

LINEAR ALGEBRA

AND MATRIX

THEORY

JIMMIE GILBERT
LINDA GILBERT

University of South Carolina at Spartanburg
Spartanburg, South Carolina



ACADEMIC PRESS

San Diego New York Boston London Sydney Tokyo Toronto

This book is printed on acid-free paper. ☺

Copyright © 1995, 1970 by ACADEMIC PRESS, INC.

All Rights Reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Academic Press, Inc.
A Division of Harcourt Brace & Company
525 B Street, Suite 1900, San Diego, California 92101-4495

United Kingdom Edition published by
Academic Press Limited
24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Gilbert, Jimmie, date.

Linear algebra and matrix theory / by Jimmie Gilbert, Linda Gilbert.

p. cm.

Includes indexes.

ISBN 0-12-282970-0

1. Algebras, Linear. I. Gilbert, Linda. II. Title.

QA184.G525 1994

512'.5--dc20

94-38932

CIP

PRINTED IN THE UNITED STATES OF AMERICA

95 96 97 98 99 00 EB 9 8 7 6 5 4 3 2 1

Preface

This text was planned for use with one of the following two options:

- The complete text would be suitable for a one-year undergraduate course for mathematics majors and would provide a strong foundation for more abstract courses at higher levels of instruction.
- A one-semester or one-quarter course could be taught from the first five chapters together with selections from the remaining chapters. The selections could be chosen so as to meet the requirements of students in the fields of business, science, economics, or engineering.

The presentation of material presumes the knowledge and maturity gained from one calculus course, and a second calculus course is a desirable prerequisite.

It is our opinion that linear algebra is well suited to provide the transition from the intuitive developments of courses at a lower level to the more abstract treatments encountered later. Throughout the treatment here, material is presented from a structural point of view: fundamental algebraic properties of the entities involved are emphasized. This approach is particularly important because the mathematical systems encountered in linear algebra furnish a wealth of examples for the structures studied in more advanced courses.

The unifying concept for the first five chapters is that of elementary operations. This concept provides the pivot for a concise and efficient development of the basic theory of vector spaces, linear transformations, matrix multiplication, and the fundamental equivalence relations on matrices.

A rigorous treatment of determinants from the traditional viewpoint is presented in Chapter 6. For a class already familiar with this material, the chapter can be omitted.

In Chapters 7 through 10, the central theme of the development is the change in the matrix representing a vector function when only certain types of basis changes are admitted. It is from this approach that the classical canonical forms for matrices are derived.

Numerous examples and exercises are provided to illustrate the theory. Exercises are included of both computational and theoretical nature. Those of a theoretical nature amplify the treatment and provide experience in constructing deductive arguments, while those of a computational nature illustrate fundamental techniques. The amount of labor in the computational problems is kept to a minimum. Even so, many of them provide opportunities to utilize current technology, if that is the wish of the instructor. Answers are provided for about half of the computational problems.

The exercises are intended to develop confidence and deepen understanding. It is assumed that students grow in maturity as they progress through the text and the proportion of theoretical problems increases in later chapters.

Since much of the interest in linear algebra is due to its applications, the solution of systems of linear equations and the study of eigenvalue problems appear early in the text. Chapters 4 and 7 contain the most important applications of the theory.

ACKNOWLEDGMENTS

We wish to express our appreciation for the support given us by the University of South Carolina at Spartanburg during the writing of this book, since much of the work was done while we were on sabbatical leave. We would especially like to thank Sharon Hahs, Jimmie Cook, and Olin Sansbury for their approval and encouragement of the project.

This entire text was produced using *Scientific Word* and *Scientific Workplace*, software packages from TCI Software Research, Inc. Special thanks are due to Christopher Casey and Fred Osborne for their invaluable assistance throughout the project.

We would like to acknowledge with thanks the helpful suggestions made by the following reviewers of the text:

Ed Dixon, Tennessee Technological University

Jack Garner, University of Arkansas at Little Rock

Edward Hinson, University of New Hampshire

Melvyn Jeter, Illinois Wesleyan University

Bob Jones, Belhaven College

We also wish to express our thanks to Dave Pallai for initiating the project at Academic Press, to Peter Renz for his editorial guidance in developing the book, and to Michael Early for his patience and encouragement while supervising production of the book.

Jimmie Gilbert
Linda Gilbert

Chapter 1

Real Coordinate Spaces

1.1 Introduction

There are various approaches to that part of mathematics known as *linear algebra*. Different approaches emphasize different aspects of the subject such as matrices, applications, or computational methods. As presented in this text, linear algebra is in essence a study of vector spaces, and this study of vector spaces is primarily devoted to finite-dimensional vector spaces. The real coordinate spaces, in addition to being important in many applications, furnish excellent intuitive models of abstract finite-dimensional vector spaces. For these reasons, we begin our study of linear algebra with a study of the real coordinate spaces. Later it will be found that many of the results and techniques employed here will easily generalize to more abstract settings.

1.2 The Vector Spaces \mathbf{R}^n

Throughout this text the symbol \mathbf{R} will denote the set of all real numbers. We assume a knowledge of the algebraic properties of \mathbf{R} , and begin with the following definition of real coordinate spaces.

Definition 1.1 *For each positive integer n , \mathbf{R}^n will denote the set of all ordered n -tuples (u_1, u_2, \dots, u_n) of real numbers u_i . Two n -tuples (u_1, u_2, \dots, u_n) and (v_1, v_2, \dots, v_n) are equal if and only if $u_i = v_i$ for $i = 1, 2, \dots, n$. The set \mathbf{R}^n is referred to as an **n -dimensional real coordinate space**. The elements of \mathbf{R}^n are called **n -dimensional real coordinate vectors, or simply vectors**. The numbers u_i in a vector (u_1, u_2, \dots, u_n) will be called the **components** of the vector. The elements of \mathbf{R} will be referred to as **scalars**.¹*

¹The terms “vector” and “scalar” are later extended to more general usage, but this will cause no confusion since the context will make the meaning clear.

The real coordinate spaces and the related terminology described in this definition are easily seen to be generalizations and extensions of the two- and three-dimensional vector spaces studied in the calculus.

When we use a single letter to represent a vector, the letter will be printed in boldface lower case Roman, such as \mathbf{v} , or written with an arrow over it, such as \vec{v} . In handwritten work with vectors, the arrow notation \vec{v} is commonly used. Scalars will be represented by letters printed in lower case italics.

Definition 1.2 *Addition in \mathbf{R}^n is defined as follows: for any $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ in \mathbf{R}^n , the sum $\mathbf{u} + \mathbf{v}$ is given by*

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n).$$

For any scalar a and any vector $\mathbf{u} = (u_1, u_2, \dots, u_n)$ in \mathbf{R}^n , the product $a\mathbf{u}$ is defined by

$$a\mathbf{u} = (au_1, au_2, \dots, au_n).$$

The operation that combines the scalar a and the vector \mathbf{u} to yield $a\mathbf{u}$ is referred to as **multiplication of the vector \mathbf{u} by the scalar a** , or simply as **scalar multiplication**. Also, the product $a\mathbf{u}$ is called a **scalar multiple** of \mathbf{u} .

The following theorem gives the basic properties of the two operations that we have defined.

Theorem 1.3 *The following properties are valid for any scalars a and b , and any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in \mathbf{R}^n :*

1. $\mathbf{u} + \mathbf{v} \in \mathbf{R}^n$. (*Closure under addition*)
2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$. (*Associative property of addition*)
3. There is a vector $\mathbf{0}$ in \mathbf{R}^n such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for all $\mathbf{u} \in \mathbf{R}^n$. (*Additive identity*)
4. For each $\mathbf{u} \in \mathbf{R}^n$, there is a vector $-\mathbf{u}$ in \mathbf{R}^n such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. (*Additive inverses*)
5. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. (*Commutative property of addition*)
6. $a\mathbf{u} \in \mathbf{R}^n$. (*Absorption under scalar multiplication*)
7. $a(b\mathbf{u}) = (ab)\mathbf{u}$. (*Associative property of scalar multiplication*)
8. $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$. (*Distributive property, vector addition*)
9. $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$. (*Distributive property, scalar addition*)
10. $1 \cdot \mathbf{u} = \mathbf{u}$.

The proofs of these properties are easily carried out using the definitions of vector addition and scalar multiplication, as well as the properties of real numbers. As typical examples, properties 3, 4, and 8 will be proved here. The remaining proofs are left as an exercise.

Proof of Property 3. The vector $\mathbf{0} = (0, 0, \dots, 0)$ is in \mathbf{R}^n , and if $\mathbf{u} = (u_1, u_2, \dots, u_n)$,

$$\mathbf{u} + \mathbf{0} = (u_1 + 0, u_2 + 0, \dots, u_n + 0) = \mathbf{u}.$$

Proof of Property 4. If $\mathbf{u} = (u_1, u_2, \dots, u_n) \in \mathbf{R}^n$, then $\mathbf{v} = (-u_1, -u_2, \dots, -u_n)$ is in \mathbf{R}^n since all real numbers have additive inverses. And since

$$\mathbf{u} + \mathbf{v} = (u_1 + (-u_1), u_2 + (-u_2), \dots, u_n + (-u_n)) = \mathbf{0},$$

\mathbf{v} is the vector $-\mathbf{u}$ as required by property (4).

Proof of Property 8. Let $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$. Then

$$\begin{aligned} a(\mathbf{u} + \mathbf{v}) &= a(u_1 + v_1, u_2 + v_2, \dots, u_n + v_n) \\ &= (a(u_1 + v_1), a(u_2 + v_2), \dots, a(u_n + v_n)) \\ &= (au_1 + av_1, au_2 + av_2, \dots, au_n + av_n) \\ &= (au_1, au_2, \dots, au_n) + (av_1, av_2, \dots, av_n) \\ &= a\mathbf{u} + a\mathbf{v}. \end{aligned} \quad \blacksquare$$

The associative property of addition in \mathbf{R}^n can be generalized so that the terms in a sum such as $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \dots + a_k\mathbf{u}_k$ can be grouped with parentheses in any way without changing the value of the result. Such a sum can be written in the compact form $\sum_{i=1}^k a_i\mathbf{u}_i$ or $\sum_i a_i\mathbf{u}_i$ if the number of terms in the sum is not important. It is always understood that only a finite number of nonzero terms is involved in such a sum.

Definition 1.4 Let \mathcal{A} be a nonempty set of vectors in \mathbf{R}^n . A vector \mathbf{v} in \mathbf{R}^n is **linearly dependent on the set \mathcal{A}** if there exist vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ in \mathcal{A} and scalars a_1, a_2, \dots, a_k such that $\mathbf{v} = \sum_{i=1}^k a_i\mathbf{u}_i$. A vector of the form $\sum_{i=1}^k a_i\mathbf{u}_i$ is called a **linear combination of the \mathbf{u}_i** .

If linear dependence on a finite set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is under consideration, the statement in Definition 1.4 is equivalent to requiring that *all* of the vectors in \mathcal{B} be involved in the linear combination. That is, a vector \mathbf{v} is linearly dependent on $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ if and only if there are scalars b_1, b_2, \dots, b_r such that $\mathbf{v} = \sum_{i=1}^r b_i\mathbf{v}_i$.

Example 1 \square As an illustration, let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = (1, 0, 2, 1), \mathbf{v}_2 = (0, -2, 2, 3), \text{ and } \mathbf{v}_3 = (1, -2, 4, 4),$$

and let $\mathbf{v} = (3, -4, 10, 9)$. Now \mathbf{v} can be written as

$$\mathbf{v} = 1 \cdot \mathbf{v}_1 + 0 \cdot \mathbf{v}_2 + 2 \cdot \mathbf{v}_3,$$

or

$$\mathbf{v} = 3 \cdot \mathbf{v}_1 + 2 \cdot \mathbf{v}_2 + 0 \cdot \mathbf{v}_3.$$

Either of these combinations shows that \mathbf{v} is linearly dependent on \mathcal{B} .

To consider a situation involving an infinite set, let \mathcal{A} be the set of all vectors in \mathbf{R}^3 that have integral components, and let $\mathbf{v} = (\sqrt{2}, \frac{2}{3}, 0)$. Now $\mathbf{u}_1 = (1, 0, 1)$, $\mathbf{u}_2 = (0, 1, 0)$, and $\mathbf{u}_3 = (0, 0, 1)$ are in \mathcal{A} , and

$$\mathbf{v} = \sqrt{2}\mathbf{u}_1 + \frac{2}{3}\mathbf{u}_2 - \sqrt{2}\mathbf{u}_3.$$

Thus \mathbf{v} is linearly dependent on \mathcal{A} . It should be noted that other choices of vectors \mathbf{u}_i can be made in order to exhibit this dependence. ■

In order to decide whether a certain vector is linearly dependent on a given set in \mathbf{R}^n , it is usually necessary to solve a system of equations. This is illustrated in the following example.

Example 2 □ Consider the question as to whether $(6, 0, -1)$ is linearly dependent on the set $\mathcal{A} = \{(2, -1, 1), (0, 1, -1), (-2, 1, 0)\}$. To answer the question, we investigate the conditions on a_1, a_2 , and a_3 that are required by the equation

$$a_1(2, -1, 1) + a_2(0, 1, -1) + a_3(-2, 1, 0) = (6, 0, -1).$$

Performing the indicated scalar multiplications and additions in the left member of this equation leads to

$$(2a_1 - 2a_3, -a_1 + a_2 + a_3, a_1 - a_2) = (6, 0, -1).$$

This vector equation is equivalent to the system of equations

$$\begin{aligned} 2a_1 &\quad - 2a_3 = 6 \\ -a_1 + a_2 + a_3 &= 0 \\ a_1 - a_2 &= -1. \end{aligned}$$

We decide to work toward the solution of this system by eliminating a_1 from two of the equations in the system. As steps toward this goal, we multiply the first equation by $\frac{1}{2}$ and we add the second equation to the third equation. These steps yield the system

$$\begin{aligned} a_1 &\quad - a_3 = 3 \\ -a_1 + a_2 + a_3 &= 0 \\ a_3 &= -1. \end{aligned}$$

Adding the first equation to the second now results in

$$\begin{aligned} a_1 &\quad - a_3 = 3 \\ a_2 &= 3 \\ a_3 &= -1. \end{aligned}$$

The solution $a_1 = 2, a_2 = 3, a_3 = -1$ is now readily obtained. Thus the vector $(6, 0, -1)$ is linearly dependent on the set \mathcal{A} . ■

Another important type of dependence is given in the definition below. This time, the phrase *linearly dependent* involves only a set instead of involving both a vector and a set.

Definition 1.5 A set \mathcal{A} of vectors in \mathbf{R}^n is **linearly dependent** if there is a collection of vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ in \mathcal{A} and scalars c_1, c_2, \dots, c_k , not all of which are zero, such that $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \dots + c_k\mathbf{u}_k = \mathbf{0}$. If no such collection of vectors exists in \mathcal{A} , then \mathcal{A} is called **linearly independent**.

Again, the case involving a finite set of vectors is of special interest. It is readily seen that a finite set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is linearly dependent if and only if there are scalars b_1, b_2, \dots, b_r , not all zero, such that $\sum_{i=1}^r b_i \mathbf{v}_i = \mathbf{0}$.

Example 3 \square The set \mathcal{B} in Example 1 is linearly dependent since $\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$. The set \mathcal{A} is also linearly dependent. For if $\mathbf{u}_1, \mathbf{u}_2$, and \mathbf{u}_3 are as before and $\mathbf{u}_4 = (-3, -2, -4)$, then $3\mathbf{u}_1 + 2\mathbf{u}_2 + \mathbf{u}_3 + \mathbf{u}_4 = \mathbf{0}$. ■

To determine the linear dependence or linear independence of a set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ of vectors in \mathbf{R}^n , it is necessary to investigate the conditions on the c_i which are imposed by requiring that $\sum_{j=1}^k c_j \mathbf{u}_j = \mathbf{0}$. If $\mathbf{u}_j = (u_{1j}, u_{2j}, \dots, u_{nj})$, we have

$$\begin{aligned}\sum_{j=1}^k c_j \mathbf{u}_j &= \sum_{j=1}^k c_j (u_{1j}, u_{2j}, \dots, u_{nj}) \\ &= (\sum_{j=1}^k c_j u_{1j}, \sum_{j=1}^k c_j u_{2j}, \dots, \sum_{j=1}^k c_j u_{nj}).\end{aligned}$$

Thus $\sum_{j=1}^k c_j \mathbf{u}_j = \mathbf{0}$ if and only if $\sum_{j=1}^k c_j u_{ij} = 0$ for each $i = 1, 2, \dots, n$. This shows that the problem of determining the conditions on the c_j is equivalent to investigating the solutions of a system of n equations in k unknowns. If $\sum_{j=1}^k c_j \mathbf{u}_j = \mathbf{0}$ implies $c_1 = c_2 = \dots = c_k = 0$, then $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is linearly independent.

The discussion in the preceding paragraph is illustrated in the next example.

Example 4 \square Consider the set of vectors

$$\{(1, 1, 8, 1), (1, 0, 3, 0), (3, 1, 14, 1)\}$$

in \mathbf{R}^4 . To determine the linear dependence or linear independence of this set, we investigate the solutions c_1, c_2, c_3 to

$$c_1(1, 1, 8, 1) + c_2(1, 0, 3, 0) + c_3(3, 1, 14, 1) = (0, 0, 0, 0).$$

Performing the indicated vector operations gives the equation

$$(c_1 + c_2 + 3c_3, c_1 + c_3, 8c_1 + 3c_2 + 14c_3, c_1 + c_3) = (0, 0, 0, 0).$$

This vector equation is equivalent to

$$\begin{aligned} c_1 + c_2 + 3c_3 &= 0 \\ c_1 &\quad + \quad c_3 = 0 \\ 8c_1 + 3c_2 + 14c_3 &= 0 \\ c_1 &\quad + \quad c_3 = 0. \end{aligned}$$

To solve this system, we first interchange the first two equations to place the equation $c_1 + c_3 = 0$ at the top.

$$\begin{aligned} c_1 &\quad + \quad c_3 = 0 \\ c_1 + c_2 + 3c_3 &= 0 \\ 8c_1 + 3c_2 + 14c_3 &= 0 \\ c_1 &\quad + \quad c_3 = 0 \end{aligned}$$

By adding suitable multiples of the first equation to each of the other equations, we then eliminate c_1 from all but the first equation. This yields the system

$$\begin{aligned} c_1 &\quad + \quad c_3 = 0 \\ c_2 + 2c_3 &= 0 \\ 3c_2 + 6c_3 &= 0 \\ 0 &= 0. \end{aligned}$$

Eliminating c_2 from the third equation gives

$$\begin{aligned} c_1 &\quad + \quad c_3 = 0 \\ c_2 + 2c_3 &= 0 \\ 0 &= 0 \\ 0 &= 0. \end{aligned}$$

It is now clear that there are many solutions to the system, and they are given by

$$\begin{aligned} c_1 &= -c_3 \\ c_2 &= -2c_3 \\ c_3 &\text{ is arbitrary.} \end{aligned}$$

In particular, it is not necessary that c_3 be zero, so the original set of vectors is linearly dependent. ■

The following theorem gives an alternative description of linear dependence for a certain type of set.

Theorem 1.6 *A set $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ in \mathbf{R}^n that contains at least two vectors is linearly dependent if and only if some \mathbf{u}_j is a linear combination of the remaining vectors in the set.*

Proof. If the set $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is linearly dependent, then there are scalars a_1, a_2, \dots, a_k such that $\sum_{i=1}^k a_i \mathbf{u}_i = \mathbf{0}$ with at least one a_i , say a_j , not zero. This implies that

$$a_j \mathbf{u}_j = -a_1 \mathbf{u}_1 - \cdots - a_{j-1} \mathbf{u}_{j-1} - a_{j+1} \mathbf{u}_{j+1} - \cdots - a_k \mathbf{u}_k$$

so that

$$\mathbf{u}_j = \left(-\frac{a_1}{a_j}\right) \mathbf{u}_1 + \cdots + \left(-\frac{a_{j-1}}{a_j}\right) \mathbf{u}_{j-1} + \left(-\frac{a_{j+1}}{a_j}\right) \mathbf{u}_{j+1} + \cdots + \left(-\frac{a_k}{a_j}\right) \mathbf{u}_k.$$

Thus \mathbf{u}_j can be written as a linear combination of the remaining vectors in the set.

Now assume that some \mathbf{u}_j is a linear combination of the remaining vectors in the set, i.e.,

$$\mathbf{u}_j = b_1 \mathbf{u}_1 + b_2 \mathbf{u}_2 + \cdots + b_{j-1} \mathbf{u}_{j-1} + b_{j+1} \mathbf{u}_{j+1} + \cdots + b_k \mathbf{u}_k.$$

Then

$$b_1 \mathbf{u}_1 + \cdots + b_{j-1} \mathbf{u}_{j-1} + (-1) \mathbf{u}_j + b_{j+1} \mathbf{u}_{j+1} + \cdots + b_k \mathbf{u}_k = \mathbf{0},$$

and since the coefficient of \mathbf{u}_j in this linear combination is not zero, the set is linearly dependent. ■

The different meanings of the word “dependent” in Definitions 1.4 and 1.5 should be noted carefully. These meanings, though different, are closely related. The preceding theorem, for example, could be restated as follows: “A set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is linearly dependent if and only if some \mathbf{u}_k is linearly dependent on the remaining vectors.” This relation is further illustrated in some of the exercises at the end of this section.

In the last section of this chapter, the following definition and theorem are of primary importance. Both are natural extensions of Definition 1.4.

Definition 1.7 *A set $\mathcal{A} \subseteq \mathbf{R}^n$ is linearly dependent on the set $\mathcal{B} \subseteq \mathbf{R}^n$ if each $\mathbf{u} \in \mathcal{A}$ is linearly dependent on \mathcal{B} .*

Thus \mathcal{A} is linearly dependent on \mathcal{B} if and only if each vector in \mathcal{A} is a linear combination of vectors that are contained in \mathcal{B} .

Other types of dependence are considered in some situations, but linear dependence is the only type that we will use. For this reason, we will frequently use the term “dependent” in place of “linearly dependent.”

In the proof of the next theorem, it is convenient to have available the notation $\sum_{i=1}^k (\sum_{j=1}^m \mathbf{u}_{ij})$ for the sum of the form $\sum_{j=1}^m \mathbf{u}_{1j} + \sum_{j=1}^m \mathbf{u}_{2j} + \cdots + \sum_{j=1}^m \mathbf{u}_{kj}$. The

associative and commutative properties for vector addition [Theorem 1.3, (2) and (5)] imply that

$$\sum_{i=1}^k \left(\sum_{j=1}^m \mathbf{u}_{ij} \right) = \sum_{j=1}^m \left(\sum_{i=1}^k \mathbf{u}_{ij} \right).$$

Theorem 1.8 *Let \mathcal{A} , \mathcal{B} , and \mathcal{C} be subsets of \mathbf{R}^n . If \mathcal{A} is dependent on \mathcal{B} and \mathcal{B} is dependent on \mathcal{C} , then \mathcal{A} is dependent on \mathcal{C} .*

Proof. Suppose that \mathcal{A} is dependent on \mathcal{B} and \mathcal{B} is dependent on \mathcal{C} , and let \mathbf{u} be an arbitrary vector in \mathcal{A} . Since \mathcal{A} is dependent on \mathcal{B} , there are vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ in \mathcal{B} and scalars a_1, a_2, \dots, a_m such that

$$\mathbf{u} = \sum_{j=1}^m a_j \mathbf{v}_j.$$

Since \mathcal{B} is dependent on \mathcal{C} , each \mathbf{v}_j can be written as a linear combination of certain vectors in \mathcal{C} . In general, for different vectors \mathbf{v}_j , different sets of vectors from \mathcal{C} would be involved in these linear combinations. But each of these m linear combinations (one for each \mathbf{v}_j) would involve only a finite number of vectors. Hence the set of all vectors in \mathcal{C} that appear in a term of at least one of these linear combinations is a finite set, say $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$, and each \mathbf{v}_j can be written in the form $\mathbf{v}_j = \sum_{i=1}^k b_{ij} \mathbf{w}_i$. Replacing the \mathbf{v}_j 's in the above expression for \mathbf{u} by these linear combinations, we obtain

$$\begin{aligned} \mathbf{u} &= \sum_{j=1}^m a_j \mathbf{v}_j \\ &= \sum_{j=1}^m a_j \left(\sum_{i=1}^k b_{ij} \mathbf{w}_i \right) \\ &= \sum_{j=1}^m \left(\sum_{i=1}^k a_j b_{ij} \mathbf{w}_i \right) \\ &= \sum_{i=1}^k \left(\sum_{j=1}^m a_j b_{ij} \mathbf{w}_i \right) \\ &= \sum_{i=1}^k \left(\sum_{j=1}^m a_j b_{ij} \right) \mathbf{w}_i. \end{aligned}$$

Letting $c_i = \sum_{j=1}^m a_j b_{ij}$ for $i = 1, 2, \dots, k$, we have $\mathbf{u} = \sum_{i=1}^k c_i \mathbf{w}_i$. Thus \mathbf{u} is dependent on \mathcal{C} .

Since \mathbf{u} was arbitrarily chosen in \mathcal{A} , \mathcal{A} is dependent on \mathcal{C} and the theorem is proved. ■

Exercises 1.2

- For any pair of positive integers i and j , the symbol δ_{ij} is defined by $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$. This symbol is known as the **Kronecker delta**.

- (a) Find the value of $\sum_{i=1}^n \left(\sum_{j=1}^m \delta_{ij} \right)$.
- (b) Find the value of $\sum_{i=1}^n \left(\sum_{j=1}^n (1 - \delta_{ij}) \right)$.
- (c) Find the value of $\sum_{i=1}^5 \left(\sum_{j=1}^4 (-1)^{\delta_{ij}} \right)$.
- (d) Find the value of $\sum_{j=1}^n \delta_{ij} \delta_{jk}$.
2. Prove the remaining parts of Theorem 1.3.
3. In each case, determine whether or not the given vector \mathbf{v} is linearly dependent on the given set \mathcal{A} .
- (a) $\mathbf{v} = (-2, 1, 4)$, $\mathcal{A} = \{(1, 1, 1), (0, 1, 1), (0, 0, 1)\}$
- (b) $\mathbf{v} = (-4, 4, 2)$, $\mathcal{A} = \{(2, 1, -3), (1, -1, 3)\}$
- (c) $\mathbf{v} = (1, 2, 1)$, $\mathcal{A} = \{(1, 0, -2), (0, 2, 1), (1, 2, -1), (-1, 2, 3)\}$
- (d) $\mathbf{v} = (2, 13, -5)$, $\mathcal{A} = \{(1, 2, -1), (3, 6, -3), (-1, 1, 0), (0, 6, -2), (2, 4, -2)\}$
- (e) $\mathbf{v} = (0, 1, 2, 0)$, $\mathcal{A} = \{(1, 0, -1, 1), (-2, 0, 2, -2), (1, 1, 1, 1), (2, 1, 0, 2)\}$
- (f) $\mathbf{v} = (2, 3, 5, 5)$, $\mathcal{A} = \{(0, 1, 1, 1), (1, 0, 1, 1), (1, -1, 0, 0), (1, 1, 2, 2)\}$
4. Assuming the properties stated in Theorem 1.3, prove the following statements.
- (a) The vector $\mathbf{0}$ in property (3) is unique.
- (b) The vector $-\mathbf{u}$ in \mathbf{R}^n which satisfies $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ is unique.
- (c) $-\mathbf{u} = (-1)\mathbf{u}$.
- (d) The vector $\mathbf{u} - \mathbf{v}$ is by definition the vector \mathbf{w} such that $\mathbf{v} + \mathbf{w} = \mathbf{u}$. Prove that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$.
5. Determine whether or not the given set \mathcal{A} is linearly dependent.
- (a) $\mathcal{A} = \{(1, 0, -2), (0, 2, 1), (-1, 2, 3)\}$
- (b) $\mathcal{A} = \{(1, 4, 3), (2, 12, 6), (5, 21, 15), (0, 2, -1)\}$
- (c) $\mathcal{A} = \{(1, 2, -1), (-1, 1, 0), (1, 3, -1)\}$
- (d) $\mathcal{A} = \{(1, 0, 1, 2), (2, 1, 0, 0), (4, 5, 6, 0), (1, 1, 1, 0)\}$
6. Show that the given set is linearly dependent and write one of the vectors as a linear combination of the remaining vectors.
- (a) $\{(2, 1, 0), (1, 1, 0), (0, 1, 1), (-1, 1, 1)\}$
- (b) $\{(1, 2, 1, 0), (3, -4, 5, 6), (2, -1, 3, 3), (-2, 6, -4, -6)\}$
7. Show that any vector in \mathbf{R}^3 is dependent on the set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, $\mathbf{e}_3 = (0, 0, 1)$.

8. Show that every vector in \mathbf{R}^3 is dependent on the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ where $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (1, 1, 0)$, and $\mathbf{u}_3 = (1, 1, 1)$.
9. Show that there is one vector in the set

$$\{(1, 1, 0), (0, 1, 1), (1, 0, -1), (1, 0, 1)\}$$
 that cannot be written as a linear combination of the other vectors in the set.
10. Prove that if the set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ of vectors in \mathbf{R}^n contains the zero vector, it is linearly dependent.
11. Prove that a set consisting of exactly one nonzero vector is linearly independent.
12. Prove that a set of two vectors in \mathbf{R}^n is linearly dependent if and only if one of the vectors is a scalar multiple of the other.
13. Prove that a set of nonzero vectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ in \mathbf{R}^n is linearly dependent if and only if some \mathbf{u}_r is a linear combination of the preceding vectors.
14. Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ be a linearly independent set of vectors in \mathbf{R}^n .
 - (a) Prove that the set $\{\mathbf{u}_1 - \mathbf{u}_2, \mathbf{u}_2 - \mathbf{u}_3, \mathbf{u}_1 + \mathbf{u}_3\}$ is linearly independent.
 - (b) Prove that the set $\{\mathbf{u}_1 - \mathbf{u}_2, \mathbf{u}_2 - \mathbf{u}_3, \mathbf{u}_1 - \mathbf{u}_3\}$ is linearly dependent.
15. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_{r-1}, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_k\}$ be a linearly independent set of k vectors in \mathbf{R}^n , and let $\mathbf{u}'_r = \sum_{j=1}^k a_j \mathbf{u}_j$ with $a_r \neq 0$. Prove that $\{\mathbf{u}_1, \dots, \mathbf{u}_{r-1}, \mathbf{u}'_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_k\}$ is linearly independent.
16. Let \emptyset denote the empty set of vectors in \mathbf{R}^n . Determine whether or not \emptyset is linearly dependent, and justify your conclusion.
17. Prove that any subset of a linearly independent set $\mathcal{A} \subseteq \mathbf{R}^n$ is linearly independent.
18. Let $\mathcal{A} \subseteq \mathbf{R}^n$. Prove that if \mathcal{A} contains a linearly dependent subset, then \mathcal{A} is linearly dependent.

1.3 Subspaces of \mathbf{R}^n

There are many subsets of \mathbf{R}^n that possess the properties stated in Theorem 1.3. A study of these subsets furnishes a great deal of insight into the structure of the spaces \mathbf{R}^n , and is of vital importance in subsequent material.

Definition 1.9 A set \mathbf{W} is a subspace of \mathbf{R}^n if \mathbf{W} is contained in \mathbf{R}^n and if the properties of Theorem 1.3 are valid in \mathbf{W} . That is,

1. $\mathbf{u} + \mathbf{v} \in \mathbf{W}$ for all \mathbf{u}, \mathbf{v} in \mathbf{W} .

2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in \mathbf{W} .
3. $\mathbf{0} \in \mathbf{W}$.
4. For each $\mathbf{u} \in \mathbf{W}$, $-\mathbf{u}$ is in \mathbf{W} .
5. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all \mathbf{u}, \mathbf{v} in \mathbf{W} .
6. $a\mathbf{u} \in \mathbf{W}$ for all $a \in \mathbf{R}$ and all $\mathbf{u} \in \mathbf{W}$.
7. $a(b\mathbf{u}) = (ab)\mathbf{u}$ for all $a, b \in \mathbf{R}$ and all $\mathbf{u} \in \mathbf{W}$.
8. $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$ for all $a \in \mathbf{R}$ and all $\mathbf{u}, \mathbf{v} \in \mathbf{W}$.
9. $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$ for all $a, b \in \mathbf{R}$ and all $\mathbf{u} \in \mathbf{W}$.
10. $1 \cdot \mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in \mathbf{W}$.

Before considering some examples of subspaces, we observe that the list of properties in Definition 1.9 can be shortened a great deal. For example, properties (2), (5), (7), (8), (9), and (10) are valid throughout \mathbf{R}^n , and hence are automatically satisfied in any subset of \mathbf{R}^n . Thus a subset \mathbf{W} of \mathbf{R}^n is a subspace if and only if properties (1), (3), (4), and (6) hold in \mathbf{W} . This reduces the amount of labor necessary in order to determine whether or not a given subset is a subspace, but an even more practical test is given in the following theorem.

Theorem 1.10 *Let \mathbf{W} be a subset of \mathbf{R}^n . Then \mathbf{W} is a subspace of \mathbf{R}^n if and only if the following conditions hold:*

- (i) \mathbf{W} is nonempty;
- (ii) for any $a, b \in \mathbf{R}$ and any $\mathbf{u}, \mathbf{v} \in \mathbf{W}$, $a\mathbf{u} + b\mathbf{v} \in \mathbf{W}$.

Proof. Suppose that \mathbf{W} is a subspace of \mathbf{R}^n . Then \mathbf{W} is nonempty, since $\mathbf{0} \in \mathbf{W}$ by property (3). Let a and b be any elements of \mathbf{R} , and let \mathbf{u} and \mathbf{v} be elements of \mathbf{W} . By property (6), $a\mathbf{u}$ and $b\mathbf{v}$ are in \mathbf{W} . Hence $a\mathbf{u} + b\mathbf{v} \in \mathbf{W}$ by property (1), and condition (ii) is satisfied.

Suppose, on the other hand, that conditions (i) and (ii) are satisfied in \mathbf{W} . From our discussion above, it is necessary only to show that properties (1), (3), (4), and (6) are valid in \mathbf{W} . Since \mathbf{W} is nonempty, there is at least one vector $\mathbf{u} \in \mathbf{W}$. By condition (ii), $1\mathbf{u} + (-1)\mathbf{u} = \mathbf{0} \in \mathbf{W}$, and property (3) is valid. Again by condition (ii), $(-1)\mathbf{u} = -\mathbf{u} \in \mathbf{W}$, so property (4) is valid. For any \mathbf{u}, \mathbf{v} in \mathbf{W} , $1 \cdot \mathbf{u} + 1 \cdot \mathbf{v} = \mathbf{u} + \mathbf{v} \in \mathbf{W}$, and property (1) is valid. For any $a \in \mathbf{R}$ and any $\mathbf{u} \in \mathbf{W}$, $a\mathbf{u} + 0 \cdot \mathbf{u} = a\mathbf{u} \in \mathbf{W}$. Thus property (6) is valid, and \mathbf{W} is a subspace of \mathbf{R}^n . ■

Example 1 □ The following list gives several examples of subspaces of \mathbf{R}^n . We will prove that the third set in the list forms a subspace and leave it as an exercise to verify that the remaining sets are subspaces of \mathbf{R}^n .

1. The set $\{\mathbf{0}\}$, called the **zero subspace** of \mathbf{R}^n .
2. The set of all scalar multiples of a fixed vector $\mathbf{u} \in \mathbf{R}^n$.

3. The set of all vectors that are dependent on a given set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ of vectors in \mathbf{R}^n .
4. The set of all vectors (x_1, x_2, \dots, x_n) in \mathbf{R}^n that satisfy a fixed equation

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

For $n = 2$ or $n = 3$, this example has a simple geometric interpretation, as we shall see later.

5. The set of all vectors (x_1, x_2, \dots, x_n) in \mathbf{R}^n that satisfy the system of equations

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

Proof for the third set. Let \mathbf{W} be the set of all vectors that are dependent on the set $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ of vectors in \mathbf{R}^n . From the discussion in the paragraph following Definition 1.4, we know that \mathbf{W} is the set of all vectors that can be written in the form $\sum_{i=1}^k a_i \mathbf{u}_i$. The set \mathbf{W} is nonempty since

$$0 \cdot \mathbf{u}_1 + 0 \cdot \mathbf{u}_2 + \cdots + 0 \cdot \mathbf{u}_k = \mathbf{0} \text{ is in } \mathbf{W}.$$

Let \mathbf{u}, \mathbf{v} be arbitrary vectors in \mathbf{W} , and let a, b be arbitrary scalars. From the definition of \mathbf{W} , there are scalars c_i and d_i such that

$$\mathbf{u} = \sum_{i=1}^k c_i \mathbf{u}_i \text{ and } \mathbf{v} = \sum_{i=1}^k d_i \mathbf{u}_i.$$

Thus we have

$$\begin{aligned} a\mathbf{u} + b\mathbf{v} &= a \left(\sum_{i=1}^k c_i \mathbf{u}_i \right) + b \left(\sum_{i=1}^k d_i \mathbf{u}_i \right) \\ &= \sum_{i=1}^k ac_i \mathbf{u}_i + \sum_{i=1}^k bd_i \mathbf{u}_i \\ &= \sum_{i=1}^k (ac_i + bd_i) \mathbf{u}_i \\ &= \sum_{i=1}^k (a+c)(d+d_i) \mathbf{u}_i. \end{aligned}$$

The last expression is a linear combination of the elements in \mathcal{A} , and consequently $a\mathbf{u} + b\mathbf{v}$ is an element of \mathbf{W} since \mathbf{W} contains all vectors that are dependent on \mathcal{A} . Both conditions in Theorem 1.10 have been verified, and therefore \mathbf{W} is a subspace of \mathbf{R}^n . ■

Our next theorem has a connection with the sets listed as 4 and 5 in Example 1 that should be investigated by the student. In this theorem, we are confronted with a situation which involves a collection that is not necessarily finite. In situations such as this, it is desirable to have available a notational convenience known as *indexing*.

Let \mathcal{L} and \mathcal{T} be nonempty sets. Suppose that with each $\lambda \in \mathcal{L}$ there is associated a unique element t_λ of \mathcal{T} , and that each element of \mathcal{T} is associated with at least one $\lambda \in \mathcal{L}$. (That is, suppose that there is given a function with domain \mathcal{L} and range \mathcal{T} .) Then we say that the set \mathcal{T} is *indexed* by the set \mathcal{L} , and refer to \mathcal{L} as an *index set*. We write $\{t_\lambda \mid \lambda \in \mathcal{L}\}$ to denote that the collection of t_λ 's is indexed by \mathcal{L} .

If $\{\mathcal{M}_\lambda \mid \lambda \in \mathcal{L}\}$ is a collection of sets \mathcal{M}_λ indexed by \mathcal{L} , then $\bigcup_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda$ indicates the union of this collection of sets. Thus $\bigcup_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda$ is the set of all elements that are contained in at least one \mathcal{M}_λ . Similarly, $\bigcap_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda$ denotes the intersection of the sets \mathcal{M}_λ , and consists of all elements that are in every \mathcal{M}_λ .

Theorem 1.11 *The intersection of any nonempty collection of subspaces of \mathbf{R}^n is a subspace of \mathbf{R}^n .*

Proof. Let $\{\mathcal{S}_\lambda \mid \lambda \in \mathcal{L}\}$ be any nonempty collection of subspaces \mathcal{S}_λ of \mathbf{R}^n , and let $\mathcal{W} = \bigcap_{\lambda \in \mathcal{L}} \mathcal{S}_\lambda$. Now $\mathbf{0} \in \mathcal{S}_\lambda$ for each $\lambda \in \mathcal{L}$, so $\mathbf{0} \in \mathcal{W}$ and \mathcal{W} is nonempty. Let $a, b \in \mathbf{R}$, and let $\mathbf{u}, \mathbf{v} \in \mathcal{W}$. Since each of \mathbf{u} and \mathbf{v} is in \mathcal{S}_λ for every $\lambda \in \mathcal{L}$, $a\mathbf{u} + b\mathbf{v} \in \mathcal{S}_\lambda$ for every $\lambda \in \mathcal{L}$. Hence $a\mathbf{u} + b\mathbf{v} \in \mathcal{W}$, and \mathcal{W} is a subspace by Theorem 1.10. ■

Thus the operation of intersection can be used to construct new subspaces from given subspaces.

There are all sorts of subsets in a given subspace \mathbf{W} of \mathbf{R}^n . Some of these have the important property of being *spanning sets* for \mathbf{W} , or sets that *span* \mathbf{W} . The following definition describes this property.

Definition 1.12 *Let \mathbf{W} be a subspace of \mathbf{R}^n . A nonempty set \mathcal{A} of vectors in \mathbf{R}^n spans \mathbf{W} if $\mathcal{A} \subseteq \mathbf{W}$ and if every vector in \mathbf{W} is a linear combination of vectors in \mathcal{A} . By definition, the empty set \emptyset spans the zero subspace.*

Intuitively, the word *span* is a natural choice in Definition 1.12 because a spanning set \mathcal{A} reaches across (hence *spans*) the entire subspace when all linear combinations of \mathcal{A} are formed.

Example 2 □ We shall show that each of the following sets spans \mathbf{R}^3 :

$$\begin{aligned}\mathcal{E}_3 &= \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}, \\ \mathcal{A} &= \{(1, 1, 1), (0, 1, 1), (0, 0, 1)\}.\end{aligned}$$

The set \mathcal{E}_3 spans \mathbf{R}^3 since an arbitrary (x, y, z) in \mathbf{R}^3 can be written as

$$(x, y, z) = x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1).$$

In calculus texts, the vectors in \mathcal{E}_3 are labeled with the standard notation

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1)$$

and this is used to write $(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

We can take advantage of the way zeros are placed in the vectors of \mathcal{A} and write

$$(x, y, z) = x(1, 1, 1) + (y - x)(0, 1, 1) + (z - y)(0, 0, 1).$$

The coefficients in this equation can be found by inspection if we start with the first component and work from left to right. ■

We shall see in Theorem 1.15 that the concept of a spanning set is closely related to the set $\langle \mathcal{A} \rangle$ defined as follows.

Definition 1.13 *For any nonempty set \mathcal{A} of vectors in \mathbf{R}^n , $\langle \mathcal{A} \rangle$ is the set of all vectors in \mathbf{R}^n that are dependent on \mathcal{A} . By definition, $\langle \emptyset \rangle$ is the zero subspace $\{\mathbf{0}\}$.*

Thus, for $\mathcal{A} \neq \emptyset$, $\langle \mathcal{A} \rangle$ is the set of all vectors \mathbf{u} that can be written as $\mathbf{u} = \sum_{j=1}^k a_j \mathbf{u}_j$ with a_j in \mathbf{R} and \mathbf{u}_j in \mathcal{A} . Since any \mathbf{u} in \mathcal{A} is dependent on \mathcal{A} , the subset relation $\mathcal{A} \subseteq \langle \mathcal{A} \rangle$ always holds.

In Example 1 of this section, the third set listed is $\langle \mathcal{A} \rangle$ where $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ in \mathbf{R}^n . When the notation $\langle \mathcal{A} \rangle$ is combined with the set notation for this \mathcal{A} , the result is a somewhat cumbersome notation:

$$\langle \mathcal{A} \rangle = \langle \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\} \rangle.$$

We make a notational agreement for situations like this to simply write

$$\langle \mathcal{A} \rangle = \langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \rangle.$$

For example, if $\mathcal{A} = \{(1, 3, 7), (2, 0, 6)\}$, we would write

$$\langle \mathcal{A} \rangle = \langle (1, 3, 7), (2, 0, 6) \rangle$$

instead of $\langle \mathcal{A} \rangle = \langle \{(1, 3, 7), (2, 0, 6)\} \rangle$ to indicate the set of all vectors that are dependent on \mathcal{A} .

It is proved in Example 1 that, for a finite subset $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ of \mathbf{R}^n , the set $\langle \mathcal{A} \rangle$ is a subspace of \mathbf{R}^n . The next theorem generalizes this result to an arbitrary subset \mathcal{A} of \mathbf{R}^n .

Theorem 1.14 *For any subset \mathcal{A} of \mathbf{R}^n , $\langle \mathcal{A} \rangle$ is a subspace of \mathbf{R}^n .*

Proof. If \mathcal{A} is empty, then $\langle \mathcal{A} \rangle$ is the zero subspace by definition.

Suppose \mathcal{A} is nonempty. Then $\langle \mathcal{A} \rangle$ is nonempty, since $\mathcal{A} \subseteq \langle \mathcal{A} \rangle$. Let $a, b \in \mathbf{R}$, and let $\mathbf{u}, \mathbf{v} \in \langle \mathcal{A} \rangle$. Now $a\mathbf{u} + b\mathbf{v}$ is dependent on $\{\mathbf{u}, \mathbf{v}\}$ and each of \mathbf{u} and \mathbf{v} is dependent on \mathcal{A} . Hence $\{a\mathbf{u} + b\mathbf{v}\}$ is dependent on $\{\mathbf{u}, \mathbf{v}\}$ and $\{\mathbf{u}, \mathbf{v}\}$ is dependent on \mathcal{A} . Therefore $\{a\mathbf{u} + b\mathbf{v}\}$ is dependent on \mathcal{A} by Theorem 1.8. Thus $a\mathbf{u} + b\mathbf{v} \in \langle \mathcal{A} \rangle$, and $\langle \mathcal{A} \rangle$ is a subspace. ■

We state the relation between Definitions 1.12 and 1.13 as a theorem, even though the proof is almost trivial.

Theorem 1.15 *Let \mathbf{W} be a subspace of \mathbf{R}^n , and let \mathcal{A} be a subset of \mathbf{R}^n . Then \mathcal{A} spans \mathbf{W} if and only if $\langle \mathcal{A} \rangle = \mathbf{W}$.*

Proof. The statement is trivial in case $\mathcal{A} = \emptyset$. Suppose, then, that \mathcal{A} is nonempty.

If \mathcal{A} spans \mathbf{W} , then every vector in \mathbf{W} is dependent on \mathcal{A} , so $\mathbf{W} \subseteq \langle \mathcal{A} \rangle$. Now $\mathcal{A} \subseteq \mathbf{W}$, and repeated application of condition (ii), Theorem 1.10, yields the fact that any linear combination of vectors in \mathcal{A} is again a vector in \mathbf{W} . Thus, $\langle \mathcal{A} \rangle \subseteq \mathbf{W}$, and we have $\langle \mathcal{A} \rangle = \mathbf{W}$.

If $\langle \mathcal{A} \rangle = \mathbf{W}$, it follows at once that \mathcal{A} spans \mathbf{W} , and this completes the proof. ■

We will refer to $\langle \mathcal{A} \rangle$ as the **subspace spanned by \mathcal{A}** . Some of the notations used in various texts for this same subspace are

$$\text{span}(\mathcal{A}), \text{ sp}(\mathcal{A}), \text{ lin}(\mathcal{A}), \text{ Sp}(\mathcal{A}), \text{ and } S[\mathcal{A}].$$

We will use $\text{span}(\mathcal{A})$ or $\langle \mathcal{A} \rangle$ in this book.

Example 3 □ With \mathcal{A} and \mathbf{W} as follows, we shall determine whether or not \mathcal{A} spans \mathbf{W} .

$$\mathcal{A} = \{(1, 0, 1, 0), (0, 1, 0, 1), (2, 3, 2, 3)\}$$

$$\mathbf{W} = \langle(1, 0, 1, 0), (1, 1, 1, 1), (0, 1, 1, 1)\rangle$$

We first check to see whether $\mathcal{A} \subseteq \mathbf{W}$. That is, we check to see if each vector in \mathcal{A} is a linear combination of the vectors listed in the spanning set for \mathbf{W} . By inspection, we see that $(1, 0, 1, 0)$ is listed in the spanning set, and

$$(0, 1, 0, 1) = (-1)(1, 0, 1, 0) + (1)(1, 1, 1, 1) + (0)(0, 1, 1, 1).$$

The linear combination is not as apparent with $(2, 3, 2, 3)$, so we place unknown coefficients in the equation

$$a_1(1, 0, 1, 0) + a_2(1, 1, 1, 1) + a_3(0, 1, 1, 1) = (2, 3, 2, 3).$$

Using the same procedure as in Example 2 of Section 1.2, we obtain the system of equations

$$\begin{aligned} a_1 + a_2 &= 2 \\ a_2 + a_3 &= 3 \\ a_1 + a_2 + a_3 &= 2 \\ a_2 + a_3 &= 3. \end{aligned}$$

It is then easy to find the solution $a_1 = -1, a_2 = 3, a_3 = 0$. That is,

$$(-1)(1, 0, 1, 0) + 3(1, 1, 1, 1) + 0(0, 1, 1, 1) = (2, 3, 2, 3).$$

Thus we have $\mathcal{A} \subseteq \mathbf{W}$.

We must now decide if every vector in \mathbf{W} is linearly dependent on \mathcal{A} , and Theorem 1.8 is of help here. We have \mathbf{W} dependent on the set

$$\mathcal{B} = \{(1, 0, 1, 0), (1, 1, 1, 1), (0, 1, 1, 1)\}.$$

If \mathcal{B} is dependent on \mathcal{A} , then \mathbf{W} is dependent on \mathcal{A} , by Theorem 1.8. On the other hand, if \mathcal{B} is not dependent on \mathcal{A} , then \mathbf{W} is not dependent on \mathcal{A} since $\mathcal{B} \subseteq \mathbf{W}$. Thus we need only check to see if each vector in \mathcal{B} is a linear combination of vectors in \mathcal{A} . We see that $(1, 0, 1, 0)$ is listed in \mathcal{A} , and

$$(1, 1, 1, 1) = (1, 0, 1, 0) + (0, 1, 0, 1).$$

Considering $(0, 1, 1, 1)$, we set up the equation

$$b_1(1, 0, 1, 0) + b_2(0, 1, 0, 1) + b_3(2, 3, 2, 3) = (0, 1, 1, 1).$$

This is equivalent to the system

$$\begin{aligned} b_1 + 2b_3 &= 0 \\ b_2 + 3b_3 &= 1 \\ b_1 + 2b_3 &= 1 \\ b_2 + 3b_3 &= 1. \end{aligned}$$

The first and third equations contradict each other, so there is no solution. Hence \mathbf{W} is not dependent on \mathcal{A} , and the set \mathcal{A} does not span \mathbf{W} . ■

We saw earlier in this section that the operation of intersection can be used to generate new subspaces from known subspaces. There is another operation, given in the following definition, that also can be used to form new subspaces from given subspaces.

Definition 1.16 Let \mathcal{S}_1 and \mathcal{S}_2 be nonempty subsets of \mathbf{R}^n . Then the **sum** $\mathcal{S}_1 + \mathcal{S}_2$ is the set of all vectors $\mathbf{u} \in \mathbf{R}^n$ that can be expressed in the form $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$, where $\mathbf{u}_1 \in \mathcal{S}_1$, and $\mathbf{u}_2 \in \mathcal{S}_2$.

Although this definition applies to nonempty subsets of \mathbf{R}^n generally, the more interesting situations are those in which both subsets are subspaces.

Theorem 1.17 *If \mathbf{W}_1 and \mathbf{W}_2 are subspaces of \mathbf{R}^n , then $\mathbf{W}_1 + \mathbf{W}_2$ is a subspace of \mathbf{R}^n .*

Proof. Clearly $\mathbf{0} + \mathbf{0} = \mathbf{0}$ is in $\mathbf{W}_1 + \mathbf{W}_2$, so that $\mathbf{W}_1 + \mathbf{W}_2$ is nonempty.

Let $a, b \in \mathbf{R}$, and let \mathbf{u} and \mathbf{v} be arbitrary elements in $\mathbf{W}_1 + \mathbf{W}_2$. From the definition of $\mathbf{W}_1 + \mathbf{W}_2$, it follows that there are vectors $\mathbf{u}_1, \mathbf{v}_1$ in \mathbf{W}_1 and $\mathbf{u}_2, \mathbf{v}_2$ in \mathbf{W}_2 such that $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ and $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$. Now $a\mathbf{u}_1 + b\mathbf{v}_1 \in \mathbf{W}_1$ and $a\mathbf{u}_2 + b\mathbf{v}_2 \in \mathbf{W}_2$ since \mathbf{W}_1 and \mathbf{W}_2 are subspaces. This gives

$$\begin{aligned} a\mathbf{u} + b\mathbf{v} &= a(\mathbf{u}_1 + \mathbf{u}_2) + b(\mathbf{v}_1 + \mathbf{v}_2) \\ &= (a\mathbf{u}_1 + b\mathbf{v}_1) + (a\mathbf{u}_2 + b\mathbf{v}_2), \end{aligned}$$

which clearly is an element of $\mathbf{W}_1 + \mathbf{W}_2$. Therefore, $\mathbf{W}_1 + \mathbf{W}_2$ is a subspace of \mathbf{R}^n by Theorem 1.10. ■

Example 4 □ Consider the subspaces \mathbf{W}_1 and \mathbf{W}_2 as follows:

$$\mathbf{W}_1 = \langle (1, -1, 0, 0), (0, 0, 0, 1) \rangle,$$

$$\mathbf{W}_2 = \langle (2, -2, 0, 0), (0, 0, 1, 0) \rangle.$$

Then \mathbf{W}_1 is the set of all vectors of the form

$$a_1(1, -1, 0, 0) + a_2(0, 0, 0, 1) = (a_1, -a_1, 0, a_2),$$

\mathbf{W}_2 is the set of all vectors of the form

$$b_1(2, -2, 0, 0) + b_2(0, 0, 1, 0) = (2b_1, -2b_1, b_2, 0),$$

and $\mathbf{W}_1 + \mathbf{W}_2$ is the set of all vectors of the form

$$\begin{aligned} a_1(1, -1, 0, 0) + a_2(0, 0, 0, 1) + b_1(2, -2, 0, 0) + b_2(0, 0, 1, 0) \\ = (a_1 + 2b_1, -a_1 - 2b_1, b_2, a_2). \end{aligned}$$

The last equation describes the vectors in $\mathbf{W}_1 + \mathbf{W}_2$, but it is not the most efficient description possible. Since $a_1 + 2b_1$ can take on any real number c_1 as a value, we see that $\mathbf{W}_1 + \mathbf{W}_2$ is the set of all vectors of the form

$$(c_1, -c_1, c_2, c_3). \blacksquare$$

Exercises 1.3

- Prove that each of the sets listed as 1, 2, 4, and 5 in Example 1 of this section is a subspace of \mathbf{R}^n .

