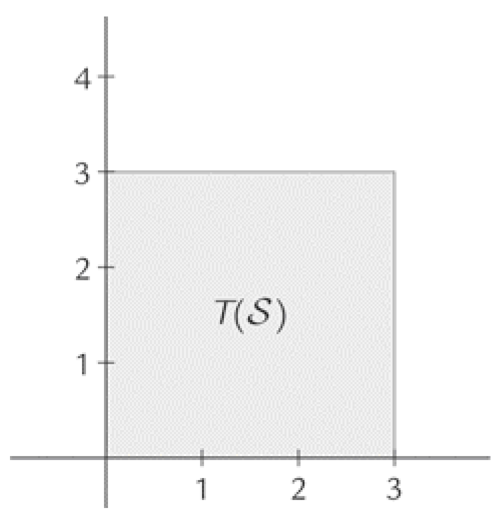
21. *T* is both one-to-one and onto.

23. *T* is not one-to-one but is onto.

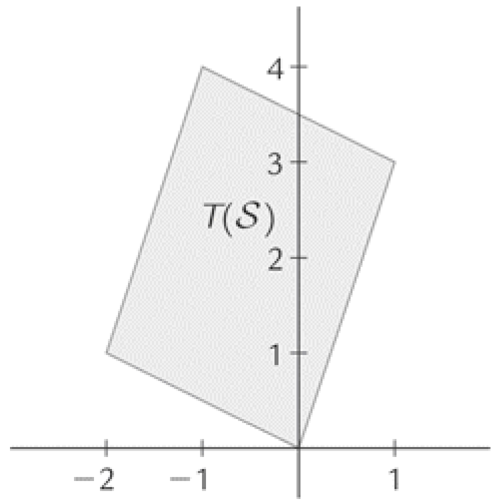
25. *T* is one-to-one but not onto.

27. *T* is neither one-to-one nor onto.

29.



31.



33. T(x)=[2030]x

35. T(x)=[7/3000]x

37. T(x)=[1-231]x

39.

(a) False

(b) True

41.

(a) True

(b) True

43.

(a) True

(b) True

45.

(a) False

(b) True

47.

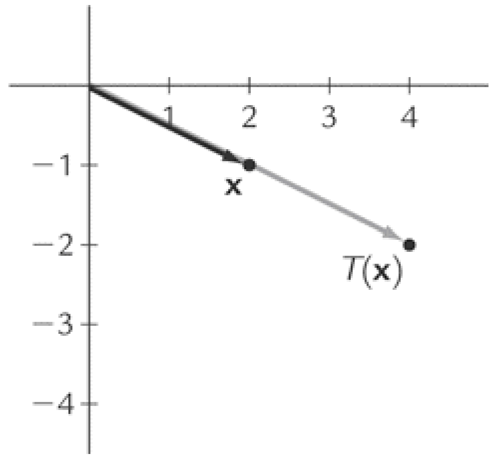
(a) False

(b) True

49.

(a) A=[r00r]

(b)



51. HINT: Show that $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ and $T(r\mathbf{x}) = rT(\mathbf{x})$.

53. HINT: Let $T(\mathbf{x}) = A\mathbf{x}$, where A is a 2×3 matrix. Explain why $A\mathbf{x} = \mathbf{0}$ must have a nontrivial solution.

55. HINT: Show that $T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$.

57. HINT: Use properties of matrix algebra.

59. HINT: Use the fact that T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution, and that $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$.

61. HINT: Start by assuming that $x_1\mathbf{u}_1 + x_2\mathbf{u}_2 = \mathbf{0}$ has a nontrivial solution, and arrive at a contradiction.

63. HINT: Use hint given with problem.

65. HINT: The unit square consists of all vectors $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$, where $\mathbf{u} = (1,0)$, $\mathbf{v} = (0,1)$, $0 \le s \le 1$, and $0 \le t \le 1$.

67.

(a) T(x)=Ax=[000200010]x

(b) T is neither one-to-one nor onto.

69.

(a) $T(x^2 + \sin(x)) = 2x + \cos(x)$

71. T([56])=[7231071349]

73. T([816])=[16242360776]

75. T is onto but not one-to-one.

77. T is neither one-to-one nor onto.

79. T is one-to-one but not onto.

Section 3.2

1.

(a) A+B=[-3504]

(b) AB+I2=[-1-724]

(c) $A + C$ is not possible.

3.

(a) (AB)T=[-22-73]

(b) CE is not defined.

(c) (A-B)D=[3-151216-30-6]

5.

(a) $(C + E)B$ is not possible.

(b) B(CT+D)=[-8368-224710]

(c) E+CD=[64-20-1121-4-317-6]

7. $a = -1$, $b = 1$, $c = -13$

9. $a = -1$, $b = 3$, $c = -3$, $d = 8$

11. $a = 2$

13.

(a) A=[-652417]

(b) A=[-2453-1035]

(c) A=[-150-2039]

(d) A=[427025]

15. $A^2 - I$

17. $ABA - A^2 + B^3A - B^2A$

19. The right side assumes that $AB = BA$, which is not true in general.
An example: A=[120-1],B=[2-110]
Then (A+B)2=[10222], but A2+2AB+B2=[12-400].

21. The right side assumes that $AB = BA$, which is not true in general.
An example: A=[120−1],B=[2-110]
Then A2-B2=[-22-22], but (A-B)(A+B)=[0-4-40].

23. AB is 4×5.

25. E=[400010001]

27. E=[010100001]

29. E=[100010201]

31. B=[100-210005]

33. B=[010130001]

35. B=[010-340001]

37.

(a) A−B=[−1−2021−1−13−130−3−2−133]

(b) AB=[145−6−1047−4−7−849−5−11−35]

(c) BA=[3−527−7112−44−19−1−1−8−212]

39.

(a) B−A=[120−2−111−31−30321−3−3]

(b) AB=[145−6−1047−4−7−849−5−11−35]

(c) BA+A=[4−7110−91130317−1−1−7013]

41.

(a) E=[010100001]

(b) E=[001010100]

(c) E=[1000-20001]

43. For example, A=[010000000],B=[100000000].

45. For example, A=[0100],B=[1000].

47. For example, A=[1111],B=[11-1-1].

49. For example, A=[1221],B=[2112],C=[1111].

51.

(a) False

(b) True

53.

(a) True

(b) False

55.

(a) False

(b) True

57.

(a) True
(b) False

59.

(a) True
(b) True

61.

(a) False
(b) True

67. HINT: Start with $(AB)^T$, and use $A^T = A$, $B^T = B$ because A, B are symmetric.

69.

(a) A^TA is $m \times m$.
(b) HINT: Show $(A^TA)^T = A^TA$.

71. HINT: Follow hint given in exercise.

73. HINT: A proof by induction works well for this one.

75.

(a) For example, A=[012-103-2-30].
(b) HINT: Look at your example for part (a).

79. After one year: [650022001300]; after two years: [537527601865]; after three years: ≈[453132082261]; after four years: ≈[389835662535].

81. Tomorrow: [742258]; the next day: ≈[734266]; the day after that : ≈[730270].

83.

(a) A+B=[-41-35-553261105132-1-2]
(b) BA−I4=[-26-154-31519-2814-1111-124-91324]
(c) $D + C$ is not possible.

85.

(a) AB=[2522-14-1-2310-121-6836-2134-31-3-120]
(b) CD=[142117742656022428262305276472 9]
(c) $(A - B)C^T$ is not possible.

87.

(a) $(C + A)B$ is not possible.

(b) C(CT+D)=[214041296112012484661461821067413 8123109]

(c) A+CD=[16201711426863214890633157734827]

Section 3.3

1. [1-3-27]

3. Inverse does not exist.

5. [9-4-21]

7. Inverse does not exist.

9. [1-2701-3000]

11. Inverse does not exist.

13. [0100000110000010]

15. [1-3-717012-4001-10001]

17. $x_1 = 35$ and $x_2 = -11$

19. x1=-154,x2=294, and x3=52

21. T-1(x)=[-2x1+3x23x1-4x2]

23. T^{-1} does not exist.

25. T^{-1} does not exist.

27.

(a) [21-10]

(b) [1-101]

(c) [0-112]

(d) [1101]

29. [10004−70−12]

31. [8−500−320000−34001−1]

33. [−830003−1000−32131−2223−901014−5001]

35. [100010001]

37. A=[1001],B=[3003]

39. A=[100010],B=[100100]

41. [0000010002100321]

43.

(a) False

(b) True

45.

(a) True

(b) True

47.

(a) True

(b) False

49.

(a) True

(b) True

51. $X = A^{-1}B$

53. $X = C^{-1}B - A$

55. A=[±100±1],A=[ab1-a2b-a](b≠0), and A=±[10c-1] (*c* can be any value).

57. HINT: Apply the Unifying Theorem.

59. $c \neq 0$ and $c \neq 1$

61. HINT: If A is $n \times n$ and not invertible, then the system $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution $\mathbf{x}_0$.

63. $B = C^{-1}AC$

65. HINT: Right-multiply by A^{-1}.

67. HINT: Since B is singular, there is a nontrivial solution to $B\mathbf{x} = \mathbf{0}$.

71. 9 j8's, 8 j8+'s, 13 j9's

73. 12 j8's, 21 j8+'s, 9 j9's

75. 3 Vigoro, 4 Parker's, 5 Bleyer's

77. 10 Vigoro, 14 Parker's, 11 Bleyer's

79. "laptop"

81. "final exam"

83. [8145-14145-23145429671456414543145-1029-2714511145-13145129-329-229529729]

85. Inverse does not exist.

Section 3.4

1. a = 2, b = −14

3. a = 4, b = 3, c = 2

5. x=[32]

7. x=[-102]

9. x=[21]

11. x=[21-10]

13. L=[10-21] and U=[1-401]

15. L=[100310-1-11] and U=[-2-11031001]

17. L=[1000-11002-31013-21],U=[-10-120310002-10001]

19. L=[100-410221] and U=[-121301-3-50012]

21. L=[1000110012100-1-11] and U=[1100-1-1002000]

23. L=[10000-110002-1100-11-1817102011701],U=[-21301110017000000]

25. L=[10-21],D=[2003],U=[1-101]

27. L=[1031],D=[100-2],U=[1-120112].

29. L=[100310-1-11],D=[-200030001],U=[112-120113001]

31. A-1=[76132313]

33. A-1=[171-2-780-1-4-6-2110101]

35. A-1=[3-53133-32144-212]

37. A=[100010001000]=LU=[1000010000100001] [100010001000]

39. A=[1001]=LU=[1001] [1001]

41. A=[1000010000100001]=LU=[1000010000100001] [1000010000100001]

43.

(a) False

(b) True

45.

(a) False

(b) False

47.

(a) False

(b) False

49. HINT: Apply properties of matrix multiplication.

53. This matrix does not have an LU factorization.

55. L=[10001210010103227-151],U=[1020-4200-14721000-500000-16]

Section 3.5

1. Stochastic

3. Stochastic

5. *a* = 0.35, *b* = 0.55

7. a=813,b=17,c=110

9. *a* = 0.7, *b* = 0.7

11. *a* = 0.5, *b* = 0.4, *c* = 0.5, *d* = 0.4

13. x3=[0.44320.5568]

15. x3=[2531675042196750]

17. x=[0.714290.28571]=[5727]

19. x=[0.398060.291260.31068]

21. Not regular.

23. Not regular.

25. A=[0.10.10.10.10.20.20.20.20.30.30.30.30.40.40.40.4]

27. A=[121120]

29. A=[100001010],x0=[010]

31.

(a) False

(b) False

33.

(a) False

(b) False

35.

(a) False

(b) True

37. HINT: Let Y=[11⋯1], show that $YA = Y$, and then show $Y(A\mathbf{x}) = 1$.

39. HINT: See exercise for hint.

43. HINT: Each column of A^{k+1} is a linear combination, with nonnegative scalars, of the columns of A^k.

45. HINT: (b) Compute A^2, then A^3, and then look for a pattern.

(a) Ak→[0011]

(b) x=[01]

47.

(a) A=[0.90.150.10.85]

(b) Probability that the sixth person in the chain hears the wrong news is 0.32881.

(c) x=[0.60.4]

49.

(a) A=[0.350.80.650.2]

(b)

(i) Probability that she will go to McDonald's two Sundays from now is 0.3575.

(ii) Probability that she will go to McDonald's three Sundays from now is 0.48913.

(c) Probability that his third fast-food experience will be at Krusty's is 0.521.

(d) x=[0.551720.44828]

51.

(a) Probability that a book is at C after two more circulations is 0.21.

(b) Probability that the book is at B after three more circulations is 0.64.

(c) x=[0.171050.631580.19737]

53. x9=x10=[0.2666670.3999990.1333330.200000]; the steady-state vector is x=[4152521515]

55. A [0100]=[0100] so [0100] has itself as its steady-state vector. Also,
A [0010]=[0010] so [0010] has itself as its steady-state vector.

Supplementary Exercises

1. T(u1)=[-712],T(u2)=[1-9]
3. T(u1)=[313],T(u2)=[15-9]
5. T(u1-u2)=[-4-5]
7. T(u1+u2-u3)=[-92]
9. Not one-to-one, not onto.
11. One-to-one but not onto.
13. B-A=[-1-2-11], *BC* is not defined, DE=[225-1112-98].
15. *DA* is not defined, E2=[1892-624-2116-3133],A3=[38-1185997].
17. A=[2123-1-27]
19. A=[07-62-76200]
21. E=[-500010001]
23. E=[100010701]
25. B=[1000-310000100005]
27. B=[000101000013-1000]
29. A-1=17[5-3-12]
31. A-1=[-232-740012-112-34]
33. No inverse.
35. A-1=[121-500-12120112-520-1-1272]
37. L=[10-31],U=[15023]
39. L=[100-110-221],U=[-23102400-5]
41. L=[10-41],D=[20019],U=[13201]
43. L=[100-210-3851],D=[-10005000-185],U=[1-200115001]

45. [5949]

47. [441223712241122]≈[0.36070.30330.3361]

49. No

51. No

Chapter 4

Section 4.1

1. This is a subspace, equal to span {[100],[001]}.

3. Not a subspace, because **0** is not in this set.

5. Not a subspace, because **0** is not in this set.

7. Not a subspace, because it is not closed under scalar multiplication.

9. Not a subspace, because it is not closed under addition.

11. Not a subspace, because it is not closed under scalar multiplication.

13. Not a subspace, because it is not closed under scalar multiplication.

15. A subspace, equal to null ([11⋯1]).

17. Not closed under scalar multiplication.

19. Not closed under addition.

21. null(A)={[00]}

23. null(A)=span{[5-21]}

25. null(A)=span{[431]}

27. null(A)={[00]}

29. null(A)={[000]}

31. null(A)=span{[3-110]}

33. **b** is not in ker(T); **c** is in range(T).

35. **b** is not in ker(T); **c** is not in range(T).

37. For example, S={[xy]:x>0}.

39. For example, S1={[x00]:x≥0} and S2={[x00]:x<0}.

41. Let $T(\mathbf{x}) = A\mathbf{x}$, where A=[1010].

43. Let $T(\mathbf{x}) = A\mathbf{x}$, where $A = I_3$.

45.

(a) True

(b) False

47.

(a) False

(b) False

49.

(a) True

(b) True

51.

(a) False

(b) True

53. HINT: If $x \neq 0$ is in a subspace S, show that every real number must be in S.

55. HINT: Determine if $\mathbf{0}$ is in the set of solutions.

57. The vector $\mathbf{0}$ alone, lines and planes through the origin, and all of $\mathbf{R}^3$.

59. HINT: Determine if $\mathbf{0}$ is in the set of solutions.

61. HINT: Show that $\mathbf{x} \neq \mathbf{0}$ and $A\mathbf{x} = \mathbf{0}$ if and only if the columns of A are linearly dependent.

63. HINT: Note that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$.

65. span{[122]}

67. span{[1212]}

69. span{[37-1375710],[-4356-556-395601]}

71. {[0000]}

Section 4.2

1. Not a basis, since $\mathbf{u}_1$ and $\mathbf{u}_2$ are not linearly independent.

3. Not a basis, because three vectors in a two-dimensional space must be linearly dependent.

5. Basis is {[1-4]}; dimension = 1.

7. Basis is {[13-3],[0-25]}; dimension = 2.

9. Basis is {[1-23-2],[02-51]}; dimension = 2.

11. Basis is {[13],[4-12]}; dimension = 2.

13. Basis is {[124],[01-3],[3-2-1]}; dimension = 3.

15. Basis is {[1-102],[2-597]}; dimension = 2.

17. Basis is {[2−6]}; dimension is 1.

19. Basis is {[111]}; dimension is 1.

21. Basis is {[300],[210],[123]}; dimension is 3.

23. One extension is {[1-3],[10]}.

25. One extension is {[-121],[100],[010]}.

27. One extension is {[13-2],[2-10],[100]}.

29. null(A)={[00]} This subspace has no basis, and nullity and nullity(A) = 0.

31. The null space has basis {[-7031],[-1100]} and nullity(A) = 2.

33. For example, {[10],[11],[01],[-11]}

35. For example, the span of the first m vectors of the n standard basis vectors of $\mathbf{R}^n$.

37. For example, S1=span{[1000],[0100]} and S2=span{[0010],[0001]}.

39. For example, u1=[100] and u2=[010].

41. span{[-1021],[1000],[0100]}

43.

(a) False

(b) True

45.

(a) False

(b) True

47.

(a) False

(b) True

49.

(a) False
(b) False

51.

(a) 1, 2, or 3
(b) 1 or 2

53. HINT: Use the Unifying eorem.

55. HINT: Show separately that the set is linearly independent and spans *S*.

57. HINT: A basis for S_1 can be expanded to a basis for S_2.

59. HINT: The entries below each pivot are equal to zero.

61. *n*

65. Subspace has basis {[2-15],[-34-2]}, with dimension 2. The vectors are not a basis for $\mathbf{R}^3$.

67. Subspace has basis {[301-2],[2-450],[-2704],[-25-54]}, with dimension 4. The vectors form a basis for $\mathbf{R}^4$.

69. Subspace has basis {[11-111],[-1012-1],[21-212]}, with dimension 3. Therefore, the vectors do not span $\mathbf{R}^5$.

Section 4.3

1. Column space basis: {[1-2-3],[-358]}
Row space basis: {[10-10],[01-4]}
Null space basis: {[1041]}
rank = 2, nullity = 1, *m* = 3

3. Column space basis: {[1-212],[2-342],[-144-4]}
Row space basis: {[1020],[01-10],[0001]}
Null space basis: {[-2110]}
rank = 3, nullity = 1, *m* = 4

5. Column space basis: {[12-1],[-2-2-2]}
Row space basis: {[1-22],[02-1]}
Null space basis: {[-1121]}
rank = 2, nullity = 1, *m* = 3

7. Column space basis: {[131],[3111],[274]}
Row space basis: {[1320],[0211],[0031]}
Null space basis: {[53-13-131]}
rank = 3, nullity = 1, m = 4

9. $x \neq 8$

11. $x = 18$

13. dim(col(A)) = 5

15. dim(row(A)) = 3, dim(col(A)) = 3, nullity(A) = 4.

17. rank(A) = 2

19. nullity(A) = 7

21. dim(range(T)) = 4

23. nullity(A) = 0

25. Maximum for rank(A) = 5, minimum for nullity(A) = 8.

27. rank(A) = 3

29. nullity(A) = 2

31. B has 3 nonzero rows.

33. A is 7 × 5.

35. For example, A=[100010].

37. For example, A=[I3×303×106×306×1].

39. For example, A=[100001000010].

41. For example, A=[1001].

43. For example, A=[100000000], so null(A)=span{[010],[001]}.

45. For example, [100010000], so col(A)=span{[100],[010]}.

47.

(a) True
(b) False

49.

(a) False
(b) False

51.

(a) False

(b) True

53. HINT: row(A) = col(A^T).

55. HINT: Apply the Rank–Nullity eorem.

57. HINT: First suppose $n < m$ and apply the Rank–Nullity eorem to A, then suppose that $m < n$ and apply the Rank–Nullity eorem to A^T.

59. rank(A) = 2, nullity(A) = 3.

61. rank(A) = 2, nullity(A) = 1.

Section 4.4

1. x=[1-7]
3. x=[02]
5. x=[112]
7. [x]$\mathcal{B}$2=[10-7]
9. [x]$\mathcal{B}$'=[21]
11. [x]$\mathcal{B}$=[4-1-4]
13. [-5-192]
15. [-19-416151-1462-5]
17. [2-1-11]
19. [-2-195]
21. [31-15-4917-7-25-6310]
23. [121-1]
25. [x]$\mathcal{B}$2=[-916]
27. [x]$\mathcal{B}$2=[-13]
29. [x]$\mathcal{B}$1=[-193-9939]
31. [ab]B1=[ba]B2
33. For example, $\mathcal{B}$={[-20],[013]}.
35. For example, $\mathcal{B}$1={[20],[0-12]} and $\mathcal{B}$2={[10],[01]}.
37. For example, $\mathcal{B}$1={[12],[37]} and $\mathcal{B}$2={[10],[01]}.
39.

(a) True
(b) True

41.

(a) True
(b) True

43. HINT: Write **u** and **v** in terms of the basis vectors of $\mathscr{B}$.

45. HINT: Focus on showing that the two properties required of a linear transformation both hold.

47. HINT: Explain why for each column $\mathbf{u}_i$ of U, the product V-1ui=[ui]ℬ2.

49. [6135485-316351778-5176235525-38235]

51. [-4953855269353836426937538211269393538340269-45269112269-129269-369269219538-416269305538-205269]

53. [122212-1-1-1]

Supplementary Exercises

1. Is a subspace: span {[010],[002]}.

3. Is a subspace: span {[110],[101],[011]}.

5. null(A) = {**0**}

7. null(A) = {**0**}

9. **b** is not in ker(T); **c** is not in range(T).

11. **b** is in ker(T); **c** is not in range(T).

13. Basis ={[-124],[312]} , dimension = 2.

15. Basis ={[21-4],[15-2]}, dimension = 2.

17. Basis ={[100],[0-10]}, dimension = 2.

19. null(A) = {**0**}, so there is no basis for the null space; nullity(A) = 0.

21. Basis ={[-2-11]}, nullity(A) = 1.

23. Basis for row(A)={[-15],[21]}Basis for col(A)={[-123],[511]} Basis for null(A): none, null(A) = {**0**}; rank(A) = 2, nullity(A) = 0, m = 2.

25. Basis for row(A)={[1-32],[1-23]}Basis for col(A)={[11],[-3-2]}Basis for null(A)={[-5-11]};
rank(A) = 2, nullity(A) = 1, m = 3.

27. Basis for ker(T)={[-27-3]}Basis for range(T)={[-123],[113]}Basis for row(A)={[-113],[037]}Basis for col(A)={[-123],[113]}Basis for null(A)={[-27-3]}

29. Basis for ker(T)={[-7513]}Basis for range(T)={[12],[4-5]}Basis for row(A)={[14-1],[2-53]}Basis for col(A)={[12],[4-5]}Basis for null(A)={[-7513]}

31. x=[-7-16]

33. x$\mathcal{B}$=[19-8]

35.

(a) True

(b) False

37.

(a) True

(b) True

39.