2. Explain the connection between the sets listed as 4 and 5 in Example 1 and Theorem 1.11.
3. Let \mathcal{L} denote the set of all real numbers λ such that $0 \leq \lambda \leq 1$. For each $\lambda \in \mathcal{L}$, let \mathcal{M}_λ be the set of all $x \in \mathbf{R}$ such that $|x| < \lambda$. Find $\bigcup_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda$ and $\bigcap_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda$.
4. Let $\mathbf{V} = \mathbf{R}^3$.
 - (a) Exhibit a set of three vectors that spans \mathbf{V} .
 - (b) Exhibit a set of four vectors that spans \mathbf{V} .
5. Formulate Definition 1.4 and Definition 1.5 for an indexed set $\mathcal{A} = \{\mathbf{u}_\lambda \mid \lambda \in \mathcal{L}\}$ of vectors in \mathbf{R}^n .
6. For each given set \mathcal{A} and subspace \mathbf{W} , determine whether or not \mathcal{A} spans \mathbf{W} .
 - (a) $\mathcal{A} = \{(1, 0, 2), (-1, 1, -3)\}, \mathbf{W} = \langle(1, 0, 2)\rangle$
 - (b) $\mathcal{A} = \{(1, 0, 2), (-1, 1, -3)\}, \mathbf{W} = \langle(1, 1, 1), (2, -1, 5)\rangle$
 - (c) $\mathcal{A} = \{(1, -2), (-1, 3)\}, \mathbf{W} = \mathbf{R}^2$
 - (d) $\mathcal{A} = \{(2, 3, 0, -1), (2, 1, -1, 2)\}, \mathbf{W} = \langle(0, -2, -1, 3), (6, 7, -1, 0)\rangle$
 - (e) $\mathcal{A} = \{(3, -1, 2, 1), (4, 0, 1, 0)\}, \mathbf{W} = \langle(3, -1, 2, 1), (4, 0, 1, 0), (0, -1, 0, 1)\rangle$
 - (f) $\mathcal{A} = \{(3, -1, 2, 1), (4, 0, 1, 0)\}, \mathbf{W} = \langle(3, -1, 2, 1), (4, 0, 1, 0), (1, 1, -1, -1)\rangle$
7. Let $\mathbf{W}_1 = \langle(2, -1, 5)\rangle$ and $\mathbf{W}_2 = \langle(3, -2, 10)\rangle$. Determine whether or not the given vector \mathbf{u} is in $\mathbf{W}_1 + \mathbf{W}_2$.
 - (a) $\mathbf{u} = (-4, 1, -5)$
 - (b) $\mathbf{u} = (3, 2, -6)$
 - (c) $\mathbf{u} = (-5, 3, -2)$
 - (d) $\mathbf{u} = (3, 0, 0)$
8. Let $\mathcal{A} = \{(1, 2, 0), (1, 1, 1)\}$, and let $\mathcal{B} = \{(0, 1, -1), (0, 2, 2)\}$. By direct verification of the conditions of Theorem 1.10 in each case, show that $\langle\mathcal{A}\rangle$, $\langle\mathcal{B}\rangle$, and $\langle\mathcal{A}\rangle + \langle\mathcal{B}\rangle$ are subspaces of \mathbf{R}^n .
9. Let $\mathcal{A}_1 = \{\mathbf{u}, \mathbf{v}\}$ and $\mathcal{A}_2 = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ be subsets of \mathbf{R}^n with $\mathbf{w} = \mathbf{u} + \mathbf{v}$. Show that $\langle\mathcal{A}_1\rangle = \langle\mathcal{A}_2\rangle$ by use of Definition 1.13.
10. Prove or disprove: $\langle\mathcal{A} + \mathcal{B}\rangle = \langle\mathcal{A}\rangle + \langle\mathcal{B}\rangle$ for any nonempty subsets \mathcal{A} and \mathcal{B} of \mathbf{R}^n .
11. Let \mathbf{W} be a subspace of \mathbf{R}^n . Use condition (ii) of Theorem 1.10 and mathematical induction to show that any linear combination of vectors in \mathbf{W} is again a vector in \mathbf{W} .

12. If $\mathcal{A} \subseteq \mathbf{R}^n$, prove that $\langle \mathcal{A} \rangle$ is the intersection of all of the subspaces of \mathbf{R}^n that contain \mathcal{A} .
13. Let \mathbf{W} be a subspace of \mathbf{R}^n . Prove that $\langle \mathbf{W} \rangle = \mathbf{W}$.
14. Prove that $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ if and only if every vector in \mathcal{A} is dependent on \mathcal{B} and every vector in \mathcal{B} is dependent on \mathcal{A} .

1.4 Geometric Interpretations of \mathbf{R}^2 and \mathbf{R}^3

For $n = 1, 2$, or 3 , the vector space \mathbf{R}^n has a useful geometric interpretation in which a vector is identified with a directed line segment. This procedure is no doubt familiar to the student from the study of the calculus. In this section, we briefly review this interpretation of vectors and relate the geometric concepts to our work. The procedure can be described as follows.

For $n = 1$, the vector $\mathbf{v} = (x)$ is identified with the directed line segment on the real line that has its initial point (tail) at the origin and its terminal point (head) at x . This is shown in Figure 1.1.

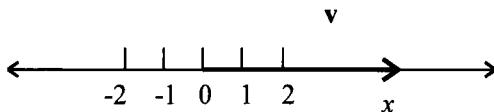


Figure 1.1

For $n = 2$ or $n = 3$, the vector $\mathbf{v} = (x, y)$ or $\mathbf{v} = (x, y, z)$ is identified with the directed line segment that has initial point at the origin and terminal point with rectangular coordinates given by the components of \mathbf{v} . This is shown in Figure 1.2.

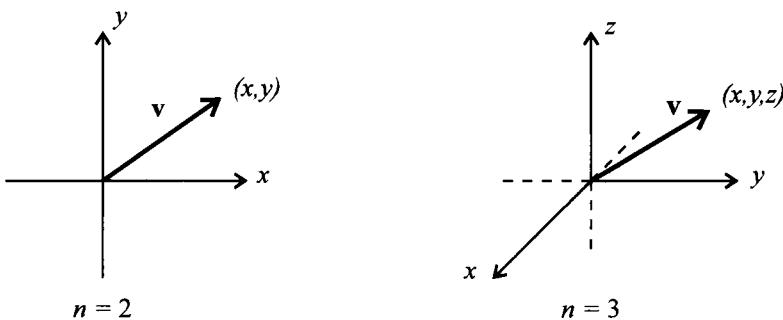


Figure 1.2

In making identifications of vectors with directed line segments, we shall follow the convention that any line segment with the same direction and the same length as the one we have described may be used to represent the same vector \mathbf{v} .

In the remainder of this section, we shall concentrate our attention on \mathbf{R}^3 . However, it should be observed that corresponding results are valid in \mathbf{R}^2 .

If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$, then $\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$. Thus, in the identification above, $\mathbf{u} + \mathbf{v}$ is the diagonal of a parallelogram which has \mathbf{u} and \mathbf{v} as two adjacent sides. This is illustrated in Figure 1.3. The vector $\mathbf{u} + \mathbf{v}$ can be drawn by placing the initial point of \mathbf{v} at the terminal point of \mathbf{u} and then drawing the directed line segment from the initial point of \mathbf{u} to the terminal point of \mathbf{v} . The “heads to tails” construction shown in Figure 1.3 is called the *parallelogram rule* for adding vectors.

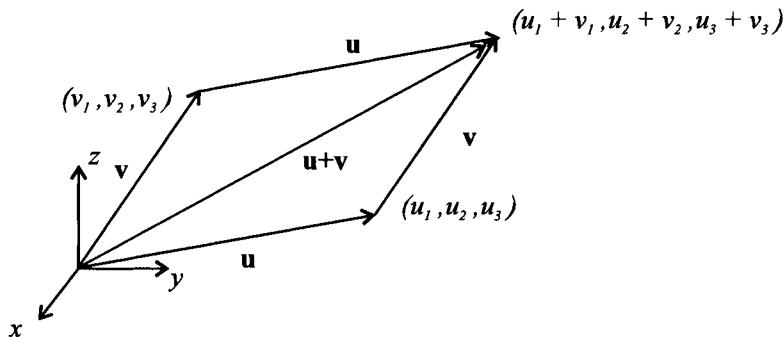


Figure 1.3

Now $\mathbf{u} - \mathbf{v}$ is the vector \mathbf{w} satisfying $\mathbf{v} + \mathbf{w} = \mathbf{u}$, so that $\mathbf{u} - \mathbf{v}$ is the directed line segment from the terminal point of \mathbf{v} to the terminal point of \mathbf{u} , as shown in Figure 1.4. Since $\mathbf{u} - \mathbf{v}$ has its head at \mathbf{u} and its tail at \mathbf{v} , this construction is sometimes referred to as the “heads minus tails” rule.

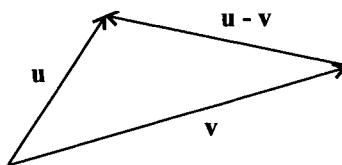


Figure 1.4

In approaching the geometric interpretation of subspaces, it is convenient to consider only those line segments with initial point at the origin. We shall do this in the following four paragraphs.

As mentioned earlier, the set of all scalar multiples of a fixed nonzero vector \mathbf{v} in \mathbf{R}^3 is a subspace of \mathbf{R}^3 . From our interpretation above, it is clear that this subspace $\langle \mathbf{v} \rangle$ consists of all vectors that lie on a line passing through the origin. This is shown in Figure 1.5.

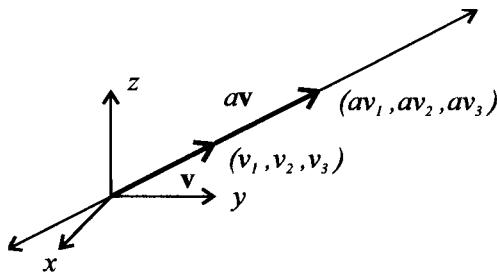


Figure 1.5

If $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2\}$ is independent, then \mathbf{v}_1 and \mathbf{v}_2 are not collinear. If P is any point in the plane determined by \mathbf{v}_1 and \mathbf{v}_2 , then the vector \overrightarrow{OP} from the origin to P is the diagonal of a parallelogram with sides parallel to \mathbf{v}_1 and \mathbf{v}_2 , as shown in Figure 1.6. In this case, the subspace $\langle \mathcal{A} \rangle$ consists of all vectors in the plane through the origin that contains \mathbf{v}_1 and \mathbf{v}_2 .

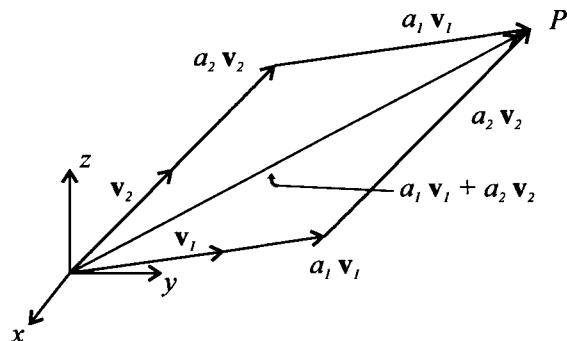


Figure 1.6

If $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent, then \mathbf{v}_1 and \mathbf{v}_2 are not collinear and \mathbf{v}_3 does not lie in the plane of \mathbf{v}_1 and \mathbf{v}_2 . Vectors $\mathbf{v}_1, \mathbf{v}_2$, and \mathbf{v}_3 of this type are shown in Figure 1.7. An arbitrary vector \overrightarrow{OP} in \mathbf{R}^3 is the diagonal of a parallelepiped with adjacent edges $a_1\mathbf{v}_1, a_2\mathbf{v}_2$, and $a_3\mathbf{v}_3$ as shown in Figure 1.7(a). The “heads to tails” construction along the edges of the parallelepiped indicated in Figure 1.7(b) shows that

$$\overrightarrow{OP} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3.$$

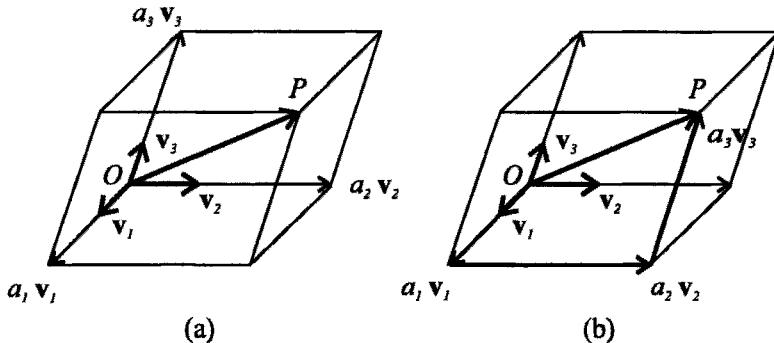


Figure 1.7

We shall prove in the next section that a subset of \mathbf{R}^3 cannot contain more than three linearly independent vectors. Thus the subspaces of \mathbf{R}^3 fall into one of four categories:

1. the origin;
2. a line through the origin;
3. a plane through the origin;
4. the entire space \mathbf{R}^3 .

It is shown in calculus courses that a plane in \mathbf{R}^3 consists of all points with rectangular coordinates (x, y, z) that satisfy a linear equation

$$ax + by + cz = d$$

in which at least one of a, b, c is not zero. A connection is made in the next example between this fact and our classification of subspaces.

Example 1 \square Consider the problem of finding an equation of the plane $\langle \mathcal{A} \rangle$ if $\mathcal{A} = \{(1, 2, 3), (3, 5, 1)\}$.

Now the line segments² extending from the origin $(0, 0, 0)$ to $(1, 2, 3)$ and from the origin to $(3, 5, 1)$ lie in the plane $\langle \mathcal{A} \rangle$, so the three points with coordinates $(0, 0, 0)$, $(1, 2, 3)$, and $(3, 5, 1)$ must all lie in the plane. This is shown in Figure 1.8.

²Note that ordered triples such as $(1, 2, 3)$ are doing double duty here. Sometimes they are coordinates of points, and sometimes they are vectors.

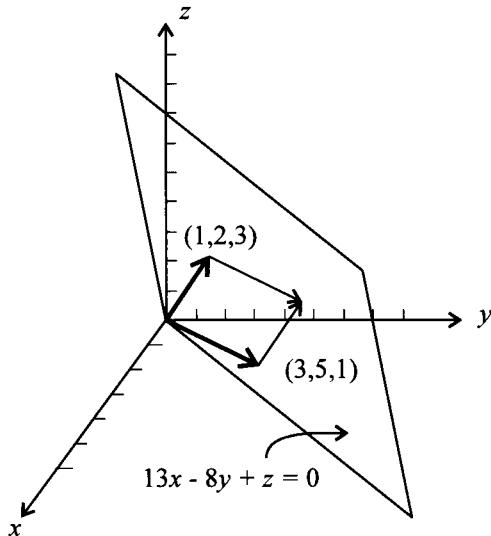


Figure 1.8

Since the points lie in the plane, their coordinates must satisfy the equation

$$ax + by + cz = d$$

of the plane. Substituting in order for $(0, 0, 0)$, $(1, 2, 3)$, and $(3, 5, 1)$, we obtain

$$\begin{aligned} 0 &= d \\ a + 2b + 3c &= d \\ 3a + 5b + c &= d . \end{aligned}$$

Using $d = 0$ and subtracting 3 times the second equation from the last one leads to

$$\begin{aligned} a + 2b + 3c &= 0 \\ -b - 8c &= 0 . \end{aligned}$$

Solving for a and b in terms of c , we obtain the following solutions

$$\begin{aligned} a &= 13c \\ b &= -8c \\ c &\text{ is arbitrary.} \end{aligned}$$

With $c = 1$, we have

$$13x - 8y + z = 0$$

as the equation of the plane $\langle \mathcal{A} \rangle$. ■

In the remainder of our discussion, we shall need the following definition, which applies to real coordinate spaces in general.

Definition 1.18 *For any two vectors $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$, the inner product (dot product, or scalar product) of \mathbf{u} and \mathbf{v} is defined by*

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{k=1}^n u_k v_k.$$

The inner product defined in this way is a natural extension of the following definitions that are used in the calculus:

$$(x_1, y_1) \cdot (x_2, y_2) = x_1 x_2 + y_1 y_2,$$

$$(x_1, y_1, z_1) \cdot (x_2, y_2, z_2) = x_1 x_2 + y_1 y_2 + z_1 z_2.$$

The distance formulas used in the calculus lead to formulas for the length $\|\mathbf{v}\|$ of a vector \mathbf{v} in \mathbf{R}^2 or \mathbf{R}^3 as follows:

$$\|(x, y)\| = \sqrt{x^2 + y^2},$$

$$\|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2}.$$

We extend these formulas for length to more general use in the next definition.

Definition 1.19 *For any $\mathbf{v} = (v_1, v_2, \dots, v_n)$ in \mathbf{R}^n , the length (or norm) of \mathbf{v} is denoted by $\|\mathbf{v}\|$ and is defined by*

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

The following properties are direct consequences of the definitions involved, and are presented as a theorem for convenient reference.

Theorem 1.20 *For any $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in \mathbf{R}^n and any a in \mathbf{R} :*

- (i) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- (ii) $(a\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (a\mathbf{v}) = a(\mathbf{u} \cdot \mathbf{v})$
- (iii) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
- (iv) $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$, or $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$
- (v) $\|a\mathbf{u}\| = |a| \|\mathbf{u}\|$.

Our next theorem gives a geometric interpretation of $\mathbf{u} \cdot \mathbf{v}$ in \mathbf{R}^2 or \mathbf{R}^3 . In the proof, we use the Law of Cosines from trigonometry: *If the sides and angles of an arbitrary triangle are labeled according to the pattern in Figure 1.9, then*

$$\cos C = \frac{a^2 + b^2 - c^2}{2ab}.$$

We state and prove the theorem for \mathbf{R}^3 , but the same result holds in \mathbf{R}^2 with a similar proof.

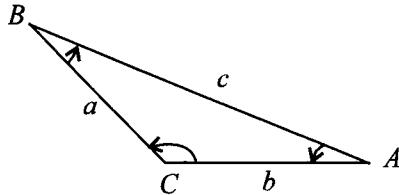


Figure 1.9

Theorem 1.21 For any two nonzero vectors $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ in \mathbf{R}^3 , $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$, where θ is the angle between the directions of \mathbf{u} and \mathbf{v} and $0^\circ \leq \theta \leq 180^\circ$.

Proof. Suppose first that $\theta = 0^\circ$ or $\theta = 180^\circ$. Then $\mathbf{v} = c\mathbf{u}$, where the scalar c is positive if $\theta = 0^\circ$ and negative if $\theta = 180^\circ$. We have

$$\begin{aligned}\|\mathbf{u}\| \|\mathbf{v}\| \cos \theta &= \|\mathbf{u}\| (|c| \cdot \|\mathbf{u}\|) \cos \theta \\ &= |c| \cos \theta \|\mathbf{u}\|^2 \\ &= c \|\mathbf{u}\|^2\end{aligned}$$

and

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} &= (u_1, u_2, u_3) \cdot (cu_1, cu_2, cu_3) \\ &= c(u_1^2 + u_2^2 + u_3^2) \\ &= c \|\mathbf{u}\|^2.\end{aligned}$$

Thus the theorem is true for $\theta = 0^\circ$ or $\theta = 180^\circ$.

Suppose now that $0^\circ < \theta < 180^\circ$. If $\mathbf{u} - \mathbf{v}$ is drawn from the head of \mathbf{v} to the head of \mathbf{u} , the vectors \mathbf{u}, \mathbf{v} and $\mathbf{u} - \mathbf{v}$ form a triangle with $\mathbf{u} - \mathbf{v}$ as the side opposite θ . (See Figure 1.10.)

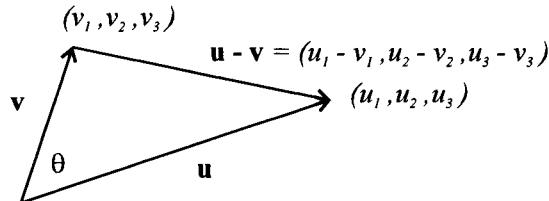


Figure 1.10

From the Law of Cosines, we have

$$\cos \theta = \frac{\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2}{2 \|\mathbf{u}\| \|\mathbf{v}\|}.$$

Thus

$$\begin{aligned}
 \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta &= \frac{1}{2}(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2) \\
 &= \frac{1}{2}\{u_1^2 + u_2^2 + u_3^2 + v_1^2 + v_2^2 + v_3^2 \\
 &\quad - [(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2]\} \\
 &= \mathbf{u} \cdot \mathbf{v}. \quad \blacksquare
 \end{aligned}$$

Corollary 1.22 In \mathbf{R}^2 or \mathbf{R}^3 , two nonzero vectors \mathbf{u} and \mathbf{v} are perpendicular (or orthogonal) if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

Proof. This follows at once from the fact that $\mathbf{u} \cdot \mathbf{v} = 0$ if and only if $\cos \theta = 0$. \blacksquare

Suppose that \mathbf{u} and \mathbf{v} are vectors in \mathbf{R}^2 or \mathbf{R}^3 represented by directed line segments with the same initial point, as shown in Figure 1.11. The vector labeled $\text{Proj}_{\mathbf{u}} \mathbf{v}$ in the figure is called the **vector projection of \mathbf{v} onto \mathbf{u}** . In order to construct $\text{Proj}_{\mathbf{u}} \mathbf{v}$, we first draw the straight line that contains \mathbf{u} . Next we draw a perpendicular segment joining the head of \mathbf{v} to the line containing \mathbf{u} . The vector from the initial point of \mathbf{u} to the foot of the perpendicular segment is $\text{Proj}_{\mathbf{u}} \mathbf{v}$. The vector $\text{Proj}_{\mathbf{u}} \mathbf{v}$ is also called the **vector component of \mathbf{v} along \mathbf{u}** .

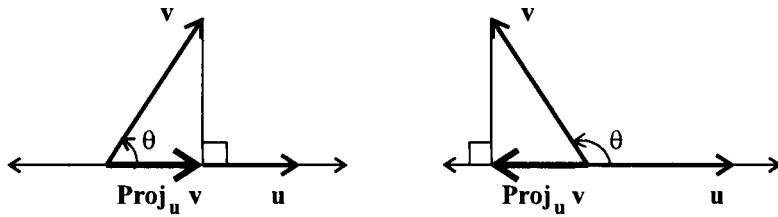


Figure 1.11

Let θ ($0^\circ \leq \theta \leq 180^\circ$) denote the angle between the directions of \mathbf{u} and \mathbf{v} as labeled in Figure 1.11. The number

$$d = \|\mathbf{v}\| \cos \theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|}$$

is called the **scalar projection of \mathbf{v} onto \mathbf{u}** or the **scalar component of \mathbf{v} along \mathbf{u}** . From Figure 1.11, it is clear that d is the length of $\text{Proj}_{\mathbf{u}} \mathbf{v}$ if $0^\circ \leq \theta \leq 90^\circ$ and d is the negative of the length of $\text{Proj}_{\mathbf{u}} \mathbf{v}$ if $\theta > 90^\circ$. Thus d can be regarded as the *directed length* of $\text{Proj}_{\mathbf{u}} \mathbf{v}$.

The geometry involved in having line segments perpendicular to each other breaks down in \mathbf{R}^n if $n > 3$. Even so, we extend the use of the word *orthogonal* to all \mathbf{R}^n . Two

vectors \mathbf{u}, \mathbf{v} in \mathbf{R}^n are called **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$. A set $\{\mathbf{u}_\lambda \mid \lambda \in \mathcal{L}\}$ of vectors in \mathbf{R}^n is an orthogonal set if $\mathbf{u}_{\lambda_1} \cdot \mathbf{u}_{\lambda_2} = 0$ whenever $\lambda_1 \neq \lambda_2$.

Exercises 1.4

1. Use Figures 1.2 and 1.3 as patterns and illustrate the parallelogram rule with the vectors $\mathbf{u} = (1, 6)$, $\mathbf{v} = (4, -4)$, and $\mathbf{u} + \mathbf{v}$ in an xy -coordinate system.
2. Use Figures 1.2 and 1.4 as patterns and sketch the vectors $\mathbf{u} = (5, 6)$, $\mathbf{v} = (2, -3)$, and $\mathbf{u} - \mathbf{v}$ in an xy -coordinate system.
3. For each $\lambda \in \mathbf{R}$, let \mathcal{M}_λ be the set of all points in the plane with rectangular coordinates (x, y) that satisfy $y = \lambda x$. Find $\bigcap_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda$ and $\bigcup_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda$.
4. Find the equation of the plane $\langle \mathcal{A} \rangle$ for the given set \mathcal{A} .
 - (a) $\mathcal{A} = \{(1, 0, 2), (2, -1, 1)\}$
 - (b) $\mathcal{A} = \{(1, 0, 2), (2, 1, 5)\}$
5. Find the lengths of the following vectors.
 - (a) $(3, -4, -12)$
 - (b) $(2, 3, 6)$
 - (c) $(1, -2, 4, 2)$
 - (d) $(2, 6, 0, -3)$
 - (e) $(1, -2, -4, 3)$
 - (f) $(3, 0, -5, 8)$
6. Determine x so that $(x, 2)$ is perpendicular to $(-3, 9)$.
7. A vector of length 1 is called a **unit vector**.
 - (a) Find a unit vector that has the same direction as $(3, -4, 12)$.
 - (b) Find a vector in the direction of $\mathbf{u} = (2, -3, 6)$ that has length 4 units.
8. Find each of the following scalar projections.
 - (a) The scalar projection of $(2, 3, 1)$ onto $(-1, -2, 4)$.
 - (b) The scalar projection of $(-3, 4, 12)$ onto $(2, 3, -6)$.
9. Find the length of the projection of the vector $(3, 4)$ onto a vector contained in the line $x - 2y = 0$.
10. Use projections to write the vector $(19, 22)$ as a linear combination of $(3, 4)$ and $(4, -3)$. (Note that $(3, 4)$ and $(4, -3)$ are perpendicular.)

- 11.** Let $\mathcal{A} = \{(1, 0, 2), (2, -1, 1)\}$ and let $\mathcal{B} = \{(1, 1, -1), (2, 1, 1)\}$.
- Find a set of vectors that spans $\langle \mathcal{A} \rangle \cap \langle \mathcal{B} \rangle$.
 - Find a set of vectors that spans $\langle \mathcal{A} \rangle + \langle \mathcal{B} \rangle$.
- 12.** Work problem 11 with $\mathcal{A} = \{(3, 1, -2), (-2, 1, 3)\}$ and $\mathcal{B} = \{(0, 1, 0), (1, 1, 1)\}$.
- 13.** Give an example of an orthogonal set of vectors in \mathbf{R}^n .
- 14.** Prove that an orthogonal set of nonzero vectors in \mathbf{R}^n is linearly independent.
- 15.** Let \mathbf{u} and \mathbf{v} be vectors in \mathbf{R}^n . Prove that $\|\mathbf{u}\| = \|\mathbf{v}\|$ if and only if $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ are orthogonal.
- 16.** Prove Theorem 1.20.
- 17.** The **cross product** $\mathbf{u} \times \mathbf{v}$ of two vectors $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ is given by

$$\begin{aligned}\mathbf{u} \times \mathbf{v} &= (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1) \\ &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{e}_1 + \begin{vmatrix} u_3 & u_1 \\ v_3 & v_1 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{e}_3,\end{aligned}$$

where $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, $\mathbf{e}_3 = (0, 0, 1)$. The symbolic determinant below is frequently used as a memory aid, since “expansion” about the first row yields the value of $\mathbf{u} \times \mathbf{v}$.

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

Prove the following facts concerning the cross product.

- $\mathbf{u} \times \mathbf{v}$ is perpendicular to each of \mathbf{u}, \mathbf{v} .
- $\mathbf{u} \times \mathbf{u} = \mathbf{0}$.
- $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$.
- $(a\mathbf{u}) \times \mathbf{v} = a(\mathbf{u} \times \mathbf{v}) = \mathbf{u} \times (a\mathbf{v})$.
- $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w})$.
- $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$.
- $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \|\mathbf{v}\| \sin \theta$, where θ is the angle between the directions of \mathbf{u} and \mathbf{v} , and $0^\circ \leq \theta \leq 180^\circ$.
- $\|\mathbf{u} \times \mathbf{v}\|$ is the area of a parallelogram with \mathbf{u} and \mathbf{v} as adjacent sides.

1.5 Bases and Dimension

We have seen in Section 1.3 that a subset \mathcal{A} of the subspace \mathbf{W} of \mathbf{R}^n may be a spanning set for \mathbf{W} and also that \mathcal{A} may be a linearly independent set. When both of these conditions are imposed, they form the requirements necessary for the subset to be a basis of \mathbf{W} .

Definition 1.23 A set \mathcal{B} of vectors is a **basis** of the subspace \mathbf{W} if (i) \mathcal{B} spans \mathbf{W} and (ii) \mathcal{B} is linearly independent.

The empty set \emptyset is regarded as being linearly independent since the condition for linear dependence in Definition 1.5 cannot be satisfied. Thus \emptyset is a basis of the zero subspace of \mathbf{R}^n .

Example 1 □ Some of our earlier work helps in providing examples concerning bases.

We saw in Example 2 of Section 1.3 that each of the sets

$$\mathcal{E}_3 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\},$$

$$\mathcal{A} = \{(1, 1, 1), (0, 1, 1), (0, 0, 1)\}$$

spans \mathbf{R}^3 . To show that the set \mathcal{A} is linearly independent, we set up the equation

$$c_1(1, 1, 1) + c_2(0, 1, 1) + c_3(0, 0, 1) = (0, 0, 0).$$

This equation leads directly to the following system of equations.

$$\begin{aligned} c_1 &= 0 \\ c_1 + c_2 &= 0 \\ c_1 + c_2 + c_3 &= 0 \end{aligned}$$

The only solution to this system is $c_1 = 0, c_2 = 0, c_3 = 0$, and therefore \mathcal{A} is linearly independent. It is even easier to see that \mathcal{E}_3 is linearly independent. Thus both \mathcal{E}_3 and \mathcal{A} are bases for \mathbf{R}^3 .

We saw in Example 4 of Section 1.2 that the set

$$\mathcal{A} = \{(1, 1, 8, 1), (1, 0, 3, 0), (3, 1, 14, 1)\}$$

is linearly dependent. It follows that this set \mathcal{A} is not a basis for the subspace

$$\mathbf{W} = \langle (1, 1, 8, 1), (1, 0, 3, 0), (3, 1, 14, 1) \rangle$$

that it spans. ■

We are concerned in much of our future work with indexed sets of vectors, and we use a restricted type of equality for this type of set. Two indexed sets \mathcal{A} and \mathcal{B} are **equal** if and only if they are indexed $\mathcal{A} = \{\mathbf{u}_\lambda \mid \lambda \in \mathcal{L}\}$ and $\mathcal{B} = \{\mathbf{v}_\lambda \mid \lambda \in \mathcal{L}\}$ by

the same index set \mathcal{L} such that $\mathbf{u}_\lambda = \mathbf{v}_\lambda$ for each $\lambda \in \mathcal{L}$. In particular, two finite sets $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ are equal if and only if they consist of the same vectors in the same order.

The equality described in the preceding paragraph is the one we shall use in the remainder of this book. For finite sets $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ of vectors, this equality is actually an equality of ordered sets. For example, if $\mathbf{u}_1 \neq \mathbf{u}_2$, then

$$\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\} \neq \{\mathbf{u}_2, \mathbf{u}_1, \dots, \mathbf{u}_k\}.$$

When we write

$$\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\},$$

this notation is meant to imply that \mathcal{A} is an ordered set with \mathbf{u}_1 as the first vector, \mathbf{u}_2 as the second vector, and so on. Moreover, we make a notational agreement for the remainder of this book that when we list the vectors in a set, this listing from left to right specifies their order. For instance, if we write

$$\mathcal{A} = \{(5, -1, 0, 2), (-4, 0, 3, 7), (1, -1, 3, 9)\}$$

this means that $(5, -1, 0, 2)$ is the first vector in \mathcal{A} , $(-4, 0, 3, 7)$ is the second vector in \mathcal{A} , and $(1, -1, 3, 9)$ is the third vector in \mathcal{A} . That is, the vectors in \mathcal{A} are automatically indexed with positive integers $1, 2, 3, \dots$ from left to right without this being stated.

Suppose now that $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is a basis of the subspace \mathbf{W} . Then \mathcal{B} spans \mathbf{W} , so that any $\mathbf{v} \in \mathbf{W}$ can be written as $\sum_{i=1}^k a_i \mathbf{v}_i$. As a matter of fact, this expression is *unique*. For if $\mathbf{v} = \sum_{i=1}^k b_i \mathbf{v}_i$ as well, we have

$$\sum_{i=1}^k a_i \mathbf{v}_i = \sum_{i=1}^k b_i \mathbf{v}_i$$

and therefore

$$\sum_{i=1}^k (a_i - b_i) \mathbf{v}_i = \mathbf{0}.$$

Since \mathcal{B} is linearly independent, this requires that $a_i - b_i = 0$ and $a_i = b_i$ for each i . In particular, we observe that if $\mathbf{u} = \mathbf{0}$, then all a_i are zero. This uniqueness of coefficients is not valid for spanning sets in general. The set \mathcal{B} in Example 1 of Section 1.2 furnishes an illustration of this fact.

Although a given subspace usually has many different bases, it happens that the number of vectors in different bases of the same subspace is always the same. The derivation of this result is the principal objective of this section.

Theorem 1.24 *Let \mathbf{W} be a subspace of \mathbf{R}^n . Suppose that a finite set $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ spans \mathbf{W} , and let \mathcal{B} be a linearly independent set of vectors in \mathbf{W} . Then \mathcal{B} contains at most r vectors.*

Proof. Let \mathbf{W} , \mathcal{A} , and \mathcal{B} be as described in the statement of the theorem. If \mathcal{B} contains less than r vectors, the theorem is true. Suppose then that \mathcal{B} contains at least r vectors, say $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\} \subseteq \mathcal{B}$.

Our proof of the theorem follows this plan: We shall show that each of the vectors \mathbf{v}_i in \mathcal{B} can be used in turn to replace a suitably chosen vector in \mathcal{A} , with \mathcal{A} dependent on the set obtained after each replacement. The replacement process finally leads to the fact that \mathcal{A} is dependent on the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$. We then prove that this set of r vectors must, in fact, be equal to \mathcal{B} .

Since \mathcal{A} spans \mathbf{W} , $\mathbf{v}_1 = \sum_{i=1}^r a_{i1} \mathbf{u}_i$ with at least one $a_{i1} \neq 0$ because $\mathbf{v}_1 \neq \mathbf{0}$. Without loss of generality, we may assume that $a_{11} \neq 0$ in the equation

$$\mathbf{v}_1 = a_{11} \mathbf{u}_1 + a_{21} \mathbf{u}_2 + \cdots + a_{r1} \mathbf{u}_r.$$

(This assumption is purely for notational convenience. We are assuming that the “suitably chosen” vector in \mathcal{A} is the first vector listed in \mathcal{A} .) The equation for \mathbf{v}_1 implies that

$$a_{11} \mathbf{u}_1 = \mathbf{v}_1 - a_{21} \mathbf{u}_2 - \cdots - a_{r1} \mathbf{u}_r$$

and therefore

$$\mathbf{u}_1 = \left(\frac{1}{a_{11}} \right) \mathbf{v}_1 + \left(-\frac{a_{21}}{a_{11}} \right) \mathbf{u}_2 + \cdots + \left(-\frac{a_{r1}}{a_{11}} \right) \mathbf{u}_r.$$

Thus \mathbf{u}_1 is dependent on $\{\mathbf{v}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$, and this clearly implies that \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$.

Assume now that \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_r\}$, where $1 \leq k < r$. Since \mathbf{W} is dependent on \mathcal{A} , then \mathbf{W} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_r\}$ by Theorem 1.8. In particular,

$$\mathbf{v}_{k+1} = \sum_{i=1}^k b_{i,k+1} \mathbf{v}_i + \sum_{i=k+1}^r a_{i,k+1} \mathbf{u}_i.$$

At least one of the coefficients $a_{i,k+1}$ of the \mathbf{u}_i must be nonzero. For if they were all zero, then \mathbf{v}_{k+1} would be a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$, and this would contradict the linear independence of \mathcal{B} . Without loss of generality, we may assume that $a_{k+1,k+1} \neq 0$. Hence we obtain

$$\mathbf{v}_{k+1} = \sum_{i=1}^k b_{i,k+1} \mathbf{v}_i + \sum_{i=k+1}^r a_{i,k+1} \mathbf{u}_i$$

where $a_{k+1,k+1} \neq 0$. This gives

$$\mathbf{u}_{k+1} = \sum_{i=1}^k \left(-\frac{b_{i,k+1}}{a_{k+1,k+1}} \right) \mathbf{v}_i + \left(\frac{1}{a_{k+1,k+1}} \right) \mathbf{v}_{k+1} + \sum_{i=k+2}^r \left(-\frac{a_{i,k+1}}{a_{k+1,k+1}} \right) \mathbf{u}_i.$$

Thus

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_r\}$$

is dependent on

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \mathbf{u}_{k+2}, \dots, \mathbf{u}_r\}.$$

Since \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_r\}$, Theorem 1.8 implies that \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \mathbf{u}_{k+2}, \dots, \mathbf{u}_r\}$.

Letting $k = 1, 2, \dots, r - 1$ in the iterative argument above, we see that each \mathbf{v}_i in \mathcal{B} can be used to replace a suitably chosen vector in \mathcal{A} until we obtain the fact that \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$. But \mathcal{B} is dependent on \mathcal{A} , so we have \mathcal{B} dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$. In particular, if \mathcal{B} had more than r elements, any \mathbf{v}_{r+1} in \mathcal{B} would be dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$. But this is impossible since \mathcal{B} is independent. Therefore, \mathcal{B} has r elements, and this completes the proof. ■

Corollary 1.25 *Any linearly independent set of vectors in \mathbf{R}^n contains at most n vectors.*

Proof. The set of n vectors $\mathbf{e}_1 = (1, 0, \dots, 0), \mathbf{e}_2 = (0, 1, \dots, 0), \dots, \mathbf{e}_n = (0, 0, \dots, 1)$ spans \mathbf{R}^n since $\mathbf{v} = (v_1, v_2, \dots, v_n)$ can be written as $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i$. The corollary follows at once from the theorem. ■

If we think in terms of geometric models as presented in Section 1.4, the next theorem seems intuitively obvious. It certainly seems obvious in \mathbf{R}^2 and \mathbf{R}^3 , and there is no reason to suspect the situation to be different in \mathbf{R}^n for other values of n . On the other hand, there is no *compelling* reason to suspect that the situation would *not* be different in \mathbf{R}^n for other values of n . At any rate, we refuse to accept such an important statement on faith or intuition, and insist that this result be validated by a logical argument based upon our development up to this point. This attitude or frame of mind is precisely what is meant when one refers to the “axiomatic method” of mathematics.

Theorem 1.26 *Every subspace of \mathbf{R}^n has a basis with a finite number of elements.*

Proof. Let \mathbf{W} be a subspace of \mathbf{R}^n . If $\mathbf{W} = \{\mathbf{0}\}$, then \emptyset is a basis of \mathbf{W} , and the theorem is true.

Suppose $\mathbf{W} \neq \{\mathbf{0}\}$. Then there is at least one nonzero \mathbf{v}_1 in \mathbf{W} . The set $\{\mathbf{v}_1\}$ is linearly independent by Problem 11 of Exercises 1.2. Thus, there are nonempty subsets of \mathbf{W} that are linearly independent, and Corollary 1.25 shows that each of these subsets contains at most n elements. Let \mathcal{T} be the set of all positive integers t such that \mathbf{W} contains a set of t linearly independent vectors. Then any t in \mathcal{T} satisfies the inequality $1 \leq t \leq n$. Let r be the largest integer in \mathcal{T} , and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ be a linearly independent set of r vectors in \mathbf{W} . We shall show that \mathcal{B} spans \mathbf{W} .

Let \mathbf{v} be any vector in \mathbf{W} . Then $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}\}$ is linearly dependent, from the choice of r . Thus, there are scalars $c_1, c_2, \dots, c_r, c_{r+1}$, not all zero, such that

$$\sum_{i=1}^r c_i \mathbf{v}_i + c_{r+1} \mathbf{v} = \mathbf{0}.$$

Now $c_{r+1} \neq 0$ since \mathcal{B} is independent. Therefore, $\mathbf{v} = \sum_{i=1}^r (-c_i/c_{r+1})\mathbf{v}_i$. This shows that \mathcal{B} spans \mathbf{W} , and hence is a basis of \mathbf{W} . ■

This brings us to the main result of this section.

Theorem 1.27 *Let \mathbf{W} be a subspace of \mathbf{R}^n , and let \mathcal{A} and \mathcal{B} be any two bases for \mathbf{W} . Then \mathcal{A} and \mathcal{B} have the same number of elements.*

Proof. If $\mathbf{W} = \{\mathbf{0}\}$, then each of \mathcal{A} and \mathcal{B} must be the empty set \emptyset , and the number of elements in both \mathcal{A} and \mathcal{B} is 0.

Suppose $\mathbf{W} \neq \{\mathbf{0}\}$. From Corollary 1.25, \mathcal{A} and \mathcal{B} are both finite. Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t\}$. Since \mathcal{A} spans \mathbf{W} and \mathcal{B} is linearly independent, $t \leq r$ by Theorem 1.24. But \mathcal{B} spans \mathbf{W} and \mathcal{A} is linearly independent, so $r \leq t$ by the same theorem. Thus, $t = r$. ■

Definition 1.28 *If \mathbf{W} is any subspace of \mathbf{R}^n , the number of vectors in a basis of \mathbf{W} is called the **dimension** of \mathbf{W} and is abbreviated as $\dim(\mathbf{W})$.*

The following theorem is somewhat trivial, but it serves to confirm that the preceding definition of dimension is consistent with our prior experience.

Theorem 1.29 *The dimension of \mathbf{R}^n is n .*

Proof. Consider the set $\mathcal{E}_n = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, where $\mathbf{e}_1 = (1, 0, \dots, 0)$, $\mathbf{e}_2 = (0, 1, \dots, 0)$, ..., $\mathbf{e}_n = (0, 0, 0, \dots, 1)$, are the same as in the proof of Corollary 1.25. It was noted in that proof that an arbitrary vector $\mathbf{v} = (v_1, v_2, \dots, v_n)$ can be written as

$$\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i,$$

and therefore \mathcal{E}_n spans \mathbf{R}^n .

The set \mathcal{E}_n is linearly independent since

$$\sum_{i=1}^n c_i \mathbf{e}_i = \mathbf{0} \text{ implies } (c_1, c_2, \dots, c_n) = (0, 0, \dots, 0)$$

and therefore all $c_i = 0$. Thus \mathcal{E}_n is a basis of \mathbf{R}^n with n elements, and it follows that \mathbf{R}^n has dimension n . ■

Definition 1.30 *The set $\mathcal{E}_n = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ used in the proof of Theorem 1.29 is called the **standard basis** of \mathbf{R}^n .*

The discussion of coefficients just before Theorem 1.24 explains why the coefficients c_1, c_2, \dots, c_n in $\mathbf{v} = \sum_{i=1}^n c_i \mathbf{v}_i$ are unique whenever $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a basis of \mathbf{R}^n . The scalars c_i are called the **coordinates** of \mathbf{v} relative to \mathcal{B} . For the special basis $\mathcal{E}_n = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, the components of \mathbf{v} are the same as the coordinates relative to \mathcal{E}_n .

Example 2 □ With the basis

$$\mathcal{E}_n = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\},$$

the coordinates of $\mathbf{v} = (x, y, z)$ relative to \mathcal{E}_3 are the same as the components x, y, z , respectively. But for the basis

$$\mathcal{B} = \{(1, 1, 1), (0, 1, 1), (0, 0, 1)\},$$

the coordinates of $\mathbf{v} = (x, y, z)$ relative to \mathcal{B} are the numbers $x, y - x, z - y$ because

$$(x, y, z) = x(1, 1, 1) + (y - x)(0, 1, 1) + (z - y)(0, 0, 1). \blacksquare$$

There are several types of problems involving “basis” and “dimension” that occur often in linear algebra. In dealing with a certain subspace \mathbf{W} , it may be necessary to find the dimension of \mathbf{W} , to find a basis of \mathbf{W} , or to determine whether or not a given set is a basis of \mathbf{W} . Frequently it is desirable to find a basis of \mathbf{W} that has certain specified properties. The fundamental techniques for attacking problems such as these are developed in the remainder of this section.

Theorem 1.31 *Every spanning set of a subspace \mathbf{W} of \mathbf{R}^n contains a basis of \mathbf{W} .*

Proof. Suppose that \mathbf{W} is a subspace of \mathbf{R}^n and that \mathcal{A} is a spanning set for \mathbf{W} . If \mathcal{A} is independent, then \mathcal{A} is a basis and the theorem is true. Consider now the possibilities when \mathcal{A} is dependent.

If $\mathcal{A} = \{\mathbf{0}\}$, then \mathbf{W} is the zero subspace. But we have already seen that \emptyset is a basis for the zero subspace and $\emptyset \subseteq \{\mathbf{0}\}$. Thus the theorem is true if $\mathcal{A} = \{\mathbf{0}\}$. If $\mathcal{A} \neq \{\mathbf{0}\}$, then there exists a \mathbf{v}_1 in \mathcal{A} that is nonzero. If \mathcal{A} is dependent on $\{\mathbf{v}_1\}$, we have a basis of \mathbf{W} . If \mathcal{A} is not dependent on $\{\mathbf{v}_1\}$, then there is a \mathbf{v}_2 in \mathcal{A} such that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent. This procedure can be repeated until we obtain a set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}, r \leq n$, that is linearly independent and spans \mathbf{W} . For if we did not obtain such a linearly independent set, we could continue until we had a linearly independent set in \mathbf{W} containing more than n vectors, and this would contradict Corollary 1.25 since $\mathbf{W} \subseteq \mathbf{R}^n$. The set \mathcal{B} thus obtained is the required basis of \mathbf{W} . ■

Although the details of the work would vary, the procedure given in the proof above provides a method for “refining” a basis from a given spanning set. This refinement procedure is demonstrated in the next example.

Example 3 □ With

$$\mathcal{A} = \{(1, 2, 1, 0), (3, -4, 5, 6), (2, -1, 3, 3), (-2, 6, -4, -6)\},$$

we shall use the procedure in the proof of Theorem 1.31 to find a basis of $\mathbf{W} = \langle \mathcal{A} \rangle$ that is contained in \mathcal{A} .

It is natural to start the procedure by choosing $\mathbf{v}_1 = (1, 2, 1, 0)$. We see that \mathcal{A} is not dependent on $\{\mathbf{v}_1\}$ because the second vector in \mathcal{A} , $(3, -4, 5, 6)$, is not a multiple of \mathbf{v}_1 . If we let $\mathbf{v}_2 = (3, -4, 5, 6)$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

We need to check now to see if \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2\}$. When we set up the equation

$$c_1(1, 2, 1, 0) + c_2(3, -4, 5, 6) = (2, -1, 3, 3),$$

this leads to the system of equations

$$\begin{aligned} c_1 + 3c_2 &= 2 \\ 2c_1 - 4c_2 &= -1 \\ c_1 + 5c_2 &= 3 \\ 6c_2 &= 3. \end{aligned}$$

The solution to this system is easily found to be $c_1 = c_2 = \frac{1}{2}$. Thus the third vector in \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2\}$. In similar fashion, we find that

$$(1)(1, 2, 1, 0) + (-1)(3, -4, 5, 6) = (-2, 6, -4, -6).$$

Thus \mathcal{A} is dependent on $\{\mathbf{v}_1, \mathbf{v}_2\}$, and

$$\{(1, 2, 1, 0), (3, -4, 5, 6)\}$$

is a basis for $\langle \mathcal{A} \rangle$.

The work we have done shows that $\langle \mathcal{A} \rangle$ has dimension 2. After checking to see that no vector in \mathcal{A} is a multiple of another vector in \mathcal{A} , we can then see that *any pair of vectors from \mathcal{A} forms a basis of $\langle \mathcal{A} \rangle$* . ■

In the proof of Theorem 1.31, we have seen how a basis of a subspace \mathbf{W} can be refined or extracted from an arbitrary spanning set. The spanning set is not *required* to be a finite set, but it could happen to be finite, or course. If a spanning set is finite, the natural refining procedure demonstrated in Example 3 can be given a simpler description: A basis of \mathbf{W} can be obtained by deleting all vectors in the spanning set that are linear combinations of the preceding vectors. Problem 13 of Exercises 1.2 assures us this will lead to an independent set, and Theorem 1.8 assures us this will lead to a spanning set. Thus a basis will result from the deletion of all vectors in a finite spanning set that are linear combinations of preceding vectors as listed in the spanning set.

Our next theorem looks at a procedure that in a sense is opposite to refining: It considers extending a linearly independent set to a basis.

Theorem 1.32 *Any linearly independent set in a subspace \mathbf{W} of \mathbf{R}^n can be extended to a basis of \mathbf{W} .*

Proof. Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ be a basis of \mathbf{W} , and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t\}$ be a linearly independent set in \mathbf{W} . If every \mathbf{u}_i is in $\langle \mathcal{B} \rangle$, then \mathcal{A} is dependent on \mathcal{B} . By Theorem 1.8, \mathbf{W} is dependent on \mathcal{B} , and \mathcal{B} is a basis of \mathbf{W} .

Suppose that some $\mathbf{u}_i \notin \langle \mathcal{B} \rangle$. Let k_1 be the smallest integer such that $\mathbf{u}_{k_1} \notin \langle \mathcal{B} \rangle$. Then $\mathcal{B}_1 = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t, \mathbf{u}_{k_1}\}$ is linearly independent. If each $\mathbf{u}_i \in \langle \mathcal{B}_1 \rangle$, then \mathcal{B}_1 spans \mathbf{W} and forms a basis. If some $\mathbf{u}_i \notin \langle \mathcal{B}_1 \rangle$, we repeat the process. After p steps ($1 \leq p \leq r$), we arrive at a set $\mathcal{B}_p = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t, \mathbf{u}_{k_1}, \mathbf{u}_{k_2}, \dots, \mathbf{u}_{k_p}\}$ such that all vectors of \mathcal{A} are dependent on \mathcal{B}_p . Thus \mathcal{B}_p spans \mathbf{W} . Since no vector in \mathcal{B}_p is a linear combination of the preceding vectors, \mathcal{B}_p is linearly independent by Problem 13 of Exercises 1.2. Therefore \mathcal{B}_p is a basis of \mathbf{W} . ■

In the next example, we follow the preceding proof to extend a linearly independent set to a basis.

Example 4 □ Given that

$$\mathcal{A} = \{(1, 2, 1, 3), (1, 0, 0, 0), (0, 0, 1, 0), (0, 1, 0, 1)\}$$

is a basis of \mathbf{R}^4 and that $\mathcal{B} = \{(1, 0, 1, 0), (0, 2, 0, 3)\}$ is linearly independent, we shall extend \mathcal{B} to a basis of \mathbf{R}^4 .

In keeping with the notational agreement made earlier in this section about indexing, we assume that \mathcal{A} and \mathcal{B} are indexed with positive integers from left to right so that the notation in the proof of Theorem 1.32 applies with

$$\mathbf{u}_1 = (1, 2, 1, 3), \mathbf{u}_2 = (1, 0, 0, 0), \mathbf{u}_3 = (0, 0, 1, 0), \mathbf{u}_4 = (0, 1, 0, 1)$$

and

$$\mathbf{v}_1 = (1, 0, 1, 0), \mathbf{v}_2 = (0, 2, 0, 3).$$

Following the proof of the theorem, we find that

$$(1, 2, 1, 3) = (1, 0, 1, 0) + (0, 2, 0, 3)$$

and thus \mathbf{u}_1 is dependent on \mathcal{B} . By inspection, we see that

$$(1, 0, 0, 0) = c_1(1, 0, 1, 0) + c_2(0, 2, 0, 3)$$

has no solution. Thus $\mathbf{u}_2 = (1, 0, 0, 0)$ is not in $\langle \mathcal{B} \rangle$, and $k_1 = 2$ is the smallest integer such that $\mathbf{u}_{k_1} \notin \langle \mathcal{B} \rangle$. Using the notation in the proof of the theorem, the set

$$\begin{aligned} \mathcal{B}_1 &= \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_2\} \\ &= \{(1, 0, 1, 0), (0, 2, 0, 3), (1, 0, 0, 0)\} \end{aligned}$$

is linearly independent. We check now for a vector \mathbf{u}_{k_2} in \mathcal{A} that is not in $\langle \mathcal{B}_1 \rangle$. We find that

$$(0, 0, 1, 0) = (1)(1, 0, 1, 0) + (0)(0, 2, 0, 3) + (-1)(1, 0, 0, 0)$$

and $\mathbf{u}_3 = (0, 0, 1, 0)$ is in $\langle \mathcal{B}_1 \rangle$, but the equation

$$(0, 1, 0, 1) = c_1(1, 0, 1, 0) + c_2(0, 2, 0, 3) + c_3(1, 0, 0, 0)$$

has no solution. Thus $\mathbf{u}_4 = (0, 1, 0, 1)$ is not in $\langle \mathcal{B}_1 \rangle$.

After two steps we have arrived at the set

$$\mathcal{B}_2 = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_2, \mathbf{u}_4\}$$

such that all vectors in \mathcal{A} are dependent on \mathcal{B}_2 . According to the proof of Theorem 1.32, this set \mathcal{B}_2 is a basis of \mathbf{R}^4 . ■

Our last two theorems in this section apply to the very special situations where the number of vectors in a set is the same as the dimension r of the subspace involved. For sets of this special type, only one of the conditions for a basis needs to be checked. This is the substance of the following two theorems.

Theorem 1.33 *Let \mathbf{W} be a subspace of \mathbf{R}^n of dimension r . Then a set of r vectors in \mathbf{W} is a basis of \mathbf{W} if and only if it is linearly independent.*

Proof. If a set of r vectors in \mathbf{W} is a basis, then it is linearly independent (and spans \mathbf{W} as well) according to Definition 1.23.

Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ be a set of r linearly independent vectors in \mathbf{W} . Then \mathcal{B} can be extended to a basis of \mathbf{W} , by Theorem 1.32. Now \mathbf{W} has a basis of r elements since it is of dimension r , and hence all bases of \mathbf{W} have r elements. In particular, the basis to which \mathcal{B} is extended has r elements, and therefore is the same as \mathcal{B} . ■

Theorem 1.34 *Let \mathbf{W} be a subspace of \mathbf{R}^n of dimension r . Then a set of r vectors in \mathbf{W} is a basis if and only if it spans \mathbf{W} .*

Proof. If a set of r vectors in \mathbf{W} is a basis, then it spans \mathbf{W} (and is linearly independent as well) by Definition 1.23.

Suppose $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is a set of r vectors which spans \mathbf{W} . According to Theorem 1.31, \mathcal{A} contains a basis of \mathbf{W} . But any basis of \mathbf{W} contains r vectors. Therefore, the basis contained in \mathcal{A} is not a proper subset of \mathcal{A} , and \mathcal{A} is a basis of \mathbf{W} . ■

Exercises 1.5

1. Given that each set \mathcal{A} below spans \mathbf{R}^3 , find a basis of \mathbf{R}^3 that is contained in \mathcal{A} .
(Hint: Follow the proof of Theorem 1.31.)

- (a) $\mathcal{A} = \{(2, 6, -3), (5, 15, -8), (3, 9, -5), (1, 3, -2), (5, 3, -2)\}$
- (b) $\mathcal{A} = \{(1, 0, 2), (0, 1, 1), (2, 1, 5), (1, 1, 3), (1, 2, 1)\}$
- (c) $\mathcal{A} = \{(1, 1, 0), (2, 2, 0), (2, 4, 1), (5, 9, 2), (7, 13, 3), (1, 2, 1)\}$
- (d) $\mathcal{A} = \{(1, 1, 2), (2, 2, 4), (1, -1, 1), (2, 0, 3), (3, 1, 5), (1, 1, 1)\}$

2. Given that each set \mathcal{A} is a basis of \mathbf{R}^4 and that each \mathcal{B} is linearly independent, follow the proof of Theorem 1.32 to extend \mathcal{B} to a basis of \mathbf{R}^4 .