(a) True

(b) True

Chapter 5

Section 5.1

1. M23=[7051],M31=[0-462]
3. M23=[615711432],M31=[1-15230312]
5. M23=[4310324451032210],M31=[3210120510032410]
7. $C_{13} = 4$, $C_{22} = -10$
9. $C_{13} = 1$, $C_{22} = 4$
11. $|A| = 60$; T is invertible.
13. $|A| = 20$; T is invertible.
15. $|A| = 51$; T is invertible.
17. $|A| = 8$; T is invertible.
19. $|A| = 14$
21. $|A|$ is not defined.
23. $|A| = -82$
25. The shortcut method does not apply.
27. $a = 9$
29. $a = 0$
31. $a = 4$
33. $a = 1$ or $a = 3$
35. $|A| = -8$ (A upper triangular)
37. $|A| = 0$ (column of zeros)
39. $|A| = 0$ (two equal rows)
41. $|A| = |A^T| = 11$
43. $|A| = |A^T| = 28$
45. $\lambda = -2$ or $\lambda = 7$
47. $\lambda = 1$

49. $\lambda = -2$, $\lambda = 1$, or $\lambda = 3$

51. $\lambda = 2$

53.

(a) $|A| = 22$, determinant after row interchange = −22.

(b) $|A| = 1$, determinant after row interchange = −1.
Conjecture: Row interchanges change the sign of the determinant.

55.

(a) $|A| = -13$, determinant after row interchange = 13.

(b) $|A| = 3$, determinant after row interchange = −3.
Conjecture: Row interchanges change the sign of the determinant.

57.

(a) $|A| = 22$, determinant after multiplying row 1 by 3 is 66.

(b) $|A| = 1$, determinant after multiplying row 1 by 3 is 3.
Conjecture: Multiplying row 1 by 3 changes the determinant by a factor of 3.

59.

(a) $|A| = -13$, determinant after multiplying row 1 by 3 is −39.

(b) $|A| = 3$, determinant after multiplying row 1 by 3 is 9.
Conjecture: Multiplying row 1 by 3 changes the determinant by a factor of 3.

61. For example, A=[12001].

63. For example, A=[1411].

65. For example, A=[5-1πe0426-3].

67. For example, A=[π058100e1].

69.

(a) False

(b) False

71.

(a) False

(b) False

73. HINT: Show that the determinant gives a linear equation in and x and y, and then plug in (x_1, y_1) and (x_2, y_2) separately to show they satisfy the equation.

75. HINT: Show that the given expression is equal to the determinant of the matrix obtained by replacing row *j* of *A* with row *i*.

77. HINT: Cofactor expansion along row or column of zeros.

79. $|A| = -26$

81. $|A| = 1215$

Section 5.2

1. $|A| = 2$
3. $|A| = 0$
5. $|A| = 1$
7. $|A| = 4$; *A* is invertible.
9. $|A| = 21$; *A* is invertible.
11. $|A| = 0$; *A* is not invertible.
13. $|A| = 8$; *A* is invertible.
15. Determinant $= -3$
17. Determinant $= -6$
19. det(AB)=det(A)det(B)=(-11)(3)=-33det(A+B)=-2≠-11+3=det(A)+det(B)
21. det(AB)=det(A)det(B)=(1)(-30)=-30det(A+B)=-76≠1-30=det(A)+det(B)
23.
    (a) $|A^2| = 9$
    (b) $|A^4| = 81$
    (c) $|A^2A^T| = 27$
    (d) |A-1|=13
25.
    (a) $|A^2B^3| = -72$
    (b) |AB-1|=-32
    (c) $|B^3A^T| = -24$
    (d) $|A^2B^3B^T| = 144$

27. $|A| = 198$

29. $|A| = 4$

31. $|A| = 0$

33. |ABCD|=-3,|A||D|-|B||C|=-18

35. Unique solution exists.

37. Unique solution exists.

39. Unique solution exists.

41. A=[1224]

43. A=[1224],B=[-1-2-2-4]

45. For example, det([111122112])=1.

47.

(a) False

(b) True

49.

(a) False

(b) True

51.

(a) True

(b) True

53. HINT: Add the opposite of one of the identical rows to the other, and then apply cofactor expansion to the resulting matrix.

55. HINT: $|A| = |A^T|$.

57. HINT: Remove a factor of (-1) from each of the rows.

59. HINT: $|A^2| = |A|^2$.

61. HINT: See hint given with problem.

63. HINT: Explain why a matrix can be transformed to echelon form without multiplying a row times a constant.

65.

(a) HINT: E is diagonal, with a for one diagonal entry and 1's for the remaining diagonal entries.

(b) HINT: E is triangular, with 1's along the diagonal.

67. HINT: See hint given with problem.

69. $|I_4 + AB| = |I_3 + BA| = -45{,}780$

Section 5.3

1. x1=218,x2=34

3. x1=9,x2=-17,x3=1

5. x1=7949,x2=2249,x3=12449

7. x2=1123

9. x2=-2521

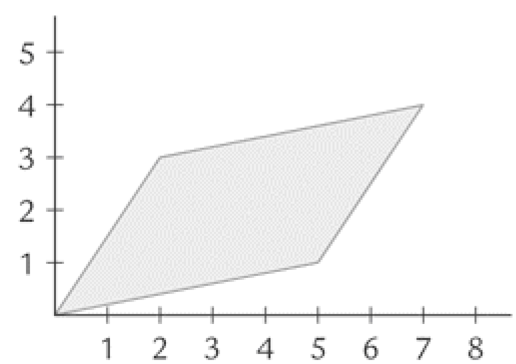
11. x2=1439

13. adj(A)=[7-5-32],A-1=[-753-2]

15. adj(A)=[001100010],A-1=[001100010]

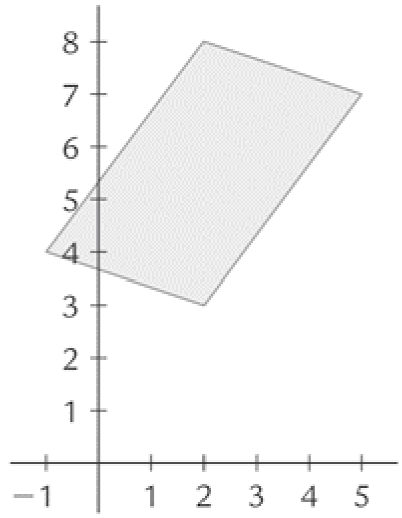
17. adj(A)=[1-2301-2001],A-1=[1-2301-2001]

19.



area = 13

21.



area = 15

23. area($T(\mathcal{D})$) = 165

25. area($T(\mathcal{D})$) = 54

27. area($T(\mathcal{D})$) = 54

29. T(x)=[5003] **x** is one possible solution.

31. T(x)=[32232-32232] **x** is one possible solution.

33. volume = 80π

35. volume = 82

37. For example,

x1+x2=12x1+2x2=2.

39. For example, let the parallelogram have vertices (0, 0), (5, 0), (5, 1), and (0, 1).

41. For example, A=[1325].

43.

(a) False

(b) True

45.

(a) False

(b) True

47. HINT: $|B| \neq 0$, so T is one-to-one. It remains to show that T is onto $\mathcal{R}$.

49. HINT: Use $|A|^{-1}\text{adj}(A) = A^{-1}$.

51. HINT: Show that the cofactor matrix of a symmetric matrix is also symmetric.

53. HINT: Consider the change in the cofactors when A is multiplied by c.

55. HINT: Start by replacing A with A^{-1} in $A = |A|^{-1}\text{adj}(A)$.

57. HINT: M_{ij} has a column (and row) of zeros when $i \neq j$.

59. x1=1221752,x2=811752,x3=13394

61. x1=704245,x2=-145,x3=247245,x4=-1749

63. adj(A)=[2753-15-72414059-2628],A-1=[2754753547-15547-725474154740547 59547-2654728547]

65. adj(A)=[-21-12660-15-36207-182161183-35-9711-7529-113],A-1=[-7141-144720141-5141-4472347-24724471184231141-35423-9742311423-2514129423-113423]

Supplementary Exercises

1. M21=[6-830],M13=[1443]

3. M21=[5-2-24-72118],M13=[-123742618]

5. det(A) = 20

7. det(A) = 110

9. a = 8

11. a=-32

13. λ=1±572

15. λ = 1, ±2

17. det(A) = −7

19. det(A) = 8

21. det(A) = −10

23. det(A) = 15

25. x1=-211,x2=-2711

27. x1=-53,x2=4,x3=1

29. adj(A)=[5-2-21],A-1=adj(A)

31. adj(A)=[24-7-2-4-3-12-42],A-1=-120adj(A)

33.

(a) False

(b) False

35.

(a) False

(b) False

37.

(a) False

(b) False

Chapter 6

Section 6.1

1. $\mathbf{x}_1$ is an eigenvector with associated eigenvalue $\lambda = -1$; $\mathbf{x}_2$ is not an is an eigenvector; $\mathbf{x}_3$ is an eigenvector with associated eigenvalue $\lambda = 4$.

3. $\mathbf{x}_1$ is an eigenvector with associated eigenvalue $\lambda = -1$; $\mathbf{x}_2$ is an eigenvector with associated eigenvalue $\lambda = 1$; $\mathbf{x}_3$ is an eigenvector with associated eigenvalue $\lambda = 2$.

5. $\mathbf{x}_1$ is an eigenvector with associated eigenvalue $\lambda = 3$; $\mathbf{x}_2$ is not an eigenvector; $\mathbf{x}_3$ is an eigenvector with associated eigenvalue $\lambda = 0$.

7. $\lambda = 3$ is not an eigenvalue of A.

9. $\lambda = -2$ is an eigenvalue of A.

11. A basis for the $\lambda = 4$ eigenspace is {[−11]}.

13. A basis for the $\lambda = 2$ eigenspace is {[52]}.

15. A basis for the $\lambda = 4$ eigenspace is {[131]}.

17. A basis for the $\lambda = 6$ eigenspace is {[111]}.

19. A basis for the $\lambda = -4$ eigenspace is {[−3−5−23]}.

21. $\det(A - \lambda I_2) = \lambda^2 + \lambda - 6$; basis for $\lambda = -3$ eigenspace is {[01]}; basis for $\lambda = 2$ eigenspace is {[54]}.

23. $\det(A - \lambda_2) = \lambda^2 + 2\lambda + 1$; basis for $\lambda = -1$ eigenspace is {[11]}.

25. $\det(A - \lambda I_3) = -(\lambda - 2)(\lambda - 3)(\lambda + 1)$; basis for $\lambda = 2$ eigenspace is {[035]}; basis for $\lambda = 3$ eigenspace is {[441]}; basis for $\lambda = -1$ eigenspace is {[001]}.

27. $\det(A - \lambda I_3) = -\lambda(\lambda - 1)(\lambda - 2)$; basis for $\lambda = 0$ eigenspace is {[1−13]}; basis for $\lambda = 1$ eigenspace is {[1−14]}; basis for $\lambda = 2$ eigenspace is {[2−15]}.

29. $\det(A - \lambda I_4) = (\lambda + 2)(\lambda + 1)(\lambda - 1)^2$; 3 basis for $\lambda = -2$ eigenspace is {[03−31]}; basis for $\lambda = -1$ eigenspace is {[420−3011]}; basis for $\lambda = 1$ eigenspace is {[0001]}.

31. For example, A=[1002].

33. For example, A=[1000−20003].

35. For example, A=[01−10]

37.

(a) False

(b) True

39.

(a) True

(b) False

41.

(a) False

(b) True

43. Characteristic polynomial = $\lambda^2 + 2\lambda + 17$

45. Characteristic polynomial = $-\lambda^3 + \lambda^2 + 19\lambda - 40$

47.

(a) A is 6×6.

(b) $\lambda = 3$, $\lambda = 2$, and $\lambda = -1$.

(c) A is invertible.

(d) The largest possible dimension of an eigenspace is 3.

49. HINT: Apply the Unifying Theorem.

51. HINT: Explain why $\det(A - I_n) = 0$.

53. HINT: What is $A\mathbf{u}$ if $\mathbf{u}$ is associated with two distinct eigenvalues?

55. HINT: Which values of λ would *not* be eigenvalues?

57. HINT: Show that $A^{-1}\mathbf{u} = \lambda^{-1}\mathbf{u}$.

59. HINT: Suppose that $\boldsymbol{\lambda}_1$ is the eigenvalue of A associated with $\mathbf{u}$ and λ_2 is the eigenvalue of B associated with $\mathbf{u}$. Determine $AB\mathbf{u}$.

61. HINT: What is $A\mathbf{u}$ when $\mathbf{u} = (1, 1, \dots, 1)$?

63. HINT: Note that $\det(A - \lambda I_n) = \det((A - \lambda I_n)^T)$.

65. HINT: What is $A\mathbf{u}$ when $\mathbf{u} = (1, 1, \dots, 1)$?

67. Basis for $\lambda = 1$ eigenspace is {[0100],[10−11]}; basis for $\lambda = 2$ eigenspace is {[−1510],[−1902]}.

69. Basis for $\lambda = 0$ eigenspace is {[0−1011]}; basis for $\lambda = 1$ eigenspace is {[−103−11]}; basis for $\lambda = 2$ eigenspace is {[−1−22−11]};basis for $\lambda = -2$ eigenspace is {[−1−33−21]}; basis for $\lambda = -1$ eigenspace is {[01000]}.

Section 6.2

1. A5=[131−39633−100]
3. A5=[1−93−1840326600−1]
5. A=[19−1230−19]
7. A=[2−340−12−11−1]
9. A is not diagonalizable.
11. P=[1211],D=[−1003]
13. P=[111−2−1−2322],D=[00001000−1]
15. P=[−11−1110101],D=[000010001]
17. P=[1001010000100003],D=[1000020000300004]
19. A1000=[1001]
21. A1000=[2(31000)−12−2(31000)31000−12−31000]
23. dimension = 2
25. For example, A=[0001].
27. For example, A=[1101].
29. For example, A=[010010002] has eigenvalues 0, 1, and 2.
31.
    (a) True
    (b) True
33.

(a) False
(b) False

35. HINT: $\mathbf{u}_1$ and $\mathbf{u}_2$ must be linearly independent.

37. HINT: Each eigenvalue has infinitely many distinct associated eigenvectors.

39. HINT: What is A^T if $A = PDP^{-1}$?

41. HINT: Let $A = PD_1P^{-1}$ and $B = PD_2P^{-1}$, and then show that $AB = BA$.

43. P=[−1−2−102211−3−6−202620],D=[−1000000000100002]

45. P=[0022440−2−2−84144840−6−4−800424],D=[−200000−1000000000001000002]

Section 6.3

1.

(a) $8 + 2i$
(b) $19 + 8i$
(c) $2 - 6i$
(d) $23 + 14i$

3. $\lambda_1 = 2 + 2i$; eigenspace basis ={[1+2i −5]}
$\lambda_2 = 2 - 2i$; eigenspace basis ={[1−2i −5]}

5. $\boldsymbol{\lambda}_1 = 2 + i$; eigenspace basis ={[1−i −1]}
$\boldsymbol{\lambda}_2 = 2 - i$; eigenspace basis ={[1+i −1]}

7. $\boldsymbol{\lambda}_1 = 3 + i$; eigenspace basis ={[1+i −1]}
$\boldsymbol{\lambda}_2 = 3 - i$; eigenspace basis ={[1−i −1]}

9. Rotation is by $\tan^{-1}(1/2) \approx 0.4636$ radians; the dilation is by 5.

11. Rotation is by tan−1(1)=π4 radians; the dilation is by 2.

13. Rotation is by tan−1(−3/4)≈−0.6435 radians; the dilation is by 5.

15. The rotation–dilation matrix is B=[2−222].

17. The rotation–dilation matrix is B=[2−112].

19. The rotation–dilation matrix is B=[3−113].

21. Other roots are $1 - 2i$ and $3 + i$, and the multiplicity of each root is 1.

23. For example, z=655+355i.

25. B=[0−220]

27. For example,

A=[1101] [1−221] [1101]−1=[3−42−1].

29. For example, A=[ii−i−i].

31.

(a) True

(b) False

33.

(a) False

(b) True

35.

(a) True

(b) True

37. HINT: Start with $z = x + iy$ and $w = u + iv$, and then apply the properties of complex conjugation.

39. HINT: Apply Exercise 37(b).

41. HINT: Start with $\lambda = x + iy$, and then apply the properties of complex conjugation.

43. HINT: $|A - \lambda I| = (a - \lambda)^2 + b^2$.

45.

(a) HINT: Write $\mathbf{u} = \mathrm{Re}(\mathbf{u}) + i\mathrm{Im}(\mathbf{u})$.

(b) HINT: Use hint given with this part of problem.

(c) HINT: Show that the real and imaginary parts of AP and PC are the same.

47. λ1,2=−2.507±1.692i⇒{[0.2373±0.3607i0.0862∓0.2878i −0.850 5]}λ3=6.013⇒{[0.58370.68890.4298]}

49. λ1,2=2.5948±0.4119i⇒{[0.66380.1906±0.2156i0.0472∓0.2804i −0.6280∓0.0368i]}λ3=13.2693⇒{[0.06390.14410.35040.9232]} λ4=−5.4589⇒{[0.6812−0.3472−0.54760.3399]}

Section 6.4

1. y1=c1e−t+c2e2ty2=c1e−t−c2e2t
3. y1=4c1e2t+c2e−2t+2c3e−2ty2=3c1e2t+2c2e−2t+3c3e−2ty3=c1e2t+c3e−2t
5. y1=c1(cos 2t−sin 2t)+c2(cos 2t+sin 2t)y2=c1(2cos 2t+sin 2t)−c2(cos 2t−2sin 2t)
7. y1=3c1e4t−c2(sin t−4cos t)et+c3(cos t+4sin t)ety2=c1e4t−2c3(cos t)et+2c2(sin t)ety3=5c1e4t−c2(sin t−3cos t)et+c3(cos t+3sin t)et
9. y1=6c1et+c2e4t+c3(3cos t−2sin t)et+c4(2cos t+3sin t)ety2=2c1et+2c2e4t+6c3(cos t)et+6c4(sin t)ety3=5c1et+3c2e4t+c3(2cos t+3sin t)et−c4(3cos t−2sin t)ety4=2c2e4t+5c3(sin t)et−5c4(cos t)et
11. y1=2c1e−t+2c2e3ty2=−c1e−t+c2e3t
13. y1=c1e−t+2c2e3ty2=c1e−t+c2e3t
15. y1=−c1(cos 2t−2sin 2t)e2t−c2(2cos 2t+sin 2t)e2ty2=5c1(cos 2t)e2t+5c2(sin 2t)e2t
17. y1=2c1+c2e−t−c3ety2=c1+3c3ety3=2c1+c2e−t
19. y1=−2e−2t+6e3ty2=−2e−2t+3e3t
21. y1=−(sin 3t)et+2(cos 3t)ety2=−(cos 3t)et−2(sin 3t)et
23. y1=2et−e2t−2e−ty2=−2et+2e2ty3=4et−2e2t−6e−t
25. y1=c1e(−0.3)t+2c2e(−0.4)ty2=−c1e(−0.3)t−3c2e(−0.4)t
27. y1=70e−0.3t−60e−0.4ty2=90e−0.4t−70e−0.3t
29. y1=−c1e−8t+5c2et,and y2=c1e−8t+4c2et. As t gets large, y1≈5c2et and y2≈4c2et, and thus the ratio y1/y2≈5/4.
31. y1=53et−23e−8t,y2=43et+23e−8t
33. For example, y1′=−3y1 and y2′=2y2.
35. For example,

    y1′=10y1+6y2y2′=−18y1−11y2.

37. For example,

    y1′=−6y1+4y2+7y3y2′=−7y1+5y2+7y3y3′=−4y1+4y2+5y3.

39.

(a) True

(b) False

41. y1≈−0.7811c1e7.065t−0.7041c2(cos 2.580t)e(−3.003t) −0.701 c3(sin 2.580t)e(−3.033t)y2≈0.4471c1e7.065t−c2(0.1528cos (2. 580t)+0.4597sin (2.580t)e(−3.033t)) −c3(0.1528sin (2.580t) −0.4597cos (2.580t))e(−3.033t)y3≈−0.4359c1e7.065t+c2(0.514 1cos (2.580t) + 0.0729sin (2.580t))e(−3.033t)+c3(0.5141sin (2.580t)) − 0.0729cos (2.580t))e(−3.033t)

43. y1≈−0.8167c1e−4.114t+1.139c2e7.297t+c3(0.2576sin (4.698t) − 0.5848(cos 4.698t))e1.408t−c4(0.2576cos (4.698t)) + 0.584 8sin (4.698t)e1.408ty2≈− 0.8101c1e−4.114t+0.1064c2e7.297t− c3(2.336sin (4.698t) −2.858cos (4.698t))e1.408t+c4(2.33cos (4.698t) + 2.858sin (4.698t)e1.408ty3≈0.5896c1e−4.114t+0.39 c2e7.297t−c3e1.408t(2.385cos (4.698t) +1.314sin (4.698t)) −c4e1.408t(2.385sin (4.698t) −1.314cos (4.698t)y4≈c1e−4.114 t+c2e7.297t+c3cos (4.698t)e1.408t+c4sin (4.698t)e1.408t

45. y1≈0.5746e8.01t−0.3412e1.106t−1.233e−7.115ty2≈−0.03121e 8.01t+0.1231e1.106t−4.092e−7.115ty3≈0.7118e8.01t+0.1924e 1.106t+2.096e−7.115t

47. y1≈11.63(cos 2.153t)e−3.179t−0.4729e12.53t−4.158e−0.1732t +7.978(sin 2.153t)e−3.179ty2≈8.266(cos 2.153t)e−3.179t−0.3 908e12.53t−5.876e−0.1732t + 6.113(sin 2.153t)e−3.179ty3≈2. 928e−0.1732t−0.1355e12.53t−4.792(cos 2.153t)e−3.179t + 3. 693(sin 2.153t)e−3.179ty4≈4.349e−0.1732t−1.249e12.53t−8.1 01(cos 2.153t)e−3.179t −7.601(sin 2.153t)e−3.179t

Section 6.5

1. x1=[11], x2=[−26], x3=[−2028]

3. x1=[644], x2=[524848], x3=[504496496]

5. x1=[30−3], x2=[90−9], x3=[270−27]

7. x1=[−11], x2=[1.00−0.33]

9. x1=[0.00−0.501.00], x2=[1.00−0.251.00]

11. x1=[0.001.00−0.67], x2=[0.001.00−0.56]

13. The Power Method will converge, with eigenvalue $\lambda = 7$.

15. The Power Method will converge, with eigenvalue $\lambda = -6$.

17. The Power Method will converge, with eigenvalue $\lambda = 6$.

19. B=[−323−2]

21. B=[−102−7−10−72−102−7]

23. B=[−715−2]

25. B=[414169262]

27. λ=11/4=4; the eigenvector is [1120].

29. For example, A=[1000] and x0=[10].

31. For example, A=[0110] and x0=[10].

33. For example, A=[0−11−1] and x0=[10].

35.

(a) False

(b) False

37.

(a) True

(b) False

39. x1=[1−1], x2=[01], x3=[1−1], x4=[01].
The sequence $\mathbf{x}_k$ does not converge; it alternates. The eigenvalues of A are $\lambda = 1$ and $\lambda = -1$, so there is no dominant eigenvalue and convergence is not assured.

41. x1=[−121], x2=[−121], … , and the sequence converges to the eigenvalue $\lambda = 1$ because x0=[−12] is an eigenvector associated with $\lambda = 1$.

43. x1=[11], x2=[−26], x3=[−2028], x4=[−104120], x5=[−464496], x6=[−19522016]

45. x1=[644], x2=[524848], x3=[504496496], x4=[500849924992], x5=[50,01649,98449,984], x6=[500,032499,968499,968]

47. x1=[30−3], x2=[90−9], x3=[270−27], x4=[810−81], x5=[2430−243], x6=[7290−729].