- (a) $\mathcal{A} = \{(1, 1, 0, 0), (0, 1, 1, 0), (0, 0, 0, 1), (0, 1, 0, 1)\}$,
 $\mathcal{B} = \{(1, 0, 2, 3), (0, 1, -2, -3)\}$
- (b) $\mathcal{A} = \{(1, 0, 0, 0), (0, 0, 1, 0), (5, 1, 11, 0), (-4, 0, -6, 1)\}$,
 $\mathcal{B} = \{(1, 0, 1, 0), (0, 2, 0, 3)\}$
- (c) $\mathcal{A} = \{(1, 1, 1, 1), (1, 1, -1, -1), (1, 0, 1, 0), (0, 1, 0, -1)\}$,
 $\mathcal{B} = \{(1, 1, 0, 0), (0, 0, 1, 1)\}$
- (d) $\mathcal{A} = \{(1, 1, 0, 0), (1, 0, 4, 6), (0, 0, 0, 1), (0, 1, 0, 1)\}$,
 $\mathcal{B} = \{(1, 0, 2, 3), (0, 1, -2, -3)\}$
3. Show that \mathcal{A} is a basis of \mathbf{R}^3 by using Theorem 1.33.
- (a) $\mathcal{A} = \{(1, -2, 3), (0, 1, -2), (1, -1, 2)\}$
- (b) $\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$
- (c) $\mathcal{A} = \{(2, 0, 0), (4, 1, 0), (3, 3, 1)\}$
- (d) $\mathcal{A} = \{(2, -1, 1), (0, 1, -1), (-2, 1, 0)\}$
4. Show that each of the sets \mathcal{A} in Problem 3 is a basis of \mathbf{R}^3 by using Theorem 1.34.
5. By direct use of the definition of a basis, show that each of the sets \mathcal{A} in Problem 3 is a basis of \mathbf{R}^3 .
6. Which of the following sets of vectors in \mathbf{R}^3 are linearly dependent?
- (a) $\{(1, 3, 1), (1, 3, 0)\}$
- (b) $\{(1, -1, 0), (0, 1, 1), (1, 1, 1), (0, 0, 1)\}$
- (c) $\{(1, 1, 0), (0, 1, 1), (1, 2, 1), (1, 0, -1)\}$
- (d) $\{(1, 0, 1), (0, 1, 1), (2, 1, 3)\}$
- (e) $\{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$
- (f) $\{(1, 1, 0), (0, 1, -1), (1, 0, 0)\}$
7. Which of the sets of vectors in Problem 6 span \mathbf{R}^3 ?
8. Which of the following sets are bases for \mathbf{R}^3 ?
- (a) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$
- (b) $\{(1, 0, 0), (0, 1, 1)\}$
- (c) $\{(1, 0, 0), (1, 0, 1), (1, 1, 1)\}$
- (d) $\{(1, 0, 0), (0, 1, 0), (1, 1, 0)\}$
9. Determine whether or not \mathcal{A} is a basis of \mathbf{R}^4 .
- (a) $\mathcal{A} = \{(1, 2, 1, 0), (3, -4, 5, 6), (2, -1, 3, 3), (-2, 6, -4, -6)\}$
- (b) $\mathcal{A} = \{(1, -1, 2, -3), (1, 1, 2, 0), (3, -1, 6, -6), (0, 2, 0, 3)\}$

- (c) $\mathcal{A} = \{(1, 0, 2, -1), (0, 1, 1, 2), (1, 2, 1, 4), (2, 2, 3, 0)\}$
(d) $\mathcal{A} = \{(1, 2, 1, -1), (0, 1, 2, 3), (1, 4, 5, 5), (2, 7, 0, 2)\}$

10. Show that the sets \mathcal{A} and \mathcal{B} span the same subspace of \mathbf{R}^4 .

- (a) $\mathcal{A} = \{(1, 1, 3, -1), (1, 0, -2, 0)\}, \mathcal{B} = \{(3, 2, 4, -2), (0, 1, 5, -1)\}$
(b) $\mathcal{A} = \{(2, 3, 0, -1), (2, 1, -1, 2)\}, \mathcal{B} = \{(0, -2, -1, 3), (6, 7, -1, 0)\}$
(c) $\mathcal{A} = \{(1, 0, 3, 0), (1, 1, 8, 4)\}, \mathcal{B} = \{(1, -1, -2, -4), (1, 1, 8, 4), (3, -1, 4, -4)\}$
(d) $\mathcal{A} = \{(1, -1, -1, -2), (1, -5, -1, 0)\}, \mathcal{B} = \{(0, 2, 0, -1), (1, -3, -1, -1), (3, -5, -3, -5)\}$

11. Find a basis of $\langle \mathcal{A} \rangle$ that is contained in \mathcal{A} .

- (a) $\mathcal{A} = \{(1, 0, 1, -1), (3, -2, 3, 5), (2, -1, 2, 2), (5, -2, 5, 3)\}$
(b) $\mathcal{A} = \{(1, 0, 1, 2), (3, 1, 0, 3), (2, 1, -1, 1), (1, -1, 4, 5)\}$
(c) $\mathcal{A} = \{(2, -1, 0, 1), (1, 2, 1, 0), (3, -4, -1, 2), (5, 3, 2, 1)\}$
(d) $\mathcal{A} = \{(1, 0, 1, 1), (0, 1, -1, 1), (1, -1, 2, 0), (2, 1, 1, 7)\}$

12. Find the dimension of $\langle \mathcal{A} \rangle$.

- (a) $\mathcal{A} = \{(1, 2, 1, 0), (3, -4, 5, 6), (2, -1, 3, 3), (-2, 6, -4, -6)\}$
(b) $\mathcal{A} = \{(4, 3, 2, -1), (5, 4, 3, -1), (-2, -2, -1, 2), (11, 6, 4, 1)\}$
(c) $\mathcal{A} = \{(1, 0, 1, 2, -1), (0, 1, -2, 1, 3), (2, 1, 0, 5, 1), (1, -1, 3, 1, -4)\}$
(d) $\mathcal{A} = \{(1, 2, 0, 1, 0), (2, 4, 1, 4, 3), (1, 2, 2, 5, -2), (-1, -2, 3, 5, 4)\}$

13. Prove that if \mathbf{W}_1 and \mathbf{W}_2 are subspaces of \mathbf{R}^n , then

$$\dim(\mathbf{W}_1 + \mathbf{W}_2) = \dim(\mathbf{W}_1) + \dim(\mathbf{W}_2) - \dim(\mathbf{W}_1 \cap \mathbf{W}_2).$$

(Hint: Let $\mathcal{C} = \{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ be a basis of $\mathbf{W}_1 \cap \mathbf{W}_2$, and extend \mathcal{C} to bases $\mathcal{A} = \{\mathbf{w}_1, \dots, \mathbf{w}_r, \mathbf{u}_1, \dots, \mathbf{u}_s\}$ and $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_r, \mathbf{v}_1, \dots, \mathbf{v}_t\}$ of \mathbf{W}_1 and \mathbf{W}_2 , respectively. Prove that $\{\mathbf{w}_1, \dots, \mathbf{w}_r, \mathbf{u}_1, \dots, \mathbf{u}_s, \mathbf{v}_1, \dots, \mathbf{v}_t\}$ is a basis of $\mathbf{W}_1 + \mathbf{W}_2$.)

14. The sum $\mathbf{W}_1 + \mathbf{W}_2 + \dots + \mathbf{W}_k = \sum_{j=1}^k \mathbf{W}_j$ of subspaces \mathbf{W}_j of \mathbf{R}^n is defined to be the set of all vectors of the form $\mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_k$, with \mathbf{v}_i in \mathbf{W}_i . The sum $\sum_{j=1}^k \mathbf{W}_j$ is called *direct* if

$$\mathbf{W}_i \cap \sum_{\substack{j=1 \\ j \neq i}}^k \mathbf{W}_j = \{\mathbf{0}\}$$

for $i = 1, 2, \dots, k$. A direct sum is written as $\mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \dots \oplus \mathbf{W}_k$. Prove that

$$\dim(\mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \dots \oplus \mathbf{W}_k) = \sum_{j=1}^k \dim(\mathbf{W}_j).$$

Chapter 2

Elementary Operations on Vectors

2.1 Introduction

The elementary operations are as fundamental in linear algebra as the operations of differentiation and integration are in the calculus. These elementary operations are indispensable both in the development of the theory of linear algebra and in the applications of this theory.

In many treatments of linear algebra, the elementary operations are introduced after the development of a certain amount of matrix theory, and the matrix theory is used as a tool in establishing the properties of the elementary operations. In the presentation here, this procedure is reversed somewhat. The elementary operations are introduced as operations on sets of vectors and many of the results in matrix theory are developed with the aid of our knowledge of elementary operations. This approach has two main advantages. The material in Chapter 1 can be used to efficiently develop several of the properties of elementary operations, and the statements of many of these properties are simpler when formulated in vector terminology.

2.2 Elementary Operations and Their Inverses

We saw in the proof of Theorem 1.29 that \mathbf{R}^n has the standard basis $\mathcal{E}_n = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ in which each vector \mathbf{e}_i has a very simple form. The Kronecker delta that was introduced in Exercises 1.2 can be used to describe the vectors \mathbf{e}_i in a concise way. Using the fact that

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

the vectors \mathbf{e}_i can be written as

$$\mathbf{e}_i = (\delta_{i1}, \delta_{i2}, \dots, \delta_{in})$$

for $i = 1, 2, \dots, n$.

Example 1 □ The vectors in the standard basis $\mathcal{E}_4 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ are given by

$$\mathbf{e}_1 = (\delta_{11}, \delta_{12}, \delta_{13}, \delta_{14}) = (1, 0, 0, 0), \quad \mathbf{e}_2 = (\delta_{21}, \delta_{22}, \delta_{23}, \delta_{24}) = (0, 1, 0, 0),$$

$$\mathbf{e}_3 = (\delta_{31}, \delta_{32}, \delta_{33}, \delta_{34}) = (0, 0, 1, 0), \quad \mathbf{e}_4 = (\delta_{41}, \delta_{42}, \delta_{43}, \delta_{44}) = (0, 0, 0, 1). \blacksquare$$

We shall see in this chapter that every subspace \mathbf{W} of \mathbf{R}^n has a basis with a certain simple form, and that particular basis is called the *standard basis* of \mathbf{W} . In order to develop the concept of this standard basis, we shall need certain operations that change one spanning set of \mathbf{W} into another spanning set of \mathbf{W} .

More specifically, we define three types of elementary operations on nonempty ordered finite sets of vectors. These types of elementary operations will be referred to hereafter as types I, II, or III. Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a set of vectors in \mathbf{R}^n .

- (i) An *elementary operation of type I* multiplies one of the \mathbf{v}_i in \mathcal{A} by a nonzero scalar.
- (ii) An *elementary operation of type II* replaces one of the vectors \mathbf{v}_s by the sum of \mathbf{v}_s and a scalar multiple of a \mathbf{v}_t ($s \neq t$) in \mathcal{A} .
- (iii) An *elementary operation of type III* interchanges two vectors in \mathcal{A} .

If the number 1 is used as the scalar in an elementary operation of type I, the resulting elementary operation is called the *identity operation*. That is, the *identity operation* on a set is that operation that leaves the set unchanged.

Example 2 □ Consider the set of vectors

$$\mathcal{A} = \{\mathbf{v}_1 = (1, 0, 2), \mathbf{v}_2 = (2, 3, 19), \mathbf{v}_3 = (0, 1, 5)\}.$$

Multiplication of the first vector in \mathcal{A} by 2 is an elementary operation of type I that yields the set

$$\mathcal{A}_1 = \{\mathbf{v}_{11} = (2, 0, 4), \mathbf{v}_{12} = (2, 3, 19), \mathbf{v}_{13} = (0, 1, 5)\}.$$

If the vector \mathbf{v}_{12} in \mathcal{A}_1 is replaced by $\mathbf{v}_{12} + (-3)\mathbf{v}_{13}$, we have an elementary operation of type II that yields the set

$$\mathcal{A}_2 = \{\mathbf{v}_{21} = (2, 0, 4), \mathbf{v}_{22} = (2, 0, 4), \mathbf{v}_{23} = (0, 1, 5)\}.$$

An interchange of \mathbf{v}_{22} and \mathbf{v}_{23} is an elementary operation of type III that produces

$$\mathcal{A}_3 = \{\mathbf{v}_{31} = (2, 0, 4), \mathbf{v}_{32} = (0, 1, 5), \mathbf{v}_{33} = (2, 0, 4)\}.$$

Application of an elementary operation of type II that replaces \mathbf{v}_{33} by $\mathbf{v}_{33} + (-1)\mathbf{v}_{31}$ gives

$$\mathcal{A}_4 = \{\mathbf{v}_{41} = (2, 0, 4), \mathbf{v}_{42} = (0, 1, 5), \mathbf{v}_{43} = (0, 0, 0)\}. \blacksquare$$

As the example above suggests, the application of a sequence of elementary operations can be used to replace a given set \mathcal{A} by a set \mathcal{A}' which has a simpler appearance. Later in this chapter, we shall turn to an investigation of the properties that \mathcal{A} and \mathcal{A}' have in common. When these properties are known, we shall see that the elementary operations can be chosen so as to make the set \mathcal{A}' display certain important information concerning \mathcal{A} . In the investigation of these properties, it is convenient to have available the concept of an *inverse* of an elementary operation.

Suppose that an elementary operation is performed on $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ to obtain a new set \mathcal{A}' , and consider the operation necessary to obtain the original set \mathcal{A} from the new set \mathcal{A}' .

- (i) If \mathcal{A}' is obtained by employing a type I elementary operation, then \mathcal{A}' is of the form

$$\mathcal{A}' = \{\mathbf{v}_1, \dots, \mathbf{v}_{s-1}, a\mathbf{v}_s, \mathbf{v}_{s+1}, \dots, \mathbf{v}_k\},$$

where $a \neq 0$. It is readily seen that \mathcal{A} is obtained from \mathcal{A}' by replacing $a\mathbf{v}_s$ by $\frac{1}{a}(a\mathbf{v}_s)$. Thus in this case \mathcal{A} is obtained from \mathcal{A}' by an elementary operation of the same type.

- (ii) If a type II operation is used to obtain \mathcal{A}' from \mathcal{A} , then \mathcal{A}' has the form

$$\mathcal{A}' = \{\mathbf{v}_1, \dots, \mathbf{v}_{s-1}, \mathbf{v}_s + b\mathbf{v}_t, \mathbf{v}_{s+1}, \dots, \mathbf{v}_k\},$$

where $s \neq t$. Now \mathbf{v}_t is in \mathcal{A}' , and if $\mathbf{v}_s + b\mathbf{v}_t$ in \mathcal{A}' is replaced by $(\mathbf{v}_s + b\mathbf{v}_t) + (-b)\mathbf{v}_t$, the original set \mathcal{A} is obtained. This replacement is an elementary operation of type II, so that, once again, \mathcal{A} is obtained from \mathcal{A}' by an elementary operation of the same type as was used in obtaining \mathcal{A}' from \mathcal{A} .

- (iii) If \mathcal{A}' is obtained from \mathcal{A} by interchanging the vectors \mathbf{v}_s and \mathbf{v}_t in \mathcal{A} , then \mathcal{A} is obtained from \mathcal{A}' by the very same operation of interchanging \mathbf{v}_s and \mathbf{v}_t .

We see, then, that once an elementary operation is applied to a set \mathcal{A} to obtain a set \mathcal{A}' , we need only apply another elementary operation of the same type to \mathcal{A}' in order to obtain \mathcal{A} .

Definition 2.1 *When an elementary operation E is applied to a set \mathcal{A} to obtain a set \mathcal{A}' , the elementary operation that must be applied to \mathcal{A}' in order to obtain \mathcal{A} is called the **inverse elementary operation** of E , and is denoted by E^{-1} .*

It is clear from our discussion above that the inverse of an elementary operation E is unique, and is of the same type as E .

Theorem 2.2 *If \mathcal{A}' is obtained from \mathcal{A} by a sequence of elementary operations, then \mathcal{A} can be obtained from \mathcal{A}' by applying the inverses of these elementary operations in reverse order.*

Proof. Suppose that \mathcal{A}' is obtained from \mathcal{A} by a sequence E_1, E_2, \dots, E_t of elementary operations. That is, the operations E_1, E_2, \dots, E_t are applied successively, obtaining a new set \mathcal{A}_i each time an E_i is applied, until we obtain $\mathcal{A}_t = \mathcal{A}'$. Now consider the sequence $E_t^{-1}, E_{t-1}^{-1}, \dots, E_2^{-1}, E_1^{-1}$ applied to \mathcal{A}' . Applying E_t^{-1} to $\mathcal{A}' = \mathcal{A}_t$, one obtains \mathcal{A}_{t-1} since E_t yields \mathcal{A}_t when applied to \mathcal{A}_{t-1} . Then applying E_{t-1}^{-1} to \mathcal{A}_{t-1} , one obtains \mathcal{A}_{t-2} . Continuing in this manner, we obtain, successively, the sets

$$\mathcal{A}_{t-3}, \mathcal{A}_{t-4}, \dots, \mathcal{A}_3, \mathcal{A}_2, \mathcal{A}_1.$$

Then applying E_1^{-1} to \mathcal{A}_1 we obtain \mathcal{A} . Thus the theorem is proved. \blacksquare

An illustration of this theorem and its proof is provided in the next example.

Example 3 \square In Example 2, the set

$$\mathcal{A}' = \mathcal{A}_4 = \{\mathbf{v}_{41} = (2, 0, 4), \mathbf{v}_{42} = (0, 1, 5), \mathbf{v}_{43} = (0, 0, 0)\}$$

is obtained from the set

$$\mathcal{A} = \{\mathbf{v}_1 = (1, 0, 2), \mathbf{v}_2 = (2, 3, 19), \mathbf{v}_3 = (0, 1, 5)\}$$

by a sequence E_1, E_2, E_3, E_4 of elementary operations that can be described as follows:

- E_1 : multiply the first vector by 2.
- E_2 : replace the second vector by the sum of the second vector and (-3) times the third vector.
- E_3 : interchange the second and third vectors.
- E_4 : replace the third vector by the sum of the third vector and (-1) times the first vector.

Utilizing the general discussion preceding Definition 2.1, we formulate the inverse elementary operations as follows.

- E_1^{-1} : multiply the first vector by $\frac{1}{2}$.
- E_2^{-1} : replace the second vector by the sum of the second vector and 3 times the third vector.
- E_3^{-1} : interchange the second and third vectors.
- E_4^{-1} : replace the third vector by the sum of the third vector and 1 times the first vector.

Applying these inverse operations to \mathcal{A}' in reverse order, we find that

E_4^{-1} applied to \mathcal{A}' yields \mathcal{A}_3 ,

E_3^{-1} applied to \mathcal{A}_3 yields \mathcal{A}_2 ,

E_2^{-1} applied to \mathcal{A}_2 yields \mathcal{A}_1 ,

and

E_1^{-1} applied to \mathcal{A}_1 yields \mathcal{A} . ■

Exercises 2.2

1. Write out the elements of the standard basis of \mathbf{R}^5 .
2. Find an elementary operation that yields

$$\{(1, 0, 2, 1), (0, 3, 0, 7), (3, 6, 4, 3)\}$$

when applied to $\{(1, 0, 2, 1), (-2, 3, -4, 5), (3, 6, 4, 3)\}$.

3. Find an elementary operation that yields

$$\{(1, 0, 2, 1), (-2, 3, -4, 5), (3, 6, 4, 3)\}$$

when applied to $\{(1, 0, 2, 1), (0, 3, 0, 7), (3, 6, 4, 3)\}$.

4. Show that the set

$$\{(2, 3, 0, -1), (2, 1, -1, 2)\}$$

can be obtained from the set $\{(0, -2, -1, 3), (6, 7, -1, 0)\}$ by a sequence of elementary operations.

5. Show that the set

$$\{(0, -2, -1, 3), (6, 7, -1, 0)\}$$

can be obtained from the set $\{(2, 3, 0, -1), (2, 1, -1, 2)\}$ by a sequence of elementary operations.

6. Assume that the set $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3\}$ is obtained from the set $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ by the sequence E_1, E_2, E_3 defined as follows.

E_1 : multiply the first vector by -3 .

E_2 : replace the second vector by the sum of the second vector and 2 times the first vector.

E_3 : replace the third vector by the sum of the third vector and (-2) times the second vector.

Write out a sequence of elementary operations that yields \mathcal{A} when applied to \mathcal{A}' .

7. Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3\}$ be sets of vectors in \mathbf{R}^n such that

$$\mathbf{v}'_1 = 2\mathbf{v}_1$$

$$\mathbf{v}'_2 = 2\mathbf{v}_2 + 3\mathbf{v}_3$$

$$\mathbf{v}'_3 = \mathbf{v}_3 + \mathbf{v}_1.$$

Write out a sequence of elementary operations that yields \mathcal{A}' when applied to \mathcal{A} .

8. Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$, and let $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3, \mathbf{v}'_4\}$ be sets of vectors in \mathbf{R}^3 such that

$$\begin{aligned}\mathbf{v}'_1 &= \mathbf{v}_1 \\ \mathbf{v}'_2 &= \mathbf{v}_1 + \mathbf{v}_2 \\ \mathbf{v}'_3 &= \mathbf{v}_2 + \mathbf{v}_3 \\ \mathbf{v}'_4 &= \mathbf{v}_3 + \mathbf{v}_4.\end{aligned}$$

Write out a sequence of elementary operations that yields \mathcal{A}' when applied to \mathcal{A} .

9. Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$, and let $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3, \mathbf{v}'_4\}$ be sets of vectors in \mathbf{R}^3 such that

$$\begin{aligned}\mathbf{v}'_1 &= \mathbf{v}_1 - \mathbf{v}_4 \\ \mathbf{v}'_2 &= \mathbf{v}_2 + 3\mathbf{v}_1 \\ \mathbf{v}'_3 &= \mathbf{v}_3 + 2\mathbf{v}_2 \\ \mathbf{v}'_4 &= 2\mathbf{v}_4.\end{aligned}$$

Write out a sequence of elementary operations that yields \mathcal{A}' when applied to \mathcal{A} .

10. With the sets \mathcal{A} and \mathcal{A}' as given in Problem 8, write out a sequence of elementary operations that yields \mathcal{A} when applied to \mathcal{A}' .
11. With the sets \mathcal{A} and \mathcal{A}' as given in Problem 9, write out a sequence of elementary operations that yields \mathcal{A} when applied to \mathcal{A}' .
12. Show that the sequence of elementary operations used to obtain \mathcal{A}_4 from \mathcal{A} in Example 2 is *not* unique by exhibiting a different sequence of elementary operations that yields \mathcal{A}' when applied to \mathcal{A} .
13. Show that the identity operation on a set with more than 1 element is an elementary operation of type II.

2.3 Elementary Operations and Linear Independence

One of the properties that is preserved by application of an elementary operation to a set is that of linear independence. This important result is established in the next theorem.

Theorem 2.3 *Suppose that \mathcal{A} and \mathcal{A}' are sets of vectors in \mathbf{R}^n such that \mathcal{A}' is obtained from \mathcal{A} by applying a single elementary operation. Then \mathcal{A}' is linearly independent if and only if \mathcal{A} is linearly independent.*

Proof. Suppose first that $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is linearly independent.

If \mathcal{A}' is obtained by a type I elementary operation, then

$$\mathcal{A}' = \{\mathbf{v}_1, \dots, \mathbf{v}_{s-1}, a\mathbf{v}_s, \mathbf{v}_{s+1}, \dots, \mathbf{v}_k\},$$

where $a \neq 0$. Now suppose that

$$\sum_{\substack{i=1 \\ i \neq s}}^k c_i \mathbf{v}_i + c_s a \mathbf{v}_s = \mathbf{0}.$$

This implies that $c_1 = \dots = c_{s-1} = c_s a = c_{s+1} = \dots = c_k = 0$ since \mathcal{A} is linearly independent. But $c_s a = 0$ implies $c_s = 0$ since $a \neq 0$. Therefore, all c_i are zero, and \mathcal{A}' is linearly independent.

If \mathcal{A}' is obtained from \mathcal{A} by a type II elementary operation, then

$$\mathcal{A}' = \{\mathbf{v}_1, \dots, \mathbf{v}_{s-1}, \mathbf{v}_s + b\mathbf{v}_t, \mathbf{v}_{s+1}, \dots, \mathbf{v}_k\},$$

where $s \neq t$. If

$$\sum_{\substack{i=1 \\ i \neq s}}^k c_i \mathbf{v}_i + c_s (\mathbf{v}_s + b\mathbf{v}_t) = \mathbf{0},$$

then

$$\sum_{\substack{i=1 \\ i \neq s, t}}^k c_i \mathbf{v}_i + c_s \mathbf{v}_s + (c_t + c_s b) \mathbf{v}_t = \mathbf{0},$$

and this implies that $c_1 = \dots = c_s = \dots = c_t + c_s b = \dots = c_k = 0$ since \mathcal{A} is linearly independent. But $c_s = 0$ and $c_t + c_s b = 0$ imply that $c_t = 0$. Hence all c_i are zero, and \mathcal{A}' is linearly independent.

If \mathcal{A}' is obtained by an elementary operation of type III, then \mathcal{A}' consists of exactly the same vectors as does \mathcal{A} , except that the order of the vectors is different. It is clear, then, that \mathcal{A}' is linearly independent.

Thus, \mathcal{A}' is linearly independent if \mathcal{A} is linearly independent.

Suppose now that \mathcal{A}' is linearly independent. Since \mathcal{A}' is obtained from \mathcal{A} by a single elementary operation, \mathcal{A} can be obtained from \mathcal{A}' by the inverse elementary operation, which is an elementary operation of the same type. It then follows from the proof of the first part of the theorem that \mathcal{A} is linearly independent. ■

Frequently, information concerning a set of vectors can be obtained by successive application of a sequence of elementary operations that cannot be obtained by use of a single elementary operation. The corollary below is extremely useful in this respect.

Corollary 2.4 *If a set \mathcal{A}' is obtained by applying a sequence of elementary operations to a set \mathcal{A} of vectors in \mathbf{R}^n , then \mathcal{A}' is linearly independent if and only if \mathcal{A} is linearly independent.*

Proof. Suppose that \mathcal{A}' is obtained from \mathcal{A} by a sequence E_1, E_2, \dots, E_t of elementary operations. Put $\mathcal{A}_0 = \mathcal{A}$, and let \mathcal{A}_i be the set obtained by applying E_i to \mathcal{A}_{i-1} for $i = 1, 2, \dots, t$. Repeated application of Theorem 2.3 yields the following information:

$$\mathcal{A}_1 \text{ is independent if and only if } \mathcal{A}_0 = \mathcal{A} \text{ is independent.}$$

$$\mathcal{A}_2 \text{ is independent if and only if } \mathcal{A}_1 \text{ is independent.}$$

$$\vdots \quad \vdots \quad \vdots$$

$$\mathcal{A}' = \mathcal{A}_t \text{ is independent if and only if } \mathcal{A}_{t-1} \text{ is independent.}$$

Thus \mathcal{A} is linearly independent if and only if \mathcal{A}' is independent. ■

As an example of the use of this corollary, consider the following.

Example 1 □ In Example 2 of Section 2.2, it is shown that the set

$$\mathcal{A}' = \{(2, 0, 4), (0, 1, 5), (0, 0, 0)\}$$

can be obtained from the set

$$\mathcal{A} = \{(1, 0, 2), (2, 3, 19), (0, 1, 5)\}$$

by a sequence of elementary operations. The set \mathcal{A}' is obviously linearly dependent since it contains the zero vector. It follows that the set \mathcal{A} is linearly dependent. ■

At this point, we have developed a somewhat crude method for investigating the linear dependence of a given set \mathcal{A} of vectors in \mathbf{R}^n . If, by application of a sequence of elementary operations to \mathcal{A} , it is possible to obtain a set that contains the zero vector (or any set that is clearly dependent), then the given set is linearly dependent. By the same token, if a set can be obtained that is clearly independent, then \mathcal{A} is linearly independent. This method is refined to a systematic procedure later in this chapter.

We conclude this section with a final corollary to Theorem 2.3.

Corollary 2.5 *A set of vectors resulting from applying a sequence of elementary operations to a basis of \mathbf{R}^n is again a basis of \mathbf{R}^n .*

Proof. Let \mathcal{A} be a basis of \mathbf{R}^n . According to Corollary 2.4, any set \mathcal{A}' obtained from \mathcal{A} by a sequence of elementary operations is a linearly independent set of n vectors, and hence is a basis by Theorem 1.33. ■

Exercises 2.3

1. Show the set $\mathcal{A} = \{(1, 2, 1, 0), (3, -4, 5, 6), (2, -1, 3, 3), (-2, 6, -4, -6)\}$ is linearly dependent by applying a sequence of elementary operations to \mathcal{A} and obtaining a set \mathcal{A}' that contains the zero vector.

2. Use the method described in Problem 1 to show that the set

$$\{(1, 1, 0), (0, 1, 1), (1, 0, -1), (1, 0, 1)\}$$

is linearly dependent.

3. Show that the set $\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$ is linearly independent by obtaining \mathcal{A} from the standard basis of \mathbf{R}^3 by a sequence of elementary operations.
4. Show that the set $\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$ is linearly independent by obtaining the standard basis of \mathbf{R}^3 from \mathcal{A} by a sequence of elementary operations.
5. Use elementary operations to determine whether or not the given set is linearly independent.
- (a) $\{(1, 0, 2), (2, -1, 1), (1, 1, -1)\}$
 - (b) $\{(1, 1, -1), (2, -1, 1), (2, 1, 1)\}$
 - (c) $\{(1, 1, 8, -1), (1, 0, 3, 0), (3, 2, 19, -2)\}$
 - (d) $\{(1, -1, -2, -4), (1, 1, 8, 4), (3, -1, 4, -4)\}$
 - (e) $\{(1, 0, 1, 0), (0, 1, 0, 1), (4, 3, 2, 3), (1, 0, 0, 0)\}$
 - (f) $\{(1, 0, 1, 0), (2, 1, 4, 3), (1, 2, 5, -2), (-1, 3, 5, 4)\}$

2.4 Standard Bases for Subspaces

In the last section, we found that linear independence is a property that is preserved by application of elementary operations. As the reader has likely anticipated, we turn next to an investigation of the application of elementary operations to spanning sets of a subspace. We then combine and sharpen our results so as to obtain a systematic method of attacking the types of problems mentioned in Section 1.5. Later, we shall see that our results have even more applications.

The following definition gives a notation that is useful in describing the result obtained when several elementary operations are applied in a sequence.

Definition 2.6 If $\mathcal{A} \subseteq \mathbf{R}^n$ and E denotes an elementary operation that may be applied to \mathcal{A} , then $E(\mathcal{A})$ denotes the set that is obtained by applying E to \mathcal{A} . If E_1, E_2, \dots, E_t is a sequence of elementary operations that may be applied to \mathcal{A} , then $E_t E_{t-1} \cdots E_2 E_1(\mathcal{A})$ is defined inductively by

$$E_t E_{t-1} \cdots E_2 E_1(\mathcal{A}) = E_t(E_{t-1} \cdots E_2 E_1(\mathcal{A})).$$

Example 1 \square Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where $\mathbf{v}_1 = (1, 0, 2)$, $\mathbf{v}_2 = (2, 1, 6)$, $\mathbf{v}_3 = (0, 3, 8)$, and consider the sequence E_1, E_2, E_3 , where the elementary operations E_i are given by:

E_1 : Replace the second vector by the sum of the second vector and (-2) times the first vector.

E_2 : Replace the third vector by the sum of the third vector and (-3) times the second vector.

E_3 : Multiply the third vector by $\frac{1}{2}$.

According to Definition 2.6,

$$\begin{aligned} E_1(\mathcal{A}) &= \{\mathbf{v}_1, \mathbf{v}_2 + (-2)\mathbf{v}_1, \mathbf{v}_3\} \\ &= \{(1, 0, 2), (0, 1, 2), (0, 3, 8)\} \\ E_2E_1(\mathcal{A}) &= E_2(E_1(\mathcal{A})) \\ &= E_2(\{(1, 0, 2), (0, 1, 2), (0, 3, 8)\}) \\ &= \{(1, 0, 2), (0, 1, 2), (0, 0, 2)\} \\ E_3E_2E_1(\mathcal{A}) &= E_3(E_2E_1(\mathcal{A})) \\ &= E_3(\{(1, 0, 2), (0, 1, 2), (0, 0, 2)\}) \\ &= \{(1, 0, 2), (0, 1, 2), (0, 0, 1)\}. \quad \blacksquare \end{aligned}$$

The next theorem in our development is fairly obvious, but it is important enough to be designated as a theorem. A restricted form of the converse is contained in the last theorem of this section, but the proof of that theorem must wait until some intermediate results are established.

Theorem 2.7 *If \mathcal{A}' is obtained by applying a sequence of elementary operations to a set \mathcal{A} of vectors in \mathbf{R}^n , then $\langle \mathcal{A}' \rangle = \langle \mathcal{A} \rangle$.*

Proof. Suppose that \mathcal{A}' is obtained by applying the sequence E_1, E_2, \dots, E_t of elementary operations E_i to \mathcal{A} . Let $\mathcal{A}_0 = \mathcal{A}$ and $\mathcal{A}_i = E_i(\mathcal{A}_{i-1})$ for $i = 1, 2, \dots, t$, so that $\mathcal{A}' = \mathcal{A}_t$.

Now \mathcal{A}_i is obtained from \mathcal{A}_{i-1} by application of E_i . If we recall the definitions of the elementary operations in Section 2.2, it is evident that each vector of \mathcal{A}_i is a linear combination of vectors in \mathcal{A}_{i-1} . Thus $\langle \mathcal{A}_i \rangle \subseteq \langle \mathcal{A}_{i-1} \rangle$. But $\mathcal{A}_{i-1} = E_i^{-1}(\mathcal{A}_i)$ so that each vector of \mathcal{A}_{i-1} is a linear combination of vectors in \mathcal{A}_i . Therefore $\langle \mathcal{A}_{i-1} \rangle \subseteq \langle \mathcal{A}_i \rangle$, and $\langle \mathcal{A}_{i-1} \rangle = \langle \mathcal{A}_i \rangle$. Letting $i = 1, 2, \dots, t$ in succession, we have

$$\langle \mathcal{A} \rangle = \langle \mathcal{A}_0 \rangle = \langle \mathcal{A}_1 \rangle = \dots = \langle \mathcal{A}_t \rangle = \langle \mathcal{A}' \rangle,$$

and the theorem is proved. \blacksquare

Combining this result and Corollary 2.4, we obtain an important corollary concerning bases of a subspace.

Corollary 2.8 *A set of vectors resulting from the application of a sequence of elementary operations to a basis of a subspace \mathbf{W} is again a basis of \mathbf{W} .*

The next theorem is our first step in standardizing the bases of subspaces. Example 2 appears just after the end of the proof of this theorem, and the work in that example illustrates the steps described in the proof. If the steps in the proof and the steps in the example are traced together, this should make each of them easier to follow.

Theorem 2.9 *Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a set of m vectors in \mathbf{R}^n that spans the subspace $\mathbf{W} = \langle \mathcal{A} \rangle$ of dimension r , where $m \geq r > 0$. Then a set $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r, \mathbf{0}, \dots, \mathbf{0}\}$ of m vectors can be obtained from \mathcal{A} by a finite sequence of elementary operations so that $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ has the following properties:*

1. *The first nonzero component from the left in \mathbf{v}'_j is a 1 in the k_j component for $j = 1, 2, \dots, r$. (This 1 is called a **leading one**.)*
2. *$k_1 < k_2 < \dots < k_r$. (In vectors listed later in \mathcal{A}' , the leading ones occur in positions that are farther to the right.)*
3. *\mathbf{v}_j is the only vector in \mathcal{A}' with a nonzero k_j component.*
4. *$\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ is a basis of \mathbf{W} .*

Proof. By Theorem 1.31, \mathcal{A} contains a basis of $\langle \mathcal{A} \rangle$. Thus, there is at least one vector in \mathcal{A} that is not zero. Let k_1 be the smallest positive integer for which some \mathbf{v}_i has nonzero k_1 component. By no more than one interchange of vectors, a spanning set for \mathbf{W} can be obtained in which the first vector has a nonzero k_1 component. Multiplication of this vector by the reciprocal of its k_1 component yields a spanning set of \mathbf{W} in which the k_1 component of the first vector is 1. Then each of the other vectors can be replaced by the sum of that vector and a suitable multiple of the new first vector to obtain a spanning set

$$\mathcal{A}_1 = \{\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \dots, \mathbf{v}_m^{(1)}\}$$

of \mathbf{W} in which

- (i) the first nonzero component in $\mathbf{v}_1^{(1)}$ is a 1 in the k_1 component, and
- (ii) $\mathbf{v}_1^{(1)}$ is the only vector in \mathcal{A}_1 with a nonzero number in any of the first k_1 positions from the left.

If $r = 1$, the theorem follows trivially at this point.

If $r > 1$, let k_2 be the least positive integer for which some $\mathbf{v}_i^{(1)}, i \neq 1$, has nonzero k_2 component. At least one such $\mathbf{v}_i^{(1)}$ exists, since all remaining vectors would otherwise be zero. By an interchange of vectors that does not involve the first vector, a spanning set for \mathbf{W} can be obtained in which the second vector has nonzero k_2 component. Multiplication of this vector by the reciprocal of its k_2 component will yield a spanning set of \mathbf{W} in which the k_2 component of the second vector is 1. Then each of the other vectors can be replaced by the sum of that vector and a suitable multiple of the new second vector to obtain a spanning set

$$\mathcal{A}_2 = \{\mathbf{v}_1^{(2)}, \mathbf{v}_2^{(2)}, \dots, \mathbf{v}_m^{(2)}\}$$

of \mathbf{W} that has the following properties:

- (i) The first nonzero component in $\mathbf{v}_1^{(2)}$ is a 1 in the k_1 component, and the first nonzero component in $\mathbf{v}_2^{(2)}$ is a 1 in the k_2 component,
- (ii) $k_1 < k_2$,
- (iii) $\mathbf{v}_1^{(2)}$ is the only vector in \mathcal{A}_2 with a nonzero k_1 component, and $\mathbf{v}_2^{(2)}$ is the only vector in \mathcal{A}_2 with a nonzero k_2 component.

That is, the first two vectors in the set \mathcal{A}_2 have the first three properties required in the statement of the theorem.

Suppose that a set \mathcal{A}_i that spans \mathbf{W} has been obtained in which the first i vectors ($i < r$) satisfy the first three properties listed in the theorem. Then k_{i+1} is chosen to be the least positive integer for which one of the last $m - i$ vectors in the set has a nonzero k_{i+1} component. Such a vector exists, for \mathcal{A}_i must contain at least r nonzero vectors. The procedure described to obtain \mathcal{A}_1 and \mathcal{A}_2 can then be repeated to obtain the set \mathcal{A}_{i+1} that spans \mathbf{W} in which the first $i + 1$ vectors satisfy the first three conditions.

It is clear, then, that a finite sequence of elementary operations can be applied to \mathcal{A} to obtain a set $\mathcal{A}_r = \{\mathbf{v}_1^{(r)}, \mathbf{v}_2^{(r)}, \dots, \mathbf{v}_m^{(r)}\}$ that spans \mathbf{W} and in which the first r vectors satisfy the conditions (1), (2), (3). Now assume that there exists an $\mathbf{v}_j^{(r)}, j > r$, with nonzero k_{r+1} component. Then it must be that $k_{r+1} > k_r$, and from this it is easily seen that $\{\mathbf{v}_1^{(r)}, \mathbf{v}_2^{(r)}, \dots, \mathbf{v}_r^{(r)}, \mathbf{v}_j^{(r)}\}$ is linearly independent, contradicting the fact that r is the dimension of \mathbf{W} . Thus the remaining $m - r$ vectors in \mathcal{A}_r are zero and $\mathcal{A}_r = \mathcal{A}'$, where \mathcal{A}' satisfies the conditions of the theorem. ■

The proof given for Theorem 2.9 is a constructive one in that it describes a method of obtaining the set \mathcal{A}' from a given set \mathcal{A} . This is illustrated in the following example.

Example 2 □ Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ for $\mathbf{v}_1 = (0, 0, 0, -1, -1)$, $\mathbf{v}_2 = (0, 3, -6, 1, -2)$, $\mathbf{v}_3 = (0, 2, -4, 2, 0)$, $\mathbf{v}_4 = (0, -1, 2, 2, 3)$, and let $\mathbf{W} = \langle \mathcal{A} \rangle$.

Following the proof of the theorem, we see that $k_1 = 2$ is the smallest positive integer for which some \mathbf{v}_i has nonzero k_1 component. By an interchange of \mathbf{v}_1 and \mathbf{v}_3 , we obtain the spanning set

$$\{(0, 2, -4, 2, 0), (0, 3, -6, 1, -2), (0, 0, 0, -1, -1), (0, -1, 2, 2, 3)\}$$

in which the first vector has nonzero k_1 component. Multiplication of this vector by $\frac{1}{2}$ yields the spanning set

$$\{(0, 1, -2, 1, 0), (0, 3, -6, 1, -2), (0, 0, 0, -1, -1), (0, -1, 2, 2, 3)\}$$

in which the k_1 component of the first vector is 1. Each of the other vectors is now replaced by the sum of that vector and a suitable multiple of the new first vector as follows.

Replace the second vector by the sum of the second vector and (-3) times the first vector.

No change is needed on the third vector. (A “suitable” multiple to add would be the zero multiple.)

Replace the fourth vector by the sum of the fourth vector and 1 times the first vector.

This yields the spanning set

$$\begin{aligned}\mathcal{A}_1 &= \{\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \mathbf{v}_3^{(1)}, \mathbf{v}_4^{(1)}\} \\ &= \{(0, 1, -2, 1, 0), (0, 0, 0, -2, -2), (0, 0, 0, -1, -1), (0, 0, 0, 3, 3)\}\end{aligned}$$

in which

- (i) the first nonzero component in $\mathbf{v}_1^{(1)}$ is a 1 in the second component, and
- (ii) $\mathbf{v}_1^{(1)} = (0, 1, -2, 1, 0)$ is the only vector in \mathcal{A}_1 with a nonzero component in either of the first two positions.

Now $k_2 = 4$ is the least positive integer for which some $\mathbf{v}_i^{(1)}, i \neq 1$, in \mathcal{A}_1 has a nonzero k_2 component, and we have $\mathbf{v}_2^{(1)} = (0, 0, 0, -2, -2)$ with a nonzero k_2 component. Multiplication of $\mathbf{v}_2^{(1)}$ by $-\frac{1}{2}$ yields

$$\{(0, 1, -2, 1, 0), (0, 0, 0, 1, 1), (0, 0, 0, -1, -1), (0, 0, 0, 3, 3)\}.$$

Each of the vectors other than the second is now replaced by the sum of that vector and a suitable multiple of the second vector as follows.

Replace the first vector by the sum of the first vector and (-1) times the second vector.

Replace the third vector by the sum of the third vector and the second vector.

Replace the fourth vector by the sum of the fourth vector and (-3) times the second vector.

This yields

$$\begin{aligned}\mathcal{A}_2 &= \{\mathbf{v}_1^{(2)}, \mathbf{v}_2^{(2)}, \mathbf{v}_3^{(2)}, \mathbf{v}_4^{(2)}\} \\ &= \{(0, 1, -2, 0, -1), (0, 0, 0, 1, 1), (0, 0, 0, 0, 0), (0, 0, 0, 0, 0)\}.\end{aligned}$$

It is evident at this point that $\mathcal{A}_2 = \mathcal{A}'$. Simultaneously with finding \mathcal{A}' , we have found that \mathbf{W} has dimension $r = 2$. ■

The conditions of Theorem 2.9 are very restrictive, and one might expect that there is only one basis of a given subspace that satisfies these conditions. The next theorem confirms that this is the case.

Theorem 2.10 *There is one and only one basis of a given subspace \mathbf{W} that satisfies the conditions of Theorem 2.9.*

Proof. Starting in Theorem 2.9 with a *basis* \mathcal{A} of \mathbf{W} , that theorem assures us that a basis \mathcal{A}' of \mathbf{W} that satisfies the conditions can be obtained from \mathcal{A} by a sequence of elementary operations. Thus there is at least one basis of the required type.

Suppose now that $\mathcal{A}' = \{\mathbf{v}'_1, \dots, \mathbf{v}'_r\}$ and $\mathcal{A}'' = \{\mathbf{v}''_1, \dots, \mathbf{v}''_r\}$ are two bases of \mathbf{W} that satisfy the conditions of Theorem 2.9. Let k'_1, \dots, k'_r and k''_1, \dots, k''_r be the sequences of positive integers described in the conditions for \mathcal{A}' and \mathcal{A}'' , respectively.

Assume $k'_1 < k''_1$. Since \mathcal{A}'' spans \mathbf{W} , there must exist scalars c_{i1} such that $\mathbf{v}'_1 = \sum_{i=1}^r c_{i1} \mathbf{v}''_i$. Since each \mathbf{v}''_i has zero j^{th} component for each $j < k''_1$, any linear combination such as $\sum_{i=1}^r c_{i1} \mathbf{v}''_i$ must have zero j^{th} component for each $j < k''_1$. But \mathbf{v}'_1 has a nonzero k'_1 component, and $k'_1 < k''_1$. This is a contradiction; hence $k'_1 \geq k''_1$. The symmetry of the conditions on \mathcal{A}' and \mathcal{A}'' implies that $k''_1 \geq k'_1$, and thus, $k'_1 = k''_1$.

Now assume $k'_2 < k''_2$. Since \mathcal{A}'' spans \mathbf{W} , there must exist scalars c_{i2} such that $\mathbf{v}'_2 = \sum_{i=1}^r c_{i2} \mathbf{v}''_i$. Now \mathbf{v}''_1 is the only vector in \mathcal{A}'' that has nonzero k'_1 component, so a linear combination $\sum_{i=1}^r c_{i2} \mathbf{v}''_i$ has zero k'_1 component if and only if $c_{i1} = 0$. Since \mathbf{v}'_2 has a zero k'_1 component, $c_{i1} = 0$, and we have $\mathbf{v}'_2 = \sum_{i=2}^r c_{i2} \mathbf{v}''_i$. For $i \geq 2$, \mathbf{v}''_i has zero j^{th} component for each $j < k''_2$, and thus the linear combination $\sum_{i=2}^r c_{i2} \mathbf{v}''_i$ has zero j^{th} component for all $j < k''_2$. But \mathbf{v}'_2 has nonzero k'_2 component, and $k'_2 < k''_2$. As before, we have a contradiction, and therefore, $k'_2 \geq k''_2$. From the symmetry of the conditions, $k''_2 \geq k'_2$, and thus, $k'_2 = k''_2$.

It is clear that this argument may be repeated to obtain $k'_j = k''_j$ for $j = 1, 2, \dots, r$.

Now \mathcal{A}'' spans \mathbf{W} , so for each \mathbf{v}'_j there must exist scalars c_{ij} such that $\mathbf{v}'_j = \sum_{i=1}^r c_{ij} \mathbf{v}''_i$. Since \mathbf{v}''_j is the only vector in \mathcal{A}'' that has nonzero k'_j component and \mathbf{v}''_i has k'_i component equal to 1, c_{ij} is the k'_j component of $\sum_{i=1}^r c_{ij} \mathbf{v}''_i$. But \mathbf{v}'_j has zero k'_i component for $i \neq j$, so $c_{ij} = 0$ for $i \neq j$, and since \mathbf{v}'_j has k'_j component equal to 1, $c_{jj} = 1$. Therefore $\mathbf{v}'_j = \mathbf{v}''_j$, and $\mathcal{A}' = \mathcal{A}''$. ■

Theorem 2.10 allows us to make the following definition.

Definition 2.11 Let \mathbf{W} be a subspace of \mathbf{R}^n of dimension r . The **standard basis** of \mathbf{W} is the basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ of \mathbf{W} that satisfies the conditions of Theorem 2.9.

That is, the standard basis of \mathbf{W} is the unique basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ that has the following properties.

1. The first nonzero component from the left in the j^{th} vector \mathbf{v}_j is a 1 in the k_j component. (This 1 is called a **leading one**.)
2. $k_1 < k_2 < \dots < k_r$.
3. The j^{th} vector \mathbf{v}_j is the only vector in the basis with a nonzero k_j component.

Clearly, if $r = m = n$ in Theorem 2.9, the standard basis thus defined is the same as that given in Definition 1.30, and our two definitions are in agreement with each other.

Example 3 □ In Example 2, we found that the standard basis of the subspace

$$\mathbf{W} = \langle (0, 0, 0, -1, -1), (0, 3, -6, 1, -2), (0, 2, -4, 2, 0), (0, -1, 2, 2, 3) \rangle$$

is the set $\{(0, 1, -2, 0, -1), (0, 0, 0, 1, 1)\}$.

If we are only interested in finding the standard basis of \mathbf{W} , the amount of work that is necessary is much less than that done in Example 2. As a first step, we might write the vectors of \mathcal{A} in a rectangular array as

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

In composing this array, we have recorded the components of \mathbf{v}_i from top to bottom in the i^{th} column from the left. (It is admittedly more natural to record these components in rows rather than columns. The reason for the use of columns will become clear in Chapter 3.) Let us use an arrow from the first array to a second to indicate that the set represented by the second array is obtained from the first array by application of one or more elementary operations.

Our work in Example 2 can then be recorded in this manner:

$$\begin{aligned} A \rightarrow & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 2 & 3 & 0 & -1 \\ -4 & -6 & 0 & 2 \\ 2 & 1 & -1 & 2 \\ 0 & -2 & -1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & -1 \\ -2 & -6 & 0 & 2 \\ 1 & 1 & -1 & 2 \\ 0 & -2 & -1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 1 & -2 & -1 & 3 \\ 0 & -2 & -1 & 3 \end{bmatrix} \\ & \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 1 & 1 & -1 & 3 \\ 0 & 1 & -1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix}. \blacksquare \end{aligned}$$

Primarily for future use, we record the following theorems.

Theorem 2.12 *The standard basis of a subspace \mathbf{W} can be obtained from any basis of \mathbf{W} by a sequence of elementary operations.*

Proof. Let \mathbf{W} be a subspace of dimension r , and let \mathcal{A}' be the standard basis of \mathbf{W} . If \mathcal{A} is any basis of \mathbf{W} , then Theorem 2.9 applies with $m = r$, asserting that \mathcal{A}' can be obtained from \mathcal{A} by a finite sequence of elementary operations. ■

Theorem 2.13 Let \mathcal{A} and \mathcal{B} be two sets of m vectors each in \mathbf{R}^n . Then $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ if and only if \mathcal{B} can be obtained from \mathcal{A} by a sequence of elementary operations.

Proof. If \mathcal{B} can be obtained from \mathcal{A} by a sequence of elementary operations, then $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ by Theorem 2.7.

Suppose now that $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle = \mathbf{W}$, where \mathbf{W} has dimension r . Let \mathcal{A}' denote the set of vectors in Theorem 2.9 that can be obtained from \mathcal{A} by a sequence E_1, E_2, \dots, E_k of elementary operations. Let \mathcal{B}' denote the corresponding set that can be obtained from \mathcal{B} by a sequence F_1, F_2, \dots, F_t of elementary operations. Then the last $m - r$ vectors in both \mathcal{A}' and \mathcal{B}' are zero, and the first r vectors in both \mathcal{A}' and \mathcal{B}' constitute the standard basis of \mathbf{W} , which is unique by Theorem 2.10. Thus $\mathcal{A}' = \mathcal{B}'$.

Now \mathcal{B} can be obtained from $\mathcal{B}' = \mathcal{A}'$ by the sequence $F_t^{-1}, \dots, F_2^{-1}, F_1^{-1}$. Thus \mathcal{B} can be obtained from \mathcal{A} by application of the sequence

$$E_1, E_2, \dots, E_k, F_t^{-1}, \dots, F_2^{-1}, F_1^{-1}$$

of elementary operations. ■

The value of Theorem 2.13 is mainly theoretical. A practical way to determine whether or not $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ is to find and compare the sets \mathcal{A}' and \mathcal{B}' described in the proof. Then $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ if and only if $\mathcal{A}' = \mathcal{B}'$.

Exercises 2.4

1. Let $\mathcal{A} = \{(-5, 11, -3, 0, 1), (0, 5, 10, 1, 6), (2, -4, 2, 1, 1), (2, 1, 12, 2, 7)\}$. Find the set \mathcal{A}' of Theorem 2.9 by the method used in the proof of Theorem 2.9 and Example 2.

2. Work Problem 1 using the set \mathcal{A} given by

$$\mathcal{A} = \{(0, 1, 1, 8, -1), (0, 1, 1, 3, 0), (0, 3, 3, 19, -2), (0, 4, 4, 22, -2), (0, 3, 3, 14, 1)\}.$$

3. For each of the sets \mathcal{A} below, find the dimension of $\langle \mathcal{A} \rangle$.

- (a) $\mathcal{A} = \{(1, 1, 3, -1), (1, 0, -2, 0), (3, 2, 4, -2)\}$
- (b) $\mathcal{A} = \{(1, 2, -1, 3), (0, 1, 0, 2), (1, 3, -1, 5), (1, 1, -1, 0)\}$
- (c) $\mathcal{A} = \{(0, 2, 0, 4, -1), (0, 5, -1, 11, 8), (0, 0, 1, -7, 9), (0, 7, 0, 8, 16)\}$
- (d) $\mathcal{A} = \{(0, 3, 0, -1, 3), (0, -2, 0, 1, -1), (0, 5, 0, 1, 13), (0, 4, 0, -2, 2)\}$

4. For each of the sets \mathcal{A} given below, use rectangular arrays as in Example 3 to find the standard basis for $\langle \mathcal{A} \rangle$.

- (a) $\mathcal{A} = \{(1, 1, 0, 0), (0, 1, 1, 0), (0, 0, 1, 1), (1, 2, 2, 1)\}$
- (b) $\mathcal{A} = \{(1, -1, 2, -3), (3, -1, 6, -6), (1, 0, 1, 0), (1, 1, 2, 0)\}$
- (c) $\mathcal{A} = \{(3, 2, 5, 1), (-1, 0, -1, -1), (2, 1, 3, 1), (1, 0, 1, 1)\}$

- (d) $\mathcal{A} = \{(0, 2, 0, 3), (1, 0, 1, 0), (3, -1, 6, -6), (1, 1, 2, 0), (1, -1, 2, -3)\}$
5. Determine whether or not each of the sets below is linearly independent by finding the dimension of the subspace spanned by the set.
- (a) $\{(1, 2, 0, 1, 0), (2, 4, 1, 4, 3), (1, 2, 2, 5, -2), (-1, -2, 3, 5, 4)\}$
 (b) $\{(1, -1, 2, -3), (3, -1, 6, -6), (1, 0, 1, 0), (1, 1, 2, 0)\}$
6. In each case, determine whether or not $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$.
- (a) $\mathcal{A} = \{(1, 1, 1, 1), (0, 0, 2, 2)\}, \mathcal{B} = \{(2, 2, 3, 3), (1, 1, 2, 2)\}$
 (b) $\mathcal{A} = \{(1, 1, 0, 0), (1, 0, 1, 1)\}, \mathcal{B} = \{(2, -1, 3, 3), (0, 1, -1, -1)\}$
 (c) $\mathcal{A} = \{(2, 2, 0, 0), (1, 2, 1, 0), (1, 1, 1, 1)\}, \mathcal{B} = \{(1, 1, 0, 0), (0, 1, 1, 0), (0, 0, 1, 2)\}$
 (d) $\mathcal{A} = \{(1, 2, 1, -1), (0, 1, 2, 3), (1, 4, 5, 5)\}, \mathcal{B} = \{(2, 4, 2, -2), (0, 3, -2, 4)\}$
 (e) $\mathcal{A} = \{(1, 1, -1), (1, 3, 1), (-1, 1, 2)\}, \mathcal{B} = \{(2, 1, 2), (1, 0, 1), (3, 2, 4), (2, 6, 3)\}$
 (f) $\mathcal{A} = \{(1, -2, -1), (3, 2, -1), (0, 4, 1), (2, 0, -1)\}, \mathcal{B} = \{(5, -2, -3), (1, 6, 1)\}$
7. For each pair $\mathbf{W}_1, \mathbf{W}_2$, find the dimension of $\mathbf{W}_1 + \mathbf{W}_2$.
- (a) $\mathbf{W}_1 = \langle (1, 1, 3, -1), (1, 0, -2, 0), (3, 2, 4, -2) \rangle, \mathbf{W}_2 = \langle (1, 0, 0, 1), (1, 1, 7, 1) \rangle$
 (b) $\mathbf{W}_1 = \langle (1, -1, 2, -3), (1, 1, 2, 0), (3, -1, 6, -6) \rangle,$
 $\mathbf{W}_2 = \langle (2, 0, 4, -3), (0, 0, 0, 1) \rangle$
8. Let \mathcal{A} and \mathcal{A}' be as given in Example 2. Write out a complete sequence of elementary operations that will yield \mathcal{A}' when applied to \mathcal{A} .
9. Given that the sets \mathcal{A} and \mathcal{B} below span the same subspace \mathbf{W} , follow the proof of Theorem 2.13 to find a sequence of elementary operations that can be used to obtain \mathcal{B} from \mathcal{A} .
- (a) $\mathcal{A} = \{(1, 1, 0, 0), (1, 0, 1, 1)\}, \mathcal{B} = \{(2, -1, 3, 3), (0, 1, -1, -1)\}$
 (b) $\mathcal{A} = \{(1, 1, 3, -1), (1, 0, -2, 0)\}, \mathcal{B} = \{(3, 2, 4, -2), (0, 1, 5, -1)\}$
 (c) $\mathcal{A} = \{(4, 5, 2, 2), (1, 1, 1, 0)\}, \mathcal{B} = \{(1, -1, 5, -4), (0, 2, -4, 4)\}$
 (d) $\mathcal{A} = \{(1, 1, 8, -1), (1, 0, 3, 0)\}, \mathcal{B} = \{(1, -1, -2, 1), (3, -1, 4, 1)\}$

Chapter 3

Matrix Multiplication

3.1 Introduction

In Example 3 of Section 2.4, it was found that rectangular arrays were a useful notational convenience in recording the results of elementary operations on a set of vectors. These rectangular arrays are specific examples of the more general concept of a *matrix* over a set \mathcal{M} , to be defined in the following section. In this chapter, we define the operation of multiplication on matrices with real numbers as elements and establish the basic properties of this operation. As mentioned earlier, the results of Chapter 2 are extremely useful in the development here, particularly in the last two sections.