49. $\lambda = 2$; eigenvector =[−10].

51. $\lambda = 4.2458$; eigenvector =[−0.05791.0000−0.6518].

53. $\lambda = 3$; eigenvector =[01−0.5].

Supplementary Exercises

1. $\lambda = 2$, basis = {[12]}

3. $\lambda = -1$, basis = {[1−1−1]}

5. $\lambda = 1$, basis = {[11]}; $\lambda = 2$, basis = {[12]}

7. $\lambda = 1$, basis = {[−111]}; $\lambda = -1$, basis = {[121]}; $\lambda = 0$, basis = {[011]}

9. x1=[11], x2=[32], x3=[75]

11. x1=[1−10], x2=[1−30], x3=[1−70]

13. x1=[1.00000.7273], x2=[1.00000.6400], x3=[1.00000.6747]

15. x1=[1.0000−0.6667−0.6667], x2=[1.0000−0.8889−0.8889], x3=[1.0000−0.9630−0.9630]

17. D=[2001],P=[5231]

19. D=[1000−10000],P=[214102111]

21. A=[−476797117]

23.

(a) $7 + i$

(b) $17 + 4i$

(c) $-2 - 5i$

(d) $14 + 5i$

25. $\lambda_1 = 1 + 2i$, basis = {[i2]}, $\lambda_2 = 1 - 2i$, basis = {[−i2]}

27. Rotation = $\tan^{-1}(1/2) \approx 0.464$ radians; dilation = 5.

29. y1=c1e−t−3c2e2ty2=−c1e−t+c2e2t

31. y1=(−2c1sin (t)+2c2cos (t))ety2=(c1sin (t)−c2cos (t))et

33. y1=0.931c1e−2.646t+0.617c2e2.646ty2=0.365c1e−2.646t+0.787c2e2.646t

35. y1=2c1et+c2e−t+4c3y2=c1et+2c3y3=c1et+c2e−t+c3

Chapter 7

Section 7.1

1. HINT: The required properties follow from the same properties of the real numbers.

3. HINT: You may assume that the sum of two continuous functions is a continuous function, as is the scalar multiple of a continuous function.

5. HINT: Adding two polynomials cannot produce a polynomial of degree greater than that of those being added. The scalar multiple of a polynomial produces a new polynomial that has the same degree or is equal to zero.

7. HINT: The hint from Exercise 3 applies here.

9. V is not a vector space under the given arithmetic operations. For instance, there is no vector $\mathbf{0}$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v}$.

11. V is not a vector space. Property 5(d) does not always hold. For instance, (1+0)[10]=[10] but(1)[10]+(0)[10]=[00]. There is also no additive identity.

13. HINT: Show that the three requirements for a subspace are met.

15. HINT: Show that the three requirements for a subspace are met.

17. HINT: Show that the three requirements for a subspace are met.

19. S is a subspace.

21. S is a subspace.

23. S is not a subspace. S is not closed under addition.

25. S is a subspace.

27. S is not a subspace. The zero vector is not in S.

29. S is a subspace.

31. S is not a subspace.

33. For example, the set of vectors in the first quadrant of $\mathbf{R}^2$, with the usual definition of addition and scalar multiplication.

35. For example, the set of vectors in the first quadrant of $\mathbf{R}^2$, with the usual definition of addition and scalar multiplication.

37. Aside from $V_1 = \mathbf{R}^n$ with the usual definition of addition and scalar multiplication, we can also have $V_2 = \mathbf{R}^n$, but we let $\mathbf{w}$ be a fixed vector and then define addition by $\mathbf{u} \oplus \mathbf{v} = \mathbf{u} + \mathbf{v} - \mathbf{w}$ and scalar multiplication by $c \odot \mathbf{u} = c(\mathbf{u} - \mathbf{w}) + \mathbf{w}$. In this case, $\mathbf{w}$ is the zero vector for V_2.

39.

(a) True

(b) False

41.

(a) True

(b) False

43. HINT: Construct $\mathbf{0}$ by using $\mathbf{u}$ in S (S is nonempty) and observing that there must be a corresponding $-\mathbf{u}$ in S.

45. HINT:

(a) Use the fact that addition of vectors is commutative.

(b) Assume that there are two zero vectors $\mathbf{0}_a$ and $\mathbf{0}_b$, and then show that $\mathbf{0}_a = \mathbf{0}_b$.

(c) Use $\mathbf{v} + 0 \cdot \mathbf{v} = (1 + 0)\mathbf{v} = \mathbf{v}$.

(d) Use $\mathbf{v} + (-1)\mathbf{v} = (1 + (-1))\mathbf{v} = 0 \cdot \mathbf{v}$ together with (c).

Section 7.2

1. $\mathbf{v}$ is in span $\{3x^2 + x - 1, x^2 - 3x + 2\}$.

3. $\mathbf{v}$ is in span$\{3x^2 + x - 1, x^2 - 3x + 2\}$.

5. $\mathbf{v} = x^3 + 2x^2 - 3x$ is not in span$\{x^3 + x - 2, x^2 + 2x + 1, x^3 - x^2 + x\}$.

7. $\mathbf{v} = x^2 + 4x + 4$ is in span$\{x^3 + x - 2, x^2 + 2x + 1, x^3 - x^2 + x\}$.

9. **v** is in span {[121013],[031−110]}.

11. **v** is in span {[121013],[031−110]}.

13. **v** is in span {[−1341],[025−3],[1421]}.

15. **v** is in span {[−1341],[025−3],[1421]}.

17. $\{x^2 - 3, 3x^2 + 1\}$ is linearly independent in $\mathbf{P}^2$.

19. $\{x^3 + 2x + 4, x^2 - x - 1, x^3 + 2x^2 + 2\}$ is not linearly independent in $\mathbf{P}^3$.

21. {[2−113],[−42−2−6]} is not linearly independent in $\mathbf{R}^{2\times 2}$.

23. {[101214],[312033]} is linearly independent in $\mathbf{R}^{2\times 3}$.

25. $\{\sin^2(x), \cos^2(x), 1\}$ is not linearly independent in $C[0, \pi]$.

27. For example,

{[1000],[0100],[0010],[0001],[1111]}

spans $\mathbf{R}^{2\times 2}$, but is not linearly independent.

29. For example, let $V = \mathbf{P}$. Then $\{1, x, x^2, x^3, \dots\}$ is an infinite linearly independent subset.

31. Let and $\mathcal{V}_1$ = {(1, 0, 0, …), (0, 0, 1, 0, 0 …), (0, 0, 0, 0, 1, 0, 0, …)} and $\mathcal{V}_2$ = {(0, 1, 0, 0, …), (0, 0, 0, 1, 0, 0 …), (0, 0, 0, 0, 0, 1, 0, 0, …)}. Then $\mathcal{V}_1$ and $\mathcal{V}_2$ are infinite linearly independent subsets of $\mathbf{R}^\infty$, and $\text{span}(\mathcal{V}_1) \cap \text{span}(\mathcal{V}_2) = \{\mathbf{0}\}$.

33.

(a) False

(b) False

35.

(a) False

(b) False

37.

(a) False

(b) False

39. HINT: Show that each polynomial is a linear combination of the given set.

41. HINT: See hint given with problem.

43. HINT: Consider cases $\mathbf{v}_1 = \mathbf{0}$ and $\mathbf{v}_1 \neq \mathbf{0}$ separately.

45. HINT: Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ spans $\mathbf{R}^\infty$. Truncate each vector to the first $m + 1$ components. Then the new vectors must also span $\mathbf{R}^{m+1}$, but cannot.

47. HINT: Apply Theorem 7.9(a).

49. HINT: $\mathbf{v}$ is a linear combination of $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$.

51. $\{x, \sin(\pi x/2), e^x\}$ is linearly independent.

53. $\{e^x, \cos^2(x), \cos(2x), 1\}$ is linearly dependent, so method shown in Example 9 will not work.

Section 7.3

1. $\mathcal{V}$ has too few vectors to be a basis for $\mathbf{P}^2$.

3. $\mathcal{V}$ could be a basis for $\mathbf{R}^{2\times 2}$, since $\dim(\mathbf{R}^{2\times 2}) = 4$ and $\mathcal{V}$ has 4 vectors.

5. $\mathcal{V}$ has too few vectors to be a basis for $\mathbf{P}^4$.

7. $\mathcal{V}$ is a basis.

9. $\mathcal{V}$ is not a basis.

11. $\mathcal{V}$ is not a basis.

13. $\dim(S) = 8$, and a basis for S is

$$[-100010000],[-100000001],[010000000],[001000000],[000100000],$$
$$[000001000],[000000100].[000000010].$$

15. $\dim(S) = 3$, and a basis for S is

$$\{[-1001],[0-101],[00-11]\}.$$

17. HINT: S is equivalent to the set of 2×2 matrices A such that $A\mathbf{v} = \mathbf{0}$.

19. $\dim(S) = \infty$. For example, $\{x(x-1)(x-2), x^2(x-1)(x-2), x^3(x-1)(x-2), \ldots\}$ is an infinite set of linearly independent vectors in $C(\mathbf{R})$, each of which vanishes at $k = 0, 1, 2$.

21. We extend $\mathcal{V}$ to $\{2x^2 + 1, 4x - 3, 1\}$ to obtain a basis for $\mathbf{P}^2$.

23. We extend $\mathcal{V}$ to

{[1001],[0110],[1110],[0100]}

to obtain a basis for $\mathbf{R}^{2\times2}$.

25. We reduce the set $\mathcal{V}$ to $\{x + 1, x + 2\}$ to obtain a basis for $\mathbf{P}^1$.

27. HINT: Show that $\{\cos(t), \sin(t)\}$ is a basis for S.

29. A basis for S is the set $\{1, x\}$.

31. For example, let $V = \mathbf{P}$, and $S = \text{span}\{1\}$.

33. For example, let $V = \mathbf{P}^1$, and $S = \text{span}\{1\}$.

35. For example, let $V = \mathbf{P}$, and $S = \text{span}\{1, x^2, x^4, x^6, \ldots\}$.

37.

(a) False

(b) True

39.

(a) False

(b) True

41.

(a) False

(b) True

43. HINT: It is enough to show that $\{\mathbf{v}_1, 2\mathbf{v}_2, \ldots, k\mathbf{v}_k\}$ is linearly independent.

45. HINT: Show that the set is linearly independent and spans $\mathbf{R}^{2\times2}$.

47. HINT: Start with a basis for V, and remove one vector at a time to obtain a basis for each of $S_{\mathrm{m}-1}, S_{m-2}, \ldots$.

49. HINT: See hint given with problem.

51. HINT: For part (a), show that a basis for V_1 must also be a basis for V_2.

53. HINT: See proof of corresponding theorem in Section 4.2.

Supplementary Exercises

1. HINT: You may assume that the sum of two continuous functions is a continuous function, as is the scalar multiple of a continuous function.

3. HINT: $\mathbf{E}^n$ is a subset of $\mathbf{P}^n$, so most of the required properties are inherited from $\mathbf{P}^n$ and need not be proved again.

5. HINT: Show that the three requirements for a subspace are met.

7. HINT: Show that the three requirements for a subspace are met.

9. $\mathbf{v}$ is not in S.

11. $\mathbf{v}$ is not in S.

13. Set is linearly independent.

15. Set is linearly independent.

17. Not a basis, because too few vectors.

19. Not a basis, because too few vectors.

21. Not a basis, because $\mathcal{V}$ is linearly dependent.

23. Not a basis, because $\mathcal{V}$ is linearly dependent.

25. Dimension = 4.

27. Dimension = 2.

Chapter 8

Section 8.1

1.

(a) $\mathbf{u}_1 \cdot \mathbf{u}_5 = -3$

(b) $\mathbf{u}_3 \cdot (-3\mathbf{u}_2) = -3$

(c) $\mathbf{u}_4 \cdot \mathbf{u}_7 = 11$

(d) $2\mathbf{u}_4 \cdot \mathbf{u}_7 = 22$

3.