3.2 Matrices of Transition

The definition of a matrix over a set \mathcal{M} is as follows.

Definition 3.1 *An **r by s matrix** over a set \mathcal{M} is a rectangular array of elements of \mathcal{M} , arranged in r rows and s columns. Such a matrix will be written in the form*

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{bmatrix}$$

where a_{ij} denotes the element in the i^{th} row and j^{th} column of the matrix. The numbers r and s are said to be the **dimensions** of A , and r by s is sometimes written as $r \times s$.

The matrix A above may be written more simply as $A = [a_{ij}]_{r \times s}$ or $A = [a_{ij}]$ if the number of rows or columns is not important.

There are several terms that are useful in describing matrices of certain types. An r by r matrix is said to be a **square matrix** or a **matrix of order r** . The elements a_{ii} of $A = [a_{ij}]$ are the **diagonal elements** of A , and a square matrix $A = [a_{ij}]$ with $a_{ij} = 0$ whenever $i \neq j$ is a **diagonal matrix**. The matrix $I_r = [\delta_{ij}]_{r \times r}$ is the **identity matrix** of order r . A matrix $A = [a_{ij}]$ is a **zero matrix** if $a_{ij} = 0$ for all pairs i, j . A matrix that has only one row is a **row matrix**, and a matrix that has only one column is a **column matrix**.

Definition 3.2 Two matrices $A = [a_{ij}]_{r \times s}$ and $B = [b_{ij}]_{p \times q}$ over a set \mathcal{M} are **equal** if and only if $r = p, s = q$, and $a_{ij} = b_{ij}$ for all pairs i, j .

Example 1 □ Consider the following matrices.

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

$$E = \begin{bmatrix} 5 & 0 \\ 0 & -3 \end{bmatrix} \quad F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

The special terms just introduced apply to these matrices in the following ways:

A, B, E , and F are square matrices.

B, E , and F are diagonal matrices.

F is an identity matrix.

Both B and D are zero matrices, but $B \neq D$.

C is a row matrix.

D is a column matrix.

None of the special terms apply to G . ■

At times, we will denote a zero matrix by the same symbol $\mathbf{0}$ that we use for a zero vector. This will not cause confusion, because the context where the symbol is used will make the meaning clear.

Now consider a set of vectors $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ in \mathbf{R}^n and a second set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$ contained in $\langle \mathcal{A} \rangle$. Since \mathcal{A} spans $\langle \mathcal{A} \rangle$, there are scalars a_{ij} in \mathbf{R} such that $\mathbf{v}_j = \sum_{i=1}^r a_{ij} \mathbf{u}_i$ for $j = 1, 2, \dots, s$. The following definition involves these scalars a_{ij} .

Definition 3.3 If $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ is a set of vectors in \mathbf{R}^n and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$ is a set of vectors in $\langle \mathcal{A} \rangle$, a **matrix of transition** (or **transition matrix**) from \mathcal{A} to \mathcal{B} is a matrix $A = [a_{ij}]_{r \times s}$ such that

$$\mathbf{v}_j = \sum_{i=1}^r a_{ij} \mathbf{u}_i \text{ for } j = 1, 2, \dots, s.$$

The term *matrix of transition* applies only to situations involving nonempty finite sets of vectors, and these sets must be ordered. Whenever a set of vectors is listed without indices, it is understood that the index j is to go with the j^{th} vector from the left. Thus the first vector from the left is to have index 1, the second from the left to have index 2, and so on. This is consistent with the notational agreement made after Example 1 in Section 1.5.

Another point needs to be emphasized in connection with Definition 3.3. The definition of the term *transition matrix* that is found in many elementary linear algebra texts is *not* equivalent to the one given here. The one stated in Definition 3.3 is what we need to present our development of matrix multiplication, and it is the one that leads to simpler proofs of major theorems later in this book.

The following example shows that the matrix of transition A is not always uniquely determined by \mathcal{A} and \mathcal{B} .

Example 2 □ Let $\mathcal{A} = \{(1, 2, 0), (2, 4, 0), (0, 0, 1)\}$ and $\mathcal{B} = \{(1, 2, 4), (2, 4, 8)\}$ in $\langle \mathcal{A} \rangle$. Now

$$(1, 2, 4) = (1)(1, 2, 0) + (0)(2, 4, 0) + (4)(0, 0, 1)$$

and

$$(2, 4, 8) = (2)(1, 2, 0) + (0)(2, 4, 0) + (8)(0, 0, 1)$$

so that

$$A_1 = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 4 & 8 \end{bmatrix}$$

is a matrix of transition from \mathcal{A} to \mathcal{B} . But

$$(1, 2, 4) = (-1)(1, 2, 0) + (1)(2, 4, 0) + (4)(0, 0, 1)$$

and

$$(2, 4, 8) = (0)(1, 2, 0) + (1)(2, 4, 0) + (8)(0, 0, 1)$$

so that

$$A_2 = \begin{bmatrix} -1 & 0 \\ 1 & 1 \\ 4 & 8 \end{bmatrix}$$

is also a matrix of transition from \mathcal{A} to \mathcal{B} . Obviously $A_1 \neq A_2$, but we note that \mathcal{A} is not linearly independent and thus is not a basis for $\langle \mathcal{A} \rangle$. ■

Note that if \mathcal{A} is a basis of a subspace \mathbf{W} and $\mathcal{B} \subseteq \mathbf{W}$, a transition matrix $A = [a_{ij}]_{r \times s}$ from \mathcal{A} to \mathcal{B} always exists. Also, the number of columns in A is the same as the number of vectors in \mathcal{B} , and the number of rows in A is the same as the number of vectors in \mathcal{A} . The scalars a_{ij} that are the coefficients in the sum

$$\mathbf{v}_j = \sum_{i=1}^r a_{ij} \mathbf{u}_i$$

are called the *coordinates* of \mathbf{v}_j relative to \mathcal{A} . It is evident that the coordinates of \mathbf{v}_j relative to \mathcal{A} are recorded in the j^{th} column of the matrix of transition from \mathcal{A} to \mathcal{B} . Consequently, the a_{ij} are unique (since \mathcal{A} is a *basis*), and the matrix $A = [a_{ij}]$ is unique. In particular, any $A = [a_{ij}]_{m \times n}$ may be considered as the matrix of transition from the standard basis \mathcal{E}_m to a set of n vectors in \mathbf{R}^m , and the components of these vectors are given by the columns of A .

Definition 3.4 If $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ is a basis of the subspace \mathbf{W} and $\mathbf{v} = \sum_{i=1}^r c_i \mathbf{u}_i$, the r by 1 matrix C given by

$$C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \end{bmatrix}$$

is called the **coordinate matrix** of \mathbf{v} relative to \mathcal{A} . The coordinate matrix of \mathbf{v} relative to \mathcal{A} is denoted by $[\mathbf{v}]_{\mathcal{A}}$, and the phrase “with respect to \mathcal{A} ” is used interchangeably with “relative to \mathcal{A} .”

For a given basis \mathcal{A} , the vector \mathbf{v} and the coordinate matrix $[\mathbf{v}]_{\mathcal{A}}$ uniquely determine each other. Also, with the notation of the paragraph preceding Definition 3.4, the j^{th} column of A , written as a column matrix

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{rj} \end{bmatrix},$$

is the coordinate matrix of \mathbf{v}_j relative to \mathcal{A} .

Example 3 □ Consider the matrix

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

This matrix is the transition matrix from \mathcal{E}_3 to the set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ where

$$\begin{aligned}\mathbf{v}_1 &= 1 \cdot \mathbf{e}_1 + 2 \cdot \mathbf{e}_2 + 3 \cdot \mathbf{e}_3 = (1, 2, 3), \\ \mathbf{v}_2 &= 2 \cdot \mathbf{e}_1 + 1 \cdot \mathbf{e}_2 + 0 \cdot \mathbf{e}_3 = (2, 1, 0), \\ \mathbf{v}_3 &= 1 \cdot \mathbf{e}_1 + (-4) \cdot \mathbf{e}_2 + 2 \cdot \mathbf{e}_3 = (1, -4, 2), \\ \mathbf{v}_4 &= (-3) \cdot \mathbf{e}_1 + 5 \cdot \mathbf{e}_2 + 1 \cdot \mathbf{e}_3 = (-3, 5, 1).\end{aligned}$$

Using the basis $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, where $\mathbf{u}_1 = (0, 1, 1)$, $\mathbf{u}_2 = (1, 0, 1)$, $\mathbf{u}_3 = (1, 1, 0)$, the same matrix A is the matrix of transition from \mathcal{A} to the set $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\}$ where

$$\begin{aligned}\mathbf{w}_1 &= 1 \cdot \mathbf{u}_1 + 2 \cdot \mathbf{u}_2 + 3 \cdot \mathbf{u}_3 = (5, 4, 3), \\ \mathbf{w}_2 &= 2 \cdot \mathbf{u}_1 + 1 \cdot \mathbf{u}_2 + 0 \cdot \mathbf{u}_3 = (1, 2, 3), \\ \mathbf{w}_3 &= 1 \cdot \mathbf{u}_1 + (-4) \cdot \mathbf{u}_2 + 2 \cdot \mathbf{u}_3 = (-2, 3, -3), \\ \mathbf{w}_4 &= (-3) \cdot \mathbf{u}_1 + 5 \cdot \mathbf{u}_2 + 1 \cdot \mathbf{u}_3 = (6, -2, 2).\end{aligned}$$

In this situation, we have an illustration of how the coordinates of a vector change from

basis to basis. For the vector $\mathbf{v}_1 = \mathbf{w}_2 = (1, 2, 3)$ has the coordinate matrix $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ relative to \mathcal{E}_3 and the coordinate matrix $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ relative to \mathcal{A} . ■

Exercises 3.2

1. If

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

is the matrix of transition from \mathcal{E}_3 to \mathcal{B} , find the standard basis of $\langle \mathcal{B} \rangle$.

2. If

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

is the matrix of transition from \mathcal{E}_4 to \mathcal{B} , find the standard basis of $\langle \mathcal{B} \rangle$.

3. Determine x and y so that $\begin{bmatrix} 2x+y & x-2y \\ 4x-8y & 3x-y \end{bmatrix}$ is a diagonal matrix.
4. If C is the matrix of transition from a set of p vectors in \mathbf{R}^n to a set of q vectors in \mathbf{R}^n , how many rows are in C ?
5. Find the transition matrix from the first set to the second.
- (a) $\{(1, 1, 0, 0), (0, 0, 1, 1)\}; \quad \{(2, 2, 3, 3), (1, 1, 2, 2)\}$
 (b) $\{(2, 2, 3, 3), (1, 1, 2, 2)\}; \quad \{(1, 1, 0, 0), (0, 0, 1, 1)\}$
 (c) $\{(1, 3), (1, 1)\}; \quad \{(3, 5), (5, 3)\}$
 (d) $\{(3, 3), (1, -1)\}; \quad \{(5, 2), (7, 10)\}$
 (e) $\{(2, 2, 0, 0), (1, 2, 1, 0), (1, 1, 1, 1), (0, 0, 0, 1)\}; \quad \{(4, 5, 2, 2), (1, 1, -1, 0)\}$
 (f) $\{(2, 3), (4, -2)\}; \quad \{(6, 1), (8, -4), (2, -5)\}$
6. In each part below, a set \mathcal{A} and a matrix A are given. In each case, find \mathcal{B} so that A is the matrix of transition from \mathcal{A} to \mathcal{B} .
- (a) $\mathcal{A} = \{(1, 2), (0, 1)\}; A = \begin{bmatrix} 4 & -1 \\ -4 & 3 \end{bmatrix}$
- (b) $\mathcal{A} = \{(1, 2), (2, 1)\}; A = \begin{bmatrix} 5 & 7 \\ 2 & 10 \end{bmatrix}$
- (c) $\mathcal{A} = \{(1, 0, 2, -1), (0, 1, 1, 2), (1, 2, 1, 4), (2, 2, 3, 0)\};$
 $A = \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 3 & -6 & 1 & -2 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & -1 & 2 & 2 & 3 \end{bmatrix}$
- (d) $\mathcal{A} = \{(3, 1, 2, 4), (1, 0, 1, -1), (0, 2, 0, -4), (0, -3, 1, -3)\}; A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -3 & 2 \\ 0 & 1 & -2 \\ 0 & 1 & 0 \end{bmatrix}$
- (e) $\mathcal{A} = \{(0, 1, -2), (-1, 1, 2), (1, -1, 0)\}; A = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & -1 \\ -1 & 2 & 0 \end{bmatrix}$

$$(f) \quad \mathcal{A} = \{(2, 1, 0, -1, 0), (0, -2, 0, 2, 0), (1, -1, 0, 1, 0)\}; \quad A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & -4 \end{bmatrix}$$

7. Determine whether or not there is a matrix of transition from the first set to the second, and find such a matrix if it exists.

- (a) $\{(1, 0, 2, 0), (0, 1, 2, 1)\}; \quad \{(3, 0, 6, 0), (-8, 4, -8, 4), (-5, 4, -2, 4)\}$
- (b) $\{(2, 4, 6, 4), (4, 6, 7, 3), (2, 0, -4, -6)\}; \quad \{(2, 2, 1, -1), (0, 2, 5, 5)\}$
- (c) $\{(2, 2, 0, 0), (1, 2, 1, 0), (1, 1, 1, 1)\}; \quad \{(1, 1, 0, 0), (0, 1, 1, 0), (0, 0, 1, 2)\}$
- (d) $\{(2, 4, 2, -2), (0, 3, -2, 4)\}; \quad \{(1, 2, 1, -1), (0, 1, 2, 3), (1, 4, 5, 5)\}$

8. Each of the sets \mathcal{A} below is a basis of $\langle \mathcal{A} \rangle$. Find the vector that has the given matrix C as its coordinate matrix with respect to \mathcal{A} .

$$(a) \quad \mathcal{A} = \{(1, 3), (-2, 1)\}; \quad C = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

$$(b) \quad \mathcal{A} = \{(2, -1, 1), (0, 1, -1), (-2, 1, 0)\}; \quad C = \begin{bmatrix} 3 \\ 1 \\ -4 \end{bmatrix}$$

$$(c) \quad \mathcal{A} = \{(3, 2, -1, 0), (0, -2, 5, 0), (2, 0, -4, 1)\}; \quad C = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$$

$$(d) \quad \mathcal{A} = \{(0, 3, 4, 6), (-1, -2, 0, 2), (4, 0, -3, 1)\}; \quad C = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

9. With \mathcal{A} as given in the corresponding part of Problem 8, the given vector \mathbf{v} is in $\langle \mathcal{A} \rangle$. Find the coordinate matrix of \mathbf{v} relative to \mathcal{A} .

- (a) $\mathbf{v} = (2, 20)$ (b) $\mathbf{v} = (4, 0, -1)$ (c) $\mathbf{v} = (4, 8, -8, -1)$ (d) $\mathbf{v} = (0, 8, 3, -9)$

10. Assume that the vector $\mathbf{v} = (x_1, x_2, x_3)$ has coordinate matrix $[\mathbf{v}]_{\mathcal{A}} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$

relative to the basis $\mathcal{A} = \{(1, 0, 1), (0, 1, 1), (1, 1, 0)\}$, and express the components x_1, x_2, x_3 of \mathbf{v} in terms of c_1, c_2 , and c_3 .

11. Suppose that the vector $\mathbf{v} = (x_1, x_2, x_3)$ has coordinate matrix $[\mathbf{v}]_{\mathcal{A}} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$ with respect to the basis $\mathcal{A} = \{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$. Express each d_i in terms of x_1, x_2 , and x_3 .
12. Assume that $\mathcal{A} = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis of \mathbf{R}^3 , where $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, and $\mathbf{w} = (w_1, w_2, w_3)$. If a vector (x_1, x_2, x_3) has coordinates d_1, d_2, d_3 relative to \mathcal{A} , express the components x_i in terms of the components of \mathbf{u}, \mathbf{v} , and \mathbf{w} .
13. Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$ be bases of \mathbf{W} , and let $P = [p_{ij}]_{r \times r}$ be the matrix of transition from \mathcal{A} to \mathcal{B} . If \mathbf{v} has coordinates c_1, c_2, \dots, c_r relative to \mathcal{A} and d_1, d_2, \dots, d_r relative to \mathcal{B} , what relation exists between the c_i and the d_i ?

3.3 Properties of Matrix Multiplication

Let $A = [a_{ij}]_{r \times s}$ be a given r by s matrix over \mathbf{R} . If \mathbf{W} is any subspace of \mathbf{R}^n and $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ is any set of r vectors in \mathbf{W} , then A is the matrix of transition from \mathcal{A} to a unique set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$, where $\mathbf{v}_j = \sum_{i=1}^r a_{ij} \mathbf{u}_i$. That is, once \mathbf{W} and \mathcal{A} are chosen, the matrix A is the matrix of transition from \mathcal{A} to one and only one subset \mathcal{B} . Also, as was pointed out in the last section, there is only one matrix of transition from \mathcal{A} to \mathcal{B} if \mathcal{A} is a basis of \mathbf{W} . With these facts in mind, we make the following definition.

Definition 3.5 Let $A = [a_{ij}]_{r \times s}$ and $B = [b_{ij}]_{s \times t}$ be matrices over \mathbf{R} , and let \mathbf{W} be a subspace of \mathbf{R}^n of dimension r . Then A is the matrix of transition from a basis $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ of \mathbf{W} to a set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$ in \mathbf{W} , and B is a matrix of transition from \mathcal{B} to a set $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t\}$ in \mathbf{W} . The product AB is defined to be the matrix of transition from \mathcal{A} to \mathcal{C} . (See Figure 3.1.)

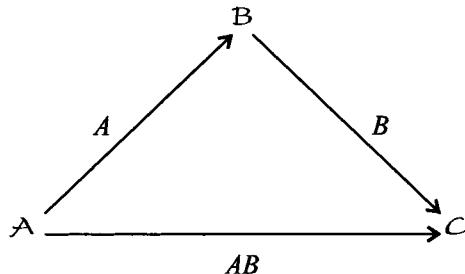
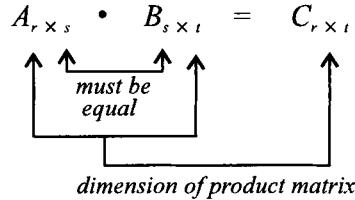


Figure 3.1

We observe that if A is r by s and B is s by t , then AB is r by t . Also, the number of rows in B must be the same as the number of columns in A in order for AB to be defined. When the product AB is defined, we say that B is *conformable* to A . The roles played by the dimensions of the matrices involved are shown in the following diagram.



The product AB as given in Definition 3.5 involves not only the matrices A and B , but choices of \mathcal{A} and \mathbf{W} as well. This means that there is a possibility that the product AB may not be unique, but may vary with the choices of \mathcal{A} and \mathbf{W} . The next theorem shows that this does not happen, and the product AB actually depends only on A and B .

Theorem 3.6 *If $A = [a_{ij}]_{r \times s}$, $B = [b_{ij}]_{s \times t}$ and $AB = [c_{ij}]_{r \times t}$, then $c_{ij} = \sum_{k=1}^s a_{ik}b_{kj}$ for all pairs i, j .*

Proof. Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ be a basis of \mathbf{W} . Then $A = [a_{ij}]_{r \times s}$ is the matrix of transition from \mathcal{A} to a set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$, and $B = [b_{ij}]_{s \times t}$ is a matrix of transition from \mathcal{B} to a set $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t\}$. That is, $\mathbf{v}_k = \sum_{i=1}^r a_{ik}\mathbf{u}_i$ for $k = 1, 2, \dots, s$ and $\mathbf{w}_j = \sum_{k=1}^s b_{kj}\mathbf{v}_k$ for $j = 1, 2, \dots, t$. Thus

$$\begin{aligned}\mathbf{w}_j &= \sum_{k=1}^s b_{kj}\mathbf{v}_k \\ &= \sum_{k=1}^s b_{kj} \left(\sum_{i=1}^r a_{ik}\mathbf{u}_i \right) \\ &= \sum_{k=1}^s \left(\sum_{i=1}^r a_{ik}b_{kj}\mathbf{u}_i \right) \\ &= \sum_{i=1}^r \left(\sum_{k=1}^s a_{ik}b_{kj} \right) \mathbf{u}_i.\end{aligned}$$

But $AB = [c_{ij}]_{r \times t}$ is the matrix of transition from \mathcal{A} to \mathcal{C} , so that $\mathbf{w}_j = \sum_{i=1}^r c_{ij}\mathbf{u}_i$. Since \mathcal{A} is a basis, \mathbf{w}_j can be written as a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$ in only one way. Therefore, $c_{ij} = \sum_{k=1}^s a_{ik}b_{kj}$ for all pairs i, j . ■

If A is a matrix with only one row, $A = [a_1, a_2, \dots, a_s]$, and B is a matrix with a

single column,

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{bmatrix},$$

then AB has only one element, given by $a_1b_1 + a_2b_2 + \cdots + a_sb_s$. This result can be committed to memory easily by mentally “standing A by B ,” forming products of pairs of corresponding elements, and then forming the sum of these products:

$$\begin{aligned} & a_1b_1 \\ & + a_2b_2 \\ & + \cdots \\ & + a_sb_s. \end{aligned}$$

It is easily seen from the formula in Theorem 3.6 that the element in the i^{th} row and j^{th} column of AB can be found by multiplying the i^{th} row of A by the j^{th} column of B , following the routine given for a single row and a single column. This aid to mental multiplication is known as the *row-by-column-rule*.

The row-by-column rule uses the same pattern as the one for computing the inner product of two vectors. (See Definition 1.18.) This pattern is illustrated in the following diagram.

$$\begin{aligned} & \underbrace{\text{column } j \text{ of } B}_{\begin{bmatrix} \cdots & b_{1j} & \cdots \\ \cdots & b_{2j} & \cdots \\ \cdots & b_{3j} & \cdots \\ \vdots & & \vdots \\ \cdots & b_{nj} & \cdots \end{bmatrix}} \\ \text{row } i \text{ of } A \} & \left[\begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & & \vdots \\ \end{array} \right] \cdot \begin{array}{l} \\ \\ \\ \end{array} \\ & = \underbrace{\begin{bmatrix} \vdots \\ \cdots & c_{ij} & \cdots \\ \vdots \end{bmatrix}}_{\{\text{row } i \text{ of } C\}} \end{aligned}$$

where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}.$$

Example 1 □ Consider the products AB and BA for the matrices

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

The number of rows in B is the same as the number of columns in A , so the product AB is defined. Performing the computations, we find that

$$\begin{aligned} AB &= \begin{bmatrix} 2 & -1 \\ 3 & 1 \\ 6 & -5 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 7 & -2 & -3 \\ 5 & -4 & 0 \end{bmatrix} \\ &= \begin{bmatrix} (2)(7) + (-1)(5) & (2)(-2) + (-1)(-4) & (2)(-3) + (-1)(0) \\ (3)(7) + (1)(5) & (3)(-2) + (1)(-4) & (3)(-3) + (1)(0) \\ (6)(7) + (-5)(5) & (6)(-2) + (-5)(-4) & (6)(-3) + (-5)(0) \\ (0)(7) + (4)(5) & (0)(-2) + (4)(-4) & (0)(-3) + (4)(0) \end{bmatrix} \\ &= \begin{bmatrix} 9 & 0 & -6 \\ 26 & -10 & -9 \\ 17 & 8 & -18 \\ 20 & -16 & 0 \end{bmatrix}. \end{aligned}$$

The product BA is not defined because the number of rows in A is 4 and the number of columns in B is 3. ■

The major result concerning properties of matrix multiplication is stated next.

Theorem 3.7 Let $A = [a_{ij}]_{r \times s}$, $B = [b_{ij}]_{s \times t}$, and $C = [c_{ij}]_{t \times v}$ be matrices over \mathbf{R} . Then $A(BC) = (AB)C$.

Proof. By Theorem 3.6, $AB = [d_{ij}]_{r \times t}$ where $d_{ij} = \sum_{m=1}^s a_{im}b_{mj}$, and $(AB)C = \left[\sum_{k=1}^t d_{ik}c_{kj} \right]_{r \times v}$ where

$$\begin{aligned} \sum_{k=1}^t d_{ik}c_{kj} &= \sum_{k=1}^t \left(\sum_{m=1}^s a_{im}b_{mk} \right) c_{kj} \\ &= \sum_{k=1}^t \left(\sum_{m=1}^s (a_{im}b_{mk})c_{kj} \right). \end{aligned}$$

Similarly, $BC = [f_{ij}]_{s \times v}$ where $f_{ij} = \sum_{k=1}^t b_{ik}c_{kj}$, and $A(BC) = [\sum_{m=1}^s a_{im}f_{mj}]_{r \times v}$ where

$$\begin{aligned}\sum_{m=1}^s a_{im}f_{mj} &= \sum_{m=1}^s a_{im} \left(\sum_{k=1}^t b_{mk}c_{kj} \right) \\ &= \sum_{m=1}^s \left(\sum_{k=1}^t a_{im}(b_{mk}c_{kj}) \right) \\ &= \sum_{k=1}^t \left(\sum_{m=1}^s a_{im}(b_{mk}c_{kj}) \right).\end{aligned}$$

But $a_{im}(b_{mk}c_{kj}) = (a_{im}b_{mk})c_{kj}$ from the associative property of multiplication of real numbers. Hence $A(BC) = (AB)C$. ■

If all necessary conformability is assumed, then matrix multiplication is associative.

Now the matrix product BA may fail to exist, even when AB is defined, so the commutative property of multiplication does not hold in general. In Example 2, matrices A and B are given for which AB and BA are both defined, and yet $AB \neq BA$.

Example 2 □ Let $A = \begin{bmatrix} 1 & 2 \\ 4 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 7 & 1 \\ 12 & -4 \end{bmatrix} \quad \text{and} \quad BA = \begin{bmatrix} -1 & 6 \\ 6 & 4 \end{bmatrix},$$

so that $AB \neq BA$. ■

There are two other fundamental properties of multiplication of real numbers that are not valid for multiplication of matrices. In general $AB = AC$ and $A \neq \mathbf{0}$ do not imply $B = C$, nor do $BA = CA$ and $A \neq \mathbf{0}$ imply that $B = C$. That is, there is no cancellation property for matrix multiplication. Also, $AB = \mathbf{0}$ does not imply that one of A, B must be zero. Thus, the product of two nonzero matrices may be a zero matrix. Examples of these situations are requested in some of the exercises.

Exercises 3.3

1. Compute the product AB for the given matrices.

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

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

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

$$(d) A = \begin{bmatrix} 5 & -2 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$$

$$(e) A = \begin{bmatrix} 2 & -5 & 3 & 0 \\ -6 & 8 & 1 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & -1 \\ -8 & -1 \\ -9 & 3 \\ 4 & 4 \end{bmatrix}$$

$$(f) A = \begin{bmatrix} 2 & 1 & -4 \\ 5 & 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 8 \\ 4 \\ -5 \end{bmatrix}$$

2. Whenever possible, compute the products BA in Problem 1.
3. Let $A = [a_{ij}]_{m \times n}$ and $B = [b_{ij}]_{r \times t}$ be matrices over \mathbf{R} .
- (a) What conditions are necessary in order that AB be defined?
 - (b) What are the dimensions of AB when it exists?
 - (c) What conditions are necessary in order that both AB and BA be defined?
 - (d) What conditions are necessary in order that the dimensions of AB be the same as the dimensions of BA ?
4. Given that $A = \begin{bmatrix} 4 & -2 & 1 & 0 \\ 5 & 0 & 2 & 1 \end{bmatrix}$ is the transition matrix from $\mathcal{A} = \{(1, 1), (0, 1)\}$ to the set \mathcal{B} , and that $B = \begin{bmatrix} 1 & 0 \\ 4 & -3 \\ 0 & -1 \\ -8 & 6 \end{bmatrix}$ is the transition matrix from \mathcal{B} to \mathcal{C} , find the matrix of transition from \mathcal{A} to \mathcal{C} .
5. Give an example of nonzero matrices A and B such that $AB = BA$.
6. Let $A = [a_{ij}]_{r \times s}$. Under what conditions does A^k exist for every positive integer k ? ($A^1 = A, A^2 = A \cdot A$, etc.)

7. Give an example of two 2×2 matrices A and B over \mathbf{R} such that $AB = \mathbf{0}$ but $A \neq \mathbf{0}$ and $B \neq \mathbf{0}$.
8. Give an example of 2×2 matrices A, B , and C over \mathbf{R} such that $AB = AC$, but $B \neq C$ and $A \neq \mathbf{0}$.
9. Let $A = [a_{ij}]_{2 \times 3}, B = [b_{ij}]_{3 \times 2}$, and $C = [c_{ij}]_{2 \times 1}$ over \mathbf{R} . For these particular matrices, verify that $A(BC) = (AB)C$ without using Theorem 3.7.
10. For each of the following pairs A, B , let $AB = [c_{ij}]$ and write a formula for c_{ij} in terms of i and j .
 - (a) $A = [a_{ij}]_{2 \times 3}$ where $a_{ij} = i + j, B = [b_{ij}]_{3 \times 4}$ where $b_{ij} = 2i - j$
 - (b) $A = [a_{ij}]_{3 \times 2}$ where $a_{ij} = i - j, B = [b_{ij}]_{2 \times 4}$ where $b_{ij} = i + j$
 - (c) $A = [a_{ij}]_{3 \times 2}$ where $a_{ij} = i - j, B = [b_{ij}]_{2 \times 3}$ where $b_{ij} = 2i + j$
 - (d) $A = [a_{ij}]_{2 \times 3}$ where $a_{ij} = i + 2j, B = [b_{ij}]_{3 \times 2}$ where $b_{ij} = i - 2j$

11. Show that the matrix equation

$$\begin{bmatrix} 1 & -2 & 1 \\ 2 & 0 & 3 \\ 1 & 4 & -1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

is equivalent to a system of linear equations in x_1, x_2 , and x_3 . (Hint: Use Definition 3.2.)

12. Reversing the procedure in Problem 11, write a matrix equation that is equivalent to the system of equations

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

13. Prove Theorem 3.7 by direct use of Definition 3.5.

3.4 Invertible Matrices

Those matrices that are matrices of transition from one basis to another basis are of particular importance.

Definition 3.8 A matrix $A = [a_{ij}]_{n \times n}$ is **nonsingular** if and only if A is a matrix of transition from one basis of \mathbf{R}^n to another basis of \mathbf{R}^n . A square matrix that is not nonsingular is called **singular**.

Suppose that $A = [a_{ij}]$ is a square matrix of order n . For each basis \mathcal{A} of \mathbf{R}^n , A is the matrix of transition from \mathcal{A} to a set \mathcal{B} of n vectors in \mathbf{R}^n . Generally, different choices of \mathcal{A} yield different sets \mathcal{B} . In order for the term *nonsingular* to be well-defined, we must show that if one of the sets \mathcal{B} is a basis, then all of them are bases.

Suppose $A = [a_{ij}]_{n \times n}$ is the matrix of transition from the basis $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ of \mathbf{R}^n to the basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{R}^n . Let $\mathcal{A}' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_n\}$ be another basis of \mathbf{R}^n , and let $\mathcal{B}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n\}$ be the set so that A is the matrix of transition from \mathcal{A}' to \mathcal{B}' .

In order to show that \mathcal{B}' is a basis of \mathbf{R}^n , it is sufficient to show that \mathcal{B}' is linearly independent (Theorem 1.33). Suppose that

$$\sum_{j=1}^n b_j \mathbf{v}'_j = \mathbf{0}.$$

Since $\mathbf{v}'_j = \sum_{i=1}^n a_{ij} \mathbf{u}'_i$, we have $\sum_{j=1}^n b_j (\sum_{i=1}^n a_{ij} \mathbf{u}'_i) = \mathbf{0}$ and

$$\sum_{i=1}^n \left(\sum_{j=1}^n b_j a_{ij} \right) \mathbf{u}'_i = \mathbf{0}.$$

But \mathcal{A}' is linearly independent, so this means that $\sum_{j=1}^n b_j a_{ij} = 0$ for $i = 1, 2, \dots, n$. Hence

$$\begin{aligned} \sum_{j=1}^n b_j \mathbf{v}_j &= \sum_{j=1}^n b_j \left(\sum_{i=1}^n a_{ij} \mathbf{u}_i \right) \\ &= \sum_{i=1}^n \left(\sum_{j=1}^n b_j a_{ij} \right) \mathbf{u}_i \\ &= \sum_{i=1}^n 0 \cdot \mathbf{u}_i \\ &= \mathbf{0}, \end{aligned}$$

and this implies that $b_1 = b_2 = \dots = b_n = 0$ since \mathcal{B} is linearly independent. Thus, \mathcal{B}' is linearly independent and is a basis of \mathbf{R}^n .

An n by n matrix A over \mathbf{R} is a nonsingular matrix if and only if the columns of A record the coordinates of one basis of \mathbf{R}^n with respect to a second (not necessarily different) basis of \mathbf{R}^n . That is, A is nonsingular if and only if the j^{th} column of A is the coordinate matrix of the j^{th} vector in a basis of \mathbf{R}^n for $j = 1, 2, \dots, n$.

In the discussion of special types of matrices just before Definition 3.2, an *identity matrix* was defined to be a matrix of the form $I_n = [\delta_{ij}]_{n \times n}$, where δ_{ij} is the Kronecker delta. There are many identity matrices, I_n , but only one for each value of n . As examples,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

From the placing of the 1's and 0's in I_n , it is easy to see that I_n is the unique matrix of transition from a basis \mathcal{A} to the same basis \mathcal{A} , and I_n is therefore nonsingular.

Using the fact that I_n is the transition matrix from \mathcal{A} to \mathcal{A} , it follows easily from Definition 3.5 that

$$I_m A = A \quad \text{and} \quad A I_n = A$$

for any $m \times n$ matrix A . In particular,

$$I_n A = A = A I_n$$

if A is an $n \times n$ matrix. Thus I_n is a *multiplicative identity* for square matrices of order n , and it is upon this fact that the term *identity matrix* is based.

The existence of a multiplicative identity for square matrices of order n leads naturally to the question as to which square matrices have multiplicative inverses. When working with matrices, it is conventional to use the term *inverse* to mean “multiplicative inverse.”

Definition 3.9 *Let A be a square matrix of order n . An $n \times n$ matrix B is an **inverse** of A if $AB = I_n = BA$. Also, a square matrix is called **invertible** if it has an inverse.*

We note that inverses of matrices occur in pairs in this sense: If B is an inverse of A , then A is also an inverse of B .

Our next theorem gives an answer to the question as to which square matrices are invertible.

Theorem 3.10 *Let A be a square matrix of order n . Then A is invertible if and only if A is nonsingular.*

Proof. Suppose that \mathcal{A} is a basis of \mathbf{R}^n and let the $n \times n$ matrix A be the transition matrix from \mathcal{A} to a set \mathcal{B} of n vectors in \mathbf{R}^n .

Assume first that A is nonsingular. Then \mathcal{B} is a basis of \mathbf{R}^n and therefore every vector in \mathcal{A} is a linear combination of vectors in \mathcal{B} . Hence there exists a transition matrix B from \mathcal{B} to \mathcal{A} . By Definition 3.5, AB is the matrix of transition from \mathcal{A} to \mathcal{A} . It follows that $AB = I_n$ since I_n is the unique matrix of transition from \mathcal{A} to \mathcal{A} . In similar fashion, we can show that BA is the matrix of transition from the basis \mathcal{B} to \mathcal{B} , and therefore $BA = I_n$.

To prove the other part of the theorem, assume that A has an inverse B , so that $AB = I_n$ and $BA = I_n$. With the same notation as in the first paragraph of this proof, let B be the transition matrix from the set \mathcal{B} of n vectors to a set \mathcal{C} of n vectors in \mathbf{R}^n . Then AB is the transition matrix from \mathcal{A} to \mathcal{C} , by Definition 3.5. But $AB = I_n$, so the set \mathcal{C} is exactly the same as \mathcal{A} . This means that B is a transition matrix from \mathcal{B} to \mathcal{A} . Therefore \mathcal{A} is dependent on \mathcal{B} , and this implies that \mathbf{R}^n is dependent on \mathcal{B} , by Theorem 1.8. That is, \mathcal{B} spans \mathbf{R}^n . It follows then from Theorem 1.34 that \mathcal{B} is a basis of \mathbf{R}^n and hence A is nonsingular. ■

Corollary 3.11 *Suppose A is the matrix of transition from a basis \mathcal{A} of \mathbf{R}^n to a basis \mathcal{B} of \mathbf{R}^n . Then the matrix B is an inverse of A if and only if B is the transition matrix from \mathcal{B} to \mathcal{A} .*

Proof. The corollary follows at once from the proof of the theorem. ■

Example 1 □ The matrix

$$A = \begin{bmatrix} 3 & 15 \\ 4 & 20 \end{bmatrix}$$

is not invertible. It is the transition matrix from the standard basis $\mathcal{E}_2 = \{(1, 0), (0, 1)\}$ to $\mathcal{B} = \{(3, 4), (15, 20)\}$, and \mathcal{B} is clearly dependent because its second vector is equal to 5 times its first vector. As an example of an invertible matrix, consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This matrix is the transition matrix from $\mathcal{E}_3 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ to $\mathcal{B} = \{1, 0, 0\}, (1, 1, 0), (1, 1, 1)\}$ and it is easy to see that \mathcal{B} is linearly independent since none of its vectors are linear combinations of the preceding vectors. Of course, it is not always so easy to tell whether a matrix is invertible or not. ■

Up to this point, we have allowed the possibility that a matrix might have two or more distinct inverses. The next theorem shows that this possibility does not actually happen.

Theorem 3.12 *The inverse of an invertible matrix is unique.*

Proof. Suppose that B and C are both inverses of the $n \times n$ matrix A . Then all of the equations

$$AB = I_n = BA \quad \text{and} \quad AC = I_n = CA$$

are valid. To prove that $B = C$, we evaluate the product BAC in two different ways. First, we have

$$\begin{aligned} B(AC) &= BI_n \quad \text{since } AC = I_n \\ &= B \quad \text{since } I_n \text{ is a multiplicative identity.} \end{aligned}$$

Second, we have

$$\begin{aligned} (BA)C &= I_n C \quad \text{since } BA = I_n \\ &= C \quad \text{since } I_n \text{ is a multiplicative identity.} \end{aligned}$$

But we proved that matrix multiplication is associative in Theorem 3.7. Therefore $B(AC) = (BA)C$, and hence $B = C$. ■

Theorem 3.12 allows us to make the following definition.

Definition 3.13 If A is an invertible matrix, its unique inverse is denoted by A^{-1} .

Example 2 □ We saw in Example 1 of this section that the matrix

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

is invertible and is the transition matrix from the basis $\mathcal{E}_3 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ to the basis $\mathcal{B} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$ of \mathbf{R}^3 . In order to find A^{-1} , it is sufficient to find the transition matrix from \mathcal{B} to \mathcal{E}_3 . Since

$$\begin{aligned} (1, 0, 0) &= (1)(1, 0, 0) + (0)(1, 1, 0) + (0)(1, 1, 1) \\ (0, 1, 0) &= (-1)(1, 0, 0) + (1)(1, 1, 0) + (0)(1, 1, 1) \\ (0, 0, 1) &= (0)(1, 0, 0) + (-1)(1, 1, 0) + (1)(1, 1, 1), \end{aligned}$$

we see that

$$A^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

If a given square matrix A is complicated, a more efficient method than the one we used here is needed to find A^{-1} . Such a method is presented in Section 3.6. ■

Some of our earlier results can be rewritten using the exponential notation in Definition 3.13. The equations $AB = I_n = BA$ now read as

$$AA^{-1} = I_n = A^{-1}A.$$

Also, we noted earlier that inverses occur in pairs: When $B = A^{-1}$, then $A = B^{-1}$. Substituting the value $B = A^{-1}$ in the equation $A = B^{-1}$ yields

$$A = (A^{-1})^{-1}.$$

In the definition and all discussion in this section, it has been required that both of the equations $AA^{-1} = I_n$ and $A^{-1}A = I_n$ be satisfied by the inverse matrix. Our next two theorems show that these equations are not independent for square matrices. In fact, for square matrices, we shall see that either of them implies the other.

Theorem 3.14 Let A be a square matrix of order n over \mathbf{R} . If there is a square matrix B over \mathbf{R} such that $AB = I_n$, then A is invertible and $B = A^{-1}$.

Proof. Suppose there is a square matrix B such that $AB = I_n$. Now A is the matrix of transition from \mathcal{E}_n to a set \mathcal{A} of n vectors in \mathbf{R}^n , B is a matrix of transition from \mathcal{A} to a set \mathcal{B} of n vectors in \mathbf{R}^n , and $AB = I_n$, so \mathcal{B} must be \mathcal{E}_n . Thus \mathcal{B} is a matrix of transition from \mathcal{A} to \mathcal{E}_n , and this means that each vector in \mathcal{E}_n is a linear combination of vectors in \mathcal{A} . That is, \mathcal{E}_n is dependent on \mathcal{A} . And since \mathbf{R}^n is dependent on \mathcal{E}_n , this means that \mathbf{R}^n is dependent on \mathcal{A} . By Theorem 1.34, \mathcal{A} is a basis of \mathbf{R}^n , and A is invertible. It follows from Corollary 3.11 that $B = A^{-1}$. ■

In view of Theorem 3.14, we see that a matrix A of order n is invertible if and only if there is a square matrix B such that $AB = I_n$.

The proof of the next theorem is quite similar to that of Theorem 3.14, and is left as an exercise.

Theorem 3.15 *Let A be a square matrix of order n over \mathbf{R} . If there is a square matrix B over \mathbf{R} such that $BA = I_n$, then A is invertible and $B = A^{-1}$.*

Theorem 3.16 *If A and B are invertible matrices of the same order, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. If A and B are invertible matrices of order n , then A^{-1} and B^{-1} exist. Since

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = I_n,$$

AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$ by Theorem 3.14. ■

Repeated application of Theorem 3.16 yields the following corollary.

Corollary 3.17 *If A_1, A_2, \dots, A_k are square matrices of order n over \mathbf{R} and each A_i is invertible, then $A_1A_2 \cdots A_k$ is invertible and $(A_1A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1}A_1^{-1}$.*

Definition 3.18 *If \mathcal{A} and \mathcal{A}' are bases of a subspace \mathbf{W} of \mathbf{R}^n such that \mathcal{A}' is obtained from \mathcal{A} by a single elementary operation, then the matrix of transition from \mathcal{A} to \mathcal{A}' is an elementary matrix.*

That is, a matrix M is an elementary matrix if and only if it is a matrix of transition from a given basis to another basis that is obtained by applying a single elementary operation E to the original basis. We shall say that the elementary matrix M and the elementary operation E are *associated* with each other. The elementary matrix is classified as type I, II, or III according to the type of elementary operation that is applied to the original basis.

Theorem 3.19 *A square matrix over \mathbf{R} is invertible if and only if it is a product of elementary matrices.*

Proof. Suppose that $A = M_1 M_2 \cdots M_t$, where each M_i is an elementary matrix. Now each M_i is invertible since it is a matrix of transition from one basis to another, and therefore A is invertible by Corollary 3.17.

Suppose now that A is of order n and invertible. Then A is the matrix of transition from \mathcal{E}_n to a basis \mathcal{A} of \mathbf{R}^n . By Theorem 2.13, \mathcal{A} can be obtained from \mathcal{E}_n by a sequence E_1, E_2, \dots, E_t of elementary operations. If M_i is the elementary matrix that is the matrix of transition for the operation E_i , then $M_1 M_2 \cdots M_t$ is a matrix of transition from \mathcal{E}_n to \mathcal{A} . But this matrix of transition is unique since \mathcal{E}_n is a basis, and therefore $A = M_1 M_2 \cdots M_t$. ■

If it is desired to write a given invertible $n \times n$ matrix A as a product of elementary matrices, there may be some difficulty in following one of the steps described in the proof of the last theorem. Usually it is not obvious how the basis \mathcal{A} can be obtained from \mathcal{E}_n by a sequence E_1, E_2, \dots, E_t of elementary operations, and an indirect approach is easier to use. If we start with the set \mathcal{A} and apply a sequence of elementary operations F_1, F_2, \dots, F_k to find the standard basis of $\langle \mathcal{A} \rangle$, that standard basis will be \mathcal{E}_n since $\langle \mathcal{A} \rangle = \mathbf{R}^n$. We will then have

$$F_k \cdots F_2 F_1(\mathcal{A}) = \mathcal{E}_n$$

and therefore

$$\mathcal{A} = F_1^{-1} F_2^{-1} \cdots F_k^{-1} (\mathcal{E}_n).$$

If M_i is the elementary matrix of transition that is associated with F_i , then this indirect procedure will yield

$$A = M_k^{-1} \cdots M_2^{-1} M_1^{-1}$$

since M_i^{-1} is the elementary matrix associated with the elementary operation F_i^{-1} . Note that we also have

$$A^{-1} = M_1 M_2 \cdots M_k$$

since $M_1 M_2 \cdots M_k$ is the transition matrix from \mathcal{A} to \mathcal{E}_n . This is diagrammed in Figure 3.2.

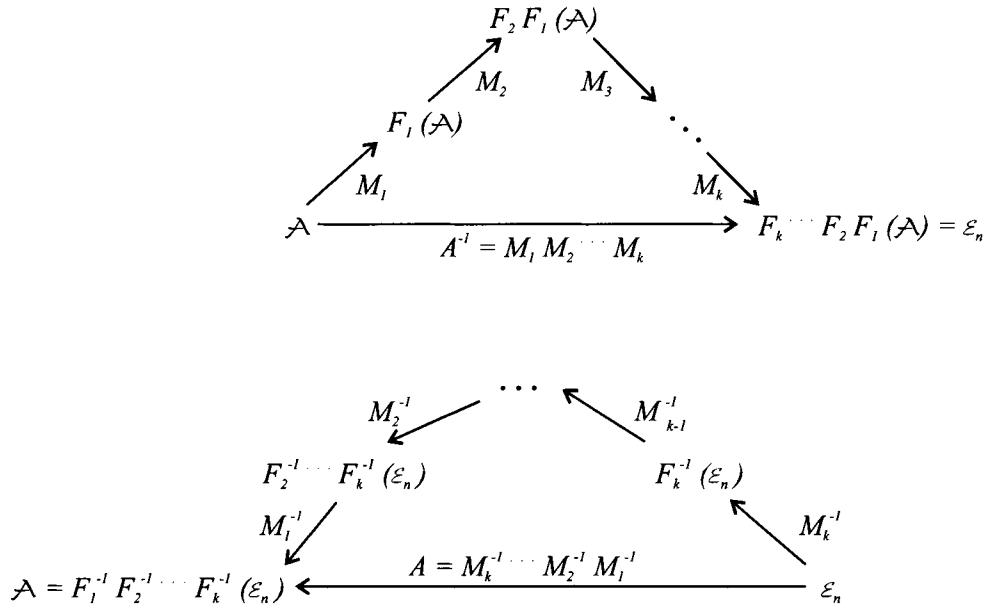


Figure 3.2

Example 3 □ Consider the problem of expressing the invertible matrix

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

as a product of elementary matrices. This matrix A is the transition matrix from \mathcal{E}_n to $\mathcal{A} = \{(2, 6), (3, 4)\}$. We start with \mathcal{A} and apply a single elementary operation at a time until we obtain \mathcal{E}_2 . Using arrows to indicate that an elementary operation F_i has been applied, this appears as follows.

$$\begin{aligned} \mathcal{A} = \{(2, 6), (3, 4)\} &\xrightarrow{F_1} \{(1, 3), (3, 4)\} \\ &\xrightarrow{F_2} \{(1, 3), (0, -5)\} \\ &\xrightarrow{F_3} \{(1, 3), (0, 1)\} \\ &\xrightarrow{F_4} \{(1, 0), (0, 1)\} = \mathcal{E}_2 \end{aligned}$$

The following display shows the elementary operations F_i , their associated elemen-

tary matrices M_i , and the inverses M_i^{-1} .

| ELEMENTARY OPERATION | ASSOCIATED MATRIX | INVERSE |
|---|---|--|
| $F_1 \left\{ \text{Multiply the first vector by } \frac{1}{2}. \right.$ | $M_1 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$ | $M_1^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ |
| $F_2 \left\{ \begin{array}{l} \text{Replace the second vector by the} \\ \text{sum of the second vector and } (-3) \\ \text{times the first vector.} \end{array} \right.$ | $M_2 = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$ | $M_2^{-1} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ |
| $F_3 \left\{ \text{Multiply the second vector by } \left(-\frac{1}{5}\right) \right.$ | $M_3 = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{5} \end{bmatrix}$ | $M_3^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & -5 \end{bmatrix}$ |
| $F_4 \left\{ \begin{array}{l} \text{Replace the first vector by the} \\ \text{sum of the first vector and } (-3) \\ \text{times the second vector.} \end{array} \right.$ | $M_4 = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$ | $M_4^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$ |

According to the discussion preceding this example, we can now express A as a product of elementary matrices by using $A = M_4^{-1}M_3^{-1}M_2^{-1}M_1^{-1}$.

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

Exercises 3.4

1. For each matrix A , determine \mathcal{A} so that A is the matrix of transition from \mathcal{E}_3 to \mathcal{A} . Then use Definition 3.8 to decide whether or not A is nonsingular.

$$\begin{array}{ll}
 \text{(a)} A = \begin{bmatrix} 2 & -6 & 6 \\ -5 & 13 & 1 \\ -2 & 4 & 10 \end{bmatrix} & \text{(b)} A = \begin{bmatrix} 4 & 4 & 4 \\ 3 & 4 & 2 \\ -6 & 1 & 7 \end{bmatrix} \\
 \text{(c)} A = \begin{bmatrix} 5 & 2 & 7 \\ 2 & 1 & 0 \\ 2 & 9 & 3 \end{bmatrix} & \text{(d)} A = \begin{bmatrix} 2 & 1 & 5 \\ 0 & 3 & 9 \\ 4 & -4 & -8 \end{bmatrix}
 \end{array}$$

2. Which of the following are elementary matrices?

$$\begin{array}{llll}
 \text{(a)} \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \text{(b)} \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \text{(c)} \begin{bmatrix} 1 & 4 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \text{(d)} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
 \end{array}$$

$$(e) \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (f) \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (g) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad (h) \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

3. Suppose that \mathcal{A} is a set of three vectors in \mathbf{R}^4 and that \mathcal{B} is obtained from \mathcal{A} by replacing the second vector by the sum of the second vector and 5 times the third vector.

- (a) Write a matrix of transition from \mathcal{A} to \mathcal{B} .
 (b) Write a matrix of transition from \mathcal{B} to \mathcal{A} .

4. Write the inverse of each elementary matrix in Problem 2.

5. Given that $A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, write A^{-1} as a product of elementary matrices.

6. Write each of the following invertible matrices as a product of elementary matrices.

$$(a) \begin{bmatrix} 2 & 4 \\ 3 & 4 \end{bmatrix} \quad (b) \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \quad (c) \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} \quad (d) \begin{bmatrix} 0 & 3 \\ -2 & 6 \end{bmatrix}$$

7. Find the inverse of each matrix in Problem 6 by use of Corollary 3.17 and the factorization obtained in Problem 6.

8. Each of the matrices A below is nonsingular. Use A as the transition matrix from the basis $\mathcal{A} = \{(1, -2), (2, 1)\}$ of \mathbf{R}^2 to the basis \mathcal{A}' of \mathbf{R}^2 . Determine A^{-1} by finding the matrix of transition from \mathcal{A}' to \mathcal{A} .

$$(a) A = \begin{bmatrix} 1 & 3 \\ -1 & -2 \end{bmatrix} \quad (b) A = \begin{bmatrix} 2 & -7 \\ -1 & 4 \end{bmatrix} \quad (c) A = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \quad (d) A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$$

9. Prove Theorem 3.15.

10. Let A, B , and C be square matrices of order n . Prove that if A is invertible and $AB = AC$, then $B = C$.

11. Prove Corollary 3.17 by mathematical induction.

12. Given the matrices $B = \begin{bmatrix} 2 & 4 \\ 7 & 8 \end{bmatrix}$ and $(AB)^{-1} = \begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix}$, find A^{-1} .

13. Given that $B^{-1} = \begin{bmatrix} 1 & 3 \\ -1 & -2 \end{bmatrix}$ and $AB = \begin{bmatrix} -4 & -5 \\ 2 & 3 \end{bmatrix}$, find the matrix A .

14. Suppose $A = \begin{bmatrix} 4 & -2 & 1 & 0 \\ 5 & 0 & 2 & 1 \end{bmatrix}$ is the transition matrix from $\mathcal{A} = \{(1, 1), (1, 0)\}$

to the set \mathcal{B} , and $B = \begin{bmatrix} 1 & 0 \\ 4 & -3 \\ 0 & -1 \\ -8 & 6 \end{bmatrix}$ is the transition matrix from \mathcal{B} to \mathcal{C} .

- (a) Find the matrix of transition from \mathcal{A} to \mathcal{C} .
- (b) Find the set \mathcal{C} .
- (c) Does $(AB)^{-1}$ exist? Explain the reason for your answer.

15. Given that the invertible matrix

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

was obtained from the invertible matrix A by adding 3 times the first column to the second column, find the elementary matrix M such that $BM = A$. (*Hint:* Let B be the matrix of transition from \mathcal{E}_3 to \mathcal{B} .)

16. Derive a formula for the inverse of a nonsingular $A = [a_{ij}]_{2 \times 2}$ by consideration of a system of equations obtained from

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

17. Prove that if $A = [a_{ij}]_{n \times n}$ is not invertible, then AB is not invertible for any $n \times p$ matrix B . (*Hint:* Use Theorem 3.14.)
18. Prove that if $A = [a_{ij}]_{n \times n}$ is singular, then BA is singular for any $p \times n$ matrix B . (*Hint:* Use Theorem 3.15.)

3.5 Column Operations and Column-Echelon Forms

The elementary matrices of each type can be conveniently described by comparing them with the identity matrix I_n . (It is clear that the identity operation has I_n as its associated elementary matrix.) The operation of multiplying a matrix by a scalar is helpful in describing these comparisons. In this section, it is especially useful in situations where the matrix is a column matrix. Other uses will be found in later sections.

Definition 3.20 For any $a \in \mathbf{R}$ and any $A = [a_{ij}]_{m \times n}$ over \mathbf{R} , the product aA is defined by $aA = [aa_{ij}]_{m \times n}$.

Note that in the product aA , each element of A is multiplied by the scalar a . A matrix of the form aI_n is called a **scalar matrix**.

Suppose now that M is an elementary matrix with E as its associated elementary operation.