(a) ‖u7‖=29

(b) ‖−u7‖=29

(c) ‖2u5‖=26

(d) ‖−3u5‖=36

5.

(a) ‖u1−u2‖=17

(b) ‖u3−u8‖=26

(c) ‖2u6−(−u3)‖=7

(d) ‖−3u2−2u5‖=311

7.

(a) $\mathbf{u}_1 \cdot \mathbf{u}_3 = -8 \neq 0$, so $\mathbf{u}_1$ and $\mathbf{u}_3$ are not orthogonal.

(b) $\mathbf{u}_3 \cdot \mathbf{u}_4 = 0$, so $\mathbf{u}_3$ and $\mathbf{u}_4$ are orthogonal.

(c) $\mathbf{u}_2 \cdot \mathbf{u}_5 = 4 \neq 0$, so $\mathbf{u}_2$ and $\mathbf{u}_5$ are not orthogonal.

(d) $\mathbf{u}_1 \cdot \mathbf{u}_8 = 8 \neq 0$, so $\mathbf{u}_1$ and $\mathbf{u}_8$ are not orthogonal.

9. a=32

11. a=283

13. Set is not orthogonal.

15. Set is not orthogonal.

17. $a = -10$

19. $a = 7$ and $b = 11$

21. ‖u1‖2=10,‖u2‖2=10,‖u1+u2‖2=20

23. ‖u1‖2=14,‖u2‖2=26,‖u1+u2‖2=40

25. ‖3u1+4u2‖=2109

27. Cauchy–Schwarz: |**u** · **v**| = 6, ‖ u‖‖ v‖=5·20=10; Triangle inequality: ‖u + v‖=37≈6.083, ‖ u‖+‖ v‖=5+20≈6.708.

29. Cauchy–Schwarz: ‖u · v‖=0,‖ u‖‖v‖=12.26≈17.664; Triangle inequality: ‖u + v‖=38≈6.164, ‖ u‖+‖ v‖=12+26≈8.563.

31. **u** is not orthogonal to *S*.

33. A basis for $S^{\perp}$ is {[31]}.

35. A basis for $S^{\perp}$ is {[−110],[201]}.

37. Let s1=[110] and s2=[1−14]. Then s=32s1−12s2.

39. For example, u=[120] and v=[10].

41. For example, u=[1525]

43. For example, [010] and [102].

45. For example, S=span{[100]}.

47. For example, {[000].[100],[010]}.

49.

(a) False

(b) True

51.

(a) True

(b) False

53.

(a) True

(b) True

55. HINT: Every vector **s** in *S* is a linear combination of a spanning set *S*.

57. HINT: Show that $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ whenever $i \neq j$.

59. HINT: Apply Theorem 8.2(c) twice.

61. HINT: Use the properties of Theorem 8.2.

63. HINT: Apply equation (2) that follows Definition 8.3.

65. HINT: Suppose that **v** is in both S and $S^\perp$. Use this to show that $\mathbf{v} \cdot \mathbf{v} = 0$.

67. HINT: If in $\mathbf{R}^n$ is a column of A and $\mathbf{x} = (x_1, \ldots, x_n)$, then $\mathbf{a}^T\mathbf{x} = \mathbf{a} \cdot \mathbf{x}$.

69. HINT:

    (a) Compare definitions of $\mathbf{u} \cdot \mathbf{v}$ and $\mathbf{u}^T\mathbf{v}$.

    (b) Start with $(A\mathbf{u}) \cdot \mathbf{v} = (A\mathbf{u})^T\mathbf{v}$.

71.

    (a) $\mathbf{u}_2 \cdot \mathbf{u}_3 = -4$

    (b) ‖u1‖=39

    (c) ‖2u1+5u3‖=826

    (d) ‖3u1−4u2−u3‖=1879

73. {[−5−710],[−92−401]}

Section 8.2

1.

    (a) proju3u2=[250−15]

    (b) proju1u2=[000]

3. projsu2=[250−15]

5.

    (a) 1‖u1‖u1=[−31414114141714]

    (b) 1‖u4‖u4=k[11414−314141714]

7. An orthogonal basis for S is {[13],[3−1]}.

9. An orthogonal basis for S is {[−221],[34−2]}.

11. An orthogonal basis for S is {[1−101],[322−1]}.

13. An orthogonal basis for S is {[−101],[242],[3−33]}.

15. projsu=[11]

17. projsu=[−329−429229]

19. projsu=[1−101]

21. projsu=[102]

23. An orthonormal basis for S is {[1101031010],[31010−11010]}.

25. An orthonormal basis for S is {[−232313],
[3292942929−22929]}.

27. An orthonormal basis for S is{[133−1330133],
[122132132−162]}.

29. An orthonormal basis for S is {[−1220122],[166136166],
[133−133133]}.

31. For example, let u=[10] and v=[10].

33. For example, let u=[10] and v=[01].

35. For example, let u=[31] and v=[12].

37.

(a) False

(b) False

39.

(a) True

(b) False

41.

(a) True

(b) True

43. HINT:

(a) Show that S_i is a subset of S_j for $i < j$.

(b) Reverse the hint for (a).

45. HINT: Show that $\mathbf{u} \cdot (\mathbf{u} + \mathbf{v}) \neq 0$ and $\mathbf{v} \cdot (\mathbf{u} + \mathbf{v}) \neq 0$.

47. HINT: Show the two required properties of a linear transformation hold.

51. HINT: Recall that $\text{proj}_S\mathbf{u}$ and $\mathbf{u}$ − are orthogonal and use the Pythagorean Theorem.

53. HINT:

(a) Use the hint given with this part of the problem.

(b) ‖projvu‖=|u · v|‖ v‖.

(c) Show ‖projvu‖=‖ u‖ only when $\mathbf{u} = c\mathbf{v}$.

55. An orthonormal basis is

{[1222211122−21122−12222] [−91738223952143452239513434522395−1
3173822395],
[8669757395883929104635539588378210463553958834209271395883]}

57. projsu=[29021477−17132954−144714771903422]

Section 8.3

1. Not symmetric

3. Symmetric

5. Not symmetric

7. Not symmetric

9. Not orthogonal

11. Orthogonal

13. Not orthogonal

15. P=[155−255255155],D=[200−3]

17. P=[133122−166133−122−1661330136],D=[00002000−1]

19. P=[−155255255155],D=[0005]

21. P=[−1221621330−136133122166133],D=[−200000003]

23. P=[−2202201022022],D=[−100010001]

25. $\lambda_1 = 1$ and $\lambda_2 = 11$.

27. $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = 5$.

29. Q−1=[15−252515]

31. Q−1=[012−12120001212]

33. Q=[31313−213132131331313],R=[130013]

35. Q=[110103101031010−11010],R=[1010010].

37. Q=[1932516291086193−8532581086593732581086],R=
[334930191086]

39. Q=[−2332929234292913−22929],R=[30029]

41. For example, A=[1002].

43. For example, A=[75−65−65−25].

45. For example, A=[0000].

47. For example, A=[−22−65].

49.

(a) True

(b) False

51.

(a) False

(b) False

53. HINT: $A^T A = I$.

55. For vectors **u** and **v**, $\mathbf{u}^T \mathbf{v} = \mathbf{u} \cdot \mathbf{v}$.

57. HINT: How is A related to A^T?

59. HINT: Show that A^2 is symmetric.

61. D≈[8.04630002.2795000−3.3258],P≈[0.3603−0.8287−0.4282−0.3790−0.54950.74460.85240.10600.5120]

63. D≈[7.6240000−1.21100005.6390000−6.051] P≈[−0.1376−0.62160.71180.2968−0.74260.51640.13910.4038−0.00780.36990.6127−0.69840.65580.45850.31400.5110]

65. Q=[−121331660133−136122133166],R=[2−2202343000]

67. Q=[1217035476790679012970 35−163395679012−37351733956790−121103579706790],R=[2322012352235350023567 90]

Section 8.4

1. σ1=8,σ2=2

3. σ1=16=4,σ2=4=2

5. σ1=3,σ2=5

7. σ1=4+5≈2.497,σ2=4−5≈1.328

9. V=[122−122122122],∑=[3001],U=[122122122−122]

11. V≈[0.22980.9732−0.97320.2298],∑≈[3.199002.40100] U≈[−0.16060.9064−0.3906−0.9845−0.11820.13020.071830.40530.9113]

13. $A = V\Sigma^T U^T$, where V=[0110],∑=[300200],
U=[23−122−16223122−162130232]

15. $A = V\Sigma^T U^T$, where V=[1001],∑[30030 00],
U=[23133−162−16623−133−16216612023200133 0136]

17. Numerical rank of A is 2.

19. Numerical rank of A is 2.

21.

(a) True

(b) False

23.

(a) True

(b) False

25. σ1u1v1T=[−0.11810.5000−0.72373.0650.0528−0.2236]
σ1u1v1T+σ2u2v2T=
[2.0001.0−0.99993.0000.99982.418×10−5]

27. σ1u1v1T=[22100000]σ1u1v1T+σ2u2v2T=[22101−101]

29. HINT: If $A = U\Sigma V^T$, then $A^T = V\Sigma^T U^T$. Compare the nonzero terms of Σ and Σ^T.

31. HINT: U and V are orthogonal, so $U^{-1} = U^T$ and $V^{-1} = V^T$.

33. HINT: Simplify $(PA)^T PA$, using P orthogonal.

35. HINTS:

(a) Note that $A^T A\mathbf{x} = A^T(A\mathbf{x})$.

(b) Recall that $(\mathrm{col}(A))^{\perp} = \mathrm{null}(A^T)$.

(c) Show that null(A) and null($A^T A$) are subsets of each other.

37. V≈ [0.45270.89160.8916−0.4528], Σ≈[5.9667001.8436],
U≈[0.97480.22280.2230−0.9748]

39. V≈[0.82240.4739
0.3147−0.24770.7963−0.5519−0.51210.3760 0.7722],
Σ≈[5.53710004.33200003.2518],
U≈[−0.9276−0.3734 −0.008949−0.08420
0.1860 0.9789−0.3639 0.9089−0.2039]

Section 8.5

1. projsy=[−1212]

3. projsy=[199189−299189218189]

5. 14x1−3x2=23−3x1+6x2=−2

7. 6x1+3x2−2x3=63x1+18x2−23x3=27−2x1+23x2+30x3=−34

9. x1=−152195 and x2=−17195

11. x1=−2t,x2=−5t−43, and $x_3 = t$

13. The normal equations are

[2242484816][c1c2c3]=[61020].

We obtain infinitely many solutions because infinitely many parabolas pass through two given points.

15. For example,

x1 =0 x2=0x1+x2=1.

17. For example,

x1+x2=02x1+2x2=03x1+3x2=04x1+4x2=1

19. For example,

x1 =0 x2 =0 x3=0.

21.

(a) False

(b) True

23.

(a) True

(b) False

25. HINT: Make the columns of A orthonormal—then this is true.

27. HINT: A^TA is an identity matrix.

29. HINT: Use hint given with problem.

31. $y = 2.2071 + 0.5214x$

33. $y = 2.081 + 0.07458x$

35. $y = 1.667 + 0.09x + 0.3833x^2$

37. $y = 2.191 - 0.06x - 0.5857x^2$

39. $y = 0.9975e^{0.9702x}$

41. $y = 15.49e^{-0.1564x}$

43. $y = 2.173x^{1.306}$

45. $y = 38.51x^{-0.5491}$

47. $p = 0.2001d^{1.499}$

49. $f(t) = 2199.8 + 2.65t - 16.75t^2$; $t = 11.539$ seconds to hit the ground.

51. $y \approx 2.307e^{-0.2211t}$. The initial size of the sample is $y \approx 2.307$ grams. The amount present at $t = 15$ is $y \approx 0.08370$ grams.

Supplementary Exercises

1.

(a) $\mathbf{u}_1 \cdot \mathbf{u}_3 = 2$

(b) $\mathbf{u}_4 \cdot \mathbf{u}_5 = 8$

(c) $\mathbf{u}_2 \cdot (-\mathbf{u}_6) = 0$

(d) $2\mathbf{u}_5 \cdot (-3\mathbf{u}_2) = -180$

3.

(a) ‖u1‖=10

(b) ‖−u4‖=14

(c) ‖3u5‖=65

(d) ‖u3+u6‖=41

5.

(a) ‖u1−u3‖=23

(b) ‖u2−u5‖=14

(c) ‖u1−u4‖=26

(d) ‖u3−u6‖=61

7.

(a) $\mathbf{u}_1 \cdot \mathbf{u}_6 = 17 \Rightarrow$ Not orthogonal

(b) $\mathbf{u}_2 \cdot \mathbf{u}_4 = 21 \Rightarrow$ Not orthogonal

(c) $\mathbf{u}_4 \cdot \mathbf{u}_1 = 0 \Rightarrow$ Orthogonal

(d) $\mathbf{u}_5 \cdot \mathbf{u}_3 = 0 \Rightarrow$ Orthogonal

9.

(a) proju2u3=154[14497]

(b) proju1u6=110[51−170]

(c) proju5u1=15[0−4−2]

(d) proju4u2=12[39−6]

11.

(a) u1‖u1‖=110[3−10]

(b) u2‖u2‖=136[271]

(c) u4‖u4‖=114[13−2]

(d) u6‖u6‖=135[5−24]

13. *A* is symmetric but not orthogonal.

15. *A* is symmetric and orthogonal.

17. *A* is not symmetric and not orthogonal.

19. D=[0005],P=[25−151525]

21. D=[000020003],P=[16−1213−26013161213]

23. Q=[25−151525],R=[252505]

25. Q=[25123015−2230015230],R=[525046230]

27. U=[121212−12],∑=[4002],V=[12−121212]

29. U=[560−16−130−25−16215−1523],∑=[600100],V=[25−151525]

31. U=[15−252515],∑=[600010],V=[560−16130−25162151523]

33. projSy=[−1525]

35. projSy=[000]

37. 6x1+x2=−1x1+6x2=1

39. 6x1+11x2−10x3=1611x1+22x2−17x3=29−10x1−17x2+18x3=−26

41. x1=2111,x2=1211

43. x1=−23,x2=53,x3=−73

Chapter 9

Section 9.1

1. $T(v2-2v1)=[-5-3]$

3. $T(2x2-4x-1)=[-45]$

5. HINT: Focus on Definition 9.1 or Theorem 9.2.

7. HINT: Focus on Definition 9.1 or Theorem 9.2.

9. HINT: Focus on Definition 9.1 or Theorem 9.2.

11. T is a linear transformation. Apply Theorem 9.2 to show this.

13. T is a linear transformation. Apply Theorem 9.2 to show this.

15. T is a linear transformation. Apply Theorem 9.2 to show this.

17. T is a linear transformation. Apply Theorem 9.2 to show this.

19. T is a linear transformation. Apply Theorem 9.2 to show this.

21. T is not a linear transformation.

23. $\ker(T) = \{p(x) : p(x) = ax + a\}$, $\text{range}(T) = \mathbf{R}$

25. $\ker(T) = \{\mathbf{0}_{\mathbf{P}^2}\}$

$$\text{range}(T)=\text{span}\{[1000],[0110],[0001]\}$$

27. T is not one-to-one, but is onto.

29. T is not one-to-one, but is onto.

31. $V = \mathbf{R}$ and $W = \mathbf{R}^2$, and define $T(a)=([a0])$.

33. $V = \mathbf{R}^2$ and $W = \mathbf{R}^2$, and define $T([ab])=[a0]$.

35. $V = \mathbf{R}^4$ and $W = \mathbf{R}^3$, and define $T([abcd])=[abc]$.

37. $V = \mathbf{R}^k$ and $W = \mathbf{R}$, and define $T(\mathbf{v}) = \mathbf{0}$.

39.

(a) True

(b) False

41.

(a) True

(b) False

43.

(a) True

(b) False

45. dim(range(T)) = 3

47. HINT: Use property (a) of a linear transformation.

49. HINT: Apply Theorem 9.2.

51. HINT: Use $\mathbf{0}_V + \mathbf{0}_V = \mathbf{0}_V$.

53. HINT: Show that ker(T) satisfies the three required properties of a subspace.

55. HINT: Use the hint with the problem.

57. HINT: Use the extended version of Theorem 9.2.

59. HINT: Recall that differentiation distributes across sums of functions.

61. HINT: Recall that differentiation distributes across sums of functions.

63. HINT: $(x^2p(x))' = 2xp(x) + x^2p'(x)$

Section 9.2

1. $\dim(V) = \dim(\mathbf{R}^8) = 8$, and $\dim(W) = \dim(\mathbf{P}^9) = 10$. Because $\dim(V) \neq \dim(W)$, the vector spaces are not isomorphic.

3. $\dim(V) = \dim(\mathbf{R}^{3\times6}) = 18$, and $\dim(W) = \dim(\mathbf{P}^{17}) = 18$. Because $\dim(V) = \dim(W)$, the vector spaces are isomorphic.

5. $\dim(V) = \dim(\mathbf{R}^{13}) = 13$, and $\dim(W) = \dim(\mathbf{C}[0,1]) = \infty$. Because $\dim(V) \neq \dim(W)$, the vector spaces are not isomorphic.

7. HINT: Apply Theorem 9.11.

9. HINT: A matrix is zero exactly when all of its entries are zero. This can be used to show that T is one-to-one. You also need to show that T is a linear transformation and is onto.

11. T is an isomorphism.

13. T is not an isomorphism.

15. T−1([cd])=(c2+d)x+c2

17. $T^{-1}(ax^2 + bx + c) = cx^2 - bx + a$, so that $T^{-1} = T$.

19. HINT: Note that all vectors in S have the form [a1a20].

21. HINT: Focus carefully on the form of a general vector in $\mathbf{P}_e$. It is helpful to consider a few concrete cases.

23. T{[abcde]}=ax4+bx3+cx2+dx+e is an isomorphism.

25. T{[abcd]}=ax3+bx2+cx+d is an isomorphism.

27. S=span{[100000],[010000],[000100],[000010]}

29. Let S be the set of all vectors of the form $(a_1, a_2, \dots a_n, 0, 0, 0, \dots)$ (i.e., all infinite vectors with entries equal to zero from some point on).

31.

(a) False

(b) True

33.

(a) False

(b) True

35.

(a) True

(b) True

37.

(a) False

(b) False

39. HINT: T must be one-to-one and onto to have an inverse.

41. HINT: The proof of Theorem 3.21 can be used as a model for showing that T^{-1} is also a linear transformation.

43. HINT: T is always onto range(T).

Section 9.3

1. v=[−11−4]

3. v=−x2−14x−9

5. [v]$\mathcal{G}$=[43]

7. [v]$\mathcal{G}$=[3−56]

9. [v]$\mathcal{G}$=[−43]

11. [v]$\mathcal{G}$=[21492−74]

13. T[v]$\mathcal{G}$=[1728]

15. T[v]$\mathcal{G}$=27x2+12x+67

17. T[v]$\mathcal{G}$=4cos x−9sin x+9e−x

19. A=[01−10]

21. A=[200010]

23. A=[−3021]

25. A=[712−18−31]

27.

(a) [a2b2c2def]

(b) [3abc3def]

29.

(a) [cabfde]

(b) [dfeacb]

31. $T^{-1}(x) = -2x - 4$

33. T−1(x+1)=[7−4]

35. $\mathbf{v} = 2x^2 - 3$, $\mathcal{G} = \{x^2, x, 1\}$

37. $\mathcal{G}$={[−730],[054]}

39. $V = \mathbf{R}^3$ and $W = \mathbf{R}^2$, and let $\mathcal{G}$={[100],[010],[001]} be the basis for V and $\mathcal{Q}$={[10],[01]} be the basis for W. Define $T(\mathbf{v}) = A\mathbf{v}$, where A=[212031].

41.

(a) False

(b) False

43.

(a) True
(b) True

45. HINT: The general results $[c\mathbf{v}]_{\mathcal{G}} = c[\mathbf{v}]_{\mathcal{G}}$ and $[\mathbf{v}_1 + \mathbf{v}_2]_{\mathcal{G}} = [\mathbf{v}_1]_{\mathcal{G}} + [\mathbf{v}_2]_{\mathcal{G}}$ are useful here.

47. HINT: A more general version of the results given in the answer to Exercise 45 can be used here.

49. HINT: The proof of part (b) follows from induction on *n*.

Section 9.4

1. S=[25−14]

3. S=[3410207610520213]

5. S=[19−1−20−1011]

7. S=[1−111]

9. A=[62204−18−59]

11. A=[362−3−8−310176]

13. A=[−3−275]

15. A=[−8891411−562892]

17. *A* and *B* are not similar matrices.

19. *A* and *B* are similar matrices.

21. $V = \mathbf{R}^2$, and let $\mathcal{G}$={[32],[43]} and $\mathcal{H}$={[10],[01]}.

23. B=[1234] and A=[3112−67−26], related by S=[5283].

25.

(a) True
(b) True

27.

(a) False
(b) False

29.

(a) False
(b) False

31. $D = S_2 S_1$

33. HINT: A and B have the same diagonal matrix D in their diagonalizations.

35. HINT: If S is invertible, then $(S^{-1})^T = (S^T)^{-1}$.

37. A and B are not similar matrices.

39. A and B are similar matrices.

Supplementary Exercises

1. $T(\mathbf{v}_1 + 3\mathbf{v}_2 + \mathbf{v}_3) = 4x^2 - x + 18$

3. HINT: Focus on Definition 9.1 or Theorem 9.2.

5. T is a linear transformation. Apply Definition 9.1 or Theorem 9.2 to show this.

7. $\ker(T) = \text{span}\{x^2 + x, x^2 - 1\}$, $\text{range}(T) = \mathbf{R}$

9. They are isomorphic, and both have dimension 24.

11. HINT: Focus on Definition 9.1 or Theorem 9.2.

13. T−1([cd])=dx+c

15. v=[710]

17. [v]$\mathcal{G}$=[−12152]

19. T([v]$\mathcal{G}$)=24x+31

21. A=[012−11]

23. A=[12−1212]

25. A=117[49−92−14−15]

27. Not similar; $\det(A) \neq \det(B)$.

Chapter 10

Section 10.1

1. $\langle \mathbf{u}, \mathbf{v} \rangle = 32$

3. $\langle p, q \rangle = -8$

5. $\langle f, g \rangle = 2$

7. $\langle A, B \rangle = -9$

9. $a = 3$

11. No value of a will make p and q orthogonal.

13. No value of b will make f and g orthogonal.

15. Norm=33

17. Norm=174

19. Norm=27

21. ‖ A‖=14

23. projuv=[321564153215]

25. projpq=−823x−1669

27. $\text{proj}_f\, g = 0$

29. projAB=[13−16160]

31. u=[10],v=[01]

33. u=[6/13−4/13],t1=t2=1

35. A=[100020003]

37. For example, let $w(x) = \cos(x)$, and define

⟨p,q⟩=∫01p(x)q(x)w(x)dx.

39. $\langle p, q \rangle = 0$ for all p and in $\mathbf{P}^2$.

41.

(a) True

(b) False

43.

(a) True

(b) False

45.