If M is of type I, then $B = E(A)$ is obtained from a basis \mathcal{A} of \mathbf{R}^n by multiplying the s^{th} vector of \mathcal{A} by a nonzero scalar a . This means that M differs from I_n only in column s , and the s^{th} column of M is the product of a and the s^{th} column of I_n .

If M is of type II, then E replaces one of the vectors \mathbf{v}_s in a basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{R}^n by $\mathbf{v}_s + b\mathbf{v}_t$, where $s \neq t$. This means that M is identical to I_n except in column s . In column s of M , 1 appears in row s and b appears in row t . Thus, M can be obtained from I_n by adding to column s of I_n the product of b and column t of I_n , where $s \neq t$.

Consider now the case where M is of type III. Then E interchanges the s^{th} and t^{th} ($s \neq t$) vectors in a basis of \mathbf{R}^n . This means that M can be obtained from I_n by interchanging the s^{th} and t^{th} columns ($s \neq t$).

We have thus found that each elementary matrix can be obtained from I_n by performing a suitable operation on the columns of I_n .

Definition 3.21 Corresponding to the operations performed on I_n above, we define three types of elementary column operations on a matrix A .

- (I) An elementary column operation of type I multiplies one of the columns of A by a nonzero scalar.
- (II) An elementary column operation of type II adds to column s of A the product of b and column t of A ($s \neq t$).
- (III) An elementary column operation of type III interchanges two columns of A .

A more unwieldy but more accurate description of an elementary column operation of type II would be this statement: an elementary column operation of type II replaces column s of A by the sum of column s and b times column t ($s \neq t$). This wording more closely matches the descriptions we used with elementary operations on sets of vectors.

Let us consider the effect on an m by n matrix A when A is multiplied on the right by an elementary matrix M of order n . This effect is diagrammed in Figure 3.3.

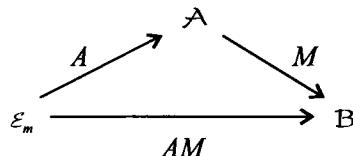


Figure 3.3

Now A is the matrix of transition from the basis \mathcal{E}_m of \mathbf{R}^m to a set \mathcal{A} of n vectors in \mathbf{R}^m , M is a matrix of transition from \mathcal{A} to a set \mathcal{B} of n vectors in \mathbf{R}^m , and AM is the matrix of transition from \mathcal{E}_m to \mathcal{B} .

If M is of type I, then \mathcal{B} is obtained from \mathcal{A} by multiplying the s^{th} vector of \mathcal{A} by a nonzero scalar a . This means that AM can be obtained from A by multiplying the s^{th} column of A by a .

If M is of type II, then \mathcal{B} is obtained from $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ by replacing \mathbf{v}_s by $\mathbf{v}_s + b\mathbf{v}_t$ ($s \neq t$). This means that AM is identical to A except in column s , and column s in AM is obtained by adding to column s of A the product of b and column t .

If M is of type III, then \mathcal{B} is obtained from \mathcal{A} by interchanging the s^{th} and t^{th} vectors ($s \neq t$) of \mathcal{A} . Thus AM is obtained from A by interchanging the s^{th} and t^{th} columns of A ($s \neq t$).

Thus, multiplication of A on the right by an elementary matrix M of a certain type performs an elementary column operation E of the same type on A . We shall say that the elementary matrix M and the elementary column operation E are *associated* with each other whenever multiplication of A on the right by M performs the elementary column operation E on A .

In each case above, AM is obtained from A by a single column operation. Hence performing an elementary operation on \mathcal{A} corresponds to performing an elementary column operation on A . By Theorem 2.7, $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$, so A and AM are matrices of transition from a given basis to set of vectors that span the same subspace.

In the remainder of this section, we shall be concerned with the reduction of a given matrix A to a certain standard form, known as the *reduced column-echelon* form.

Definition 3.22 A matrix $A = [a_{ij}]_{m \times n}$ over \mathbf{R} that satisfies the following conditions is a matrix in **reduced column-echelon form**, or a **reduced column-echelon matrix**.

1. The first nonzero element in column j is a 1 in row k_j for $j = 1, 2, \dots, r$. (This 1 is called a **leading one**.)
2. $k_1 < k_2 < \dots < k_r \leq m$ (That is, for each change in columns from left to right, the leading one appears in a lower row.)
3. For $j = 1, 2, \dots, r$, the leading one in column j is the only nonzero element in row k_j .
4. Each of the last $n - r$ columns consists entirely of zeros.

For future use, we note that conditions (1) and (3) can be reworded in the following ways.

1. $a_{ij} = 0$ for $i < k_j$, and $a_{k_j j} = 1$ for $j = 1, 2, \dots, r$.
3. Column j is the only column with a nonzero element in row k_j .

Thus a matrix is in reduced column-echelon form if and only if its nonzero columns record the components of the standard basis of a subspace, or if its nonzero columns form the matrix of transition from \mathcal{E}_m to the standard basis of a subspace.

Example 1 □ Consider the question as to which of the following matrices are in reduced column-echelon form.

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

The matrix A is not in reduced column-echelon form. It fails to satisfy condition (1) in the second and third columns because the first nonzero element in each of these columns is not a 1. The matrix B is in reduced column-echelon form since it satisfies all four conditions. The matrix C fails on condition (3) because the leading 1 in column 3 is not the only nonzero element in row $k_3 = 4$. The matrix D fails on condition (2) because $k_2 = 3$ and $k_3 = 2$ violate $k_2 < k_3$. That is, the leading 1 in column 3 of D fails to be in a lower row than the leading 1 in column 2. ■

Theorem 3.23 *If A is any $m \times n$ matrix over \mathbf{R} , there is an invertible matrix Q over \mathbf{R} such that $AQ = A'$ is a matrix in reduced column-echelon form. The matrix A' is uniquely determined by A .*

Proof. Let $A = [a_{ij}]$ be an $m \times n$ matrix over \mathbf{R} and consider the set $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of n vectors in \mathbf{R}^m where $\mathbf{v}_j = (a_{1j}, a_{2j}, \dots, a_{mj})$. (That is, A is the matrix of transition from \mathcal{E}_m to \mathcal{A} .) By Theorem 2.9 (with m and n interchanged), there is a sequence E_1, E_2, \dots, E_t of elementary operations that can be applied to \mathcal{A} to obtain a set $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r, \mathbf{0}, \dots, \mathbf{0}\}$ in which

1. The first nonzero component from the left in \mathbf{v}_j is a 1 in the k_j component for $j = 1, 2, \dots, r$.
2. $k_1 < k_2 < \dots < k_r \leq m$.
3. \mathbf{v}_j is the only vector in \mathcal{A}' with a nonzero k_j component.
4. $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ is a basis of $\langle \mathcal{A} \rangle$.

Now each of the elementary operations E_i has an associated elementary matrix Q_i , and when E_i is applied to a set, a matrix of transition from that set to the new set is Q_i . As the diagram in Figure 3.4 shows, this means that a matrix of transition from \mathcal{A} to \mathcal{A}' is $Q_1 Q_2 \cdots Q_t$, and that $AQ_1 Q_2 \cdots Q_t = A'$ is the matrix of transition from \mathcal{E}_m to \mathcal{A}' . Thus the i^{th} element in column j of $A' = [a'_{ij}]$ is equal to the i^{th} component of \mathbf{v}'_j and therefore:

1. The first nonzero element in column j of A' is a 1 in row k_j , for $j = 1, 2, \dots, r$.
2. $k_1 < k_2 < \dots < k_r \leq m$.

3. Column j is the only column with a nonzero element in row k_j .
4. The last $n - r$ columns consist entirely of zeros.

The matrix $Q = Q_1 Q_2 \cdots Q_t$ is invertible since each Q_i is invertible, and the matrix A' is uniquely determined by A since \mathcal{A}' is uniquely determined by \mathcal{A} . ■

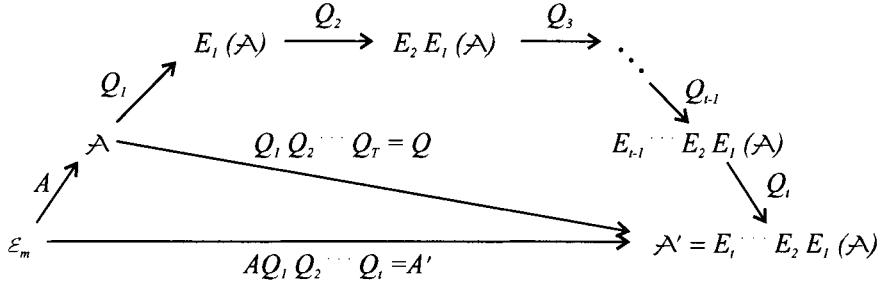


Figure 3.4

We saw in Chapter 2 that the set \mathcal{A}' used in the proof of Theorem 3.23 is uniquely determined by the set \mathcal{A} . It follows from this fact that each matrix A over \mathbf{R} has an associated unique reduced column-echelon form A' as described in the proof. However, we have seen in Chapter 2 that the sequence of elementary operations used to obtain \mathcal{A}' from \mathcal{A} is *not unique*. The invertible matrix Q is similarly *not unique* in spite of the fact that $AQ = A'$ is unique for A . An example that demonstrates this lack of uniqueness is given at the end of this section.

Definition 3.24 *The matrix A' in Theorem 3.23 is called the **reduced column-echelon form** for A .*

We shall now show how the proof of Theorem 3.23 can be interpreted so as to give a systematic procedure for finding an invertible Q such that AQ is in reduced column-echelon form. This procedure is closely related to that used in Chapter 2 in finding the standard basis of a subspace.

Suppose that $A = [a_{ij}]$ is a given $m \times n$ matrix over \mathbf{R} . Interpreting A as a matrix of transition from E_m to \mathcal{A} is equivalent to obtaining A by recording the components of vectors in \mathcal{A} , as was done in Chapter 2. We have seen that performing an elementary operation on \mathcal{A} corresponds to performing an elementary column operation on A . With these interpretations, the procedure in Example 3 of Section 2.4 can be regarded as a method for obtaining the reduced column-echelon form

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

for the matrix

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

This is an efficient method for finding the reduced column-echelon form, but it does not yield the matrix Q of Theorem 3.23.

In order to obtain the matrix Q , one needs a method for recording the products $Q_1 Q_2 \cdots Q_i$ for $i = 1, 2, \dots, t$. Consider the sequence of products

$$\begin{aligned} AI_n &= A \\ AI_n Q_1 &= AQ_1 \\ AI_n Q_1 Q_2 &= AQ_1 Q_2 \\ &\vdots \\ AI_n Q_1 Q_2 \cdots Q_t &= AQ_1 Q_2 \cdots Q_t = A'. \end{aligned}$$

Let E_i be the elementary column operation that has Q_i as its associated matrix. Now $AQ_1 = E_1(A)$, and $AQ_1 Q_2 \cdots Q_i = E_i E_{i-1} \cdots E_1(A)$ in general. Thus the right members of the equations above may be found by applying the sequence E_1, E_2, \dots, E_t to A . The left members have as factors the products $I_n Q_1 Q_2 \cdots Q_i = E_i E_{i-1} \cdots E_1(I_n)$. Thus Q can be found by applying the same sequence of elementary operations to I_n . What is desired, then, is an efficient method of recording the results $E_i \cdots E_2 E_1(A)$ and $E_i \cdots E_2 E_1(I_n)$. This can be done effectively by recording both A and I_n in a single matrix as

$$\begin{bmatrix} A \\ I_n \end{bmatrix}.$$

Instead of applying the elementary column operations separately to A and I_n ,

$$\begin{bmatrix} A \\ I_n \end{bmatrix}, \quad \begin{bmatrix} E_1(A) \\ E_1(I_n) \end{bmatrix}, \quad \begin{bmatrix} E_2 E_1(A) \\ E_2 E_1(I_n) \end{bmatrix}, \quad \text{etc.,}$$

one can simply apply the operation to the entire matrices

$$\begin{bmatrix} A \\ I_n \end{bmatrix}, \quad E_1 \left(\begin{bmatrix} A \\ I_n \end{bmatrix} \right), \quad E_2 E_1 \left(\begin{bmatrix} A \\ I_n \end{bmatrix} \right), \quad \text{etc.}$$

This procedure is quite valid since the same operations in the same order are to be applied to each of A and I_n .

Example 2 □ Using the matrix A in Example 3 of Section 2.4, we have

$$\left[\begin{array}{c|c} A & I_n \\ \hline \end{array} \right] = \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 3 & 2 & -1 \\ 0 & -6 & -4 & 2 \\ -1 & 1 & 2 & 2 \\ -1 & -2 & 0 & 3 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 2 & 3 & 0 & -1 \\ -4 & -6 & 0 & 2 \\ 2 & 1 & -1 & 2 \\ 0 & -2 & -1 & 3 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \rightarrow$$

$$\rightarrow \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & -1 \\ -2 & -6 & 0 & 2 \\ 1 & 1 & -1 & 2 \\ 0 & -2 & -1 & 3 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 1 & -2 & -1 & 3 \\ 0 & -2 & -1 & 3 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & -\frac{3}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right].$$

Thus

$$A' = \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{array} \right]$$

and

$$Q = \begin{bmatrix} 0 & 0 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{3}{2} \\ -\frac{1}{4} & \frac{3}{4} & \frac{3}{4} & -\frac{7}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is an invertible matrix such that $AQ = A'$. It is easy to show that Q is not unique by performing elementary operations that involve only the last two columns of Q . For instance, adding the third column of Q to its fourth column yields the matrix

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

such that $AB = AQ = A'$. ■

Exercises 3.5

1. Describe the elementary column operation that is associated with the given elementary matrix M .

$$(a) M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (b) M = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$(c) M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \quad (d) M = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2. Which of the following are in reduced column-echelon form?

$$(a) \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \quad (b) \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (c) \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$(d) \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (e) \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (f) \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3. Write the elementary matrix of transition from \mathcal{A} to \mathcal{B} .

(a) $\mathcal{A} = \{(1, 2, -1), (2, 1, 3)\}, \mathcal{B} = \{(1, 2, -1), (0, -3, 5)\}$

(b) $\mathcal{A} = \{(1, 0, 1), (0, 1, 1), (1, 1, 0)\}, \mathcal{B} = \{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}$

4. Given that $B = \begin{bmatrix} 1 & 4 & 2 \\ 2 & 0 & 1 \\ -2 & 1 & 3 \end{bmatrix}$ is obtained from the matrix A by multiplying A on the right by $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$, find A .

5. If M is an elementary matrix, are A and MA matrices of transition from a given basis to sets of vectors that span the same subspace? Illustrate with an example.

6. Find the reduced column-echelon form for each matrix.

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

$$(d) \begin{bmatrix} 0 & 0 & 0 & 0 \\ 2 & 3 & 0 & -1 \\ -4 & -6 & 0 & 2 \\ 2 & 1 & -1 & 2 \\ 0 & -2 & -1 & 3 \end{bmatrix} \quad (e) \begin{bmatrix} 1 & 2 & 1 & -1 \\ 2 & 4 & 2 & -2 \\ 0 & 1 & 2 & 3 \\ 1 & 4 & 5 & 5 \\ 0 & 3 & -2 & 4 \\ 1 & 6 & 1 & 6 \end{bmatrix} \quad (f) \begin{bmatrix} 1 & 2 & 1 & -1 \\ 2 & 4 & 2 & -2 \\ 0 & 1 & 2 & 3 \\ 1 & 4 & 5 & 5 \\ 0 & 3 & -2 & 4 \end{bmatrix}$$

$$(g) \begin{bmatrix} 1 & 1 & 3 & 2 & 0 \\ -1 & 1 & -1 & 0 & 0 \\ 2 & 2 & 6 & 4 & 0 \\ -3 & 0 & -6 & -3 & 1 \end{bmatrix} \quad (h) \begin{bmatrix} 1 & 1 & 3 & 1 & 1 \\ 1 & 0 & 2 & 0 & 1 \\ 3 & -2 & 4 & 0 & 7 \\ -1 & 0 & -2 & 1 & 1 \end{bmatrix}$$

7. In each part of Problem 6, let A be the given matrix and find an invertible matrix Q such that AQ is in reduced column-echelon form.

3.6 Row Operations and Row-Echelon Forms

In this section we shall develop elementary row operations that are analogous to the column operations of the preceding section. In later chapters, these row operations will be equally as important as column operations.

The connection that was made between columns and vectors led to a natural and simple correspondence between elementary operations on sets of vectors and elementary column operations. The interpretation of elementary row operations is not as simple, and a difference in the method of investigation is necessary. The fundamental technique is much the same for the different types of operations, with the most complicated situations occurring with operations of type II. For this reason, the derivation is carried out for operations of type II, and those for the other types are left as exercises.

Let us return to the comparisons of elementary matrices with I_n that were made earlier. If we direct our attention to the rows of the elementary matrices rather than the columns, the resulting descriptions are as follows.

If M is the elementary matrix of type I on page 83, then M is obtained from I_n by multiplying row s of I_n by a nonzero scalar a .

If M is the elementary matrix of type II on page 83, then M is obtained from I_n by adding to row t of I_n the product of b and row s of I_n , where $s \neq t$.

If M is the elementary matrix of type III on page 83, then M is obtained from I_n by interchanging row s and t ($s \neq t$).

Thus, each elementary matrix can be obtained from I_n by performing a suitable operation on the rows of I_n .

Definition 3.25 *There are three types of elementary row operations on a matrix A .*

- (I) *An elementary row operation of type I multiplies one of the rows of A by a nonzero scalar a .*
- (II) *An elementary row operation of type II adds to row t in A the product of b and row s in A , where $s \neq t$.*
- (III) *An elementary row operation of type III interchanges two rows in A .*

The descriptions of the products MA , where M is an elementary matrix, are very much like those obtained for the products AM in Section 3.5, even though the derivations are fundamentally different. As mentioned earlier, we consider a matrix M of type II here and leave the other derivations as exercises.

Suppose that M is an $m \times m$ elementary matrix of type II and $A = [a_{ij}]$ is $m \times n$ over \mathbf{R} . The matrix M is the matrix of transition from \mathcal{E}_m to a basis $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ of \mathbf{R}^m , where $\mathbf{u}_i = \mathbf{e}_i$ for $i \neq s$ and $\mathbf{u}_s = \mathbf{e}_s + b\mathbf{e}_t$ for $s \neq t$. Also, A is the matrix of transition from \mathcal{A} to a set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, and MA is the matrix of transition from \mathcal{E}_m to \mathcal{B} . This is shown in Figure 3.5.

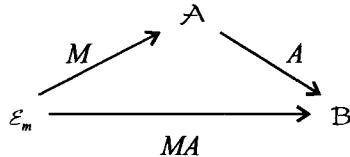


Figure 3.5

We have

$$\begin{aligned}
 \mathbf{v}_j &= \sum_{i=1}^m a_{ij} \mathbf{u}_i \\
 &= \sum_{\substack{i=1 \\ i \neq s}}^m a_{ij} \mathbf{u}_i + a_{sj} \mathbf{u}_s \\
 &= \sum_{\substack{i=1 \\ i \neq s}}^m a_{ij} \mathbf{e}_i + a_{sj} (\mathbf{e}_s + b\mathbf{e}_t) \\
 &= \sum_{\substack{i=1 \\ i \neq t}}^m a_{ij} \mathbf{e}_i + (a_{tj} + ba_{sj}) \mathbf{e}_t
 \end{aligned}$$

so that the coordinates of \mathbf{v}_j relative to \mathcal{E}_m are the same as the coordinates of \mathbf{v}_j relative to \mathcal{A} except for the t^{th} coordinate, and the t^{th} coordinate of \mathbf{v}_j relative to \mathcal{E}_m is obtained by adding to the t^{th} coordinate of \mathbf{v}_j relative to \mathcal{A} the product of b and the s^{th} coordinate relative to \mathcal{A} . Hence multiplying A on the left by M simply adds the product of b and the s^{th} row to the t^{th} row of A . That is, row t of A is replaced by the sum of row t of A and b times row s of A .

The descriptions of MA for elementary matrices M of types I and III are as follows.

- If M is obtained from I_m by multiplying row s of I_m by a , then MA is obtained from A by multiplying row s of A by a .
- If M is obtained from I_m by interchanging rows s and t of I_m ($s \neq t$), then MA is obtained from A by interchanging rows s and t of A .

The concept of the transpose of a matrix is extremely useful in obtaining the row analogue of the reduced column-echelon form. This analogous form is called the *reduced row-echelon form* of a matrix.

Definition 3.26 If $A = [a_{ij}]$ is any $m \times n$ matrix over \mathbf{R} , the **transpose** of A is the $n \times m$ matrix $B = [b_{ij}]$ with $b_{ij} = a_{ji}$ for $i = 1, 2, \dots, n; j = 1, 2, \dots, m$. The transpose of A is denoted by A^T . If A is a matrix such that $A^T = A$, then A is called **symmetric**. If $A^T = -A$, then A is **skew-symmetric**.

Thus row i of A^T is composed of the elements from column i of A , in order from left to right rather than from top to bottom. We say that A^T is obtained from A “by interchanging rows and columns.”

Example 1 □ Consider the matrices

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

It is easy to see that $B^T = B$ and

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

Thus B is a symmetric matrix and C is a skew-symmetric matrix. ■

Our next theorem states that the transpose of a product is equal to the product of the transposes *in reverse order*.

Theorem 3.27 *If $A = [a_{ij}]_{m \times n}$ and $B = [b_{ij}]_{n \times r}$ over \mathbf{R} , then $(AB)^T = B^T A^T$.*

Proof. Since $AB = [c_{ij}]_{m \times r}$ with $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$, then we have $(AB)^T = [\sum_{k=1}^n a_{jk}b_{ki}]_{r \times m}$. Also, $B^T = [d_{ij}]_{r \times n}$ with $d_{ij} = b_{ji}$ and $A^T = [f_{ij}]_{n \times m}$ with $f_{ij} = a_{ji}$. Thus $B^T A^T = [\sum_{k=1}^n d_{ik}f_{kj}]_{r \times m} = [\sum_{k=1}^n b_{ki}a_{jk}]_{r \times m} = (AB)^T$. ■

Definition 3.28 *An $m \times n$ matrix $A = [a_{ij}]$ over \mathbf{R} is in reduced row-echelon form if and only if the following conditions are satisfied:*

1. *The first nonzero element in row i is a 1 in column k_i for $i = 1, 2, \dots, r$. (This 1 is called a leading one.)*
2. *$k_1 < k_2 < \dots < k_r \leq n$. (That is, for each change in rows from upper to lower, the leading one appears farther to the right.)*
3. *For $i = 1, 2, \dots, r$, the leading one in row i is the only nonzero element in column k_i .*
4. *Each of the last $m - r$ rows consists entirely of zeros.*

Example 2 □ Consider the question as to which of the following matrices are in reduced row-echelon form.

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

The matrix A is not in reduced row-echelon form since the row of zeros is placed above a nonzero element, violating condition (4). The matrix B is not in reduced row-echelon form because the leading 1 in row 3 is not the only nonzero element in column $k_3 = 4$. The matrix C fails on condition (2) because the leading 1 in the second row does *not* appear farther to the right than the leading 1 in the first row. That is, $k_1 = 4$ and $k_2 = 1$ do not satisfy $k_1 < k_2$. The matrix D satisfies all four conditions, and D is in reduced row-echelon form. ■

Conditions (1) and (3) in Definition 3.28 can be reworded in the following ways that parallel the rewording we used with the definition of reduced column-echelon form:

1. $a_{ij} = 0$ for $j < k_i$, and $a_{ik_i} = 1$ for $i = 1, 2, \dots, r$.
3. Row i is the only row with a nonzero element in column k_i for $i = 1, 2, \dots, r$.

These alternative wordings are more useful in proving our next theorem.

Theorem 3.29 *A matrix A is in reduced column-echelon form if and only if A^T is in reduced row-echelon form.*

Proof. Let $A = [a_{ij}]_{m \times n}$ and let $B = [b_{ij}]_{n \times m} = A^T$. If we let j denote the row numbers and i denote the column numbers of elements of A in Definition 3.22, then A is in reduced column-echelon form if and only if these conditions hold:

1. $a_{ji} = 0$ for $j < k_i$, and $a_{k_i i} = 1$ for $i = 1, 2, \dots, r$.
2. $k_1 < k_2 < \dots < k_r \leq m$.
3. Column i is the only column of A with a nonzero element in row k_i .
4. Each of the last $n - r$ columns of A consists entirely of zeros.

The elements in column i of A are the elements in rows i of A^T , so the conditions on A are satisfied if and only if:

1. $b_{ij} = a_{ji} = 0$ for $j < k_i$, and $b_{ik_i} = 1$ for $i = 1, 2, \dots, r$.
2. $k_1 < k_2 < \dots < k_r \leq m$.
3. Row i is the only row of A^T with a nonzero element in column k_i .
4. Each of the last $n - r$ rows of A^T consists entirely of zeros.

Since A^T is an $n \times m$ matrix, these conditions are equivalent to the assertion that A^T is in reduced row-echelon form. ■

From this theorem and the fact that $(A^T)^T = A$, it follows that A^T is in reduced column-echelon form if and only if A is in reduced row-echelon form.

Theorem 3.30 *If A is invertible, then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.*

Proof. If A is invertible, then $A^{-1}A = I_n$ and therefore $(A^T) \cdot (A^{-1})^T = I_n^T = I_n$ by Theorem 3.27. According to Theorem 3.14, A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$. ■

Theorem 3.31 *If A is an $m \times n$ matrix over \mathbf{R} , there exists an invertible matrix P over \mathbf{R} such that PA is in reduced row-echelon form.*

Proof. By Theorem 3.23, there is an invertible matrix Q such that A^TQ is in reduced column-echelon form, and this means that $(A^TQ)^T = Q^T(A^T)^T = Q^TA$ is in reduced row-echelon form. But $P = Q^T$ is an invertible matrix, so the theorem is proved. ■

In the proof of Theorem 3.31, the reduced column-echelon form for A^T is uniquely determined by A^T , and therefore the reduced row-echelon form PA is uniquely determined by A . We make the following definition.

Definition 3.32 *The unique matrix PA in the statement of Theorem 3.31 is called the reduced row-echelon form for A .*

Theorem 3.33 *A square matrix $A = [a_{ij}]_{n \times n}$ is invertible if and only if the reduced row-echelon form for A is I_n .*

Proof. Assume that A is invertible, and let P be an invertible matrix such that PA is in reduced row-echelon form. There must be no rows of zeros in PA , for otherwise, A^T would obviously be singular. This means that $k_i = i$ of each i , and $PA = I_n$.

If the reduced row-echelon form for A is I_n , there is an invertible matrix P such that $PA = I_n$. Then A is invertible by Theorem 3.15. ■

We have seen that a matrix is invertible if and only if it is a product of elementary matrices. Thus the last theorem implies that there are elementary matrices P_1, P_2, \dots, P_s such that $P_s P_{s-1} \cdots P_2 P_1 A$ is in reduced row-echelon form. But multiplication of a given matrix on the left by P_i performs an elementary row operation on the given matrix. Thus a sequence of elementary row operations may be applied to A in order to obtain a reduced row-echelon matrix. In much the same manner as with column operations, this gives rise to a systematic method for finding the invertible matrix P .

With the notation of the preceding paragraph, consider the sequence of equations

$$\begin{aligned} I_m A &= A \\ P_1 I_m A &= P_1 A \\ &\vdots \\ P_s \cdots P_2 P_1 I_m A &= P_s \cdots P_2 P_1 A = PA. \end{aligned}$$

These equations indicate that if the reduced row-echelon form is obtained by application of a certain sequence of elementary row operations to A , the matrix P may be obtained by application of the same sequence of operations, in the same order, to I_m . This can be done efficiently by recording A and I_m in a single matrix as $[A, I_m]$ and performing the row operations simultaneously on A and I_m .

Example 3 \square To illustrate the procedure, let

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

and consider the problem of finding an invertible matrix P such that PA is in reduced row-echelon form.

$$\begin{aligned} [A, I_3] &= \left[\begin{array}{ccc|ccc} 2 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & -3 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -3 & 0 & 1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 1 \end{array} \right] \\ &\rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -3 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 & -1 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & -\frac{3}{2} & -\frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{array} \right] \end{aligned}$$

Thus, I_3 is the reduced row-echelon form for A , and

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

is an invertible matrix such that $PA = I_3$. ■

In Example 3, we found P such that $PA = I_3$. According to Theorem 3.33, A is invertible when its reduced row-echelon form is an identity matrix. However, the equation $PA = I_n$ implies more: It implies that $P = A^{-1}$, by Theorem 3.15. Thus the procedure illustrated in Example 3 is an efficient and systematic way of finding A^{-1} when it exists.

Suppose now that P is invertible and PA is in reduced row-echelon form but different from I_n . Then PA must contain at least one row of zeros, and therefore PA is not invertible. This implies that A is not invertible, by Theorem 3.16.

Summarizing our discussion of the procedure in Example 3, we can say that one of the following possibilities must happen:

1. If the procedure leads to $PA = I_n$, then A is invertible and $P = A^{-1}$.
2. If the procedure leads to a row of zeros in PA , then A is not invertible.

Exercises 3.6

- 1.** Describe the elementary row operation that is associated with the given elementary matrix M .

$$(a) M = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(b) M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}$$

$$(c) M = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$(d) M = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- 2.** Which of the following are in reduced row-echelon form?

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

$$(b) \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(c) \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$(d) \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$(e) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

$$(f) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- 3.** Find the reduced row-echelon form for each of the matrices given in Problem 6 of Exercises 3.5.

- 4.** In each part of Problem 6 in Exercises 3.5, let A be the given matrix and find an invertible matrix P such that PA is in reduced row-echelon form.

- 5.** Find the inverse of each of the following invertible matrices.

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

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

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

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

- 6.** Given $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$, write A^T as a product of elementary matrices.

7. If $A = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$ is the matrix of transition from \mathcal{E}_3 to \mathcal{A} , (a) find \mathcal{A} and (b) find the transition matrix from \mathcal{A} to \mathcal{E}_3 .

8. Prove that if M is obtained from I_m by multiplying row s of I_m by a , then MA is obtained from A by multiplying row s by a .
9. Prove that if M is obtained from I_m by interchanging row s and t of I_m , then MA is obtained from A by interchanging row s and t .

10. Given that the matrix

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

is obtained from A by the following elementary operations:

- i. First, the second and third rows of A are interchanged.
ii. Next, the first column is multiplied by 2 and added to the second column.

Find the matrix A .

11. Prove that $(A^T)^T = A$.
12. Which types of elementary matrices are symmetric?
13. Prove that any two diagonal elements of a skew-symmetric matrix over \mathbf{R} are equal.
14. Prove or disprove.
- (a) The set of all symmetric $n \times n$ matrices over \mathbf{R} is closed under multiplication.
(b) The set of all skew-symmetric $n \times n$ matrices over \mathbf{R} is closed under multiplication.

15. Prove that, for any matrix A over \mathbf{R} , AA^T is defined and is a symmetric matrix.
16. Extend Theorem 3.27 to any product with a finite number of factors:

$$(A_1 A_2 \cdots A_k)^T = A_k^T \cdots A_2^T A_1^T.$$

17. Prove that if A is symmetric, then A^k is symmetric for any positive integer k .
18. Let $A = [a_{ij}]_{n \times n}$ with $a_{ik} = a_{ij} + a_{jk}$ for all i, j, k .
- (a) Prove that A is skew-symmetric.
(b) Show that A is determined by its first row.

3.7 Row and Column Equivalence

Sometimes elements of sets are associated with each other in ways that have some of the basic properties of equality. These associations between elements are called *relations*, and certain types of relations are called *equivalence relations*. A knowledge of the various equivalence relations on a set adds depth to our understanding of the structure of the set.

The formal definition of a relation can be made as follows: A **relation** (more precisely, a **binary relation**) on a set T is a nonempty set “ \sim ”, of ordered pairs (a, b) of elements a and b of T . If (a, b) is in “ \sim ”, then we say that a has the relation \sim to b . We write $a \sim b$ to indicate that a has the relation \sim to b .

If the relation under consideration is ordinary equality, then the set \sim consists of all ordered pairs (a, a) , and we write $a = b$ for $a \sim b$.

Definition 3.34 A relation \sim on a set T is an **equivalence relation** if and only if the following conditions are satisfied for arbitrary a, b, c in T :

1. $a \sim a$ is true for all $a \in T$.
2. If $a \sim b$ is true, then $b \sim a$ is true.
3. If $a \sim b$ is true and $b \sim c$ is true, then $a \sim c$ is true.

The properties (1), (2), and (3) are known as the **reflexive**, **symmetric**, and **transitive** properties, respectively.

We are interested here in equivalence relations on matrices. There are two such relations that are intimately connected with the two preceding sections.

Definition 3.35 Let A and B be matrices over \mathbf{R} . The matrix B is **column-equivalent** to A if and only if there exists an invertible matrix Q such that $B = AQ$.

The term “column-equivalent” defines a relation on the set of all matrices over \mathbf{R} . The relation consists of the set of all ordered pairs (B, A) such that $B = AQ$ for some invertible matrix Q . It is clear that if A is $m \times n$, then any matrix that is column-equivalent to A is $m \times n$.

Theorem 3.36 The relation defined as column-equivalence is an equivalence relation on the set of all matrices over \mathbf{R} . That is:

1. For any matrix A , A is column-equivalent to A .
2. If A is column-equivalent to B , then B is column-equivalent to A .
3. If A is column-equivalent to B and B is column-equivalent to C , then A is column-equivalent to C .

Proof. The statement (1) follows from $A = AI_n$.

If $A = BQ$, where Q is invertible, then $B = AQ^{-1}$ and Q^{-1} is invertible. Thus (2) is valid.

Assume $A = BQ$ and $B = CP$, where Q and P are invertible. Then $A = (CP)Q = C(PQ)$, and PQ is invertible. Thus (3) is valid. ■

The relation of column-equivalence can be described in several alternate ways. Four of these are stated in the next theorem, and examples illustrating these alternate descriptions follow Theorem 3.38.

Theorem 3.37 *Let A and B be $m \times n$ matrices over \mathbf{R} . For any given basis of \mathbf{R}^m , A will be the matrix of transition from the given basis to a set \mathcal{A} of n vectors, and B will be the matrix of transition from the given basis to a set \mathcal{B} of n vectors. Each of the following statements implies the other three:*

1. *B is column-equivalent to A .*
2. *B may be obtained from A by a sequence of elementary column operations.*
3. *$\langle \mathcal{B} \rangle = \langle \mathcal{A} \rangle$.*
4. *\mathcal{B} may be obtained from \mathcal{A} by a sequence of elementary operations.*

Proof. Statement (3) is true if and only if (4) is true, according to Theorem 2.13. In view of Theorem 3.19, (1) means the same as the assertion that $B = AQ_1Q_2 \cdots Q_t$, where each Q_i is elementary. It thus follows from the discussion following Definition 3.21 that (1) and (2) are equivalent. Thus the theorem will be proved if we show that each of (1) and (4) implies the other.

Suppose first that B is column-equivalent to A . Then $B = AQ_1Q_2 \cdots Q_t$, where each Q_i is elementary. This means that $AQ_1Q_2 \cdots Q_t$ is the matrix of transition from the given basis to \mathcal{B} . And since A is the matrix of transition from the given basis to \mathcal{A} , $Q_1Q_2 \cdots Q_t$ is a matrix of transition from \mathcal{A} to \mathcal{B} . Each Q_i has an associated elementary operation E_i , and we have $E_t \cdots E_2E_1(\mathcal{A}) = \mathcal{B}$. Thus (1) implies (4).

Assume now that \mathcal{B} may be obtained from \mathcal{A} by a sequence E_1, E_2, \dots, E_t of elementary operations E_i . If Q_i is the elementary matrix associated with E_i , then $Q_1Q_2 \cdots Q_t$ is the matrix of transition from \mathcal{A} to $E_t \cdots E_2E_1(\mathcal{A}) = \mathcal{B}$. Since A is the matrix of transition from the given basis to \mathcal{A} , $AQ_1Q_2 \cdots Q_t$ is the matrix of transition from the given basis to \mathcal{B} . But the matrix of transition from the given basis to \mathcal{B} is B , and this matrix is unique. Hence $B = AQ_1Q_2 \cdots Q_t$, and B is column-equivalent to A . ■

Theorem 3.38 *Any matrix over \mathbf{R} is column-equivalent to a unique matrix in reduced column-echelon form.*

Proof. This is a restatement of Theorem 3.23. ■

We consider now some examples that illustrate the quantities involved in Theorem 3.37 and its proof.

Example 1 □ Let the matrices A and B be given by

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 1 & 0 \\ 8 & 4 & 2 \\ 6 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & 2 \\ 2 & -2 & 8 \\ 3 & -9 & 0 \end{bmatrix}.$$

Then A is the transition matrix from \mathcal{E}_4 to

$$\mathcal{A} = \{(2, 1, 8, 6), (0, 1, 4, 0), (1, 0, 2, 3)\}$$

and B is the transition matrix from \mathcal{E}_4 to

$$\mathcal{B} = \{(1, 0, 2, 3), (-3, 1, -2, -9), (0, 2, 8, 0)\}.$$

Theorem 2.13 assures us that \mathcal{B} and \mathcal{A} span the same subspace of \mathbf{R}^4 if and only if \mathcal{B} can be obtained from \mathcal{A} by a sequence of elementary operations. By the end of the next example, we can see that $\langle \mathcal{B} \rangle = \langle \mathcal{A} \rangle$. ■

Example 2 □ For the matrices A and B in Example 1, we shall show that B is column-equivalent to A and find an invertible matrix Q such that $B = AQ$. To show that B is column-equivalent to A we need only confirm that their reduced column-echelon forms are equal. But in order to find an invertible Q such that $B = AQ$, we need first to find invertible matrices M_1 and M_2 such that $AM_1 = A'$ and $BM_2 = B'$ are both in reduced column-echelon form. Using the same procedure as in Example 2 of Section 3.5, we obtain the following results.

$$\left[\begin{array}{c|ccc} A & 2 & 0 & 1 \\ \hline I_3 & 1 & 1 & 0 \\ & 8 & 4 & 2 \\ & 6 & 0 & 3 \end{array} \right] \xrightarrow{\quad} \left[\begin{array}{c|ccc} & 1 & 0 & 2 \\ \hline & 0 & 1 & 1 \\ & 2 & 4 & 8 \\ & 3 & 0 & 6 \end{array} \right] \xrightarrow{\quad} \left[\begin{array}{c|ccc} & 1 & 0 & 0 \\ \hline & 0 & 1 & 1 \\ & 2 & 4 & 4 \\ & 3 & 0 & 0 \end{array} \right] \xrightarrow{\quad} \left[\begin{array}{c|ccc} & 1 & 0 & 0 \\ \hline & 0 & 1 & 0 \\ & 2 & 4 & 0 \\ & 3 & 0 & 0 \end{array} \right]$$

Thus $M_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & -2 \end{bmatrix}$ is an invertible matrix such that $AM_1 = A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 4 & 0 \\ 3 & 0 & 0 \end{bmatrix}$.

Working in similar fashion with B , we have

$$\left[\begin{array}{c} B \\ I_3 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc|ccc} 1 & -3 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 0 \\ 2 & -2 & 8 & 2 & 4 & 8 & 2 & 4 & 0 \\ \hline 3 & -9 & 0 & 3 & 0 & 0 & 3 & 0 & 0 \\ 1 & 0 & 0 & 1 & 3 & 0 & 1 & 3 & -6 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right].$$

Thus $M_2 = \begin{bmatrix} 1 & 3 & -6 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$ is an invertible matrix such that

$$BM_2 = B' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 4 & 0 \\ 3 & 0 & 0 \end{bmatrix}.$$

As mentioned earlier, the fact that $A' = B'$ shows that B is column-equivalent to A . The equation $BM_2 = AM_1$ implies that $B = AM_1M_2^{-1}$, and thus $Q = M_1M_2^{-1}$ is an invertible matrix such that $B = AQ$. When these computations are performed, we find

$$\text{that } M_2^{-1} = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \text{ and}$$

$$Q = M_1M_2^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -3 & -2 \end{bmatrix}$$

is an invertible matrix such that $B = AQ$. According to Theorem 3.37, the sets \mathcal{B} and \mathcal{A} in Example 1 do indeed span the same subspace. ■

Example 3 □ Suppose now that we wish to find elementary matrices Q_1, Q_2, \dots, Q_t such that $B = AQ_1Q_2 \cdots Q_t$ as described in the proof of Theorem 3.37. These elementary matrices can be obtained from the work in Example 1 because both A' and B'

were found by performing a single elementary column operation in each step. It is easy to see that the work done in Example 1 with A can be represented by

$$\begin{bmatrix} A \\ I_3 \end{bmatrix} \rightarrow \begin{bmatrix} AE_1 \\ I_3 E_1 \end{bmatrix} \rightarrow \begin{bmatrix} AE_1 E_2 \\ I_3 E_1 E_2 \end{bmatrix} \rightarrow \begin{bmatrix} AE_1 E_2 E_3 \\ I_3 E_1 E_2 E_3 \end{bmatrix}$$

where E_1, E_2 , and E_3 are elementary matrices given by

$$E_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

and $M_1 = E_1 E_2 E_3$. Similarly, the work with B can be represented by

$$\begin{bmatrix} B \\ I_3 \end{bmatrix} \rightarrow \begin{bmatrix} BF_1 \\ I_3 F_1 \end{bmatrix} \rightarrow \begin{bmatrix} BF_1 F_2 \\ I_3 F_1 F_2 \end{bmatrix}$$

where F_1 and F_2 are elementary matrices given by

$$F_1 = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad F_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

and $M_2 = F_1 F_2$. To find the desired Q_1, Q_2, \dots, Q_t , we write

$$\begin{aligned} B &= AM_1 M_2^{-1} \\ &= AE_1 E_2 E_3 (F_1 F_2)^{-1} \\ &= AE_1 E_2 E_3 F_2^{-1} F_1^{-1}. \end{aligned}$$

Thus $Q_1 = E_1, Q_2 = E_2, Q_3 = E_3, Q_4 = F_2^{-1}, Q_5 = F_1^{-1}$ are elementary matrices such that $B = A Q_1 Q_2 Q_3 Q_4 Q_5$. Writing out all the elementary matrices involved, we have

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

As might be expected, there is an equivalence relation connected with row operations that parallels the one connected with column operations.

Definition 3.39 Let A and B be matrices over \mathbf{R} . The matrix B is **row-equivalent** to A if and only if there exists an invertible matrix P such that $B = PA$.

Similar to the situation with Definition 3.35 the term “row-equivalent” defines a relation on the set of all matrices over \mathbf{R} . It is left as an exercise to prove that this relation is an equivalence relation.

We have seen in Section 3.6 that multiplication of a given matrix on the left by a product of elementary matrices yields the same result as the application of a sequence of elementary row operations to the matrix. In combination with Theorem 3.19 this shows that the following conditions are equivalent:

1. B is row-equivalent to A ;
2. B may be obtained from A by applying a sequence of elementary row operations.

Theorem 3.40 *Any matrix over \mathbf{R} is row-equivalent to a unique matrix in reduced row-echelon form.*

Proof. This theorem follows from Theorem 3.31 and the discussion of uniqueness just before Definition 3.32. ■

Exercises 3.7

1. Determine whether or not the given matrices are column-equivalent.

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

$$(b) A = \begin{bmatrix} 3 & 2 \\ 6 & 4 \\ 13 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 1 \\ 10 & 2 \\ 9 & 2 \end{bmatrix}$$

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

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

2. Which of the pairs of matrices in Problem 1 are row-equivalent?

3. In the following list of matrices, A and B are row-equivalent. Also, P_1 and P_2 are invertible and such that P_1A and P_2B are both in reduced row-echelon form. Find an invertible P such that $B = PA$.

$$A = \begin{bmatrix} 3 & 2 \\ 6 & 4 \\ 13 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 1 \\ 10 & 2 \\ 9 & 2 \end{bmatrix},$$

$$P_1 = \begin{bmatrix} 1 & 4 & -2 \\ -1 & -6 & 3 \\ -2 & 1 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & 1 & -1 \\ 1 & -5 & 5 \\ -2 & 1 & 0 \end{bmatrix}$$

4. In the following list of matrices, A and B are row-equivalent and P_1 is an invertible matrix such that P_1A is in reduced row-echelon form. Find an invertible matrix P such that $PA = B$.

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 7 & 12 \\ 3 & 4 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 1 & 2 & 3 \end{bmatrix}, \quad P_1 = \begin{bmatrix} 7 & -3 & 0 \\ -2 & 1 & 0 \\ -13 & 5 & 1 \end{bmatrix}$$

5. For each pair of matrices in Problem 1 that are column-equivalent, find an invertible matrix Q such that $B = AQ$.
6. For each pair of matrices in Problem 1 that are row-equivalent, find an invertible matrix P such that $B = PA$.
7. For each pair of matrices in Problem 1 that are column-equivalent, find elementary matrices Q_1, Q_2, \dots, Q_t such that $B = AQ_1Q_2 \cdots Q_t$.
8. Prove that the relation defined by “row-equivalent” is an equivalence relation on the set of all matrices over \mathbf{R} .
9. Justify your answer for each of the following questions.
- (a) Which square matrices over \mathbf{R} are column-equivalent to I_n ?
- (b) Which square matrices over \mathbf{R} are row-equivalent to I_n ?
10. Prove that B is column-equivalent to A if and only if B^T is row-equivalent to A^T .
11. Prove that A and B are column-equivalent if and only if they have the same reduced column-echelon form.
12. Assuming that A and B are row-equivalent, describe a procedure for finding elementary matrices P_1, \dots, P_t such that $B = P_1 \cdots P_t A$.

3.8 Rank and Equivalence

There is one more equivalence relation on matrices that we wish to consider at this time. This relation is of fundamental importance in the study of linear transformations, and it has traditionally been referred to by the term “equivalent,” although other terms have been used.

Definition 3.41 *Let A and B be matrices over \mathbf{R} . The matrix B is equivalent to A if and only if there are invertible matrices P and Q such that $B = PAQ$.*

Thus, B is equivalent to A if B is row-equivalent to A or if B is column-equivalent to A . We shall see in Problem 3 of the exercises for this section that B may be equivalent to A , and yet be neither row-equivalent nor column-equivalent to A .

It is readily verified that the relation in Definition 3.41 is a true equivalence relation. From this point on, “equivalence of matrices” will refer to this relation. Equivalence of matrices is intimately connected with the *rank* of a matrix.

Definition 3.42 *If A is any $m \times n$ matrix over \mathbf{R} , the rank of A is the number r of nonzero columns in the reduced column-echelon form of A . The rank of A will be denoted by $\text{rank}(A)$.*

Definition 3.43 *Let $A = [a_{ij}]_{m \times n}$ over \mathbf{R} , and let $\mathbf{v}_j = (a_{1j}, a_{2j}, \dots, a_{mj})$ for $j = 1, 2, \dots, n$. The subspace of \mathbf{R}^m spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is called the column space of A , and the vectors \mathbf{v}_j are called the column vectors of A .*

It is easy to see that A is the matrix of transition from \mathcal{E}_m to a spanning set of the column space of A .

Theorem 3.44 *The rank of A is the dimension of the column space of A .*

Proof. Let $A = [a_{ij}]_{m \times n}$ have rank r , and let $A' = [a'_{ij}]_{m \times n}$ be the reduced column-echelon form of A . Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, where $\mathbf{v}_j = (a_{1j}, a_{2j}, \dots, a_{mj})$, and let $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n\}$, where $\mathbf{v}'_j = (a'_1j, a'_2j, \dots, a'_mj)$. Then $\langle \mathcal{A} \rangle$ and $\langle \mathcal{A}' \rangle$ are the column spaces of A and A' , respectively. It is clear from conditions (1) and (2) of Definition 3.22 that $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ is linearly independent and therefore is a basis of $\langle \mathcal{A}' \rangle$. But $\langle \mathcal{A} \rangle = \langle \mathcal{A}' \rangle$ by Theorem 3.37 since A and A' are column-equivalent. Thus $\langle \mathcal{A} \rangle$ has dimension r . ■

Corollary 3.45 *If A is $m \times n$ over \mathbf{R} and Q is any invertible $n \times n$ matrix, then $\text{rank}(A) = \text{rank}(AQ)$.*

Proof. Since A and AQ are column-equivalent, their column spaces are equal. Hence their ranks are equal, by the theorem. ■

Theorem 3.46 *Let A be an $m \times n$ matrix over \mathbf{R} , and let r be the rank of A . There exist invertible matrices P and Q such that PAQ has the first r diagonal elements equal to 1, and all other elements zero.*

Proof. Let Q be an invertible matrix such that $AQ = A' = [a_{ij}]_{m \times n}$ is in reduced column-echelon form. As in the proof of the preceding theorem, the set $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ is a basis of the column space $\langle \mathcal{A} \rangle = \langle \mathcal{A}' \rangle$ of A .

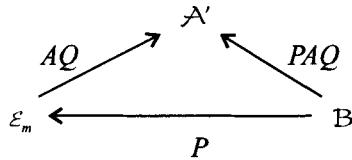


Figure 3.6

According to Theorem 1.32, the linearly independent set $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ can be extended to a basis $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}$ of \mathbf{R}^m , where $\mathbf{w}_i = \mathbf{v}'_i$ for $i = 1, 2, \dots, r$. Let P be the invertible $m \times m$ matrix of transition from \mathcal{B} to \mathcal{E}_m (see Figure 3.6). Then PAQ is the matrix of transition from \mathcal{B} to \mathcal{A}' . Since the first r vectors of \mathcal{A}' are the same as the first r vectors of \mathcal{B} , the first r columns of PAQ are the same as the first r columns of I_m . And since the last $n - r$ vectors of \mathcal{A}' are zero, the last $n - r$ columns of PAQ are zero. Thus

$$PAQ = \left[\begin{array}{ccc|ccc} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{array} \right] = \left[\begin{array}{c|c} I_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right],$$

and the theorem is proved. ■

Let

$$D_r = \left[\begin{array}{c|c} I_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right]$$

denote the matrix PAQ in the proof of Theorem 3.46 above. The proof of the theorem can be followed literally step by step so as to obtain invertible matrices P and Q such that $PAQ = D_r$. The only difficulty is in finding the matrix of transition P from \mathcal{B} to \mathcal{E}_m . This is probably most easily done by writing out the matrix P^{-1} (which is the matrix of transition from \mathcal{E}_m to \mathcal{B}), and then finding the inverse of P^{-1} . However, we have seen earlier that multiplication on the right by Q is equivalent to applying a sequence of column operations, and multiplication on the left by P is equivalent to

applying a sequence of row operations. Thus one may proceed to use column operations on $\begin{bmatrix} A \\ I_n \end{bmatrix}$ to obtain $\begin{bmatrix} A' \\ Q \end{bmatrix}$ as we did in Section 3.5, and then use row operations on $[A', I_m]$ to obtain $[D_r, P]$, where $D_r = PA' = PAQ$. This is illustrated in the following example.

Example 1 \square For the matrix

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

we shall find invertible matrices P and Q such that $PAQ = D_r$.

We first use column operations to transform $\begin{bmatrix} A \\ I_3 \end{bmatrix}$ into $\begin{bmatrix} A' \\ Q \end{bmatrix}$, where $A' = AQ$ is in reduced column-echelon form.

$$\begin{aligned} \begin{bmatrix} A \\ I_3 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & -1 \\ 3 & 6 & -3 \\ -1 & 1 & 0 \\ 2 & 4 & -2 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 3 & 0 & 0 \\ -1 & 3 & -1 \\ 2 & 0 & 0 \\ \hline 1 & -2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &\rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 3 & 0 & 0 \\ -1 & -1 & 3 \\ 2 & 0 & 0 \\ \hline 1 & 1 & -2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 3 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 0 \\ \hline 0 & -1 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & 3 \end{bmatrix} = \begin{bmatrix} A' \\ Q \end{bmatrix} \end{aligned}$$

Next we use row operations to transform $[A', I_4]$ into $[D_r, P]$, where $D_r = PA' = PAQ$.

$$\begin{aligned} [A', I_4] &= \left[\begin{array}{ccc|ccccc} 1 & 0 & 0 & | & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & | & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & | & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & | & 0 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccccc} 1 & 0 & 0 & | & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & | & -3 & 1 & 0 & 0 \\ 0 & 1 & 0 & | & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & | & -2 & 0 & 0 & 1 \end{array} \right] \\ &\rightarrow \left[\begin{array}{ccc|ccccc} 1 & 0 & 0 & | & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & | & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & | & -3 & 1 & 0 & 0 \\ 0 & 0 & 0 & | & -2 & 0 & 0 & 1 \end{array} \right] = [D_r, P] \end{aligned}$$

Thus

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 1 & 0 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 0 & -1 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & 3 \end{bmatrix}$$

are invertible matrices such that

$$PAQ = \left[\begin{array}{ccc|c} 1 & 0 & | & 0 \\ 0 & 1 & | & 0 \\ \hline 0 & 0 & + & 0 \\ 0 & 0 & | & 0 \end{array} \right] = \left[\begin{array}{cc|c} I_2 & | & \mathbf{0} \\ \hline 0 & | & 0 \end{array} \right] = D_2. \blacksquare$$

Theorem 3.47 Let A and B be $m \times n$ matrices over \mathbf{R} . Then B is equivalent to A if and only if B and A have the same rank.

Proof. Let r denote the rank of A , and let r' denote the rank of B .

Suppose that $r = r'$. Then there are invertible matrices P, Q, P' , and Q' such that $PAQ = D_r = D_{r'} = P'BQ'$. The matrices P and P' are $m \times m$, and the matrices Q and Q' are $n \times n$. Hence the equation $PAQ = P'BQ'$ implies that

$$(P')^{-1}PAQ(Q')^{-1} = B,$$

where $(P')^{-1}P$ and $Q(Q')^{-1}$ are invertible matrices. Therefore B is equivalent to A .

Assume now that B is equivalent to A . Let D_r be the diagonal matrix equivalent to A that has the form described in Theorem 3.46 and let $D_{r'}$ be the corresponding matrix that is equivalent to B . Since A and B are equivalent, D_r and $D_{r'}$ are equivalent. Therefore there are invertible matrices P and Q such that $PD_rQ = D_{r'}$ and

$$PD_r = D_{r'}Q^{-1}.$$

Now PD_r has $n - r$ columns of zeros so that $\text{rank}(PD_r) \leq r$. Thus

$$r' = \text{rank}(D_{r'}) = \text{rank}(D_{r'}Q^{-1}) = \text{rank}(PD_r) \leq r.$$

Similarly,

$$r = \text{rank}(D_r) = \text{rank}(D_rQ) = \text{rank}(P^{-1}D_{r'}) \leq r'$$

so that $r = r'$. ■

The following theorem and corollary are extremely useful in connection with the solution of systems of linear equations in Chapter 4.

Theorem 3.48 *Let $A = [a_{ij}]_{m \times n}$ over \mathbf{R} . Then A and A^T have the same rank.*

Proof. Let r denote the rank of A , and let P and Q be invertible matrices such that

$$PAQ = \begin{bmatrix} I_r & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \\ \mathbf{0} & | & \mathbf{0} \end{bmatrix} = D_r.$$

Now Q^T and P^T are invertible (Theorem 3.30), and $Q^T A^T P^T = D_r^T$. The dimension of the column space of D_r^T is clearly r , so the D_r^T has rank r . But A^T and D_r^T are equivalent, so A^T must have rank r also. ■

Corollary 3.49 *If A has rank r , then r is the number of nonzero rows in the reduced row-echelon form of A .*

Proof. Suppose that A has rank r , and let P be an invertible matrix such that PA is in reduced row-echelon form. By Theorem 3.29, $(PA)^T = A^T P^T$ is in reduced column-echelon form. Since A , A^T and $A^T P^T$ all have the same rank r , the number of nonzero columns in $(PA)^T = A^T P^T$ is r . But the number of nonzero columns in $(PA)^T$ is the same as the number of nonzero rows in PA . Hence the corollary is proved. ■

This result leads to the next corollary. The proof is requested in Problem 14 of the exercises.

Corollary 3.50 *An $n \times n$ matrix A is invertible if and only if A has rank n .*

Exercises 3.8

1. Find the rank of each of the following matrices.

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

$$(c) \begin{bmatrix} 1 & 0 & 2 \\ 4 & 1 & 3 \\ 3 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix} \quad (d) \begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

2. Find the standard basis of the column space of each matrix A in Problem 1.
3. Which of the pairs A, B of matrices in Problem 1 of Exercises 3.7 are equivalent? (Compare this with the results of Problems 1 and 2 of Exercises 3.7.)
4. Which of the matrices in Problem 1 above are equivalent?
5. Which of the following matrices are equivalent?

$$A = \begin{bmatrix} 1 & 3 & 2 \\ -2 & -6 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -6 & 7 \\ 2 & -4 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 2 \\ 6 & 4 \\ 13 & 9 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & -1 \\ 4 & 1 & -2 \\ 3 & 1 & -1 \\ 2 & 1 & 0 \end{bmatrix}$$

6. Answer the following questions for the matrices

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

and give a reason for each answer.

- (a) Are A and B column-equivalent?
 (b) Are A and B row-equivalent?
 (c) Are A and B equivalent?

7. For $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 0 & 1 & 2 \end{bmatrix}$, it is given that $Q = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$ is an invertible matrix such that $AQ = A' = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Let $\mathcal{A}' = \{(1, 2, 0), (0, 0, 1), (0, 0, 0)\}$. Find a basis \mathcal{B} of \mathbf{R}^3 such that the matrix of transition from \mathcal{B} to \mathcal{A}' is $D_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and an invertible P such that $PAQ = D_2$. (*Hint:* See the proof of Theorem 3.46.)

8. For each matrix A below, follow the proof of Theorem 3.46 step by step to find invertible matrices P and Q such that $PAQ = D_r$. In your development, write out the sets \mathcal{A} , \mathcal{A}' , and \mathcal{B} .

$$(a) A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad (b) A = \begin{bmatrix} 3 & 2 & 1 & 0 \\ -4 & -3 & -2 & 0 \\ 1 & 0 & 3 & 1 \\ -3 & -3 & 1 & 1 \end{bmatrix}$$

9. For each matrix A , find invertible matrices P and Q such that PAQ has the first r elements of the main diagonal equal to 1, and all other elements 0.

$$(a) A = \begin{bmatrix} 0 & 2 & 4 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 3 & 6 & 1 \end{bmatrix} \quad (b) A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & 2 \\ -3 & -6 & -6 \end{bmatrix}$$

10. In Problem 1 above, let A be the matrix in part (c), and let B be the matrix in part (d). Given that A and B are equivalent, find invertible matrices P and Q such that $B = PAQ$. (*Hint:* See the proof of Theorem 3.47.)
11. In Problem 1 above, let A be the matrix in part (a), and let B be the matrix in part (b). Given that A and B are equivalent, find invertible matrices P and Q such that $B = PAQ$.
12. Prove that the relation “equivalence of matrices” is an equivalence relation on the set of all matrices over \mathbf{R} .
13. Prove that if B is conformable to A , then $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$.
14. Prove Corollary 3.50.

Chapter 4

Vector Spaces, Matrices, and Linear Equations

4.1 Introduction

As promised earlier, the preceding results will now be extended to more general situations. This extension is followed by an application of these results to the solution of systems of linear equations.

4.2 Vector Spaces

The theory developed thus far depended basically on the fact that the set of all real numbers forms a field. Our first objective is to extend this theory to other fields and other vector spaces.

Definition 4.1 Suppose that \mathcal{F} is a set of elements in which a relation of equality and operations of addition and multiplication, denoted by $+$ and \cdot , respectively, are defined. Then \mathcal{F} is a field with respect to these operations if the conditions below are satisfied for all a, b, c in \mathcal{F} :

1. $a + b$ is in \mathcal{F} . (Closure property for addition)
2. $(a + b) + c = a + (b + c)$. (Associative property of addition)
3. There is an element 0 in \mathcal{F} such that $a + 0 = a$ for every a in \mathcal{F} . (Additive identity)
4. For each a in \mathcal{F} , there is an element $-a$ in \mathcal{F} such that $a + (-a) = 0$. (Additive inverses)
5. $a + b = b + a$. (Commutative property of addition)
6. $a \cdot b$ is in \mathcal{F} . (Closure property for multiplication)
7. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. (Associative property of multiplication)

8. There is an element 1 in \mathcal{F} such that $a \cdot 1 = a$ for every a in \mathcal{F} . (Multiplicative identity)
9. For each $a \neq 0$ in \mathcal{F} , there is an element a^{-1} in \mathcal{F} such that $a \cdot (a^{-1}) = 1$. (Multiplicative inverses)
10. $a \cdot b = b \cdot a$. (Commutative property of multiplication)
11. $(a + b) \cdot c = a \cdot c + b \cdot c$. (Distributive property)

The notation ab will be used interchangeably with $a \cdot b$ to indicate multiplication. Elements of a field will be referred to as **scalars**.

There are many fields other than the real numbers. Two familiar examples are provided by the set of all rational numbers or the set of all complex numbers. Fields that are subsets of the complex numbers are called **number fields**. There are many fields other than the number fields, and some of them may be known to the student. In the material from here on, the results obtained may be interpreted in as much generality as the student's background permits. If that background included no knowledge of fields other than number fields, \mathcal{F} may be regarded as being a number field throughout the development.

The following definition should be compared with Definition 1.2 and Theorem 1.3 of Chapter 1.

Definition 4.2 Let \mathcal{F} be a field, and suppose that \mathbf{V} is a set of elements such that for each a in \mathcal{F} and each \mathbf{v} in \mathbf{V} , there is a product $a\mathbf{v}$ defined. The operation that yields this product is called **scalar multiplication**. Moreover, suppose that an operation of addition is defined in \mathbf{V} . Then \mathbf{V} is a **vector space over \mathcal{F}** with respect to these operations if the conditions below are satisfied for any a, b in \mathcal{F} and any $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in \mathbf{V} :

1. $\mathbf{u} + \mathbf{v}$ is in \mathbf{V} . (Closure of \mathbf{V} under addition)
2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$. (Associative property of addition in \mathbf{V})
3. There is an element $\mathbf{0}$ in \mathbf{V} such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all \mathbf{v} in \mathbf{V} . (Additive identity)
4. For each \mathbf{v} in \mathbf{V} , there is an element $-\mathbf{v}$ in \mathbf{V} such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$. (Additive inverses in \mathbf{V})
5. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. (Commutative property of addition in \mathbf{V})
6. $a\mathbf{v}$ is in \mathbf{V} . (Closure of \mathbf{V} under scalar multiplication)
7. $a(b\mathbf{v}) = (ab)\mathbf{v}$. (Associative property of scalar multiplication)
8. $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$. (Distributive property, addition in \mathbf{V})
9. $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$. (Distributive property, addition in \mathcal{F})
10. $1 \cdot \mathbf{v} = \mathbf{v}$.

The elements of \mathbf{V} are called **vectors** if \mathbf{V} is a vector space over \mathcal{F} .

With the exception of Section 1.4, all of the definitions, theorems, and proofs in Chapter 1 through Definition 1.23 apply to arbitrary vector spaces \mathbf{V} . Each corresponding statement may be obtained by replacing \mathbf{R} by an arbitrary field \mathcal{F} and \mathbf{R}^n by an arbitrary vector space \mathbf{V} over \mathcal{F} . Some restrictions are necessary in the remainder of Chapter 1.

Definition 4.3 Let \mathbf{V} be a vector space over \mathcal{F} . Then \mathbf{V} is a **finite-dimensional vector space** (or a **vector space of finite dimension over \mathcal{F}**) if there exists a basis of \mathbf{V} with a finite number of elements. If no such basis exists, \mathbf{V} is **infinite-dimensional**, or of **infinite dimension over \mathcal{F}** .

The results in Chapter 1 after Definition 1.23 apply only to finite-dimensional vector spaces. If \mathbf{R} is replaced by \mathcal{F} and \mathbf{R}^n is replaced by a finite-dimensional vector space \mathbf{V} over \mathcal{F} , Theorems 1.24, 1.26, and 1.27 and Definition 1.28 remain valid *with the proofs unchanged* except for notation. In particular, any two bases of a finite-dimensional vector space have the same number of elements, and the number of elements in a basis is the **dimension** of the vector space. In Theorems 1.31 through 1.34, \mathbf{R}^n is replaced by an n -dimensional vector space \mathbf{V} and \mathbf{W} is a subspace of \mathbf{V} .

Let us consider now some examples of vector spaces. In each case, \mathcal{F} denotes a field.

Example 1 □ For a fixed positive integer n , let \mathcal{F}^n denote the set of all n -tuples (u_1, u_2, \dots, u_n) with u_i in \mathcal{F} . Two elements $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ are **equal** if and only if $u_i = v_i$ for $i = 1, 2, \dots, n$. With addition defined in \mathcal{F}^n by

$$(u_1, u_2, \dots, u_n) + (v_1, v_2, \dots, v_n) = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n)$$

and scalar multiplication by

$$a(u_1, u_2, \dots, u_n) = (au_1, au_2, \dots, au_n)$$

the same techniques used to prove Theorem 1.3 in Section 1.2 can be used to prove that \mathcal{F}^n is a vector space over \mathcal{F} . Denoting the multiplicative identity in \mathcal{F} by 1, the vectors

$$(1, 0, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, 0, \dots, 1)$$

can be shown to form a basis of \mathcal{F}^n , and therefore \mathcal{F}^n has dimension n over \mathcal{F} . For $n = 1$, we can identify (a_1) with a_1 . That is, \mathcal{F} is a vector space of dimension one over \mathcal{F} . ■

Example 2 □ For a fixed nonnegative integer n , let \mathcal{P}_n denote the set of all polynomials of the form

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

with each a_i in \mathcal{F} . We shall refer to this set as “ \mathcal{P}_n over \mathcal{F} .” That is, \mathcal{P}_n over \mathcal{F} is the set of all polynomials $\sum_{i=0}^n a_i x^i$ in the variable x with coefficients in \mathcal{F} and degree less than or equal to n .¹ Let

$$p(x) = a_0 + a_1x + \cdots + a_nx^n$$

and

$$q(x) = b_0 + b_1x + \cdots + b_nx^n$$

¹Note that we are using a_0 interchangeably with a_0x^0 in the sigma notation.

denote two elements of \mathcal{P}_n . Then $p(x)$ and $q(x)$ are **equal** if and only if $a_i = b_i$ for $i = 0, 1, 2, \dots, n$. With addition and scalar multiplication defined in the usual ways by

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$$

and

$$cp(x) = ca_0 + ca_1x + \cdots + ca_nx^n,$$

it is easy to verify that \mathcal{P}_n is a vector space over \mathcal{F} with the zero polynomial

$$\mathbf{0} = 0 + 0x + 0x^2 + \cdots + 0x^n$$

as its additive identity. With $p(x)$ as given above, its additive inverse is the polynomial

$$\begin{aligned}-p(x) &= (-a_0) + (-a_1)x + \cdots + (-a_n)x^n \\ &= (-1)p(x).\end{aligned}$$

The set \mathcal{B} of $n + 1$ polynomials given by

$$\mathcal{B} = \{1, x, x^2, \dots, x^n\}$$

spans \mathcal{P}_n since an arbitrary $p(x) = \sum_{i=0}^n a_i x^i$ is automatically a linear combination of vectors in \mathcal{B} :

$$p(x) = a_0(1) + a_1(x) + a_2(x^2) + \cdots + a_n(x^n).$$

Also, a linear combination of vectors in \mathcal{B} yields the zero polynomial $\mathbf{0}$ if and only if all coefficients are zero. Thus \mathcal{B} is linearly independent and forms a basis of \mathcal{P}_n . It follows from this fact that all bases of \mathcal{P}_n have $n + 1$ elements, and \mathcal{P}_n is of dimension $n + 1$. ■

For our next two examples, we draw on topics from the calculus.

Example 3 □ Let \mathbf{V} denote the set of all infinite sequences

$$\{a_n\} = a_1, a_2, \dots, a_n, \dots$$

of real numbers a_n . Two sequences $\{a_n\}$ and $\{b_n\}$ in \mathbf{V} are **equal** if and only if $a_n = b_n$ for all positive integers n . This set \mathbf{V} is a vector space over \mathbf{R} with respect to the operation of addition

$$\{a_n\} + \{b_n\} = \{a_n + b_n\}$$

and scalar multiplication

$$c\{a_n\} = \{ca_n\}$$

for c in \mathbf{R} . The zero vector in \mathbf{V} is the sequence $\{c_n\}$ with $c_n = 0$ for all n , and the additive inverse of $\{a_n\}$ in \mathbf{V} is

$$\begin{aligned}-\{a_n\} &= \{-a_n\} \\ &= (-1)\{a_n\}.\end{aligned}$$

We are not equipped in this text to prove it, but this vector space \mathbf{V} is of infinite dimension over \mathbf{R} . ■

Example 4 □ Let \mathbf{V} be the set of all real-valued functions of the real variable t with domain the set \mathbf{R} of all real numbers. Two functions f and g in \mathbf{V} are **equal** if and only if $f(t) = g(t)$ for all real numbers t . With respect to the ordinary operations of addition

$$(f + g)(t) = f(t) + g(t)$$

and scalar multiplication

$$(cf)(t) = c(f(t))$$

used in the calculus, the set \mathbf{V} is a vector space over \mathbf{R} . The zero vector in \mathbf{V} is the constant function that is identically zero for all values of the variable t . The additive inverse of f is the function $-f$ given by

$$\begin{aligned} (-f)(t) &= -(f(t)) \\ &= (-1)(f(t)). \end{aligned}$$

We are not equipped to prove it here, but this \mathbf{V} is an infinite-dimensional vector space over \mathbf{R} . ■

Example 5 □ Let $\mathcal{F}_{m \times n}$ denote the set of all m by n matrices $A = [a_{ij}]_{m \times n}$ with elements a_{ij} in \mathcal{F} . With addition defined by

$$A + B = [a_{ij}]_{m \times n} + [b_{ij}]_{m \times n} = [a_{ij} + b_{ij}]_{m \times n}$$

and scalar multiplication defined by

$$cA = c[a_{ij}]_{m \times n} = [ca_{ij}]_{m \times n},$$

$\mathcal{F}_{m \times n}$ is a finite-dimensional vector space over \mathcal{F} . The additive identity in $\mathcal{F}_{m \times n}$ is the **zero matrix**

$$O_{m \times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} = [0]_{m \times n},$$

and the additive inverse of $A = [a_{ij}]_{m \times n}$ is the matrix

$$\begin{aligned} -A &= [-a_{ij}]_{m \times n} \\ &= (-1)A. \end{aligned}$$

In many instances, we shall also use the zero vector symbol $\mathbf{0}$ to indicate a zero matrix. ■

As illustrations of the operations defined in Example 5, we have

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

and

$$3 \begin{bmatrix} 5 & -4 & 7 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 15 & -12 & 21 \\ 3 & 0 & 6 \end{bmatrix}$$

in the vector space $\mathbf{R}_{2 \times 3}$ of all 2×3 matrices over \mathbf{R} .

Our last example in this section demonstrates that the operations used in a set to form a vector space are not unique and do not have to be defined in a certain way. Some operations that can be used as addition and scalar multiplication may look somewhat strange when first encountered.

Example 6 □ Let \mathbf{V} be the set of all ordered pairs of real numbers with the usual equality:

$$(x_1, x_2) = (y_1, y_2) \text{ if and only if } x_1 = y_1 \text{ and } x_2 = y_2.$$

Addition and scalar multiplication are defined in \mathbf{V} as follows:

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1 + 1, x_2 + y_2 + 1),$$

$$c(x_1, x_2) = (c + cx_1 - 1, c + cx_2 - 1).$$

We shall systematically check the ten conditions required by Definition 4.2 in order that \mathbf{V} be a vector space over \mathbf{R} . To this end, let $\mathbf{u} = (u_1, u_2)$, $\mathbf{v} = (v_1, v_2)$, and $\mathbf{w} = (w_1, w_2)$ be arbitrary elements in \mathbf{V} , and let a and b represent arbitrary real numbers.

(1) The sum

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1 + 1, u_2 + v_2 + 1)$$

is in \mathbf{V} since both $u_1 + v_1 + 1$ and $u_2 + v_2 + 1$ are real numbers.

(2) We have

$$\begin{aligned} (\mathbf{u} + \mathbf{v}) + \mathbf{w} &= (u_1 + v_1 + 1, u_2 + v_2 + 1) + (w_1, w_2) \\ &= ((u_1 + v_1 + 1) + w_1 + 1, (u_2 + v_2 + 1) + w_2 + 1) \\ &= (u_1 + v_1 + w_1 + 2, u_2 + v_2 + w_2 + 2) \end{aligned}$$

and

$$\begin{aligned} \mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (u_1, u_2) + (v_1 + w_1 + 1, v_2 + w_2 + 1) \\ &= (u_1 + (v_1 + w_1 + 1) + 1, u_2 + (v_2 + w_2 + 1) + 1) \\ &= (u_1 + v_1 + w_1 + 2, u_2 + v_2 + w_2 + 2). \end{aligned}$$

Thus addition in \mathbf{V} is associative.

(3) The element $(-1, -1)$ in \mathbf{V} is an additive identity since

$$\begin{aligned} (v_1, v_2) + (-1, -1) &= (v_1 - 1 + 1, v_2 - 1 + 1) \\ &= (v_1, v_2) \end{aligned}$$

for all $\mathbf{v} = (v_1, v_2)$ in \mathbf{V} . Thus we write $\mathbf{0} = (-1, -1)$.

(4) The additive inverse of $\mathbf{v} = (v_1, v_2)$ is $(-v_1 - 2, -v_2 - 2)$ in \mathbf{V} since

$$\begin{aligned}(v_1, v_2) + (-v_1 - 2, -v_2 - 2) &= (v_1 - v_1 - 2 + 1, v_2 - v_2 - 2 + 1) \\&= (-1, -1) \\&= \mathbf{0}.\end{aligned}$$

(5) Addition in \mathbf{V} is commutative since

$$\begin{aligned}\mathbf{u} + \mathbf{v} &= (u_1 + v_1 + 1, u_2 + v_2 + 1) \\&= (v_1 + u_1 + 1, v_2 + u_2 + 1) \\&= \mathbf{v} + \mathbf{u}.\end{aligned}$$

(6) The product

$$a(v_1, v_2) = (a + av_1 - 1, a + av_2 - 1)$$

is always in \mathbf{V} since both $a + av_1 - 1$ and $a + av_2 - 1$ are real numbers.

(7) We have

$$\begin{aligned}a(b\mathbf{v}) &= a(b + bv_1 - 1, b + bv_2 - 1) \\&= (a + a(b + bv_1 - 1) - 1, a + a(b + bv_2 - 1) - 1) \\&= (a + ab + a(bv_1) - a - 1, a + ab + a(bv_2) - a - 1) \\&= ((ab) + (ab)v_1 - 1, ab + (ab)v_2 - 1) \\&= (ab)\mathbf{v},\end{aligned}$$

so scalar multiplication has the associative property.

(8) Verifying the distributive property with addition in \mathbf{V} , we find that

$$\begin{aligned}a(\mathbf{u} + \mathbf{v}) &= a(u_1 + v_1 + 1, u_2 + v_2 + 1) \\&= (a + au_1 + av_1 + a - 1, a + au_2 + av_2 + a - 1) \\&= (a + au_1 - 1, a + au_2 - 1) + (a + av_1 - 1, a + av_2 - 1) \\&= a\mathbf{u} + a\mathbf{v}.\end{aligned}$$

(9) The distributive property with addition in \mathbf{R} is valid since

$$\begin{aligned}(a + b)\mathbf{v} &= ((a + b) + (a + b)v_1 - 1, (a + b) + (a + b)v_2 - 1) \\&= (a + b + av_1 + bv_1 - 1, a + b + av_2 + bv_2 - 1) \\&= (a + av_1 - 1, a + av_2 - 1) + (b + bv_1 - 1, b + bv_2 - 1) \\&= a\mathbf{v} + b\mathbf{v}.\end{aligned}$$

(10) We have

$$\begin{aligned} 1 \cdot \mathbf{v} &= (1 + 1(v_1) - 1, 1 + 1(v_2) - 1) \\ &= (v_1, v_2). \end{aligned}$$

We have verified all the conditions in Definition 4.2, so this set \mathbf{V} is a vector space over \mathbf{R} with respect to these operations, even though these operations are dramatically different from the standard operations in the familiar vector space \mathbf{R}^2 . ■

Exercises 4.2

1. Verify that the set in the indicated example from this section is actually a vector space.
 - (a) Example 1
 - (b) Example 2
 - (c) Example 3
 - (d) Example 4
 - (e) Example 5
2. Write out a basis for the vector space $\mathbf{R}_{3 \times 2}$.
3. What is the dimension of $\mathcal{F}_{m \times n}$?
4. Find a set that forms a basis for the vector space \mathbf{V} in Example 6 of this section, and prove that your set is a basis for \mathbf{V} .

In Problems 5-16, assume that equality in the given set is the same as in the example of this section that involves the same elements, and determine if the given set is a vector space over \mathbf{R} with respect to the operations defined in the problem. If it is not, list all conditions in Definition 4.2 that fail to hold.

5. The set \mathbf{V} of all ordered pairs of real numbers with operations defined by

$$\begin{aligned} (x_1, x_2) + (y_1, y_2) &= (x_1 + y_1, 0), \\ c(x_1, x_2) &= (cx_1, cx_2). \end{aligned}$$

6. The set \mathbf{V} of all ordered pairs of real numbers with operations defined by

$$\begin{aligned} (x_1, x_2) + (y_1, y_2) &= (x_1 + y_1, x_2 + y_2), \\ c(x_1, x_2) &= (cx_1, x_2). \end{aligned}$$

7. The set \mathbf{V} of all ordered pairs of *positive* real numbers with operations defined by

$$\begin{aligned} (x_1, x_2) + (y_1, y_2) &= (x_1 y_1, x_2 y_2), \\ c(x_1, x_2) &= (x_1^c, x_2^c). \end{aligned}$$

8. The set \mathbf{W} of all $p(x)$ in Example 2 of this section that have zero constant term, with operations as defined in Example 2.

9. The set \mathbf{V} of all ordered triples of real numbers with operations defined by

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3), \\ c(x_1, x_2, x_3) = (cx_1, 0, 0).$$

10. The set \mathbf{V} of all ordered triples of real numbers with operations defined by

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3), \\ c(x_1, x_2, x_3) = (x_1^c, x_2^c, x_3^c).$$

11. The set \mathbf{V} of all diagonal matrices in $\mathbf{R}_{n \times n}$ with operations the same as those defined in Example 5.
12. The set \mathbf{V} of all skew-symmetric matrices in $\mathbf{R}_{n \times n}$ with operations the same as those defined in Example 5.
13. The set \mathbf{V} of all ordered pairs of real numbers with operations defined by

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1 + 1, x_2 + y_2 + 1), \\ c(x_1, x_2) = (cx_1, cx_2).$$

14. The set \mathbf{V} of all ordered pairs of real numbers with operations defined by

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2), \\ c(x_1, x_2) = (x_1, x_2).$$

15. The set \mathbf{W} of all f in Example 4 that are differentiable (that is, have a derivative at every real number), with operations as defined in Example 4.
16. The set \mathbf{W} of all f in Example 4 such that $f(0) = 0$, with operations as defined in Example 4.
17. Prove that the additive identity in a vector space is unique.
18. Prove that the additive inverse $-\mathbf{v}$ of an element \mathbf{v} in a vector space is unique.
19. Prove that $-\mathbf{v} = (-1)\mathbf{v}$ for an arbitrary vector \mathbf{v} in a vector space.
20. The vector $\mathbf{u} - \mathbf{v}$ is defined as the vector \mathbf{w} that satisfies the equation $\mathbf{v} + \mathbf{w} = \mathbf{u}$.
Prove that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$.
21. Let \mathbf{V} be a vector space over \mathcal{F} .
- (a) Prove that $0\mathbf{v} = \mathbf{0}$ for an arbitrary vector $\mathbf{v} \in \mathbf{V}$.

- (b) If $\mathbf{0}$ denotes the additive identity in \mathbf{V} , prove that $c\mathbf{0} = \mathbf{0}$ for all c in \mathcal{F} .
22. Suppose \mathbf{V} is a vector space over \mathcal{F} . Prove that if $c\mathbf{v} = \mathbf{0}$ for c in \mathcal{F} and \mathbf{v} in \mathbf{V} , then either $c = 0$ or $\mathbf{v} = \mathbf{0}$.
23. As mentioned just after Definition 4.1, the set \mathcal{C} of all complex numbers is a field. Assuming this fact, prove that each of the sets below is a field.
- (a) The set of all complex numbers c that can be written in the form $c = a + b\sqrt{2}$ with a and b rational numbers.
 - (b) The set of all complex numbers c that can be written as $c = a + bi$ with a and b rational numbers. (The symbol i denotes the complex number such that $i^2 = -1$.)
24. Let $\mathbf{R}_{2 \times 2}$ be as defined in Example 5, and let \mathcal{F} be the subset of $\mathbf{R}_{2 \times 2}$ that consists of all invertible matrices in $\mathbf{R}_{2 \times 2}$, together with the zero matrix of order 2. With multiplication as usual and addition as given in Example 5, determine whether or not \mathcal{F} is a field, and justify your answer.

4.3 Subspaces and Related Concepts

Some of the discussion in the last section outlined the way that the theory in Chapter 1 extends to general vector spaces. In order to add depth and meaning to this outline, we shall illustrate here the concepts of subspace, spanning set, basis, and transition matrix in vector spaces other than \mathbf{R}^n . We begin with the concept of a subspace.

If \mathbf{W} is a subset of the vector space \mathbf{V} over \mathcal{F} , then \mathbf{W} may also be a vector space over \mathcal{F} . The terminology used in \mathbf{R}^n generalizes to arbitrary vector spaces, and such a subset is called a **subspace** of \mathbf{V} .

Theorem 1.10 also extends to an arbitrary vector space \mathbf{V} over \mathcal{F} . That is, a subset \mathbf{W} of the vector space \mathbf{V} is a subspace of \mathbf{V} if and only if the following conditions hold:

- i. \mathbf{W} is nonempty;
- ii. for any $a, b \in \mathcal{F}$ and any $\mathbf{u}, \mathbf{v} \in \mathbf{W}$, $a\mathbf{u} + b\mathbf{v} \in \mathbf{W}$.

Example 1 \square Let \mathbf{W} be the set of all symmetric matrices in the vector space $\mathbf{R}_{n \times n}$ of all square matrices of order n over \mathbf{R} . That is, for an $n \times n$ matrix A over \mathbf{R} ,

$$A \in \mathbf{W} \text{ if and only if } A^T = A.$$

We shall show that \mathbf{W} is a subspace of $\mathbf{R}_{n \times n}$. The set \mathbf{W} is nonempty since the zero matrix $\mathbf{0}_{n \times n} = [0]_{n \times n}$ is in \mathbf{W} . Let r and s be arbitrary real numbers, and let $A = [a_{ij}]_{n \times n}$ and $B = [b_{ij}]_{n \times n}$ be arbitrary elements in \mathbf{W} . Then A and B are symmetric, so that $a_{ji} = a_{ij}$ and $b_{ji} = b_{ij}$ for $i = 1, 2, \dots, n$; $j = 1, 2, \dots, n$. Evaluating $rA + sB$, we have

$$\begin{aligned} rA + sB &= [ra_{ij}]_{n \times n} + [sb_{ij}]_{n \times n} \\ &= [c_{ij}]_{n \times n} \\ &= C, \end{aligned}$$

where $c_{ij} = ra_{ij} + sb_{ij}$ for $i = 1, 2, \dots, n; j = 1, 2, \dots, n$. Thus

$$\begin{aligned} c_{ji} &= ra_{ji} + sb_{ji} \\ &= ra_{ij} + sb_{ij} \quad \text{since } a_{ji} = a_{ij} \text{ and } b_{ji} = b_{ij} \\ &= c_{ij} \end{aligned}$$

for all pairs i, j . This means that $rA + sB$ is symmetric and hence a member of \mathbf{W} . Therefore \mathbf{W} is a subspace of $\mathbf{R}_{n \times n}$. ■

For a nonempty subset \mathcal{A} of a general vector space \mathbf{V} , the set $\langle \mathcal{A} \rangle$ is the set of all linear combinations of vectors in \mathcal{A} . The same development used in Section 1.3 applies in \mathbf{V} , and $\langle \mathcal{A} \rangle$ is the **subspace spanned** by \mathcal{A} . As in Section 1.3, $\langle \emptyset \rangle$ is the zero subspace $\{\mathbf{0}\}$ of \mathbf{V} .

Example 2 □ In the vector space $\mathbf{R}_{2 \times 2}$, consider the sets of vectors $\mathcal{A} = \{A_1, A_2\}$ and $\mathcal{B} = \{B_1, B_2\}$, where

$$A_1 = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix}, B_2 = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}.$$

We shall show that $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$.

By inspection, we see that A_2 is not a multiple of A_1 . Hence $\mathcal{A} = \{A_1, A_2\}$ is linearly independent and $\langle \mathcal{A} \rangle$ has dimension 2. Similarly, B_2 is not a multiple of B_1 and therefore $\langle \mathcal{B} \rangle$ has dimension 2. Knowing that $\langle \mathcal{A} \rangle$ and $\langle \mathcal{B} \rangle$ have the same dimension, it is sufficient to show that $\langle \mathcal{B} \rangle \subseteq \langle \mathcal{A} \rangle$, in order to prove that $\langle \mathcal{B} \rangle = \langle \mathcal{A} \rangle$. But $\langle \mathcal{B} \rangle$ is dependent on \mathcal{B} , so we need only show that $\mathcal{B} \subseteq \langle \mathcal{A} \rangle$, by Theorem 1.8. That is, we need only demonstrate that each of B_1 and B_2 is a linear combination of A_1 and A_2 .

Setting up the equation $c_1 A_1 + c_2 A_2 = B_1$, we have

$$c_1 \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix}.$$

This leads to the system of equations

$$\begin{aligned} 2c_1 + c_2 &= 3 \\ -c_1 + 2c_2 &= -4 \\ c_2 &= -1 \\ c_1 &= 2, \end{aligned}$$

so $B_1 = 2A_1 - A_2$. Similarly, we find that $B_2 = A_1 + 2A_2$. Thus we have proved that $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$. ■

Just after Example 3 in Section 1.5, we noted that a basis of a subspace can be refined from a finite spanning set for the subspace by deleting all vectors in the spanning set that are linear combinations of preceding vectors. This procedure is illustrated in the following example.

Example 3 □ Let

$$\begin{aligned} p_1(x) &= 1 + x + 2x^2, & p_2(x) &= 6 + 6x + 12x^2, & p_3(x) &= 1 + x^2 + x^3, \\ p_4(x) &= 2 + x + 3x^2 + x^3, & p_5(x) &= 1 + x^2 \end{aligned}$$

in the vector space \mathcal{P}_3 over \mathbf{R} , and let

$$\mathcal{A} = \{p_1(x), p_2(x), p_3(x), p_4(x), p_5(x)\}.$$

We shall find a subset of \mathcal{A} that forms a basis of $\langle \mathcal{A} \rangle$.

Employing the procedure described just before this example, we see that $p_2(x) = 6p_1(x)$, so $p_2(x)$ can be deleted from \mathcal{A} to obtain the spanning set

$$\{p_1(x), p_3(x), p_4(x), p_5(x)\}$$

of $\langle \mathcal{A} \rangle$. We see that $p_3(x)$ is not a multiple of $p_1(x)$, so we then check to see if $p_4(x)$ is a linear combination of $p_1(x)$ and $p_3(x)$. It is easy to discover that

$$p_4(x) = p_1(x) + p_3(x),$$

so $p_4(x)$ can be deleted from the last spanning set of $\langle \mathcal{A} \rangle$ we obtained, leaving

$$\{p_1(x), p_3(x), p_5(x)\}$$

as a spanning set of $\langle \mathcal{A} \rangle$. Setting up the equation

$$p_5(x) = c_1 p_1(x) + c_2 p_3(x)$$

leads to

$$\begin{aligned} 1 + x^2 &= c_1(1 + x + 2x^2) + c_2(1 + x^2 + x^3) \\ &= (c_1 + c_2) + c_1 x + (2c_1 + c_2)x^2 + c_2 x^3 \end{aligned}$$

and the resulting system of equations

$$\begin{aligned} c_1 + c_2 &= 1 \\ c_1 &= 0 \\ 2c_1 + c_2 &= 1 \\ c_2 &= 0, \end{aligned}$$

which clearly has no solution. Thus $\{p_1(x), p_3(x), p_5(x)\}$ is linearly independent and therefore forms a basis of $\langle \mathcal{A} \rangle$ that is contained in \mathcal{A} . ■

The concept of a transition matrix from one finite set of vectors to another in an arbitrary vector space applies in every vector space. We consider an example now in the vector space \mathcal{P}_2 of all polynomials in x with degree ≤ 2 and coefficients in \mathbf{R} .

Example 4 □ Consider the bases $\mathcal{A} = \{p_1(x), p_2(x), p_3(x)\}$ and $\mathcal{B} = \{q_1(x), q_2(x), q_3(x)\}$ of \mathcal{P}_2 over \mathbf{R} , where

$$\begin{aligned} p_1(x) &= 1 + x, & p_2(x) &= x, & p_3(x) &= x + x^2, \\ q_1(x) &= 1, & q_2(x) &= 1 + x, & q_3(x) &= 1 + x + x^2. \end{aligned}$$

We shall find the matrix of transition A from \mathcal{A} to \mathcal{B} . In the matrix $A = [a_{ij}]_{3 \times 3}$, the j^{th} column is the coordinate matrix of $q_j(x)$ with respect to \mathcal{A} . Thus we must write each $q_j(x)$ as a linear combination of the polynomials in \mathcal{A} . The required linear combinations are given by

$$\begin{aligned} 1 &= (1)(1+x) + (-1)(x) + (0)(x+x^2), \\ 1+x &= (1)(1+x) + (0)(x) + (0)(x+x^2), \\ 1+x+x^2 &= (1)(1+x) + (-1)(x) + (1)(x+x^2). \end{aligned}$$

That is,

$$\begin{aligned} q_1(x) &= (1)p_1(x) + (-1)p_2(x) + (0)p_3(x), \\ q_2(x) &= (1)p_1(x) + (0)p_2(x) + (0)p_3(x), \\ q_3(x) &= (1)p_1(x) + (-1)p_2(x) + (1)p_3(x). \end{aligned}$$

Thus the transition matrix from \mathcal{A} to \mathcal{B} is

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

We note that the coordinates of the $q_j(x)$ are entered as *columns* of A , not as *rows* of A . ■

Exercises 4.3

- Determine whether the given set of vectors is linearly dependent or linearly independent in the vector space \mathcal{P}_2 over \mathbf{R} .
 - $\{1 - 2x, x + x^2, 1 - x + x^2\}$
 - $\{1 + x, x + x^2, 1 + 2x + x^2\}$
 - $\{1 - x^2, x + 2x^2, 1 + 3x^2\}$
 - $\{1 + x, 1 + x^2, 1 + x + x^2\}$
 - $\{1 - 2x, 2 - x^2, 4 + x^2, 1 + x + x^2\}$
 - $\{1 + x, 1 - x\}$
- Determine whether the given set of polynomials spans \mathcal{P}_2 over \mathbf{R} .
 - $\{1 - x, 1 + x, 1 - x^2\}$
 - $\{1 - x, x + x^2, 1 + x^2\}$
 - $\{1 + x, x + x^2, 1 + 2x + x^2, 1 - x^2\}$
 - $\{1 - x, x + x^2, 1 + x + x^2, 1 + x^2\}$

3. Determine whether the given set of vectors in $\mathbf{R}_{2 \times 2}$ is linearly dependent or linearly independent.

(a) $\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \right\}$

(b) $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right\}$

(c) $\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$

(d) $\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$

4. Which of the following sets of vectors are bases for the vector space \mathcal{P}_2 over \mathbf{R} ?

(a) $\{1 - x, 1 - x + x^2, 1 + x^2, 1 - x - x^2\}$ (b) $\{1 - x, 1 - x^2, x - x^2\}$
 (c) $\{1 - x, x - x^2, 1 + x^2\}$ (d) $\{1 + x, 1 + x^2\}$

5. Find the transition matrix from the basis \mathcal{A} to the basis \mathcal{B} in \mathcal{P}_2 over \mathbf{R} .

(a) $\mathcal{A} = \{p_1(x), p_2(x), p_3(x)\}, \mathcal{B} = \{q_1(x), q_2(x), q_3(x)\}$, where
 $p_1(x) = 1 - x, p_2(x) = x + x^2, p_3(x) = x,$
 $q_1(x) = 1, q_2(x) = x + x^2, q_3(x) = 1 + x^2$

(b) $\mathcal{A} = \{p_1(x), p_2(x), p_3(x)\}, \mathcal{B} = \{q_1(x), q_2(x), q_3(x)\}$, where
 $p_1(x) = x + x^2, p_2(x) = x, p_3(x) = 1 + x^2,$
 $q_1(x) = x, q_2(x) = x - x^2, q_3(x) = 1 + x - x^2$

6. Find a subset of the given set of vectors that forms a basis for the subspace spanned by the given set.

(a) $p_1(x) = x + x^2, p_2(x) = 1 + x + x^2, p_3(x) = 1, p_4(x) = 1 + 2x + 2x^2$, in \mathcal{P}_2 over \mathbf{R}

(b) $p_1(x) = 1 + x^2, p_2(x) = x + x^2, p_3(x) = 1 - x, p_4(x) = 1 + x + 2x^2$, in \mathcal{P}_2 over \mathbf{R}

(c) $\left\{ \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \right\}$ in $\mathbf{R}_{2 \times 2}$

(d) $\left\{ \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 5 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix} \right\}$ in $\mathbf{R}_{2 \times 2}$

- (e) $p_1(x) = x^2 + x + 1, p_2(x) = x^2 - x - 2, p_3(x) = x^3 + x - 1, p_4(x) = x - 1$ in \mathcal{P}_3 over \mathbf{R}
(f) $p_1(x) = x^2 - 3x + 2, p_2(x) = x^2 + x - 2, p_3(x) = x^3 - 1, p_4(x) = x - 1$ in \mathcal{P}_3 over \mathbf{R}

7. Verify that the given set \mathbf{W} is a subspace of \mathbf{R}^4 and find a basis for \mathbf{W} .

- (a) \mathbf{W} is the set of all vectors in \mathbf{R}^4 of the form $(a, 0, b, 0)$.
(b) \mathbf{W} is the set of all vectors in \mathbf{R}^4 of the form (a, b, c, d) with $c = a$ and $d = a + b$.

In Problems 8-19, determine which of the given sets \mathbf{W} are subspaces of the indicated vector space. If a set is not a subspace, state a reason.

8. The set \mathbf{W} of all vectors in \mathbf{R}^3 of the form (a, b, c) , where $a + b = 1$.
9. The set \mathbf{W} of all vectors in \mathbf{R}^3 of the form (a, b, c) , where $b = a^2$.
10. The set \mathbf{W} of all vectors in \mathbf{R}^3 of the form (a, b, c) , where $c = 2a + b$.
11. The set \mathbf{W} of all vectors in \mathbf{R}^3 of the form (a, b, c) , where $c = ab$.
12. The set \mathbf{W} of all polynomials $a_0 + a_1x + a_2x^2$ in \mathcal{P}_2 over \mathbf{R} that have $a_1 = 0$.
13. The set \mathbf{W} of all polynomials $a_0 + a_1x + a_2x^2$ in \mathcal{P}_2 over \mathbf{R} that have $a_0 + a_1 = 0$.

14. The set \mathbf{W} of all matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ in $\mathbf{R}_{2 \times 2}$ that have $a + b = 0$.
15. The set \mathbf{W} of all matrices of the form $\begin{bmatrix} 1 & a \\ b & c \end{bmatrix}$ in $\mathbf{R}_{2 \times 2}$.
16. The set \mathbf{W} of all matrices in $\mathbf{R}_{2 \times 2}$ that have the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with a, b, c , and d integers.
17. The set \mathbf{W} that consists of the zero matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ together with all invertible matrices in $\mathbf{R}_{2 \times 2}$.
18. The set \mathbf{W} of all matrices of the form $\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ in $\mathbf{R}_{2 \times 2}$.
19. The set \mathbf{W} of all matrices in $\mathbf{R}_{2 \times 2}$ that have the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ where $a^2 = d^2$.

4.4 Isomorphisms of Vector Spaces

The objective of this section is to derive the most important single result concerning finite-dimensional vector spaces. The concept of isomorphism between two vector spaces is essential to an understanding of this fundamental result, which is contained in Theorem 4.5. For this reason it is necessary to first consider the concept of an isomorphism. An isomorphism is a certain type of mapping, so we begin with a discussion of mappings in general.

A **mapping** (or **function**, or **transformation**) from the set \mathcal{S} into the set \mathcal{T} is a set f of ordered pairs (s, t) of elements $s \in \mathcal{S}, t \in \mathcal{T}$ that has the following property: for each s in \mathcal{S} , there is exactly one element t in \mathcal{T} such that $(s, t) \in f$. The notation $t = f(s)$ indicates that t is the unique element of \mathcal{T} that is associated with s by the rule that $(s, t) \in f$. We say that $f(s)$ is the **image** of s , and that s is an **inverse image** of $f(s)$.

If f is a mapping of \mathcal{S} into \mathcal{T} , then it may happen that $f(s_1) = f(s_2)$, even though $s_1 \neq s_2$. If it is true that $f(s_1) = f(s_2)$ always implies $s_1 = s_2$, then f is called **injective** or **one-to-one**. Another point of interest is that it is not required that every element of \mathcal{T} be an image of an element in \mathcal{S} . If it happens that every t in \mathcal{T} is the image of at least one s in \mathcal{S} under f , we say that f is a **surjective** mapping of \mathcal{S} into \mathcal{T} , or that f maps \mathcal{S} onto \mathcal{T} . If f is both injective and surjective, then f is called **bijective**.

Two mappings f and g of \mathcal{S} into \mathcal{T} are **equal** if and only if $f(s) = g(s)$ for every s in \mathcal{S} .

The rule $f(x) = \sin x$ defines a mapping of the set \mathbf{R} of real numbers into \mathbf{R} . This mapping f is clearly not injective (for example, $\sin \frac{\pi}{3} = \sin \frac{2\pi}{3} = \frac{\sqrt{3}}{2}$). Also, f is not surjective since there is no x in \mathbf{R} such that $f(x) = 2$.

If $\mathcal{S} = \{x \in \mathbf{R} \mid 0 \leq x \leq \pi\}$ and $\mathcal{T} = \{t \in \mathbf{R} \mid 0 \leq t \leq 1\}$, then the rule $f(x) = \sin x$ defines a mapping of \mathcal{S} into \mathcal{T} that is surjective but is not injective.

If $\mathcal{S} = \{x \in \mathbf{R} \mid 0 \leq x \leq \frac{\pi}{2}\}$ and $\mathcal{T} = \{t \in \mathbf{R} \mid 0 \leq t \leq 1\}$, the rule $f(x) = \sin x$ defines a mapping of \mathcal{S} into \mathcal{T} that is both surjective and injective.

These examples illustrate the fact that the surjective and injective properties depend on the sets \mathcal{S} and \mathcal{T} as well as the rule that defines the mapping.

Definition 4.4 Let \mathbf{U} and \mathbf{V} be vector spaces over the same field \mathcal{F} . An **isomorphism** from \mathbf{U} to \mathbf{V} is a bijective mapping f of \mathbf{U} into \mathbf{V} that has the property that

$$f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v})$$

for all a, b in \mathcal{F} and all \mathbf{u}, \mathbf{v} in \mathbf{U} . If an isomorphism from \mathbf{U} to \mathbf{V} exists, \mathbf{U} and \mathbf{V} are said to be **isomorphic** vector spaces.

Since an isomorphism f from \mathbf{U} to \mathbf{V} is a bijective mapping, it pairs the elements of \mathbf{U} and \mathbf{V} in a one-to-one fashion. In view of the property $f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v})$, f is said to “preserve linear combinations.” As particular instances of this property, $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$ and $f(a\mathbf{u}) = af(\mathbf{u})$. Thus f preserves sums and scalar products, and the vector spaces \mathbf{U} and \mathbf{V} are structurally the same.

We are now in a position to prove the principal result of this section. This result shows that the first example in Section 4.2 furnishes a completely typical pattern for n -dimensional vector spaces over a field \mathcal{F} .

Theorem 4.5 *Any n -dimensional vector space over the field \mathcal{F} is isomorphic to \mathcal{F}^n .*

Proof. Let \mathbf{V} be an n -dimensional vector space over the field \mathcal{F} , and let

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$$

be a basis of \mathbf{V} . Each $\mathbf{u} \in \mathbf{V}$ can be written *uniquely* as $\mathbf{u} = \sum_{i=1}^n a_i \mathbf{v}_i$, so the rule

$$f(\mathbf{u}) = (a_1, a_2, \dots, a_n)$$

defines an injective mapping of \mathbf{V} into \mathcal{F}^n . It is clear that f is surjective.

Let $\mathbf{u} = \sum_{i=1}^n a_i \mathbf{v}_i$ and $\mathbf{v} = \sum_{i=1}^n b_i \mathbf{v}_i$ be arbitrary vectors in \mathbf{V} , and let a and b be arbitrary scalars in \mathcal{F} . Then

$$\begin{aligned} f(a\mathbf{u} + b\mathbf{v}) &= f\left(\sum_{i=1}^n (aa_i + bb_i)\mathbf{v}_i\right) \\ &= (aa_1 + bb_1, aa_2 + bb_2, \dots, aa_n + bb_n) \\ &= a(a_1, a_2, \dots, a_n) + b(b_1, b_2, \dots, b_n) \\ &= af(\mathbf{u}) + bf(\mathbf{v}), \end{aligned}$$

so that f is an isomorphism from \mathbf{V} to \mathcal{F}^n . ■

Exercises 4.4

- Let \mathbf{V} be the vector space in the second example in Section 4.2. Exhibit an isomorphism from \mathbf{V} to \mathcal{F}^{n+1} .
- Define a mapping f of $\mathbf{R}_{3 \times 2}$ into \mathbf{R}^6 that is an isomorphism from $\mathbf{R}_{3 \times 2}$ to \mathbf{R}^6 , and prove that your mapping is an isomorphism.
- For each subspace \mathbf{W} below, determine the dimension r of \mathbf{W} and find an isomorphism from \mathbf{W} to \mathbf{R}^r .
 - $\mathbf{W} = \langle (1, -1, 1), (2, -2, 2), (0, 1, 0), (1, 0, 1) \rangle$ in \mathbf{R}^3
 - $\mathbf{W} = \langle (4, -2, 5), (0, -4, 0), (12, -18, 15), (4, 0, 5) \rangle$ in \mathbf{R}^3
 - $\mathbf{W} = \langle (2, 0, 4, -1), (5, -1, 11, 8), (0, 1, -7, 9), (7, 0, 8, 16) \rangle$ in \mathbf{R}^4
 - $\mathbf{W} = \langle (4, -2, 5, 5), (0, -4, 0, 4), (12, -18, 15, 27), (4, 0, 5, 3) \rangle$ in \mathbf{R}^4
 - $\mathbf{W} = \langle p_1(x), p_2(x), p_3(x), p_4(x) \rangle$ in \mathcal{P}_3 over \mathbf{R} , where

$$p_1(x) = x^3 + x^2 + 1, \quad p_2(x) = x^2 + x + 1,$$

$$p_3(x) = 2x^3 + 3x^2 + x + 3, \quad p_4(x) = x^3 - x.$$

- (f) $\mathbf{W} = \langle p_1(x), p_2(x), p_3(x), p_4(x) \rangle$ in \mathcal{P}_3 over \mathbf{R} , where

$$p_1(x) = x^3 + x + 1, \quad p_2(x) = x^2,$$

$$p_3(x) = 2x^3 + 3x^2 + 2x + 2, \quad p_4(x) = x + 1.$$

(g) $\mathbf{W} = \left\langle \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} -2 & 0 \\ 2 & -2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix} \right\rangle$ in $\mathbf{R}_{2 \times 2}$

(h) $\mathbf{W} = \left\langle \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 3 \\ 6 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} \right\rangle$ in $\mathbf{R}_{2 \times 2}$

4. Let \mathbf{V} be the subspace spanned by the set $\{p_1(x), p_2(x), p_3(x), p_4(x)\}$ of polynomials in Problem 6(e) of Exercises 4.3. Find an isomorphism from \mathbf{V} to a subspace of \mathbf{R}^4 .
5. Let \mathbf{V} be the subspace spanned by the set $\{p_1(x), p_2(x), p_3(x), p_4(x)\}$ of polynomials in Problem 6(f) of Exercises 4.3. Find an isomorphism from \mathbf{V} to a subspace of \mathbf{R}^4 .
6. Prove that the relation of being isomorphic is an equivalence relation on the set of all vector spaces over \mathcal{F} .

4.5 Standard Bases for Subspaces

The definitions of the elementary operations on sets of vectors as given in Chapter 2 apply unchanged to arbitrary vector spaces. The following statements and proofs also apply unchanged, and may be used in arbitrary vector spaces: Definition 2.1, Theorem 2.2, Theorem 2.3, Corollary 2.4, Definition 2.6, Theorem 2.7, and Corollary 2.8. These results may be summarized in the following statement: If \mathcal{A} and \mathcal{A}' are sets of vectors in a vector space \mathbf{V} such that \mathcal{A}' is obtained from \mathcal{A} by a sequence of elementary operations, then (i) \mathcal{A} and \mathcal{A}' span the same subspace of \mathbf{V} and (ii) \mathcal{A}' is linearly independent if and only if \mathcal{A} is linearly independent. The significance of this statement lies in the fact that elementary operations may be used to obtain simpler forms of spanning sets for a subspace and also to investigate linear independence.

From Corollary 2.8 to the end of Chapter 2, the statements and proofs of results are intimately connected with \mathbf{R}^n . Thus more changes are necessary in formulating the corresponding development for subspaces of arbitrary vector spaces. However, the proofs of the theorems are basically unchanged. For this reason these theorems are stated here with only indications as to the changes in notation necessary in order to obtain the proofs.

Theorem 4.6 *Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ be a fixed basis of the vector space \mathbf{V} over \mathcal{F} , and let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a set of m vectors in \mathbf{V} that spans the subspace $\mathbf{W} = \langle \mathcal{A} \rangle$ of dimension $r > 0$. Then a set $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r, \mathbf{0}, \dots, \mathbf{0}\}$ of m vectors can be obtained*

from \mathcal{A} by a finite sequence of elementary operations so that $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ has the following properties:

1. The first nonzero coordinate of \mathbf{v}'_j with respect to \mathcal{B} is a 1 for the k_j coordinate for $j = 1, 2, \dots, r$. That is, $\mathbf{v}'_j = \sum_{i=k_j}^n a'_{ij} \mathbf{u}_i$ with $a'_{k_j j} = 1$.
2. $k_1 < k_2 < \dots < k_r$
3. \mathbf{v}'_j is the only vector in \mathcal{A}' with a nonzero k_j coordinate relative to \mathcal{B} .
4. $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ is a basis of \mathbf{W} .

Proof. The proof can be obtained from the proof of Theorem 2.9 by replacing \mathbf{R}^n by \mathbf{V} , \mathcal{E}_n by \mathcal{B} , and \mathbf{e}_i by \mathbf{u}_i so the a'_{ij} represents the i^{th} coordinate of \mathbf{v}'_j relative to \mathcal{B} instead of the i^{th} component of \mathbf{v}'_j . ■

Theorem 4.7 For a fixed basis \mathcal{B} , there is one and only one basis of a given subspace \mathbf{W} that satisfies the conditions of Theorem 4.6.

Proof. It follows from Theorem 4.6 that there is at least one such basis of \mathbf{W} . Let $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r\}$ and $\mathcal{A}'' = \{\mathbf{v}''_1, \mathbf{v}''_2, \dots, \mathbf{v}''_r\}$ be two bases of \mathbf{W} that satisfy the conditions. Then the same replacements used in the proof of Theorem 4.6 can be used in the proof of Theorem 2.10 to obtain a proof that $\mathcal{A}' = \mathcal{A}''$. ■

Definition 4.8 Let \mathcal{B} be a fixed basis of the n -dimensional vector space \mathbf{V} over \mathcal{F} , and let \mathbf{W} be a subspace of \mathbf{V} of dimension r . The basis of \mathbf{W} that satisfies the conditions of Theorem 4.6 is called the **standard basis** of \mathbf{W} relative to \mathcal{B} .

That is, the standard basis of \mathbf{W} relative to $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is the unique basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ that has the following properties:

1. $\mathbf{v}_j = \sum_{i=k_j}^n a_{ij} \mathbf{u}_i$ with $a_{k_j j} = 1$
2. $k_1 < k_2 < \dots < k_r$
3. \mathbf{v}_j is the only vector in \mathcal{A} that has a nonzero k_j coordinate relative to \mathcal{B} .

When referring to standard bases, we use the phrase “with respect to \mathcal{B} ” interchangeably with “relative to \mathcal{B} .”

The proofs of the following two theorems can be obtained from those of Theorems 2.12 and 2.13 by making the same changes as were indicated in Theorems 4.6 and 4.7.

Theorem 4.9 Let \mathcal{B} be a fixed basis of the finite-dimensional vector space \mathbf{V} over \mathcal{F} . For any subspace \mathbf{W} of \mathbf{V} , the standard basis of \mathbf{W} relative to \mathcal{B} can be obtained from any basis of \mathbf{W} by a sequence of elementary operations.

Theorem 4.10 Let \mathcal{A} and \mathcal{B} be two sets of m vectors each in the finite-dimensional vector space \mathbf{V} over \mathcal{F} . Then $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ if and only if \mathcal{B} can be obtained from \mathcal{A} by a sequence of elementary operations.

Example 1 □ The set $\mathcal{B} = \{\mathbf{u}_1 = (1, 0, 2), \mathbf{u}_2 = (2, 1, 6), \mathbf{u}_3 = (0, 3, 8)\}$ is a basis of \mathbb{R}^3 . Let $\mathcal{A} = \{(4, -2, 2), (6, 26, 78), (5, 12, 40), (14, 22, 82)\}$ and consider the problem of finding the standard basis of $\langle \mathcal{A} \rangle$ relative to \mathcal{B} .