(a) False

(b) False

47. HINT: Use the distributive property of the real numbers to establish (b) and (c) of the definition of inner product.

49. HINT: The solution to Exercise 47 can be used as a model for this problem.

51. HINT: Example 5 can serve as a guide for this proof.

53. HINT: Review the properties of matrix transposes.

55. HINT: $||c\mathbf{v}||^2 = \langle c\mathbf{v}, c\mathbf{v}\rangle$

57. HINT: Use induction on k.

59. HINT: ‖1‖ v‖v‖=|1‖ v‖|‖ v‖ by Exercise 55.

61. HINT: Use properties of inner products to verify the required properties of a linear transformation.

63. HINT: $||\mathbf{u} - \mathbf{v}||^2 = \langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v}\rangle$

65. HINT: See Theorem 8.8 in Section 8.1.

67. HINT: Apply Exercise 63.

Section 10.2

1. {[29351831183],[−1226613366−11166],[89933−719833−2319833]}

3. a=−5,{[133−163163],[1316−56],[162132132]}

5. a=−1,{16x2+16x,12x2−12x−1,13x2−23x}

7. v=(5)[451]+(1)[−32−6]+(−1)[16−7−23]

9. projSv=[727351087108]

11. $\text{proj}_S\, f = 2\sin(x)$

13. {[1−10],[65451]}

15. {1,x2−13}

17. {x,−15x+1}

19. For example, let u2=[4−3].

21. For example, let $p_2(x) = 5x - 3$.

23. For example, $f_1(x) = 0$, $f_2(x) = 1$, and $f_3(x) = \cos(x)$.

25.

(a) False

(b) True

27.

(a) False

(b) False

29. HINT: Apply hint given with problem.

31. HINT: The suggested approach works well.

33. HINT: $\text{proj}_S\mathbf{u}$ is in S.

35.

(a) $\mathbf{u}$

(b) $\mathbf{0}$

37. HINT: ‖ v‖2=⟨v,v⟩ and Theorem 10.12.

Section 10.3

1. y=52+1x

3. The slope of ℓ_1 would be greater than the slope of ℓ_2.

5. The resulting line will be the same.

7. f2(x)=12+2πcos (x)

9. f2(x)=12+2πsin (x)

11. f2(x)=1+2sin (x)−sin (2x)

13. f2(x)=13π2−4cos (x)+cos (2x)

15. $a_2 = 1$ and $b_3 = -1$, with all other Fourier coefficients 0.

17. a0=32 and a8=−12, with all other Fourier coefficients 0.

19. g1(x)=32−cos (x)

21. g1(x)=12+32cos (x)+32sin (x)

23. For example, consider the data set {(−1, −1), (0, 1), (1, −1)}. is set has an ordinary least squares regression line y=−13 and a weighted least squares regression line with triple the weight on the right-most point y=−12−14x.

25. For example, let $f(x) = 1$. Then, $a_0 = 1$, and all other Fourier coefficients are 0.

27. $f(x) = 1 + x$

29.

   (a) True
   (b) False

31.

   (a) False
   (b) False

33. HINT: ⟨1, sin (kx)⟩=1π∫−ππsin (kx)dx

35. HINT: ‖sin (kx)‖2=1π∫−ππsin2 (kx)dx

37. HINT: Take $u = x$ and $dv = \cos(kx)dx$ in the integration by parts formula.

39. HINT: $\cos(k\pi) = (-1)^k$ and $\sin(k\pi) = 0$ for all integers k.

41. g5(x)≈2.7215+0.5445 cos (x)+0.05 cos (2x)+0.1555 cos (3x) + 0.075 cos (4x)+0.1555 cos (5x)
−0.06768 sin (x) −0.025 sin (2x)+0.03232 sin (3x)
−0.03232 sin (5x)

43. f3(x)≈3.6761−3.6761 cos (x)+1.4704 cos (2x)
−0.7352 cos (3x) +3.6761 sin (x)
−2.9409 sin (2x)+2.2056 sin (3x)

Supplementary Exercises

1. 〈**u**, **v**〉 = 23

3. ⟨f,g⟩=23

5. ‖[32−2]‖=34

7. ‖3x2‖=325

9. projuv=[102]

11. projfg=12(3x−1)

13. True

15. True

17. No choice of *a* will make the set orthogonal.

19. $\text{proj}_S f = 1 + 4\sin(x)$

21. {[111],16[−517],117[6−812]}

23. yˆ=498+3942x

25. f2(x)=1+32π−12πcos (x)

27. g2(x)=32−7cos(x)+3cos (2x)

Chapter 11

Section 11.1

1. $Q(\mathbf{x}_0) = 22$

3. $Q(\mathbf{x}_0) = 8$

5. Q(x)=4x12+x22

7. Q(x)=x12+2x22+6x1x2

9. Q(x)=x12+3x22−2x32

11. Q(x)=2x12+x22+2x32+3x42

13. A=[133−5]

15. A=[30301030−1]

17. A=[5030−1−63−63]

19. *A* is indefinite.

21. *A* is indefinite.

23. *A* is indefinite.

25. *A* is indefinite.

27. Q(x)=x12+x22,c=−1

29. For example, let Q(x)=x12−x22 and $c = 0$. Then Q(x)=x12−x22=0⇒(x1−x2)(x1+x2)=0, and the graph consists of the intersecting lines $x_1 - x_2 = 0$ and $x_1 + x_2 = 0$.

31. The only quadratic form, which is also a linear transformation, is $Q(\mathbf{x}) = 0$ for all $\mathbf{x}$.

33. Q(x)=x12+x22

35.

(a) True

(b) True

37.

(a) False

(b) False

39. HINT: What form must such a quadratic form take?

41. $Q(\mathbf{x}) = \mathbf{x}^T I \mathbf{x} = \mathbf{x}^T \mathbf{x} = ||\mathbf{x}||^2$, so the identity matrix I_n is the matrix of $Q(\mathbf{x}) = ||\mathbf{x}||^2$.

43. HINT: Evaluate $\mathbf{0}^T A \mathbf{0}$.

Section 11.2

1. A1=[3], A2=[3557]

3. A1=[1], A2=[1440], A3=[14−3402−325]

5. A1=[2], A2=[2113], A3=[210134040], A4=[210−11341040−2−11−21]

7. $\det(A_1) = 2 > 0$ and $\det(A_2) = -5 < 0$, so A is not positive definite.

9. $\det(A_1) = 1 > 0$, $\det(A_2) = 1 > 0$, and $\det(A_3) = 1 > 0$, so A is positive definite.

11. $\det(A_1) = 1 > 0$, $\det(A_2) = 1 > 0$, $\det(A_3) = 1 > 0$, and $\det(A_4) = 4 > 0$, so A is positive definite.

13. $\det(A_1) = 1 > 0$ and $\det(A_2) = 1 > 0$, so A is positive definite.

L=[10−21], U=[1−201]

15. $\det(A_1) = 1 > 0$, $\det(A_2) = 1 > 0$ and $\det(A_3) = 1 > 0$, so A is positive definite.

L=[100−2102−11], U=[0−2201−1001]

17. $\det(A_1) = 1 > 0$ and $\det(A_2) = 4 > 0$, so A is positive definite.

L=[10−21], D=[1004], U=[1−201]

19. $\det(A_1) = 1 > 0$, $\det(A2) = 1 > 0$, and $\det(A_3) = 1 > 0$, so A is positive definite.

L=[100310221], D=[100010001], U=[132012001]

21. $\det(A_1) = 1 > 0$ and $\det(A_2) = 1 > 0$, so A is positive definite.

Lc=[1021]

23. $\det(A_1) = 1 > 0$, $\det(A_2) = 1 > 0$, and $\det(A_3) = 1 > 0$, so A is positive definite.

Lc=[100210301]

25. A=[−100−1]

27. A=[−1000−1000−1]

29. A=[−1000−10001]

31.

(a) False

(b) False

33.

(a) False

(b) True

Section 11.3

1. max = 2, min = −3

3. max = 3, min = −5

5. max = 2, min = −4

7. max = 5, min = 0

9. max = 1, min = −1

11. max = 3, min = 0

13. max = 20, min = 0

15. max = 100, min = −100

17. max = 104, min = 0

19. max=20+337,min=20−337

21. max = 3, min = 1

23. max = 1, min = 0

25. Q(x)=x12+5x22

27. Q(x)=x12+6x22

29. Q(x)=−19x12+49x22

31.

(a) True

(b) False

33.

(a) False

(b) True

35. HINT: $q_j \le q_k$ for all k =1, 2, … , n, and $q_k \le q_i$ for all k = 1, 2, … , n.

37. HINT: $Q(c\mathbf{x}) = (c\mathbf{x})^T A(c\mathbf{x})$

Section 11.4

1.

(a) [3i−4+i2−i]

(b) [8+i11−4i7+12i]

(c) [−14+i−12+9i−2−21i]

3. c does not exist.

5. Yes

7.

(a) 24 − 6i

(b) 44

(c) 35

9. Divide **u** by 39, and divide **v** by 210.

11.

(a) [2+i4−3i−3−i3+2i]

(b) [−2−3i9−12i−13+5i4+i]

13. c does not exist.

15. Yes

17.

(a) 18 + 2i

(b) 14 − 50i

(c) 42

19. Divide A by 29, and divide C by 27.

21.

(a) $(2 + 4i) - (4 - 2i)x$

(b) $9 - (3 + 3i)x$

23. c does not exist.

25. (2+i)+(3−2i)x=(−217−4317i)(1+ix)+(1217+2017i)(3−(1+i)x)

27.

(a) 136+53i

(b) 133+103i

(c) 233

29. Divide h_1 by 233, divide h_2 by 233.

31. $\mathbf{u} = 1$, $\mathbf{v} = i$ in $\mathbf{C}$.

33. ⟨u, v⟩=2u1v¯1+⋯+2unv¯n

35. $V = \mathbf{C}$, and $S = \mathbf{R}$

37.

(a) True

(b) False

39.

(a) True

(b) True

41. HINT: Apply the properties of complex numbers and arithmetic.

43. HINT: Consider property (2).

45. HINT: Combine properties of complex numbers with definition of inner product space.

47. HINT: Apply property (d) of definition of inner product space.

49. HINT: Note that 0≤‖u−⟨u, v⟩‖ v‖2v‖2.

Section 11.5

1. A*=[1−i2+i−3i1−4i]

3. A*=[3−i1+4i2−2i−5i−801+i6−i7i]

5. A*=[1−2i34i2i5−6i1+i36i73+2i−4i1−i3−2i11]

7. *A* is not Hermitian.

9. *A* is Hermitian.

11. *A* is Hermitian.

13. *A* is normal.

15. *A* is normal.

17. *A* is normal.

19. A=[0iii0iii0]

21. A=[i000i000i]

23.

(a) True

(b) False

25.

(a) True

(b) False

27. HINT: If A has real entries, then so does A^T.

29. HINT: Apply the result in Exercise 28.

31. HINT: Apply the result in Exercise 28.

33. HINT: A unitary implies that $A^{-1} = A^*$, so $A^*A = I_n$.

35. HINT: Compute A^*A for A normal and upper triangular.

Supplementary Exercises

1. $Q(\mathbf{x}_0) = -108$

3. Q(x)=−3x12+7x22−8x1x2

5. A=[26065000−3]

7. Quadratic form is indefinite.

9. A1=[4], A2=[455−2]

11. *A* is positive definite.

13. maximum = 5, minimum = − 2

15. maximum=(3+37)/2≈4.541,minimum=(3−37)/2≈−1.541

17. maximum = 40, minimum = 0

19. maximum=2518(13+2329)≈85.083,minimum=2518(13−2329)≈ −48.972

21. u+3w=[11+11i11−19i15−6i],−v+u−4w= [−18−25i−19+32i−21+15i]

23. 〈**v**, **w**〉 = 56 − 28*i*

25. u‖ u‖=134[2−4i−1+2i3i],w‖ w‖=1133[3+5i4−7i5−3i]

27. A*=[2+i8−3i2−i6+i]

29. *A* is not Hermitian.

INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z

APPLICATIONS INDEX

Astronomy

Biology

Business and Economics

Chemistry

Engineering

Environmental Science

Mathematics

Numerical Linear Algebra

Physics

Politics

Statistics