In order to obtain the coordinates a'_{ij} as in Theorem 4.6, we determine the coordinates of the vectors in \mathcal{A} with respect to \mathcal{B} and then record these coordinates in the columns of a matrix. (That is, we follow the same procedure as in Chapter 2, using coordinates rather than components.) The vectors of \mathcal{A} are found to be given by

$$\begin{aligned}(4, -2, 2) &= 2\mathbf{u}_1 + \mathbf{u}_2 - \mathbf{u}_3, \\ (6, 26, 78) &= -4\mathbf{u}_1 + 5\mathbf{u}_2 + 7\mathbf{u}_3, \\ (5, 12, 40) &= -\mathbf{u}_1 + 3\mathbf{u}_2 + 3\mathbf{u}_3, \\ (14, 22, 82) &= 7\mathbf{u}_2 + 5\mathbf{u}_3,\end{aligned}$$

so that the matrix of coordinates is

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

By using elementary column operations on A (which corresponds to using elementary operations on \mathcal{A}), we find the reduced column-echelon form for A is

$$A' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{6}{7} & \frac{5}{7} & 0 & 0 \end{bmatrix}.$$

Thus the standard basis for $\langle \mathcal{A} \rangle$ with respect to \mathcal{B} is $\{\mathbf{v}'_1, \mathbf{v}'_2\}$, where

$$\begin{aligned}\mathbf{v}'_1 &= \mathbf{u}_1 - \frac{6}{7}\mathbf{u}_3 = \left(1, -\frac{18}{7}, -\frac{34}{7}\right), \\ \mathbf{v}'_2 &= \mathbf{u}_2 + \frac{5}{7}\mathbf{u}_3 = \left(2, \frac{22}{7}, \frac{82}{7}\right).\end{aligned}$$
■

Exercises 4.5

- Using $\mathcal{B} = \{(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)\}$ as the fixed basis of \mathbb{R}^4 , find the standard basis of $\langle \mathcal{A} \rangle$ relative to \mathcal{B} for each set \mathcal{A} below. Write your answers in component form.
 - $\mathcal{A} = \{(4, 4, 3, 1), (7, 7, 5, 1), (4, 4, 3, 1)\}$
 - $\mathcal{A} = \{(5, 4, 2, 3), (3, 3, 2, 2), (8, 7, 4, 5), (2, 1, 0, 1)\}$
 - $\mathcal{A} = \{(-1, -2, -1, -3), (4, 3, 2, 0), (2, -1, 0, -6), (3, 1, 1, -3), (1, 1, 1, 1)\}$

- (d) $\mathcal{A} = \{(4, 3, 2, -1), (-1, -2, -2, 0), (7, 4, 2, -2), (2, 1, 1, 1), (10, 9, 8, 1)\}$
2. Let $p_1(x) = 2x^2 + 2, p_2(x) = x + 1, p_3(x) = 2x^2 - 3x + 1$ in the vector space \mathcal{P}_2 over \mathbf{R} . Given that $\mathcal{B} = \{p_1(x), p_2(x), p_3(x)\}$ is a basis of \mathcal{P}_2 over \mathbf{R} , find the standard basis of $\langle \mathcal{A} \rangle$ with respect to \mathcal{B} if \mathcal{A} is given by
- (a) $\mathcal{A} = \{q_1(x), q_2(x), q_3(x)\}$, where $q_1(x) = 2p_1(x) - 3p_2(x) - p_3(x)$,
 $q_2(x) = p_1(x) - p_2(x) - p_3(x)$, and $q_3(x) = 2p_1(x) - p_2(x) - p_3(x)$.
- (b) $\mathcal{A} = \{q_1(x), q_2(x), q_3(x), q_4(x)\}$, where $q_1(x) = -p_3(x)$,
 $q_2(x) = 3p_1(x) - 6p_2(x) + p_3(x)$, $q_3(x) = 2p_1(x) - 4p_2(x) + 2p_3(x)$, and
 $q_4(x) = -p_1(x) + 2p_2(x) + 2p_3(x)$.
3. Let $\mathbf{u}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \mathbf{u}_4 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
in the vector space $\mathbf{R}_{2 \times 2}$. It is given that $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is a basis of $\mathbf{R}_{2 \times 2}$. For each set \mathcal{A} , find the standard basis of $\langle \mathcal{A} \rangle$ relative to \mathcal{B} .
- (a) $\mathcal{A} = \{A_1, A_2, A_3\}$, where
 $A_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 4 & 5 \\ 4 & 3 \end{bmatrix}$.
- (b) $\mathcal{A} = \{A_1, A_2, A_3, A_4\}$, where
 $A_1 = \begin{bmatrix} -2 & -2 \\ -2 & -1 \end{bmatrix}, A_2 = \begin{bmatrix} -4 & -7 \\ -1 & -2 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}, A_4 = \begin{bmatrix} 6 & 7 \\ 5 & 3 \end{bmatrix}$.
4. Determine in each case whether or not the given sets span the same subspace of \mathcal{P}_2 over \mathbf{R} .
- (a) $\mathcal{A} = \{2x^2 + 2, 2x^2 - 2x + 2\}, \mathcal{B} = \{4x^2 - 6x + 4, 2x, x^2 + 1\}$.
- (b) $\mathcal{A} = \{2x^2 + 4x + 4, 4x^2 - 7x + 7, 18x^2 - 10x + 18\}$,
 $\mathcal{B} = \{4x^2 + 8x + 8, 14x^2 - 14x + 8\}$.
5. Let \mathbf{V} be the vector space \mathcal{P}_2 over \mathbf{R} .
- (a) Find the matrix of transition from the basis $\{1, x, x^2\}$ of \mathbf{V} to the basis \mathcal{B} in Problem 2.
- (b) Find the matrix of transition from the basis \mathcal{B} in Problem 2 to the basis $\{1, x, x^2\}$ of \mathbf{V} .

4.6 Matrices over an Arbitrary Field

At this point the generalization of Chapters 1 and 2 to arbitrary vector spaces is complete. This makes available the basis for the theory of multiplication of matrices over an arbitrary field \mathcal{F} . An examination of the theory in Chapter 3 reveals that the entire development rests only on the facts that \mathbf{R} is a field and \mathbf{R}^n is an n -dimensional vector space over \mathbf{R} . With a single exception, in order to obtain the general form of a result, it is necessary only to replace \mathbf{R} by \mathcal{F} and \mathbf{R}^n by an n -dimensional vector space over \mathcal{F} . The single exception is in the definition of the column space of a matrix in Definition 3.43. In this definition, \mathbf{R}^n is replaced by \mathcal{F}^n .

From this point on, we shall make free use of the general forms of all the properties of matrix multiplication derived in Chapter 3.

4.7 Systems of Linear Equations

Let \mathbf{Q}_n denote the set of all linear equations in n unknowns x_1, x_2, \dots, x_n with coefficients in a field \mathcal{F} . That is,

$$\mathbf{Q}_n = \{a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \mid a_i \in \mathcal{F} \text{ and } b \in \mathcal{F}\},$$

and an element of \mathbf{Q}_n is an equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$. Two elements of \mathbf{Q}_n are **equal** if their corresponding coefficients are equal and their constant terms are equal. **Addition** in \mathbf{Q}_n is defined by

$$\begin{aligned} & [a_1x_1 + a_2x_2 + \cdots + a_nx_n = b] + [a'_1x_1 + a'_2x_2 + \cdots + a'_nx_n = b'] \\ &= [(a_1 + a'_1)x_1 + (a_2 + a'_2)x_2 + \cdots + (a_n + a'_n)x_n = (b + b')]. \end{aligned}$$

Scalar multiplication is given by

$$a \cdot [a_1x_1 + a_2x_2 + \cdots + a_nx_n = b] = [aa_1x_1 + aa_2x_2 + \cdots + aa_nx_n = ab].$$

It is readily verified that \mathbf{Q}_n is a vector space over \mathcal{F} .

A system of linear equations in x_1, x_2, \dots, x_n with coefficients in \mathcal{F} such as

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

can be regarded as a set of vectors $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ in \mathbf{Q}_n , where \mathbf{v}_i is the i^{th} equation in the system. A *solution* of the system \mathcal{A} is a set of values for x_1, x_2, \dots, x_n that satisfies each equation in \mathcal{A} . The set of all solutions is the *solution set* of \mathcal{A} . In some

cases, it is more convenient to write these solutions in vector form as $\mathbf{v} = (x_1, x_2, \dots, x_n)$, while in others, the matrix form

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is more convenient to use.

Theorem 4.11 *With the notation of the preceding paragraph, let $\langle \mathcal{A} \rangle$ denote the subspace of \mathbf{Q}_n that is spanned by \mathcal{A} . Then each solution of the system \mathcal{A} is a solution of every equation in the subspace $\langle \mathcal{A} \rangle$.*

Proof. Any equation \mathbf{v} in $\langle \mathcal{A} \rangle$ is a linear combination of the equations in \mathcal{A} , $\mathbf{v} = \sum_{i=1}^m c_i \mathbf{v}_i$. Thus, \mathbf{v} has the form

$$\begin{aligned} & c_1[a_{11}x_1 + \cdots + a_{1n}x_n = b_1] \\ & + c_2[a_{21}x_1 + \cdots + a_{2n}x_n = b_2] \\ & + \cdots \\ & + c_m[a_{m1}x_1 + \cdots + a_{mn}x_n = b_m] \end{aligned}$$

or

$$\left(\sum_{j=1}^m c_j a_{j1} \right) x_1 + \left(\sum_{j=1}^m c_j a_{j2} \right) x_2 + \cdots + \left(\sum_{j=1}^m c_j a_{jn} \right) x_n = \left(\sum_{j=1}^m c_j b_j \right).$$

If $x_1 = d_1, x_2 = d_2, \dots, x_n = d_n$ is a solution of the system \mathcal{A} , then

$$a_{i1}d_1 + a_{i2}d_2 + \cdots + a_{in}d_n = b_i \text{ for } i = 1, 2, \dots, m.$$

Therefore

$$\begin{aligned} & \left(\sum_{j=1}^m c_j a_{j1} \right) d_1 + \left(\sum_{j=1}^m c_j a_{j2} \right) d_2 + \cdots + \left(\sum_{j=1}^m c_j a_{jn} \right) d_n \\ & = c_1(a_{11}d_1 + \cdots + a_{1n}d_n) + c_2(a_{21}d_1 + \cdots + a_{2n}d_n) + \cdots \\ & \quad + c_m(a_{m1}d_1 + \cdots + a_{mn}d_n) \\ & = c_1b_1 + c_2b_2 + \cdots + c_mb_m, \end{aligned}$$

and $x_1 = d_1, x_2 = d_2, \dots, x_n = d_n$ is a solution of \mathbf{v} . ■

Definition 4.12 *Two systems \mathcal{A} and \mathcal{B} contained in \mathbf{Q}_n are equivalent if and only if they have the same solutions.*

The concept of elementary operations on sets of vectors applies in \mathbf{Q}_n and is a useful tool in solving systems of equations.

Theorem 4.13 *Let \mathcal{A} be a system of m equations contained in \mathbf{Q}_n . If the set \mathcal{B} in \mathbf{Q}_n is obtained from \mathcal{A} by a sequence of elementary operations, then \mathcal{B} and \mathcal{A} are equivalent systems of equations.*

Proof. Suppose that \mathcal{B} is obtained from \mathcal{A} by a sequence of elementary operations. Then $\mathcal{B} \subseteq \langle \mathcal{A} \rangle$, so that every solution of \mathcal{A} is a solution of \mathcal{B} by Theorem 4.11.

But since \mathcal{B} is obtained from \mathcal{A} by a sequence of elementary operations, \mathcal{A} can be obtained from \mathcal{B} by a sequence of elementary operations (Theorem 2.2). Therefore, $\mathcal{A} \subseteq \langle \mathcal{B} \rangle$, and every solution of \mathcal{B} is a solution of \mathcal{A} . This completes the proof. ■

Matrices are valuable tools in the solution of systems of linear equations. Consider the system \mathcal{A} given by

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

This system can be written compactly as a single matrix equation $AX = B$, where

$$A = [a_{ij}]_{m \times n}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Example 1 □ The system of linear equations

$$\begin{aligned} x_1 + 3x_2 + 2x_3 - x_5 &= 7 \\ 2x_1 + 6x_2 + 5x_3 + 6x_4 + 4x_5 &= 0 \\ x_1 + 3x_2 + 2x_3 + 2x_4 + x_5 &= 9 \end{aligned}$$

is equivalent to the single matrix equation

$$\begin{bmatrix} x_1 + 3x_2 + 2x_3 - x_5 \\ 2x_1 + 6x_2 + 5x_3 + 6x_4 + 4x_5 \\ x_1 + 3x_2 + 2x_3 + 2x_4 + x_5 \end{bmatrix} = \begin{bmatrix} 7 \\ 0 \\ 9 \end{bmatrix},$$

and this matrix equation can be written in factored form as

$$\begin{bmatrix} 1 & 3 & 2 & 0 & -1 \\ 2 & 6 & 5 & 6 & 4 \\ 1 & 3 & 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 7 \\ 0 \\ 9 \end{bmatrix}.$$

Thus we have a matrix equation of the form $AX = B$, with

$$A = \begin{bmatrix} 1 & 3 & 2 & 0 & -1 \\ 2 & 6 & 5 & 6 & 4 \\ 1 & 3 & 2 & 2 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 7 \\ 0 \\ 9 \end{bmatrix}. \blacksquare$$

Definition 4.14 In the system $AX = B$, the matrix A is called the **coefficient matrix**, X is the **matrix of unknowns**, and B is the **matrix of constants**. The matrix

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

is called the **augmented matrix** of the system.

Each system in a certain set of variables such as \mathcal{A} above has a unique augmented matrix. That is, each system has an augmented matrix, and different systems have different augmented matrices. Any elementary operation performed on \mathcal{A} is reflected in the augmented matrix as an elementary row operation, and any elementary row operation on the augmented matrix produces a corresponding elementary operation on \mathcal{A} .

Theorem 4.15 Two systems $AX = B$ and $A'X = B'$ of m linear equations in n unknowns x_1, x_2, \dots, x_n are equivalent systems if their augmented matrices $[A, B]$ and $[A', B']$ are row-equivalent.

Proof. Let \mathcal{A} and \mathcal{A}' be the sets of vectors in \mathbf{Q}_n that consist of the equations in the systems $AX = B$ and $A'X = B'$, respectively.

If the augmented matrices $[A, B]$ and $[A', B']$ are row-equivalent, then $[A', B']$ can be obtained from $[A, B]$ by a sequence of elementary row operations. Hence A' can be obtained from A by a sequence of elementary operations and A and A' have the same solutions, by Theorem 4.13. ■

Corollary 4.16 *Let A be $m \times n$ in the system $AX = B$, and let P be any invertible $m \times m$ matrix. Then $AX = B$ and $PAX = PB$ have the same solutions.*

Proof. This follows from the fact that the augmented matrices $[A, B]$ and $[PA, PB] = P[A, B]$ are row-equivalent. ■

Theorem 4.17 *The system $AX = B$ has a solution if and only if*

$$\text{rank}([A, B]) = \text{rank}(A).$$

Proof. The system $AX = B$ can be rewritten in the form

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

Thus the system has a solution if and only if there are scalars x_1, x_2, \dots, x_n in \mathcal{F} such that

$$\begin{aligned} x_1(a_{11}, a_{21}, \dots, a_{m1}) + x_2(a_{12}, a_{22}, \dots, a_{m2}) + \cdots \\ + x_n(a_{1n}, a_{2n}, \dots, a_{mn}) = (b_1, b_2, \dots, b_m). \end{aligned}$$

Let \mathbf{S} denote the column space of A , let \mathbf{S}^* denote the column space of $[A, B]$, and let $\mathbf{b} = (b_1, b_2, \dots, b_m) \in \mathcal{F}^m$. According to the last statement of the preceding paragraph, $AX = B$ has a solution if and only if \mathbf{b} is in \mathbf{S} . But \mathbf{b} is in \mathbf{S} if and only if \mathbf{S} and \mathbf{S}^* have the same dimension, i.e., if and only if A and $[A, B]$ have the same rank. ■

Theorem 4.18 *If A is an $m \times n$ matrix over \mathcal{F} and $\text{rank}([A, B]) = \text{rank}(A) = r$, then the solutions to $AX = B$ can be expressed in terms of $n - r$ parameters.*

Proof. Let P be an invertible matrix such that $PA = A'$ is in reduced row-echelon form. The original system is equivalent to $A'X = B'$ where $PB = B'$ (Theorem 4.15).

By Corollary 3.49, A' has r nonzero rows. Thus the system $A'X = B'$ has the form

$$\begin{aligned}
 x_{k_1} + a'_{1,k_1+1}x_{k_1+1} + \cdots + a'_{1,k_2-1}x_{k_2-1} + \cdots \\
 + a'_{1,k_2+1}x_{k_2+1} + \cdots + a'_{1,n}x_n &= b'_1 \\
 x_{k_2} + a'_{2,k_2+1}x_{k_2+1} + \cdots + a'_{2,n}x_n &= b'_2 \\
 &\vdots \\
 x_{k_r} + \cdots + a'_{r,n}x_n &= b'_r \\
 0 &= b'_{r+1} \\
 &\vdots \\
 0 &= b'_m.
 \end{aligned}$$

In this system each variable $x_{k_i}, i = 1, 2, \dots, r$ occurs just once in the i^{th} equation with a coefficient 1. Hence each of x_{k_1}, \dots, x_{k_r} can be expressed in terms of the remaining $n - r$ variables. ■

The variables $x_{k_1}, x_{k_2}, \dots, x_{k_r}$ in the last paragraph are called the **leading variables**, and the remaining $n - r$ variables are called the **parameters** in the solution of the system.

The proof of Theorem 4.18 furnishes at one and the same time a method for determining the existence of solutions and a method for obtaining them. To solve the system $AX = B$, we can use elementary row operations to transform the augmented matrix $[A, B]$ into reduced row-echelon form $[A', B']$. The condition that $\text{rank}([A, B]) = \text{rank}(A)$ is reflected in the conditions $0 = b'_{r+1}, \dots, 0 = b'_m$ since

$$\text{rank}(A) = \text{rank}(PA) = \text{rank}(A')$$

and

$$\text{rank}([A, B]) = \text{rank}(P[A, B]) = \text{rank}([A', B']).$$

If $\text{rank}([A, B]) > \text{rank}(A)$, then $\text{rank}([A', B']) > r$ and at least one of the equations $0 = b'_{r+1}, \dots, 0 = b'_m$ will be contradictory. If $\text{rank}([A, B]) = \text{rank}(A)$, then there are solutions, and they can be obtained by solving for the leading variables in terms of the parameters. This method of solution is called **Gauss-Jordan elimination**. A system is solved by Gauss-Jordan elimination in Example 1.

Example 2 □ Consider the system of equations

$$\begin{aligned}
 x_1 + 2x_2 + x_3 - 4x_4 + x_5 &= 1 \\
 x_1 + 2x_2 - x_3 + 2x_4 + x_5 &= 5 \\
 2x_1 + 4x_2 + x_3 - 5x_4 &= 2 \\
 x_1 + 2x_2 + 3x_3 - 10x_4 + x_5 &= -3.
 \end{aligned}$$

The augmented matrix $[A, B]$ can be transformed to reduced row-echelon form as follows.

$$\begin{aligned}
 [A, B] &= \left[\begin{array}{cccccc} 1 & 2 & 1 & -4 & 1 & 1 \\ 1 & 2 & -1 & 2 & 1 & 5 \\ 2 & 4 & 1 & -5 & 0 & 2 \\ 1 & 2 & 3 & -10 & 1 & -3 \end{array} \right] \rightarrow \left[\begin{array}{cccccc} 1 & 2 & 1 & -4 & 1 & 1 \\ 0 & 0 & -2 & 6 & 0 & 4 \\ 0 & 0 & -1 & 3 & -2 & 0 \\ 0 & 0 & 2 & -6 & 0 & -4 \end{array} \right] \\
 &\rightarrow \left[\begin{array}{cccccc} 1 & 2 & 0 & -1 & 1 & 3 \\ 0 & 0 & 1 & -3 & 0 & -2 \\ 0 & 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \rightarrow \left[\begin{array}{cccccc} 1 & 2 & 0 & -1 & 0 & 2 \\ 0 & 0 & 1 & -3 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] = [A', B']
 \end{aligned}$$

The reduced row-echelon form $[A', B']$ corresponds to the system

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

When we solve for the leading variables x_1, x_3, x_5 in terms of the parameters x_2, x_4 , we obtain the solutions to the system:

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

where x_2 and x_4 are arbitrary. ■

There is another point of view to Theorem 4.18 and its proof that is useful in Chapter 5. The fifth set in Example 1 of Section 1.3 generalizes immediately from \mathbf{R}^n to \mathcal{F}^n . That is, the set of all vectors $\mathbf{v} = (x_1, x_2, \dots, x_n)$ in \mathcal{F}^n with components x_i that satisfy a given system of equations $A\mathbf{X} = \mathbf{0}$ is a subspace \mathbf{W} of \mathcal{F}^n . The $n - r$ parameters in the proof of Theorem 4.18 represent the components of \mathbf{v} that can be assigned values arbitrarily. We can solve for the leading variables $x_{k_1}, x_{k_2}, \dots, x_{k_r}$ in the equation $A'\mathbf{X} = \mathbf{0}$ and express them in terms of the $n - r$ parameters. If we then replace these leading variables by their values in terms of the parameters, the vector (x_1, x_2, \dots, x_n) that represents the general solution of the system can be obtained as a linear combination of the vectors in a basis of \mathbf{W} . There will be $n - r$ vectors in this basis since there are $n - r$ parameters present. This is illustrated in the following example.

Example 3 □ Consider the system of equations

$$\begin{aligned}x_1 + 2x_2 + x_3 - 4x_4 + x_5 &= 0 \\x_1 + 2x_2 - x_3 + 2x_4 + x_5 &= 0 \\2x_1 + 4x_2 + x_3 - 5x_4 &= 0 \\x_1 + 2x_2 + 3x_3 - 10x_4 + x_5 &= 0.\end{aligned}$$

The coefficient matrix in this system is the same as the one in Example 1, and the reduced row-echelon form for this system can be obtained simply by replacing the constants B' in $[A', B']$ by a column of zeros to obtain $[A', \mathbf{0}]$. This gives

$$[A', \mathbf{0}] = \left[\begin{array}{cccccc} 1 & 2 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

which corresponds to the system

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

Solving for the leading variables, we get

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

Replacing the leading variables by their values in terms of the parameters yields

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

Thus the subspace \mathbf{W} of all solutions to $AX = \mathbf{0}$ is given by

$$\mathbf{W} = \langle (-2, 1, 0, 0, 0), (1, 0, 3, 1, 0) \rangle,$$

and $\{(-2, 1, 0, 0, 0), (1, 0, 3, 1, 0)\}$ is a basis for the solution space \mathbf{W} . As a final remark concerning these solutions, we note that the matrix form

$$X = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ 3 \\ 1 \\ 0 \end{bmatrix}$$

can be easily predicted by listing *all* the variables in a column with the *leading* variables expressed in terms of the parameters:

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

Exercises 4.7

In Problems 1-14, find all solutions of the given system of equations.

- | | |
|-----------------------------------|------------------------------------|
| 1. $x_1 + 2x_2 + x_3 = 5$ | 2. $x_1 + x_2 - x_3 = 4$ |
| $-x_1 - x_2 + x_3 = 2$ | $3x_1 + 3x_2 - 2x_3 = 11$ |
| $x_2 + 3x_3 = 1$ | $4x_1 + 5x_2 - 3x_3 = 17$ |
| 3. $2x_1 + x_3 = 1$ | 4. $x_1 - x_2 - x_3 = 0$ |
| $x_1 + x_2 = -2$ | $2x_1 - x_3 = 1$ |
| $2x_1 + 4x_2 - x_3 = -3$ | $3x_1 + x_2 - x_3 = 2$ |
| 5. $2x_1 + x_2 = -4$ | 6. $9x_1 - 6x_2 = 15$ |
| $x_1 - x_2 = 4$ | $15x_1 - 10x_2 = 25$ |
| $-3x_1 + 3x_2 = 2$ | $6x_1 - 4x_2 = 10$ |
| 7. $x_1 + 2x_2 + 5x_3 = 0$ | 8. $4x_1 + 4x_2 - 7x_3 + 3x_4 = 0$ |
| $4x_1 + 12x_2 + 21x_3 + 2x_4 = 0$ | $3x_1 + 3x_2 - 5x_3 + 2x_4 = 0$ |
| $3x_1 + 6x_2 + 15x_3 - 3x_4 = 0$ | |

9. $x_1 - 2x_2 - 2x_3 - 3x_4 = 1$

$$2x_1 - 4x_2 + 2x_3 = 2$$

$$3x_1 - 6x_2 + x_3 - 2x_4 = 3$$

$$x_3 + x_4 = 0$$

10. $4x_1 + 3x_2 + 2x_3 - x_4 = 4$

$$5x_1 + 4x_2 + 3x_3 - x_4 = 4$$

$$-2x_1 - 2x_2 - x_3 + 2x_4 = -3$$

$$11x_1 + 6x_2 + 4x_3 + x_4 = 11$$

11. $x_1 + 3x_2 - x_3 + 2x_5 = 2$

$$2x_1 + 6x_2 + x_3 + 6x_4 + 4x_5 = 13$$

$$-x_1 - 3x_2 - 2x_4 - 2x_5 = -5$$

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

$$x_1 + x_2 + 2x_3 + 3x_4 + 4x_5 = 4$$

$$2x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 6$$

13. $x_1 + 2x_2 + x_4 = 1$

$$2x_1 + 4x_2 + x_3 + 4x_4 + 3x_5 = 6$$

$$x_1 + 2x_2 + 2x_3 + 4x_4 - 2x_5 = 1$$

$$-x_1 - 2x_2 + 3x_3 + 5x_4 + 4x_5 = 6$$

14. $x_1 + x_2 + 3x_3 + 2x_4 = 0$

$$x_1 - x_2 + x_3 = 2$$

$$2x_1 + 2x_2 + 6x_3 + 4x_4 = 0$$

$$3x_1 + 6x_3 + 3x_4 - x_5 = 3$$

In Problems 15 and 16, (a) find the rank of the coefficient matrix, (b) find the rank of the augmented matrix, and (c) determine if the system is consistent by comparing these ranks. It is not necessary to solve the systems.

15. $x_1 + 2x_2 - x_3 = 3$

$$x_2 + x_3 = 1$$

$$x_1 - 3x_3 = 0$$

16. $x_1 + 2x_2 + 3x_3 + x_4 = 0$

$$x_2 + x_3 + x_4 = 0$$

$$x_1 + x_3 - x_4 = 1$$

17. For the following matrix A , (a) find a basis of the column space of A , and (b) express each column vector of A as a linear combination of the vectors in your basis.

$$A = \begin{bmatrix} 3 & 2 & -2 & 7 \\ 6 & 4 & -4 & 14 \\ 1 & 1 & 2 & 3 \\ -5 & -4 & -2 & -13 \\ 11 & 8 & -2 & 27 \end{bmatrix}$$

18. Find all real numbers a and b for which the system of equations below does not have a solution.

$$x_1 + x_3 = 1$$

$$ax_1 + x_2 + 2x_3 = 0$$

$$3x_1 + 4x_2 + bx_3 = 2$$

In Problems 19 and 20, find the values of the real number a for which the given system (a) has no solution, (b) has exactly one solution, (c) has infinitely many solutions.

19. $x_1 + 2x_2 - x_3 = 2$

$$2x_1 + 6x_2 + 3x_3 = 4$$

$$3x_1 + 8x_2 + (a^2 - 2)x_3 = a + 8$$

20. $x_1 + x_2 + x_3 = 3$

$$2x_1 + 3x_2 + 3x_3 = 8$$

$$3x_1 + 3x_2 + (a^2 - 6)x_3 = a + 6$$

21. Let $\mathcal{B} = \{[x_1 = 0], [x_2 = 0], \dots, [x_n = 0], [x_1 = 1]\}$.

(a) Prove that \mathcal{B} is a basis of \mathbf{Q}_n .

(b) What is the dimension of \mathbf{Q}_n over \mathcal{F} ?

22. For each of Problems 1-14, let A be the coefficient matrix of the given system. Let \mathbf{W} be the subspace consisting of all (x_1, x_2, \dots, x_n) in \mathbf{R}^n that satisfy the system $AX = \mathbf{0}$. Find a basis of \mathbf{W} in each case.
23. Let \mathbf{W} be the subspace of all (x_1, x_2, \dots, x_n) in \mathcal{F}^n that satisfy the system $AX = \mathbf{0}$, and let $\mathbf{c} = (c_1, c_2, \dots, c_n)$, where $x_1 = c_1, x_2 = c_2, \dots, x_n = c_n$ is a particular solution to the system $AX = B$. Prove that $\mathbf{c} + \mathbf{W}$ is the complete set of solutions to $AX = B$.

Chapter 5

Linear Transformations

5.1 Introduction

In this chapter, the important concept of a linear transformation of a vector space is introduced. Matrices prove to be a powerful tool in the study of linear transformations of finite-dimensional vector spaces. They can be used to classify linear transformations according to certain equivalence relations that are based on fundamental properties common to different linear transformations.

5.2 Linear Transformations

Suppose that f is a mapping of \mathcal{S} into \mathcal{T} , and let \mathcal{A} be an arbitrary subset of \mathcal{S} . The set

$$f(\mathcal{A}) = \{t \mid t = f(s) \text{ for some } s \in \mathcal{A}\}$$

is called the **image** of \mathcal{A} under f . Thus, f is a surjective mapping of \mathcal{S} into \mathcal{T} if and only if $f(\mathcal{S}) = \mathcal{T}$. For any subset \mathcal{B} of \mathcal{T} , the set

$$f^{-1}(\mathcal{B}) = \{s \in \mathcal{S} \mid f(s) \in \mathcal{B}\}$$

is called the **inverse image** of \mathcal{B} . In particular, if \mathcal{B} consists of a single element t ,

$$f^{-1}(t) = \{s \in \mathcal{S} \mid f(s) = t\}.$$

Thus, f is injective if, for every $t \in f(\mathcal{S})$, $f^{-1}(t)$ consists of exactly one element.

We write $f : \mathcal{S} \rightarrow \mathcal{T}$ or $\mathcal{S} \xrightarrow{f} \mathcal{T}$ to indicate that f is a mapping of \mathcal{S} into \mathcal{T} . Our interest in this chapter is with those mappings of one vector space into another. Throughout the chapter, \mathbf{U} and \mathbf{V} will denote vector spaces over the *same* field \mathcal{F} .

Definition 5.1 A linear transformation T is a mapping $T : \mathbf{U} \rightarrow \mathbf{V}$ which has the property that

$$T(a\mathbf{u} + b\mathbf{w}) = aT(\mathbf{u}) + bT(\mathbf{w})$$

for all \mathbf{u}, \mathbf{w} in \mathbf{U} and all a, b in \mathcal{F} .

We recall from Section 4.4 that an isomorphism is a bijective mapping f that has the property $f(a\mathbf{u} + b\mathbf{w}) = af(\mathbf{u}) + bf(\mathbf{w})$ required in Definition 5.1. Hence every isomorphism is a linear transformation. However, a linear transformation of \mathbf{U} into \mathbf{V} may be neither injective nor surjective, even though it preserves linear combinations just as an isomorphism does.

In addition to the isomorphisms, another family of examples of linear transformations is provided by the *zero transformations*. For a given pair of vector spaces \mathbf{U} and \mathbf{V} , the **zero linear transformation** is the mapping $Z : \mathbf{U} \rightarrow \mathbf{V}$ defined by $Z(\mathbf{u}) = \mathbf{0}$ for all \mathbf{u} in \mathbf{U} .

The following examples provide some more detailed illustrations concerning linear transformations.

Example 1 □ Consider the mapping $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ defined by¹

$$T(x, y) = (4x + 5y, 6x - y).$$

For arbitrary $\mathbf{u} = (u_1, u_2), \mathbf{w} = (w_1, w_2)$ in \mathbf{R}^2 and arbitrary a, b in \mathbf{R} , we have

$$\begin{aligned} T(a\mathbf{u} + b\mathbf{w}) &= T(au_1 + bw_1, au_2 + bw_2) \\ &= (4(au_1 + bw_1) + 5(au_2 + bw_2), 6(au_1 + bw_1) - (au_2 + bw_2)) \\ &= (4au_1 + 5au_2, 6au_1 - au_2) + (4bw_1 + 5bw_2, 6bw_1 - bw_2) \\ &= aT(\mathbf{u}) + bT(\mathbf{w}), \end{aligned}$$

and therefore T is a linear transformation of \mathbf{R}^2 into \mathbf{R}^2 . ■

A linear transformation of a vector space \mathbf{V} into itself is called a **linear operator**. The mapping T in Example 1 provides an example of a linear operator.

Example 2 □ Let \mathcal{P}_n denote the vector space that consists of all polynomials in x with degree n or less and coefficients in \mathbf{R} , and consider the mapping $T : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ defined by

$$T(p(x)) = (1+x)p(x)$$

for all $p(x)$ in \mathcal{P}_1 . A specific computation of a value of T is provided by

$$\begin{aligned} T(3 + 2x) &= (1+x)(3+2x) \\ &= 3 + 5x + 2x^2. \end{aligned}$$

For arbitrary $p(x), q(x)$ in \mathcal{P}_1 and arbitrary a, b in \mathbf{R} , the polynomials $T(p(x))$ and $T(q(x))$ are in \mathcal{P}_2 and

$$\begin{aligned} T(ap(x) + bq(x)) &= (1+x)(ap(x) + bq(x)) \\ &= a(1+x)p(x) + b(1+x)q(x) \\ &= aT(p(x)) + bT(q(x)). \end{aligned}$$

¹For notational convenience, the outer parentheses in $T((x_1, x_2, \dots, x_n))$ are dropped for mappings defined on \mathbf{R}^n .

Thus T is a linear transformation of \mathcal{P}_1 into \mathcal{P}_2 . ■

Example 3 □ Let the mapping $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be defined by

$$T(x, y) = (x + 1, 2x + y).$$

For any $\mathbf{u} = (u_1, u_2), \mathbf{w} = (w_1, w_2)$ in \mathbf{R}^2 and scalars a, b in \mathbf{R} ,

$$\begin{aligned} T(a\mathbf{u} + b\mathbf{w}) &= T(au_1 + bw_1, au_2 + bw_2) \\ &= (au_1 + bw_1 + 1, 2au_1 + 2bw_1 + au_2 + bw_2) \end{aligned}$$

and

$$\begin{aligned} aT(\mathbf{u}) + bT(\mathbf{w}) &= a(u_1 + 1, 2u_1 + u_2) + b(w_1 + 1, 2w_1 + w_2) \\ &= (au_1 + bw_1 + a + b, 2au_1 + au_2 + 2bw_1 + bw_2). \end{aligned}$$

We see that the equality

$$T(a\mathbf{u} + b\mathbf{w}) = aT(\mathbf{u}) + bT(\mathbf{w})$$

holds if and only if $a + b = 1$. Since this equation is not always true for a and b in \mathbf{R} , we conclude that T is *not* a linear transformation. ■

In order to indicate some of the variety in linear transformations, we consider some more examples.

Example 4 □ Let $A = [a_{ij}]_{m \times n}$ be a fixed (that is, constant) matrix over \mathbf{R} and define $T : \mathbf{R}_{n \times 1} \rightarrow \mathbf{R}_{m \times 1}$ by

$$T(X) = AX$$

for all

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

in $\mathbf{R}_{n \times 1}$. Let X, Y be arbitrary vectors in $\mathbf{R}_{n \times 1}$, and let a, b be arbitrary real numbers. We have both $T(X) = AX$ and $T(Y) = AY$ in $\mathbf{R}_{m \times 1}$, and

$$\begin{aligned} T(aX + bY) &= A(aX + bY) \\ &= A(aX) + A(bY) \\ &= aAX + bAY \\ &= aT(X) + bT(Y). \end{aligned}$$

Thus T is a linear transformation of $\mathbf{R}_{n \times 1}$ into $\mathbf{R}_{m \times 1}$. This type of linear transformation is called a **matrix transformation**. ■

Example 5 □ We saw in Example 4 and Problem 15 of Section 4.2 that the set \mathbf{V} of all real-valued functions of t with domain \mathbf{R} and the set \mathbf{W} of all differentiable functions of t with domain \mathbf{R} form vector spaces over \mathbf{R} with respect to the usual operations of addition and scalar multiplication. Consider the mapping $T : \mathbf{W} \rightarrow \mathbf{V}$ defined by

$$T(f(t)) = \frac{df}{dt} = f'(t).$$

That is, T maps each differentiable function onto its derivative. Using familiar facts from the calculus, we get

$$\begin{aligned} T(af(t) + bg(t)) &= \frac{d}{dt}(af(t) + bg(t)) \\ &= \frac{d}{dt}(af(t)) + \frac{d}{dt}(bg(t)) \\ &= a\frac{df}{dt} + b\frac{dg}{dt} \\ &= aT(f(t)) + bT(g(t)). \end{aligned}$$

Thus the process of differentiating functions in \mathbf{W} is a linear transformation from \mathbf{W} to \mathbf{V} . ■

In this chapter, we study linear transformations as mathematical quantities themselves. In our study, we need to define three operations on them. Addition and multiplication by a scalar are considered first.

Definition 5.2 If S and T are linear transformations of \mathbf{U} into \mathbf{V} , the sum $S + T$ is defined by

$$(S + T)(\mathbf{u}) = S(\mathbf{u}) + T(\mathbf{u})$$

for all \mathbf{u} in \mathbf{U} . Also, for each linear transformation T of \mathbf{U} into \mathbf{V} and each scalar a in \mathcal{F} , we define the product of a and T to be the mapping aT of \mathbf{U} into \mathbf{V} given by

$$(aT)(\mathbf{u}) = a(T(\mathbf{u})).$$

Example 6 □ Let $\mathbf{U} = \mathbf{R}^3$ and $\mathbf{V} = \mathbf{R}^2$, and let S and T be defined by

$$\begin{aligned} S(x_1, x_2, x_3) &= (2x_1 + x_3, 3x_1 - x_2), \\ T(x_1, x_2, x_3) &= (-x_1 + 3x_2, 2x_1 - x_2 + x_3). \end{aligned}$$

It is easy to verify that S and T are linear transformations. For example, if $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{w} = (w_1, w_2, w_3)$ and a, b are in \mathbf{R} , then

$$\begin{aligned} S(a\mathbf{u} + b\mathbf{w}) &= S(au_1 + bw_1, au_2 + bw_2, au_3 + bw_3) \\ &= (2(au_1 + bw_1) + au_3 + bw_3, 3(au_1 + bw_1) - (au_2 + bw_2)) \\ &= (2au_1 + au_3, 3au_1 - au_2) + (2bw_1 + bw_3, 3bw_1 - bw_2) \\ &= a(2u_1 + u_3, 3u_1 - u_2) + b(2w_1 + w_3, 3w_1 - w_2) \\ &= aS(\mathbf{u}) + bS(\mathbf{w}). \end{aligned}$$

The sum $S + T$ is given by

$$\begin{aligned}(S + T)(x_1, x_2, x_3) &= (2x_1 + x_3, 3x_1 - x_2) + (-x_1 + 3x_2, 2x_1 - x_2 + x_3) \\ &= (x_1 + 3x_2 + x_3, 5x_1 - 2x_2 + x_3),\end{aligned}$$

and the product aT is given by

$$\begin{aligned}(aT)(x_1, x_2, x_3) &= a(-x_1 + 3x_2, 2x_1 - x_2 + x_3) \\ &= (-ax_1 + 3ax_2, 2ax_1 - ax_2 + ax_3). \blacksquare\end{aligned}$$

With the definitions of addition and scalar multiplication given in Definition 5.2, the linear transformations of \mathbf{U} into \mathbf{V} can be regarded as possible vectors. The next theorem shows that they are indeed vectors.

Theorem 5.3 *Let \mathbf{U} and \mathbf{V} be vector spaces over the same field \mathcal{F} . Then the set of all linear transformations of \mathbf{U} into \mathbf{V} is a vector space² over \mathcal{F} .*

Proof. For a complete proof, each of the ten conditions of Definition 4.2 must be verified. We verify the first six here, leaving the others as exercises. Let T_1, T_2 , and T_3 denote arbitrary linear transformation of \mathbf{U} into \mathbf{V} , let \mathbf{u} and \mathbf{w} be arbitrary vectors in \mathbf{U} , and let a, b , and c be scalars.

Since

$$\begin{aligned}(T_1 + T_2)(a\mathbf{u} + b\mathbf{w}) &= T_1(a\mathbf{u} + b\mathbf{w}) + T_2(a\mathbf{u} + b\mathbf{w}) \\ &= aT_1(\mathbf{u}) + bT_1(\mathbf{w}) + aT_2(\mathbf{u}) + bT_2(\mathbf{w}) \\ &= a[T_1(\mathbf{u}) + T_2(\mathbf{u})] + b[T_1(\mathbf{w}) + T_2(\mathbf{w})] \\ &= a(T_1 + T_2)(\mathbf{u}) + b(T_1 + T_2)(\mathbf{w}),\end{aligned}$$

$T_1 + T_2$ is a linear transformation of \mathbf{U} into \mathbf{V} .

Addition is associative, since

$$\begin{aligned}(T_1 + (T_2 + T_3))(\mathbf{u}) &= T_1(\mathbf{u}) + (T_2 + T_3)(\mathbf{u}) \\ &= T_1(\mathbf{u}) + [T_2(\mathbf{u}) + T_3(\mathbf{u})] \\ &= [T_1(\mathbf{u}) + T_2(\mathbf{u})] + T_3(\mathbf{u}) \\ &= (T_1 + T_2)(\mathbf{u}) + T_3(\mathbf{u}) \\ &= ((T_1 + T_2) + T_3)(\mathbf{u}).\end{aligned}$$

The zero linear transformation Z is an additive identity, since

$$(T_1 + Z)(\mathbf{u}) = T_1(\mathbf{u}) + Z(\mathbf{u}) = T_1(\mathbf{u}) + \mathbf{0} = T_1(\mathbf{u})$$

²The notation $L(\mathbf{U}, \mathbf{V})$ is often used to denote this vector space.

for all \mathbf{u} in \mathbf{U} .

The additive inverse of T_1 is the linear transformation $-T_1$ of \mathbf{U} into \mathbf{V} defined by $(-T_1)(\mathbf{u}) = -T_1(\mathbf{u})$ since

$$(T_1 + (-T_1))(\mathbf{u}) = T_1(\mathbf{u}) + (-T_1(\mathbf{u})) = \mathbf{0}$$

for all \mathbf{u} in \mathbf{U} .

For any \mathbf{u} in \mathbf{U} ,

$$(T_1 + T_2)(\mathbf{u}) = T_1(\mathbf{u}) + T_2(\mathbf{u}) = T_2(\mathbf{u}) + T_1(\mathbf{u}) = (T_2 + T_1)(\mathbf{u}),$$

so $T_1 + T_2 = T_2 + T_1$.

Since

$$\begin{aligned} (cT_1)(a\mathbf{u} + b\mathbf{w}) &= c(T_1(a\mathbf{u} + b\mathbf{w})) \\ &= c(aT_1(\mathbf{u}) + bT_1(\mathbf{w})) \\ &= a(cT_1(\mathbf{u})) + b(cT_1(\mathbf{w})) \end{aligned}$$

cT_1 is a linear transformation of \mathbf{U} into \mathbf{V} . ■

There is a third operation involving transformations that we will consider later.

Theorem 5.4 *Let T be a linear transformation of \mathbf{U} into \mathbf{V} . If \mathbf{U}_1 is any subspace of \mathbf{U} , then $T(\mathbf{U}_1)$ is a subspace of \mathbf{V} .*

Proof. Let \mathbf{U}_1 be a subspace of \mathbf{U} . Then $\mathbf{0} \in \mathbf{U}_1$ and $T(\mathbf{0}) = \mathbf{0}$, so $T(\mathbf{U}_1)$ is nonempty.

Let $\mathbf{v}_1, \mathbf{v}_2$ be in $T(\mathbf{U}_1)$, and a_1, a_2 be in \mathcal{F} . There exist vectors $\mathbf{u}_1, \mathbf{u}_2$ in \mathbf{U}_1 such that $T(\mathbf{u}_1) = \mathbf{v}_1$ and $T(\mathbf{u}_2) = \mathbf{v}_2$. Since \mathbf{U}_1 is a subspace of \mathbf{U} , $a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ is in \mathbf{U}_1 and

$$\begin{aligned} T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2) &= a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) \\ &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2. \end{aligned}$$

Thus, $a_1\mathbf{v}_1 + a_2\mathbf{v}_2$ is in $T(\mathbf{U}_1)$, and $T(\mathbf{U}_1)$ is a subspace of \mathbf{V} . ■

Definition 5.5 *The subspace $T(\mathbf{U})$ of \mathbf{V} is called the **range** of T . The dimension of $T(\mathbf{U})$ is called the **rank** of T . The rank of T will be denoted by $\text{rank}(T)$.*

Near the end of this section we will devise a method for finding the rank of T when \mathbf{U} and \mathbf{V} are finite-dimensional.

Theorem 5.6 *Let T be a linear transformation of \mathbf{U} into \mathbf{V} . If \mathbf{W} is any subspace of \mathbf{V} , the inverse image $T^{-1}(\mathbf{W})$ is a subspace of \mathbf{U} .*

Proof. Let \mathbf{W} be a subspace of \mathbf{V} . Then $\mathbf{0} \in \mathbf{W}$ and hence $\mathbf{0} \in T^{-1}(\mathbf{W})$ since $T(\mathbf{0}) = \mathbf{0}$. Thus $T^{-1}(\mathbf{W})$ is nonempty. Let $\mathbf{u}_1, \mathbf{u}_2$ be in $T^{-1}(\mathbf{W})$, and let a_1, a_2 be in \mathcal{F} . Then $\mathbf{v}_1 = T(\mathbf{u}_1)$ and $\mathbf{v}_2 = T(\mathbf{u}_2)$ are in \mathbf{W} , and this means that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 = a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) = T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2)$$

is in \mathbf{W} . Therefore, $a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ is in $T^{-1}(\mathbf{W})$, and $T^{-1}(\mathbf{W})$ is a subspace of \mathbf{U} . ■

Definition 5.7 The subspace $T^{-1}(\mathbf{0})$ is called the **kernel** of the linear transformation T . The dimension of $T^{-1}(\mathbf{0})$ is the **nullity** of T , denoted by $\text{nullity}(T)$.

We concentrate our attention now to the case where \mathbf{U} is finite-dimensional.

Theorem 5.8 If T is a linear transformation of \mathbf{U} into \mathbf{V} and $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis of \mathbf{U} , then $T(\mathcal{A})$ spans $T(\mathbf{U})$.

Proof. Suppose that $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis of \mathbf{U} , and consider the set $T(\mathcal{A}) = \{T(\mathbf{u}_1), T(\mathbf{u}_2), \dots, T(\mathbf{u}_n)\}$. For any vector \mathbf{v} in $T(\mathbf{U})$, there is a vector \mathbf{u} in \mathbf{U} such that $T(\mathbf{u}) = \mathbf{v}$. The vector \mathbf{u} can be written as $\mathbf{u} = \sum_{i=1}^n a_i \mathbf{u}_i$ since \mathcal{A} is a basis of \mathbf{U} . This gives $\mathbf{v} = T(\sum_{i=1}^n a_i \mathbf{u}_i) = \sum_{i=1}^n a_i T(\mathbf{u}_i)$, and $T(\mathcal{A})$ spans $T(\mathbf{U})$. ■

The essence of a method for finding the rank of a linear transformation T is contained in the proof of Theorem 5.8. For the set $T(\mathcal{A})$ contains a basis of $T(\mathbf{U})$, and the number of elements in the basis is $\text{rank}(T)$. Our next example demonstrates the use of this method to find a basis for the range of a linear transformation T , and we also find a basis for the kernel of T . More efficient and systematic methods for finding these bases are developed in the next section.

Example 7 □ It is given that the mapping $T : \mathbf{R}_{2 \times 3} \rightarrow \mathbf{R}^5$ defined by

$$T \left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \right) = (a_{11} + a_{21}, 2a_{11} + a_{21} - a_{12}, a_{22} + a_{13}, 0, a_{23})$$

is a linear transformation of $\mathbf{R}_{2 \times 3}$ into \mathbf{R}^5 . We shall (a) find a basis for the range of T , (b) find a basis for the kernel of T , and (c) state the rank and nullity of T .

(a) In order to find a basis for $T(\mathbf{R}_{2 \times 3})$, we first obtain a spanning set $T(\mathcal{A})$ as described in the proof of Theorem 5.8. The set $\mathcal{A} = \{A_1, A_2, A_3, A_4, A_5, A_6\}$ forms a basis of $\mathbf{R}_{2 \times 3}$, where

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The set $T(\mathcal{A})$ is given by

$$\begin{aligned} T(\mathcal{A}) = \{ & (1, 2, 0, 0, 0), (0, -1, 0, 0, 0), (0, 0, 1, 0, 0), (1, 1, 0, 0, 0), \\ & (0, 0, 1, 0, 0), (0, 0, 0, 0, 1) \}, \end{aligned}$$

and we know that $T(\mathcal{A})$ contains a basis of $T(\mathbf{R}_{2 \times 3})$. Using the refinement process from Section 1.5 and Example 3 in Section 4.3, we find that the first three vectors in $T(\mathcal{A})$ are linearly independent and the fourth vector can be written as

$$(1, 1, 0, 0, 0) = (1)(1, 2, 0, 0, 0) + (1)(0, -1, 0, 0, 0) + (0)(0, 0, 1, 0, 0).$$

Thus the fourth vector can be deleted from the spanning set $T(\mathcal{A})$. The fifth vector in $T(\mathcal{A})$ is a repetition of the third vector, so it can also be deleted. The last vector is clearly not a linear combination of the preceding vectors since it is the only one with a nonzero fifth component. Thus

$$\{(1, 2, 0, 0, 0), (0, -1, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 0, 1)\}$$

is a basis for the range of T .

(b) To find a basis for $T^{-1}(\mathbf{0})$, we set $T(\mathbf{v}) = \mathbf{0}$ and obtain the system of equations

$$\begin{aligned} a_{11} + a_{21} &= 0 \\ 2a_{11} + a_{21} - a_{12} &= 0 \\ a_{22} + a_{13} &= 0 \\ 0 &= 0 \\ a_{23} &= 0. \end{aligned}$$

Using Gauss-Jordan elimination, we find that the reduced row-echelon form for the augmented matrix is

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

Solving for the leading variables in terms of the parameters, we get

$$\begin{aligned} a_{11} &= a_{12} \\ a_{21} &= -a_{12} \\ a_{12} &= a_{12} \\ a_{22} &= -a_{13} \\ a_{13} &= a_{13} \\ a_{23} &= 0. \end{aligned}$$

Substituting for the leading variables leads to

$$\begin{aligned} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} &= \begin{bmatrix} a_{12} & a_{12} & a_{13} \\ -a_{12} & -a_{13} & 0 \end{bmatrix} \\ &= a_{12} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} + a_{13} \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}. \end{aligned}$$

Thus the set

$$\left\{ \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \right\}$$

forms a basis for the kernel $T^{-1}(\mathbf{0})$.

(c) From the number of vectors in our bases, we see that

$$\text{rank}(T) = \dim(T(\mathbf{R}_{2 \times 3})) = 4$$

and

$$\text{nullity}(T) = \dim(T^{-1}(\mathbf{0})) = 2. \blacksquare$$

We note that the sum of the rank and nullity of T is equal to the dimension of the domain of T in the example above. Our next theorem states that this equality always holds.

Theorem 5.9 *Let T be a linear transformation of \mathbf{U} into \mathbf{V} . If \mathbf{U} has finite dimension, then*

$$\text{rank}(T) + \text{nullity}(T) = \dim(\mathbf{U}).$$

Proof. Suppose that \mathbf{U} has dimension n , and let k be the nullity of T . Choose $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ to be a basis of the kernel $T^{-1}(\mathbf{0})$. This linearly independent set can be extended to a basis

$$\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$$

of \mathbf{U} . According to Theorem 5.8, the set $T(\mathcal{A})$ spans $T(\mathbf{U})$. But $T(\mathbf{u}_1) = T(\mathbf{u}_2) = \cdots = T(\mathbf{u}_k) = \mathbf{0}$, so this means that the set of $n - k$ vectors $\{T(\mathbf{u}_{k+1}), T(\mathbf{u}_{k+2}), \dots, T(\mathbf{u}_n)\}$ spans $T(\mathbf{U})$. To show that this set is linearly independent, suppose that

$$c_{k+1}T(\mathbf{u}_{k+1}) + c_{k+2}T(\mathbf{u}_{k+2}) + \cdots + c_nT(\mathbf{u}_n) = \mathbf{0}.$$

Then

$$T(c_{k+1}\mathbf{u}_{k+1} + c_{k+2}\mathbf{u}_{k+2} + \cdots + c_n\mathbf{u}_n) = \mathbf{0},$$

and $\sum_{i=k+1}^n c_i\mathbf{u}_i$ is in $T^{-1}(\mathbf{0})$. Thus there are scalars d_1, d_2, \dots, d_k such that $\sum_{i=1}^k d_i\mathbf{u}_i = \sum_{i=k+1}^n c_i\mathbf{u}_i$ and

$$\sum_{i=1}^k d_i\mathbf{u}_i - \sum_{i=k+1}^n c_i\mathbf{u}_i = \mathbf{0}.$$

Since \mathcal{A} is a basis, each c_i and each d_i must be zero. Hence $\{T(\mathbf{u}_{k+1}), T(\mathbf{u}_{k+2}), \dots, T(\mathbf{u}_n)\}$ is a basis of $T(\mathbf{U})$. Since $\text{rank}(T)$ is the dimension of $T(\mathbf{U})$, $n - k = \text{rank}(T)$ and

$$\text{rank}(T) + \text{nullity}(T) = n = \dim(\mathbf{U}). \blacksquare$$

In the next section, a systematic method for finding a basis of $T^{-1}(\mathbf{0})$ is described.

Exercises 5.2

- Let S and T be the mappings of \mathbf{R}^2 into \mathbf{R}^2 defined by $S(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$ and $T(x_1, x_2) = (-x_2, -x_1)$.
 - Prove that each of S and T is a linear operator.
 - Find $S + T$ and $2S - 3T$.
- Determine whether the given mapping $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is a linear operator.
 - $T(x, y) = (x - y, 0)$
 - $T(x, y) = (xy, x)$
 - $T(x, y) = (x + 1, y - 1)$
 - $T(x, y) = (x + y, x - y)$
 - $T(x, y) = x(2, 1)$
 - $T(x, y) = (x - y)(x + y, 0)$
- Determine whether the given mapping $T : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ is a linear transformation.
 - $T(x, y, z) = (2x + y, x + z)$
 - $T(x, y, z) = (x - y, x^2 - y^2)$
 - $T(x, y, z) = (x + y + 1, x + y - 1)$
 - $T(x, y, z) = (x + y + z, 0)$
- Determine which of the following mappings $T : \mathcal{P}_1 \rightarrow \mathcal{P}_1$ are linear operators.
 - $T(a_0 + a_1x) = a_0x$
 - $T(a_0 + a_1x) = a_0 + a_1(x + 1)$
 - $T(a_0 + a_1x) = a_0a_1 + a_0x$
 - $T(a_0 + a_1x) = a_0 + a_1 + a_0x$
- Determine which of the following mappings $T : \mathbf{R}_{2 \times 2} \rightarrow \mathbf{R}_{2 \times 2}$ are linear operators. In parts (a)-(c), A denotes a constant nonzero 2×2 matrix over \mathbf{R} .
 - $T(X) = A + AX$
 - $T(X) = AX - XA$
 - $T(X) = AXA$
 - $T(X) = 2X$

- 6.** Determine whether the given mapping is a linear transformation.
- $T : \mathbf{R}_{m \times n} \rightarrow \mathbf{R}_{n \times m}$, $T(X) = X^T$, where X^T is the transpose of X
 - $T : \mathbf{R}_{m \times n} \rightarrow \mathbf{R}_{n \times n}$, $T(X) = X^T X$, where X^T is the transpose of X
 - $T : \mathbf{R}_{n \times n} \rightarrow \mathbf{R}$, $T(X) = \sum_{i=1}^n x_{ii}$, where $X = [x_{ij}]_{n \times n}$ (The value of this mapping is called the **trace** of X .)
 - $T : \mathcal{P}_2 \rightarrow \mathcal{P}_1$, $T(a_0 + a_1x + a_2x^2) = (a_0 + a_1) + (a_1 + a_2)x$
 - $T : \mathcal{P}_2 \rightarrow \mathcal{P}_2$, $T(p(x)) = p(x - 1)$
 - $T : \mathcal{P}_2 \rightarrow \mathcal{P}_3$, $T(p(x)) = xp(x) + p(1)$
- 7.** Let T be the linear transformation of \mathbf{R}^4 into \mathbf{R}^3 given by
- $$\begin{aligned} T(x_1, x_2, x_3, x_4) \\ = (3x_1 - 2x_2 - x_3 - 4x_4, x_1 + x_2 - 2x_3 - 3x_4, 2x_1 - 3x_2 + x_3 - x_4). \end{aligned}$$
- Find a basis of $T(\mathbf{R}^4)$.
 - Find two linearly independent vectors in the kernel of T .
- 8.** Let T be the mapping of $\mathbf{R}_{2 \times 2}$ into $\mathbf{R}_{2 \times 2}$ defined by
- $$T\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{bmatrix} 2a_{11} + a_{12} & 2a_{21} - a_{22} \\ a_{11} + 3a_{22} & a_{21} - 3a_{12} \end{bmatrix}.$$
- Prove that T is a linear transformation.
 - Find a basis for the range of T .
- 9.** Find the rank and nullity of the given linear transformation of \mathbf{U} into \mathbf{V} .
- $\mathbf{U} = \mathbf{V} = \mathbf{R}^4$,
- $$\begin{aligned} T(x_1, x_2, x_3, x_4) = (x_1 + x_2 + x_3 - x_4, x_1 + 2x_2 + x_3 - x_4, \\ x_1 - x_2 + 3x_3, -x_1 + 5x_2 - 5x_3 - x_4) \end{aligned}$$
- $\mathbf{U} = \mathbf{R}^4$, $\mathbf{V} = \mathbf{R}^3$,
- $$T(x_1, x_2, x_3, x_4) = (x_1 + 2x_2 + x_3 + 3x_4, 2x_3 - 4x_4, x_1 + 2x_2 + 3x_3 - x_4)$$
- 10.** In each part of Problem 9, find the standard basis of the kernel of T by solving the system of equations that results from setting $T(\mathbf{u}) = \mathbf{0}$.
- 11.** Let T be a linear transformation of \mathbf{U} into \mathbf{V} . Prove the following statements.
- $T(\mathbf{0}) = \mathbf{0}$

- (b) $T(-\mathbf{u}) = -T(\mathbf{u})$ for all $\mathbf{u} \in \mathbf{U}$
 (c) $T(\mathbf{u} - \mathbf{w}) = T(\mathbf{u}) - T(\mathbf{w})$ for all \mathbf{u}, \mathbf{w} in \mathbf{U}
 (d) $T\left(\sum_{i=1}^n a_i \mathbf{u}_i\right) = \sum_{i=1}^n a_i T(\mathbf{u}_i)$ for all scalars a_i in \mathcal{F} and vectors \mathbf{u}_i in \mathbf{U}
12. Let T be a linear transformation of \mathbf{U} into \mathbf{V} , and let \mathbf{U}_1 be a subspace of \mathbf{U} . Prove that if \mathcal{A} spans \mathbf{U}_1 , then $T(\mathcal{A})$ spans $T(\mathbf{U}_1)$. (Note: \mathcal{A} is not necessarily finite.)
13. In each part below, T is a linear transformation of \mathbf{R}^4 into \mathbf{R}^3 . For the given subspace \mathbf{U}_1 and the basis \mathcal{B} of \mathbf{R}^3 , find the standard basis of $T(\mathbf{U}_1)$ relative to \mathcal{B} .
- (a) $T(x_1, x_2, x_3, x_4) = (x_1 + 3x_2 + x_4, 3x_1 + 5x_2 - x_3 + 2x_4, 5x_1 + 2x_2 - 2x_3 - 2x_4)$,
 $\mathbf{U}_1 = \langle(1, 0, 1, 0), (1, 0, 2, 0), (0, 1, 0, -1)\rangle, \mathcal{B} = \{(1, 0, 1), (0, 1, 1), (1, 1, 1)\}$
- (b) $T(x_1, x_2, x_3, x_4) = (x_1 - x_3 + 3x_4, x_2 + 2x_4, 2x_1 - x_2 - 2x_3 + 4x_4)$,
 $\mathbf{U}_1 = \langle(2, 0, 1, 1), (-1, 1, 0, 0), (4, 0, -2, 0)\rangle, \mathcal{B} = \{(1, 2, 0), (1, 0, 2), (0, 0, 1)\}$
14. Complete the proof of Theorem 5.3.
15. Let T be a linear transformation of \mathbf{U} into \mathbf{V} with nullity 0. Prove that if $T(\mathbf{u}_4)$ is dependent on $\{T(\mathbf{u}_1), T(\mathbf{u}_2), T(\mathbf{u}_3)\}$, then \mathbf{u}_4 is dependent on $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.
16. Let T be a linear transformation of \mathbf{U} into \mathbf{V} , and let \mathbf{U}_1 and \mathbf{U}_2 be subspaces of \mathbf{U} . Prove that $T(\mathbf{U}_1 + \mathbf{U}_2) = T(\mathbf{U}_1) + T(\mathbf{U}_2)$.
17. Let T be a linear transformation of \mathbf{U} into \mathbf{V} , and let \mathbf{U}_1 and \mathbf{W} denote subspaces of \mathbf{U} and \mathbf{V} , respectively. Prove or disprove the statements below.
- (a) $T(T^{-1}(\mathbf{W})) = \mathbf{W}$ (b) $T^{-1}(T(\mathbf{U}_1)) = \mathbf{U}_1$

5.3 Linear Transformations and Matrices

In this and the following section, \mathbf{U} and \mathbf{V} will denote vector spaces of dimension n and m , respectively, over the same field \mathcal{F} , and T will denote a linear transformation of \mathbf{U} into \mathbf{V} .

Suppose that $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis of \mathbf{U} . Any \mathbf{u} in \mathbf{U} can be written uniquely in the form $\mathbf{u} = \sum_{j=1}^n x_j \mathbf{u}_j$, and

$$T(\mathbf{u}) = \sum_{j=1}^n x_j T(\mathbf{u}_j).$$

This shows that the value of T at every \mathbf{u} in \mathbf{U} is determined by the values of T at the basis vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$. If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ is a basis of \mathbf{V} , then each $T(\mathbf{u}_j)$ can be written uniquely as

$$T(\mathbf{u}_j) = \sum_{i=1}^m a_{ij} \mathbf{v}_i.$$

Thus, with each choice of bases \mathcal{A} and \mathcal{B} , a linear transformation T of \mathbf{U} into \mathbf{V} determines a unique indexed set $\{a_{ij}\}$ of mn elements of \mathcal{F} . These elements make up the matrix of T relative to the bases \mathcal{A} and \mathcal{B} .

Definition 5.10 Suppose that $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ are bases of \mathbf{U} and \mathbf{V} , respectively. Let T be a linear transformation of \mathbf{U} into \mathbf{V} . The **matrix of T relative to the bases \mathcal{A} and \mathcal{B}** is the matrix

$$A = [a_{ij}]_{m \times n} = [T]_{\mathcal{B}, \mathcal{A}}$$

where the a_{ij} are determined by the conditions

$$\begin{aligned} T(\mathbf{u}_j) &= \sum_{i=1}^m a_{ij} \mathbf{v}_i \\ &= a_{1j} \mathbf{v}_1 + a_{2j} \mathbf{v}_2 + \cdots + a_{mj} \mathbf{v}_m \end{aligned}$$

for $j = 1, 2, \dots, n$.

The symbols $A = [a_{ij}]_{m \times n}$ and $[T]_{\mathcal{B}, \mathcal{A}}$ in Definition 5.10 denote the same matrix, but the first one places notational emphasis on the elements of the matrix, while the second one places emphasis on T and the bases \mathcal{A} and \mathcal{B} . This matrix A is also referred to as the **matrix of T with respect to \mathcal{A} and \mathcal{B}** , and we say that T is **represented by the matrix A** .

As mentioned earlier, the elements a_{ij} are uniquely determined by T for given bases \mathcal{A} and \mathcal{B} . Another way to describe A is to observe that the j^{th} column of A is the coordinate matrix of $T(\mathbf{u}_j)$ with respect to \mathcal{B} . That is,

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} = [T(\mathbf{u}_j)]_{\mathcal{B}}$$

and

$$A = [T]_{\mathcal{B}, \mathcal{A}} = [[T(\mathbf{u}_1)]_{\mathcal{B}}, [T(\mathbf{u}_2)]_{\mathcal{B}}, \dots, [T(\mathbf{u}_n)]_{\mathcal{B}}].$$

Example 1 \square With \mathbf{R} as the field of scalars, let $T : \mathcal{P}_2 \rightarrow \mathcal{P}_3$ be the linear transformation defined by

$$T(a_0 + a_1x + a_2x^2) = (2a_0 + 2a_2) + (a_0 + a_1 + 3a_2)x + (a_1 + 2a_2)x^2 + (a_0 + a_2)x^3.$$

We shall find the matrix A of T relative to the bases

$$\mathcal{A} = \{1, 1 - x, x^2\} \text{ of } \mathcal{P}_2$$

and

$$\mathcal{B} = \{1, x, 1 - x^2, 1 + x^3\} \text{ of } \mathcal{P}_3.$$

To find the first column of A , we compute $T(1)$ and write it as a linear combination of the vectors in \mathcal{B} .

$$\begin{aligned} T(1) &= 2 + x + x^3 \\ &= (1)(1) + (1)(x) + (0)(1 - x^2) + (1)(1 + x^3). \end{aligned}$$

Thus the first column of A is

$$[T(1)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

For the remaining columns of A , we follow the same procedure with the second and third base vectors in \mathcal{A} .

$$\begin{aligned} T(1-x) &= 2 - x^2 + x^3 \\ &= (0)(1) + (0)(x) + (1)(1 - x^2) + (1)(1 + x^3) \\ T(x^2) &= 2 + 3x + 2x^2 + x^3 \\ &= (3)(1) + (3)(x) + (-2)(1 - x^2) + (1)(1 + x^3). \end{aligned}$$

Thus the matrix of T with respect to \mathcal{A} and \mathcal{B} is given by

$$A = [T]_{\mathcal{B}, \mathcal{A}} = [[T(1)]_{\mathcal{B}}, [T(1-x)]_{\mathcal{B}}, [T(x^2)]_{\mathcal{B}}] = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 0 & 3 \\ 0 & 1 & -2 \\ 1 & 1 & 1 \end{bmatrix}. \blacksquare$$

If T is a linear transformation of \mathbf{R}^n into \mathbf{R}^m , it is usually easier to find the matrix of T relative to \mathcal{E}_n and \mathcal{E}_m than with any other choice of bases, because the coordinates of a vector are the same as the components when working with the standard bases. However, we shall see in the next section that other choices of bases may give a much simpler matrix for T .

Example 2 □ Consider the linear transformation $T : \mathbf{R}^4 \rightarrow \mathbf{R}^3$ given by

$$T(x_1, x_2, x_3, x_4) = (x_1 - x_3 + x_4, 2x_1 + x_2 + 3x_4, x_1 + 2x_2 + 3x_3 + 3x_4).$$

Using the standard bases \mathcal{E}_4 and \mathcal{E}_3 , we compute

$$\begin{aligned} T(1, 0, 0, 0) &= (1, 2, 1) & T(0, 1, 0, 0) &= (0, 1, 2) \\ T(0, 0, 1, 0) &= (-1, 0, 3) & T(0, 0, 0, 1) &= (1, 3, 3). \end{aligned}$$

Thus the matrix of T relative to \mathcal{E}_4 and \mathcal{E}_3 is

$$[T]_{\mathcal{E}_3, \mathcal{E}_4} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 2 & 1 & 0 & 3 \\ 1 & 2 & 3 & 3 \end{bmatrix}. \blacksquare$$

The mapping $f : \mathbf{R}^n \rightarrow \mathbf{R}_{n \times 1}$ defined by

$$f(x_1, x_2, \dots, x_n) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is a natural isomorphism. It is so natural, in fact, that some texts make no distinction between

$$(x_1, x_2, \dots, x_n) \text{ and } \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

and consider them as being the same entity. This isomorphism leads to a natural connection between the preceding example and Example 4 of Section 5.2. From the example in Section 5.2, we know that the matrix transformation $S : \mathbf{R}_{4 \times 1} \rightarrow \mathbf{R}_{3 \times 1}$ defined by

$$S \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 2 & 1 & 0 & 3 \\ 1 & 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 - x_3 + x_4 \\ 2x_1 + x_2 + 3x_4 \\ x_1 + 2x_2 + 3x_3 + 3x_4 \end{bmatrix}$$

is a linear transformation. It is clear at a glance that the matrix transformation S and the mapping T in Example 2 are the same except for notation. That is, the two mappings differ only by an isomorphism.

The work in Examples 1 and 2 of this section illustrates that, for given bases \mathcal{A} and \mathcal{B} , the matrix $[T]_{\mathcal{B}, \mathcal{A}}$ is uniquely determined by T . On the other hand, for a given matrix $A = [a_{ij}]_{m \times n}$ and fixed bases \mathcal{A} and \mathcal{B} , there is only one linear transformation that has A as its matrix relative to \mathcal{A} and \mathcal{B} . For if A is the matrix of both S and T , then we have $S(\mathbf{u}_j) = \sum_{i=1}^m a_{ij} \mathbf{v}_i = T(\mathbf{u}_j)$ and, for any $\mathbf{u} = \sum_{j=1}^n x_j \mathbf{u}_j$ in \mathbf{U} ,

$$S(\mathbf{u}) = S \left(\sum_{j=1}^n x_j \mathbf{u}_j \right) = \sum_{j=1}^n x_j S(\mathbf{u}_j) = \sum_{j=1}^n x_j T(\mathbf{u}_j) = T \left(\sum_{j=1}^n x_j \mathbf{u}_j \right) = T(\mathbf{u}).$$

But $S(\mathbf{u}) = T(\mathbf{u})$ for all \mathbf{u} in \mathbf{U} means $S = T$. Thus, for fixed bases \mathcal{A} and \mathcal{B} , T and $A = [T]_{\mathcal{B}, \mathcal{A}}$ determine each other uniquely by the rule

$$T(\mathbf{u}_j) = \sum_{i=1}^m a_{ij} \mathbf{v}_i, \text{ or } \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} = [T(\mathbf{u}_j)]_{\mathcal{B}}.$$

Generally speaking, a change in either or both of the bases \mathcal{A} and \mathcal{B} is reflected in a change in the matrix of the linear transformation. With each particular choice of \mathcal{A} and \mathcal{B} , we say that the matrix A of T relative to \mathcal{A} and \mathcal{B} represents T relative to these bases. Thus *different* matrices may represent the *same* linear transformation with different choices of bases. These matrices, though different, have certain properties in common. The exact relationship between these matrices is revealed in Section 5.4, but the next theorem gives some useful information on this subject.

Theorem 5.11 *If A is any matrix that represents the linear transformation T of \mathbf{U} into \mathbf{V} , then $\text{rank}(A) = \text{rank}(T)$. That is, $\text{rank}([T]_{\mathcal{B}, \mathcal{A}}) = \text{rank}(T)$.*

Proof. Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be bases of \mathbf{U} and \mathbf{V} , respectively, and suppose that $A = [a_{ij}]_{m \times n}$ represents T relative to these bases. Then $T(\mathbf{u}_j) = \sum_{i=1}^m a_{ij} \mathbf{v}_i$, so that A is the matrix of transition from \mathcal{B} to $T(\mathcal{A})$. According to Theorem 5.8, $T(\mathcal{A})$ spans $T(\mathbf{U})$. This means that the columns of A record the coordinates relative to \mathcal{B} of a spanning set for $T(\mathbf{U})$. By Theorem 3.37, the reduced column-echelon form A' of A also records the coordinates relative to \mathcal{B} of a spanning set for $T(\mathbf{U})$. Hence the number of nonzero columns in A' is the dimension of $T(\mathbf{U})$, and we have $\text{rank}(A) = \text{rank}(T)$. ■

Thus, a convenient method for finding the rank of a linear transformation T is to find the rank of a matrix that represents T relative to a pair of bases \mathcal{A} and \mathcal{B} . As a matter of fact, the reduced column-echelon form of such a matrix will disclose not only the rank of T , but the standard basis of $T(\mathbf{U})$ relative to \mathcal{B} .

Example 3 □ In Example 1 of this section, we found that $T : \mathcal{P}_2 \rightarrow \mathcal{P}_3$ defined by

$$T(a_0 + a_1 x + a_2 x^2) = (2a_0 + 2a_2) + (a_0 + a_1 + 3a_2)x + (a_1 + 2a_2)x^2 + (a_0 + a_2)x^3$$

has the matrix

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

relative to $\mathcal{A} = \{1, 1-x, x^2\}$ and $\mathcal{B} = \{1, x, 1-x^2, 1+x^3\}$. The matrix A can be transformed to reduced column-echelon form as follows.

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 0 & 3 \\ 0 & 1 & -2 \\ 1 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & -2 \\ 1 & 1 & -2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = A'$$

Thus $\text{rank}(T) = \text{rank}(A) = 2$, and the computations

$$\begin{aligned} (1)(1) + (1)(x) + (0)(1-x^2) + (1)(1+x^3) &= 2 + x + x^3 \\ (0)(1) + (0)(x) + (1)(1-x^2) + (1)(1+x^3) &= 2 - x^2 + x^3 \end{aligned}$$

show that the standard basis of $T(\mathcal{P}_2)$ relative to \mathcal{B} is

$$\{2 + x + x^3, 2 - x^2 + x^3\}. \blacksquare$$

Theorem 5.12 Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be bases of \mathbf{U} and \mathbf{V} , respectively, and let $A = [a_{ij}]_{m \times n} = [T]_{\mathcal{B}, \mathcal{A}}$ be the matrix of T relative to the bases \mathcal{A} and \mathcal{B} . If \mathbf{u} is an arbitrary vector in \mathbf{U} , then

$$[T(\mathbf{u})]_{\mathcal{B}} = [T]_{\mathcal{B}, \mathcal{A}}[\mathbf{u}]_{\mathcal{A}}.$$

Proof. Let

$$[\mathbf{u}]_{\mathcal{A}} = X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad [T(\mathbf{u})]_{\mathcal{B}} = Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

so that $\mathbf{u} = \sum_{j=1}^n x_j \mathbf{u}_j$ and $T(\mathbf{u}) = \sum_{i=1}^m y_i \mathbf{v}_i$. Since $A = [a_{ij}]_{m \times n}$ is the matrix of T relative to \mathcal{A} and \mathcal{B} , we have

$$\begin{aligned}\sum_{i=1}^m y_i \mathbf{v}_i &= T(\mathbf{u}) \\ &= T\left(\sum_{j=1}^n x_j \mathbf{u}_j\right) \\ &= \sum_{j=1}^n x_j T(\mathbf{u}_j) \\ &= \sum_{j=1}^n x_j \left(\sum_{i=1}^m a_{ij} \mathbf{v}_i\right) \\ &= \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j\right) \mathbf{v}_i.\end{aligned}$$

But the coordinates of $T(\mathbf{u})$ relative to \mathcal{B} are unique, so this means that $y_i = \sum_{j=1}^n a_{ij} x_j$, and

$$\begin{aligned}Y &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = AX.\end{aligned}$$

That is,

$$[T(\mathbf{u})]_{\mathcal{B}} = [T]_{\mathcal{B}, \mathcal{A}} [\mathbf{u}]_{\mathcal{A}}. \blacksquare$$

Theorem 5.12 shows that the matrix transformation from $\mathbf{R}_{n \times 1}$ to $\mathbf{R}_{m \times 1}$ defined in Section 5.2 by $T(X) = AX$ generalizes to coordinates with arbitrary linear transformations of finite-dimensional vector spaces. However, it should be kept in mind that X , A , and Y in the equation $Y = AX$ are taken relative to the bases \mathcal{A} and \mathcal{B} , and consequently change when \mathcal{A} and \mathcal{B} are changed.

Example 4 □ Let T be the linear transformation of \mathbf{R}^4 into \mathbf{R}^3 that has the matrix

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

with respect to the bases $\mathcal{A} = \{(1, 1, 1, 1), (1, 1, 1, 0), (1, 1, 0, 0), (1, 0, 0, 0)\}$ of \mathbf{R}^4 and $\mathcal{B} = \{(0, 1, 1), (1, 0, 0), (0, 0, 1)\}$ of \mathbf{R}^3 . Suppose we wish to use this matrix A to find the value of $T(3, 4, -1, -4)$.

To find the coordinates of $\mathbf{u} = (3, 4, -1, -4)$ relative to \mathcal{A} , we write \mathbf{u} as a linear combination of the vectors in \mathcal{A} . The strategic placement of zeros in the vectors of \mathcal{A} allows us to obtain the coefficients of these vectors by inspection: The first vector in \mathcal{A} is the only one with a nonzero fourth component, the first and second vectors in \mathcal{A} are the only ones with nonzero third components, and so on. We obtain

$$(3, 4, -1, -4) = (-4)(1, 1, 1, 1) + (3)(1, 1, 1, 0) + 5(1, 1, 0, 0) + (-1)(1, 0, 0, 0).$$

Thus

$$[\mathbf{u}]_{\mathcal{A}} = \begin{bmatrix} -4 \\ 3 \\ 5 \\ -1 \end{bmatrix}$$

and therefore

$$[T(\mathbf{u})]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 2 & 1 & 0 & 3 \\ 1 & 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} -4 \\ 3 \\ 5 \\ -1 \end{bmatrix} = \begin{bmatrix} -10 \\ -8 \\ 14 \end{bmatrix}.$$

Using these coordinates with the vectors in the basis \mathcal{B} , we find that

$$\begin{aligned} T(3, 4, -1, -4) &= (-10)(0, 1, 1) + (-8)(1, 0, 0) + 14(0, 0, 1) \\ &= (-8, -10, 4). \end{aligned}$$
■

Theorem 5.12 reveals a practical approach to the problem of finding a basis for the kernel $T^{-1}(\mathbf{0})$. For $\mathbf{u} \in T^{-1}(\mathbf{0})$ if and only if the coordinates X of \mathbf{u} satisfy $AX = \mathbf{0}$. Thus, the solutions to the system of equations $AX = \mathbf{0}$ furnish the coordinates of the vectors in $T^{-1}(\mathbf{0})$.

Example 5 □ With \mathbf{R} as the field of scalars for \mathcal{P}_4 , let $T : \mathcal{P}_4 \rightarrow \mathbf{R}^4$ be the linear transformation defined by

$$\begin{aligned} T(a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4) &= \\ (a_0 + a_1 + a_2 + 2a_4, a_1 + 2a_2 + a_3 - a_4, a_3 - a_4, a_0 + a_1 + a_2 + 2a_4). \end{aligned}$$

A natural choice of bases is $\mathcal{A} = \{1, x, x^2, x^3, x^4\}$ for \mathcal{P}_4 and the standard basis $\mathcal{E}_4 =$

$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ for \mathbf{R}^4 . Straightforward computations yield

$$\begin{aligned} T(1) &= (1, 0, 0, 1), & T(x) &= (1, 1, 0, 1) \\ T(x^2) &= (1, 2, 0, 1), & T(x^3) &= (0, 1, 1, 0), \\ T(x^4) &= (2, -1, -1, 2). \end{aligned}$$

Thus the matrix of T relative to \mathcal{A} and \mathcal{B} is

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

and we need to solve the linear system $AX = \mathbf{0}$. Using the augmented matrix $[A, \mathbf{0}]$, we have

$$[A, \mathbf{0}] \rightarrow \begin{bmatrix} 1 & 1 & 1 & 0 & 2 & 0 \\ 0 & 1 & 2 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 1 & 1 & 1 & 0 & 2 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 0 & 2 & 0 \\ 0 & 1 & 2 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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

The solutions to $AX = \mathbf{0}$ are given by

$$x_1 = x_3 - 2x_5$$

$$x_2 = -2x_3$$

$$x_3 = x_3$$

$$x_4 = x_5$$

$$x_5 = x_5$$

and in matrix form by

$$X = x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Using the coordinate matrices $[1, -2, 1, 0, 0]^T$ and $[-2, 0, 0, 1, 1]^T$ with the basis \mathcal{A} , we obtain the basis

$$\{1 - 2x + x^2, -2 + x^3 + x^4\}$$

for the kernel of T . ■

Theorem 5.13 Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be bases of \mathbf{U} and \mathbf{V} , respectively. If $A = [a_{ij}]_{m \times n}$ is a matrix such that the equation

$$[T(\mathbf{u})]_{\mathcal{B}} = A[\mathbf{u}]_{\mathcal{A}}$$

is satisfied for all \mathbf{u} in \mathbf{U} , then A is the matrix of the linear transformation T relative to \mathcal{A} and \mathcal{B} .

Proof. For $\mathbf{u} = \mathbf{u}_j$, we have

$$[\mathbf{u}_j] = \begin{bmatrix} \delta_{1j} \\ \delta_{2j} \\ \vdots \\ \delta_{nj} \end{bmatrix}$$

and

$$T[(\mathbf{u}_j)]_{\mathcal{B}} = A[\mathbf{u}_j]_{\mathcal{A}} = \begin{bmatrix} \sum_{k=1}^n a_{1k} \delta_{kj} \\ \sum_{k=1}^n a_{2k} \delta_{kj} \\ \vdots \\ \sum_{k=1}^n a_{mk} \delta_{kj} \end{bmatrix} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

Thus $T(\mathbf{u}_j) = \sum_{i=1}^m a_{ij} \mathbf{v}_i$, and A is the matrix of T relative to \mathcal{A} and \mathcal{B} . ■

Exercises 5.3

- Let $\mathbf{v}_1 = (2, -1)$ and $\mathbf{v}_2 = (1, 0)$ in \mathbf{R}^2 . Find a formula for $T(x, y)$ if T is the linear transformation of \mathbf{R}^2 into \mathbf{R}^3 for which $T(\mathbf{v}_1) = (1, 0, 1)$ and $T(\mathbf{v}_2) = (0, 1, 1)$.

2. Suppose $T : \mathbf{R}_{2 \times 1} \rightarrow \mathbf{R}_{3 \times 1}$ is a linear transformation such that

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}.$$

$$\text{Find } T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right).$$

3. Consider the basis $\{p_1(x), p_2(x), p_3(x)\}$ for \mathcal{P}_2 over \mathbf{R} , where $p_1(x) = 1 + x + x^2$, $p_2(x) = x + x^2$, $p_3(x) = x^2$. Find a formula for $T(a_0 + a_1x + a_2x^2)$ if $T : \mathcal{P}_2 \rightarrow \mathbf{R}^2$ is the linear transformation such that $T(p_1(x)) = (1, 0)$, $T(p_2(x)) = (1, 0)$, and $T(p_3(x)) = (0, 1)$.

4. Let $T : \mathbf{R}_{2 \times 2} \rightarrow \mathbf{R}_{3 \times 1}$ be a linear transformation for which it is known that

$$T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, T\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix},$$

$$T\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, T\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix}.$$

$$\text{Find } T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right).$$

In Problems 5–8, find the matrix of the given linear transformation T relative to the bases \mathcal{A} and \mathcal{B} .

5. $T : \mathbf{R}^3 \rightarrow \mathbf{R}^4$, $\mathcal{A} = \mathcal{E}_3$, $\mathcal{B} = \mathcal{E}_4$

$$T(x_1, x_2, x_3) = (x_1 + x_2 - x_3, 5x_2 - 2x_3, 4x_1 + x_3, 2x_1 + 3x_2 + x_3)$$

6. $T : \mathbf{R}^4 \rightarrow \mathbf{R}^3$, $\mathcal{A} = \mathcal{E}_4$, $\mathcal{B} = \mathcal{E}_3$

$$T(x_1, x_2, x_3, x_4) = (x_1 + x_2 + x_3 - x_4, x_1 + 2x_2 + x_3 - x_4, x_1 - x_2 + x_3)$$

7. $T : \mathcal{P}_2 \rightarrow \mathcal{P}_3$ over \mathbf{R} , $\mathcal{A} = \{1, 1 - x, x - x^2\}$, $\mathcal{B} = \{1, 1 + x, 1 - x^2, x + x^3\}$

$$T(p(x)) = (1 - x)p(x)$$

8. $T : \mathcal{P}_2 \rightarrow \mathcal{P}_1$ over \mathbf{R} , $\mathcal{A} = \{1 + x^2, 1 + x, 1\}$, $\mathcal{B} = \{1 - x, x\}$

$$T(a_0 + a_1x + a_2x^2) = (a_0 + a_1 + a_2) + (a_1 - 2a_2)x$$

9. If T is the linear transformation that has the matrix $\begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}$ relative to the bases $\{(1, 1), (-3, 1)\}$ of \mathbf{R}^2 and \mathcal{E}_3 of \mathbf{R}^3 , find $T(-2, 2)$.

10. Let T be the linear transformation of \mathbf{R}^3 into \mathbf{R}^2 that has the matrix $\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix}$ relative to the bases $\{(1, -1, 1), (0, 1, 0), (1, 0, 0)\}$ of \mathbf{R}^3 and $\{(3, 2), (2, 1)\}$ of \mathbf{R}^2 . Find $T(2, 0, 1)$.

11. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \\ 4 & 3 & 1 \end{bmatrix}$ be the matrix of the linear transformation $T : \mathcal{P}_2 \rightarrow \mathcal{P}_3$ over \mathbf{R} with respect to the bases $\{4, 1+x, 1+x^2\}$ and $\{1, x, x^2, x^3\}$. Find $T(2-2x+x^2)$.

12. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}$ be the matrix of the linear transformation $T : \mathcal{P}_2 \rightarrow \mathcal{P}_1$ with respect to the bases $\{1+x+x^2, x+x^2, x^2\}$ and $\{1, 2+3x\}$. Find $T(2+5x+x^2)$.

13. The linear operator T on \mathbf{R}^2 has the matrix $\begin{bmatrix} 4 & -1 \\ -4 & 3 \end{bmatrix}$ relative to the basis $\mathcal{A} = \mathcal{B} = \{(1, 2), (0, 1)\}$. A vector \mathbf{u} has coordinates $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ relative to this basis. Find $T(\mathbf{u})$ in component form (x, y) .

14. Suppose the bases \mathcal{A} of $\mathbf{R}_{2 \times 2}$ and \mathcal{B} of \mathcal{P}_1 over \mathbf{R} are given by

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\},$$

$$\mathcal{B} = \{1+x, 1-x\}.$$

- Let T be the linear transformation $T : \mathbf{R}_{2 \times 2} \rightarrow \mathcal{P}_1$ with matrix $\begin{bmatrix} 2 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 \end{bmatrix}$ relative to \mathcal{A} and \mathcal{B} . Find $T\left(\begin{bmatrix} 0 & 0 \\ 3 & 2 \end{bmatrix}\right)$.

15. Let T be the linear transformation of \mathbf{R}^3 into \mathbf{R}^2 whose matrix is

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

relative to the standard bases \mathcal{E}_3 and \mathcal{E}_2 . Find the matrix of T relative to the bases $\{(1, 2, 0), (1, 1, 1), (1, 1, 0)\}$ of \mathbf{R}^3 and $\{(1, 0), (1, 1)\}$ of \mathbf{R}^2 .

16. Let T be a linear operator on \mathbf{R}^2 that maps $(2, 1)$ onto $(5, 2)$ and $(1, 2)$ onto $(7, 10)$. Determine the matrix of T with respect to the bases $\mathcal{A} = \mathcal{B} = \{(3, 3), (1, -1)\}$.

17. Find the matrix of T in Problem 9 relative to the bases $\{(-1, 3), (-1, 1)\}$ of \mathbf{R}^2 and \mathcal{E}_3 of \mathbf{R}^3 .

18. A linear operator T on \mathbf{R}^4 has the matrix

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

with respect to the bases $\mathcal{A} = \mathcal{B} = \mathcal{E}_4$. Find a basis for $T(\mathbf{R}^4)$.

19. Suppose T is the linear transformation of \mathbf{R}^n into \mathbf{R}^m that has the given matrix A relative to the standard bases \mathcal{E}_n and \mathcal{E}_m . Find a basis for the range of T .

$$(a) A = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 2 & 3 & 1 & 0 \\ 1 & 3 & 2 & 3 \end{bmatrix} \quad (b) A = \begin{bmatrix} 3 & -2 & -1 & -4 \\ 1 & 1 & -2 & -3 \\ -2 & 3 & -1 & 1 \end{bmatrix}$$

$$(c) A = \begin{bmatrix} 1 & 2 & -1 & 2 & 1 \\ 1 & 4 & 4 & -3 & -1 \\ 2 & 6 & 3 & -1 & 0 \\ 3 & 8 & 2 & 1 & 2 \end{bmatrix} \quad (d) A = \begin{bmatrix} 1 & 2 & 0 & 1 & 0 \\ 2 & 4 & 1 & 4 & 3 \\ 1 & 2 & 2 & 5 & -2 \\ -1 & -2 & 3 & 5 & 4 \end{bmatrix}$$

20. Find a basis for the kernel of T in each part of Problem 19.

21. Let T be the linear transformation of \mathbf{R}^5 into \mathbf{R}^3 that has the matrix A relative to the bases $\{(1, 1, 1, 1, 1), (1, 1, 1, 1, 0), (1, 1, 0, 0, 0), (1, 0, 0, 0, 0), (0, 0, 0, 0, 1)\}$ of \mathbf{R}^5 and $\{(1, 1, 1), (0, 1, 0), (1, 0, 0)\}$ of \mathbf{R}^3 . Find a basis for the range of T .

$$(a) A = \begin{bmatrix} 1 & 3 & 2 & 0 & -1 \\ 2 & 6 & 4 & 6 & 4 \\ 1 & 3 & 2 & 2 & 1 \end{bmatrix} \quad (b) A = \begin{bmatrix} 1 & -2 & 0 & 1 & -4 \\ 2 & -4 & 1 & 3 & -5 \\ 1 & -2 & 0 & 0 & -2 \end{bmatrix}$$

22. Find a basis for the kernel of T in each part of Problem 21.

23. Let T be the linear operator on \mathbf{R}^3 that has the matrix $\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \\ 3 & 3 & 0 \end{bmatrix}$ relative to the basis $\mathcal{A} = \mathcal{B} = \{(1, 0, 1), (1, 1, 0), (0, 1, 1)\}$ of \mathbf{R}^3 . Find a basis for the kernel of T .

24. Let T be a linear transformation of \mathbf{U} into \mathbf{V} , and let \mathbf{U}_1 and \mathbf{U}_2 be subspaces of \mathbf{U} . Prove or disprove that $T(\mathbf{U}_1 \cap \mathbf{U}_2) = T(\mathbf{U}_1) \cap T(\mathbf{U}_2)$.
25. Let T be a linear transformation of \mathbf{U} into \mathbf{V} , and let \mathbf{W}_1 and \mathbf{W}_2 be subspaces of \mathbf{V} . Prove or disprove the statements below.

- (a) $T^{-1}(\mathbf{W}_1 + \mathbf{W}_2) = T^{-1}(\mathbf{W}_1) + T^{-1}(\mathbf{W}_2)$.
(b) $T^{-1}(\mathbf{W}_1 \cap \mathbf{W}_2) = T^{-1}(\mathbf{W}_1) \cap T^{-1}(\mathbf{W}_2)$.

5.4 Change of Basis

It is the purpose of this section to give a complete description of the relation between those matrices that represent the same linear transformation. This description is found by examining the effect that a change in the bases of \mathbf{U} and \mathbf{V} has on the matrix of T .

Theorem 5.14 *Let $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$ and $\mathcal{C}' = \{\mathbf{w}'_1, \mathbf{w}'_2, \dots, \mathbf{w}'_k\}$ be two bases of the vector space \mathbf{W} over \mathcal{F} . For an arbitrary vector \mathbf{w} in \mathbf{W} , let*

$$[\mathbf{w}]_{\mathcal{C}} = C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} \text{ and } [\mathbf{w}]_{\mathcal{C}'} = C' = \begin{bmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_k \end{bmatrix}$$

denote the coordinate matrices of \mathbf{w} relative to \mathcal{C} and \mathcal{C}' , respectively. If P is the matrix of transition from \mathcal{C} to \mathcal{C}' , then $C = PC'$. That is,

$$[\mathbf{w}]_{\mathcal{C}} = P[\mathbf{w}]_{\mathcal{C}'}.$$

Proof. Let $P = [p_{ij}]_{k \times k}$, and assume that the hypotheses of the theorem are satisfied. Then

$$\mathbf{w} = \sum_{i=1}^k c_i \mathbf{w}_i = \sum_{j=1}^k c'_j \mathbf{w}'_j$$

and

$$\mathbf{w}'_j = \sum_{i=1}^k p_{ij} \mathbf{w}_i.$$

Combining these equalities, we have

$$\begin{aligned}\mathbf{w} &= \sum_{i=1}^k c_i \mathbf{w}_i \\ &= \sum_{j=1}^k c'_j \left(\sum_{i=1}^k p_{ij} \mathbf{w}_i \right) \\ &= \sum_{i=1}^k \left(\sum_{j=1}^k p_{ij} c'_j \right) \mathbf{w}_i.\end{aligned}$$

Therefore $c_i = \sum_{j=1}^k p_{ij} c'_j$ and

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} p_{11}c'_1 + p_{12}c'_2 + \cdots + p_{1k}c'_k \\ p_{21}c'_1 + p_{22}c'_2 + \cdots + p_{2k}c'_k \\ \vdots \\ p_{k1}c'_1 + p_{k2}c'_2 + \cdots + p_{kk}c'_k \end{bmatrix} = PC'. \blacksquare$$

Example 1 □ Consider the bases $\mathcal{C} = \{x, 2+x\}$ and $\mathcal{C}' = \{4+x, 4-x\}$ of the vector space $\mathbf{W} = \mathcal{P}_1$ over \mathbf{R} . Since

$$\begin{aligned}4+x &= (-1)(x) + (2)(2+x), \\ 4-x &= (-3)(x) + (2)(2+x)\end{aligned}$$

the matrix of transition P from \mathcal{C} to \mathcal{C}' is given by

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

The vector $\mathbf{w} = 4+3x$ can be written as

$$4+3x = 2(4+x) + (-1)(4-x)$$

so

$$[\mathbf{w}]_{\mathcal{C}'} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

According to Theorem 5.14, the coordinate matrix $[\mathbf{w}]_{\mathcal{C}}$ may be found from

$$[\mathbf{w}]_{\mathcal{C}} = P[\mathbf{w}]_{\mathcal{C}'} = \begin{bmatrix} -1 & -3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

This result can be checked by using the base vectors in \mathcal{C} . We get

$$(1)(x) + 2(2 + x) = 4 + 3x = \mathbf{w},$$

so the value for $[\mathbf{w}]_{\mathcal{C}}$ is correct. ■

The following theorem gives a full description of the effect of a change in the bases \mathcal{A} and \mathcal{B} .

Theorem 5.15 Suppose that T has matrix $A = [a_{ij}]_{m \times n}$ relative to the bases \mathcal{A} of \mathbf{U} and \mathcal{B} of \mathbf{V} . If Q is the matrix of transition from \mathcal{A} to the basis \mathcal{A}' of \mathbf{U} and P is the matrix of transition from \mathcal{B} to the basis \mathcal{B}' of \mathbf{V} , then the matrix of T relative to \mathcal{A}' and \mathcal{B}' is $P^{-1}AQ$. (See Figure 5.1.)

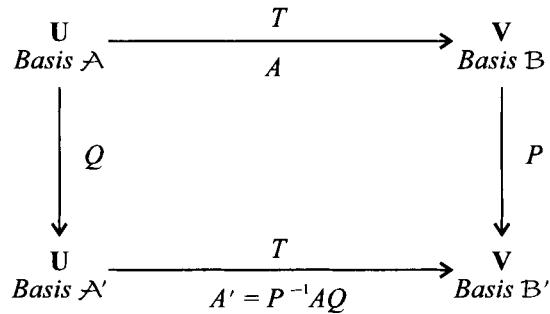


Figure 5.1

Proof. Assume that the hypotheses of the theorem are satisfied. Let \mathbf{u} be an arbitrary vector in \mathbf{U} , let X and X' denote the coordinate matrices of \mathbf{u} relative to \mathcal{A} and \mathcal{A}' , respectively, and let Y and Y' denote the coordinate matrices of $T(\mathbf{u})$ relative to \mathcal{B} and \mathcal{B}' , respectively. According to Theorems 5.12 and 5.14, we have $Y = AX$, where $Y = PY'$ and $X = QX'$. Substituting for Y and X , we have

$$PY' = AQQX',$$

and therefore

$$Y' = (P^{-1}AQ)X'.$$

By Theorem 5.13, $P^{-1}AQ$ is the matrix of T relative to \mathcal{A}' and \mathcal{B}' . ■

Theorem 5.16 Two $m \times n$ matrices A and B represent the same linear transformation T of \mathbf{U} into \mathbf{V} if and only if A and B are equivalent.

Proof. If A and B represent T relative to set of bases \mathcal{A}, \mathcal{B} and $\mathcal{A}', \mathcal{B}'$, respectively, then $B = P^{-1}AQ$, where Q is the matrix of transition from \mathcal{A} to \mathcal{A}' and P is the matrix of transition from \mathcal{B} to \mathcal{B}' . Hence A and B are equivalent.

If B is equivalent to A , then $B = P^{-1}AQ$ for invertible P and Q . If A represents T relative to \mathcal{A} and \mathcal{B} , then B represents T relative to \mathcal{A}' and \mathcal{B}' , where Q is the matrix of transition from \mathcal{A} to \mathcal{A}' and P is the matrix of transition from \mathcal{B} to \mathcal{B}' . ■

The proof of Theorem 5.16 can be modified so as to obtain two similar results concerning row equivalence and column equivalence. For requiring $\mathcal{B}' = \mathcal{B}$ is the same as requiring $P = I_m$, and requiring $\mathcal{A}' = \mathcal{A}$ is the same as requiring $Q = I_n$. Thus we have the following theorems.

Theorem 5.17 *Two $m \times n$ matrices A and B represent the same linear transformation T of \mathbf{U} into \mathbf{V} relative to the same basis of \mathbf{U} if and only if they are row-equivalent.*

Theorem 5.18 *Two $m \times n$ matrices A and B represent the same linear transformation T of \mathbf{U} into \mathbf{V} relative to the same basis of \mathbf{V} if and only if they are column-equivalent.*

The three preceding theorems give a full exposition of the connection between linear transformations and the equivalence relations on matrices that were studied in Chapter 3. However, there is one more major application of matrix theory to the study of linear transformations. This application is contained in the following theorem.

Theorem 5.19 *Let T be an arbitrary linear transformation of \mathbf{U} into \mathbf{V} , and let r be the rank of T . Then there exist bases \mathcal{A}' of \mathbf{U} and \mathcal{B}' of \mathbf{V} such that the matrix of T relative to \mathcal{A}' and \mathcal{B}' has the first r diagonal elements equal to 1, and all other elements zero.*

Proof. With the stated hypotheses, suppose that \mathcal{A} and \mathcal{B} are bases of \mathbf{U} and \mathbf{V} , respectively, and that T has matrix A relative to \mathcal{A} and \mathcal{B} . By Theorem 3.46, there exist invertible matrices P^{-1} and Q such that $P^{-1}AQ$ has the first r diagonal elements equal to 1, and all other elements zero. Let \mathcal{A}' and \mathcal{B}' be bases such that Q is the matrix of transition from \mathcal{A} to \mathcal{A}' and P is the matrix of transition from \mathcal{B} to \mathcal{B}' . Then $P^{-1}AQ$ is the matrix of T relative to \mathcal{A}' and \mathcal{B}' , and the theorem is proved. ■

Thus, with suitable choice of bases, each linear transformation T of \mathbf{U} into \mathbf{V} can be represented by a matrix of the form

$$D_r = \begin{bmatrix} I_r & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \end{bmatrix},$$

where r is the rank of T . From a different point of view, this means that two linear transformations of \mathbf{U} into \mathbf{V} can be represented by the same matrix if and only if they have the same rank. It is easy to see that the relation of having the same rank is an equivalence relation on the set of all linear transformations of \mathbf{U} into \mathbf{V} .

Example 2 □ Let T be the linear transformation from \mathbf{R}^4 to P_2 over \mathbf{R} defined by

$$T(a_1, a_2, a_3, a_4) = (a_1 - a_2 + 2a_4) + (a_1 - a_2 + a_3 + a_4)x + (2a_1 - 2a_2 + a_3 + 3a_4)x^2.$$

We shall find bases \mathcal{A}' of \mathbf{R}^4 and \mathcal{B}' of P_2 such that the matrix of T relative to \mathcal{A}' and \mathcal{B}' has the form

$$D_r = \begin{bmatrix} I_r & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \end{bmatrix}.$$

We first find the matrix A of T with respect to \mathcal{E}_4 and $\mathcal{B} = \{1, x, x^2\}$. Since

$$T(1, 0, 0, 0) = (1)(1) + (1)(x) + (2)(x^2),$$

$$T(0, 1, 0, 0) = (-1)(1) + (-1)(x) + (-2)(x^2),$$

$$T(0, 0, 1, 0) = (0)(1) + (1)(x) + (1)(x^2),$$

$$T(0, 0, 0, 1) = (2)(1) + (1)(x) + (3)(x^2),$$

the matrix A is given by

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

Following the proof of Theorem 5.19 and using the same procedure as in Section 3.8, we first find an invertible matrix Q such that $AQ = A'$ is in reduced column-echelon form.

$$\left[\begin{array}{c|cccc} A & \hline I_4 \end{array} \right] = \left[\begin{array}{cccc} 1 & -1 & 0 & 2 \\ 1 & -1 & 1 & 1 \\ 2 & -2 & 1 & 3 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 \\ 2 & 0 & 1 & -1 \\ \hline 1 & 1 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & -2 \\ 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right] = \left[\begin{array}{c|cc} AQ & \hline Q \end{array} \right]$$

Next we use row operations to transform $[A', I_3]$ into $[D_r, P^{-1}]$, where $D_r = P^{-1}A' = P^{-1}AQ$.

$$[A', I_3] = \left[\begin{array}{ccccc|ccc} 1 & 0 & 0 & 0 & | & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & | & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & | & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccccc|ccc} 1 & 0 & 0 & 0 & | & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & | & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & | & -1 & -1 & 1 \end{array} \right] = [D_r, P^{-1}].$$

Thus

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 1 & 0 & 1 & -2 \\ 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

are invertible matrices such that

$$P^{-1}AQ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I_2 & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \end{bmatrix} = D_2.$$

According to the proof of Theorem 5.19, the desired basis \mathcal{A}' and \mathcal{B}' can be found by using Q as the transition matrix from \mathcal{A} to \mathcal{A}' and P as the transition matrix from \mathcal{B} to \mathcal{B}' . Using Q to find \mathcal{A}' , we get

$$\mathcal{A}' = \{(1, 0, -1, 0), (0, 0, 1, 0), (1, 1, 0, 0), (-2, 0, 1, 1)\}.$$

To find \mathcal{B}' , we first obtain P by taking the inverse of P^{-1} and then use P as the transition matrix from \mathcal{B} to \mathcal{B}' . We find

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

and $\mathcal{B}' = \{1 + x^2, x + x^2, x^2\}$. The original defining equation for $T(a_1, a_2, a_3, a_4)$ can be used to check that D_2 is in fact the matrix of T with respect to \mathcal{A}' and \mathcal{B}' . ■

Exercises 5.4

- Find \mathbf{v} if $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ and $\mathcal{B} = \{x + x^2, 1 + x^2, x\}$ in \mathcal{P}_2 over \mathbf{R} .
- Let $\mathcal{B} = \{(2, -1, 1), (0, 1, -1), (-2, 1, 0)\}$. Find the coordinate matrix of $(6, 0, -1)$ relative to the basis \mathcal{B} of \mathbf{R}^3 .
- Suppose the basis \mathcal{A} of \mathcal{P}_2 over \mathbf{R} is given by $\mathcal{A} = \{x + x^2, 2 + x, 1\}$. Given that $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ is the transition matrix from \mathcal{A} to \mathcal{B} , find the basis \mathcal{B} of \mathcal{P}_2 .

4. Suppose $\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$ is the transition matrix from the basis $\mathcal{A} = \{(-4, 2), (10, -6)\}$ of \mathbf{R}^2 to the basis \mathcal{B} .

(a) Find \mathcal{B} . (b) Find the transition matrix from \mathcal{B} to \mathcal{A} .

5. Let $\mathcal{A} = \{(1, 1), (2, 0)\}$ and $\mathcal{B} = \{(0, 2), (2, 1)\}$ in \mathbf{R}^2 .

(a) Find $[\mathbf{u}]_{\mathcal{A}}$ if $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. (b) Find \mathbf{v} if $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

6. Let \mathcal{A} be the basis of \mathcal{P}_2 over \mathbf{R} given by $\mathcal{A} = \{x^2, 1 + x, x + x^2\}$. Given that

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

is the transition matrix from \mathcal{B} to \mathcal{A} and that $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, find \mathbf{u} .

7. Let $\mathcal{A} = \mathcal{E}_4$ in \mathbf{R}^4 and $\mathcal{B} = \{x^2, x, 1\}$ in \mathcal{P}_2 over \mathbf{R} . If T is the linear transformation

that is represented by $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$ relative to \mathcal{A} and \mathcal{B} , find the matrix that represents T with respect to \mathcal{A}' and \mathcal{B}' where

$$\begin{aligned} \mathcal{A}' &= \{(1, 0, 0, 0), (0, 0, 1, 0), (1, -1, 0, 0), (0, -1, 1, 1)\}, \\ \mathcal{B}' &= \{x^2 + 1, x, 1\}. \end{aligned}$$

8. Suppose $S : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ is the linear transformation with matrix $\begin{bmatrix} 1 & -3 & 1 \\ 2 & -6 & 2 \end{bmatrix}$ relative to the bases \mathcal{E}_3 and \mathcal{E}_2 . Find the matrix of S with respect to the bases $\{(1, 0, 1), (1, 0, 0), (1, 1, 0)\}$ and $\{(1, -1), (2, 0)\}$.

9. Let S be the linear transformation from \mathbf{R}^3 to \mathbf{R}^2 that has the matrix $\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix}$

relative to \mathcal{E}_3 and \mathcal{E}_2 . Given that $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$ is the transition matrix from \mathcal{E}_3

to \mathcal{A}' and $\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$ is the transition matrix from \mathcal{E}_3 to \mathcal{B}' , find the matrix of S relative to \mathcal{A}' and \mathcal{B}' .

10. Let T be the linear transformation of \mathbf{R}^3 into \mathbf{R}^2 that has the matrix $\begin{bmatrix} 1 & -1 & 1 \\ 2 & 1 & 1 \end{bmatrix}$ relative to the bases $\{(1, 2, 0), (1, 1, 1), (1, 1, 0)\}$ of \mathbf{R}^3 and $\{(1, 1), (1, -1)\}$ of \mathbf{R}^2 . Find the matrix of T relative to the bases $\{(2, 3, 0), (1, 1, 1), (2, 3, 1)\}$ of \mathbf{R}^3 and $\{(3, -1), (1, -1)\}$ of \mathbf{R}^2 .
11. Let T be the linear operator on \mathcal{P}_2 over \mathbf{R} that has the matrix $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 2 & 3 & 1 \end{bmatrix}$ relative to the bases $\mathcal{A} = \mathcal{B} = \{x + x^2, -1 + x, x\}$. Find the matrix of T relative to the basis $\mathcal{A}' = \mathcal{B}' = \{2 - x + x^2, -6x - 2x^2, x\}$.
12. Suppose the linear operator T on \mathcal{P}_1 over \mathbf{R} has the matrix $\begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$ with respect to $\mathcal{A} = \mathcal{B} = \{1 - x, x\}$. Find the matrix of T with respect to $\mathcal{A}' = \mathcal{B}' = \{2 - x, -1\}$.
13. The linear operator $S : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ has the matrix $\begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$ with respect to the bases $\mathcal{A} = \mathcal{B} = \{(1, 1), (0, 1)\}$. Find the matrix of S with respect to $\mathcal{A}' = \mathcal{B}' = \{(2, 1), (1, 2)\}$.
14. Let T be the linear operator in Problem 13 of Exercises 5.3. Use matrices of transition to find the matrix of T relative to \mathcal{E}_2 and \mathcal{E}_2 . Then use the new matrix to compute $T(\mathbf{u})$, thus making a check on the answer previously obtained.
15. Work Problem 17 of Exercises 5.3 using matrices of transition.
16. Suppose the linear transformation $S : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ has the matrix $A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 1 & 0 \end{bmatrix}$ relative to \mathcal{E}_3 and \mathcal{E}_2 . Find bases \mathcal{A}' of \mathbf{R}^3 and \mathcal{B}' of \mathbf{R}^2 such that the matrix \tilde{A}' of S relative to \mathcal{A}' and \mathcal{B}' is the reduced row-echelon form for A .
17. Suppose the linear transformation $T : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ has the matrix $A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix}$

relative to \mathcal{E}_3 and \mathcal{E}_2 . Given that $B_1 = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$ and $B_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ are invertible matrices such that $B_1AB_2 = D_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, find bases \mathcal{A}' of \mathbf{R}^3 and \mathcal{B}' of \mathbf{R}^2 such that T has matrix D_2 relative to \mathcal{A}' and \mathcal{B}' .

18. Suppose T is the linear transformation of \mathbf{R}^n into \mathbf{R}^m that has the given matrix A relative to \mathcal{E}_n and \mathcal{E}_m . Find bases \mathcal{A}' of \mathbf{R}^n and \mathcal{B}' of \mathbf{R}^m that satisfy the conditions given in Theorem 5.19.

$$(a) A = \begin{bmatrix} 1 & 3 & 2 & 0 & -1 \\ 2 & 6 & 5 & 6 & 1 \\ 1 & 3 & 2 & 2 & 0 \end{bmatrix} \quad (b) A = \begin{bmatrix} 3 & -2 & -1 & -4 \\ 1 & 1 & -2 & -3 \\ -2 & 3 & -1 & 1 \end{bmatrix}$$

5.5 Composition of Linear Transformations

In this section we conclude our study of the relation between a linear transformation and its associated matrix. We shall see that there are simple and direct connections between performing binary operations on linear transformations and performing binary operations on their associated matrices.

At various points, we have considered operations of addition and scalar multiplication on both matrices and linear transformations (see Example 5 of Section 4.2 and Definition 5.2). The intimate relation between these operations is described in the following theorem.

Theorem 5.20 *Let \mathbf{U} and \mathbf{V} be vector spaces over \mathcal{F} of dimensions n and m , respectively, and let S and T be linear transformations of \mathbf{U} into \mathbf{V} . If S has matrix A and T has matrix B relative to certain bases of \mathbf{U} and \mathbf{V} , then $S + T$ has matrix $A + B$ relative to these same bases. Also, for any a in \mathcal{F} , aT has matrix aB relative to these bases.*

Proof. Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be bases of \mathbf{U} and \mathbf{V} , respectively. If S has matrix $A = [a_{ij}]_{m \times n}$ and T has matrix $B = [b_{ij}]_{m \times n}$ relative to \mathcal{A} and \mathcal{B} , then $S(\mathbf{u}_j) = \sum_{i=1}^m a_{ij} \mathbf{v}_i$ and $T(\mathbf{u}_j) = \sum_{i=1}^m b_{ij} \mathbf{v}_i$. Hence

$$(S + T)(\mathbf{u}_j) = \sum_{i=1}^m (a_{ij} + b_{ij}) \mathbf{v}_i,$$

and $S + T$ has matrix $A + B$ relative to \mathcal{A} and \mathcal{B} . Also,

$$(aT)(\mathbf{u}_j) = \sum_{i=1}^m (ab_{ij}) \mathbf{v}_i,$$

and aT has matrix aB relative to \mathcal{A} and \mathcal{B} . ■

We turn our attention now to the “third operation” on linear transformations that was mentioned on page 150. This third operation is the composition of linear transformations, defined by the same kind of rule as used for composite functions in the calculus.

In the calculus, two given functions f and g are combined to produce the composite function $f \circ g$ by the rule

$$(f \circ g)(x) = f(g(x)).$$

The domain of $f \circ g$ is the set of all x in the domain of g such that f is defined at $g(x)$.

In linear algebra, it is common to refer to the composite of two linear transformations as their *product*. We adopt this usage here, stated formally in the following definition.

Definition 5.21 Let \mathbf{U} , \mathbf{V} , and \mathbf{W} be vector spaces over the same field \mathcal{F} , and suppose that S is a linear transformation of \mathbf{U} into \mathbf{V} and that T is a linear transformation of \mathbf{V} into \mathbf{W} . Then the **product** TS is the mapping of \mathbf{U} into \mathbf{W} defined by

$$TS(\mathbf{u}) = (T \circ S)(\mathbf{u}) = T(S(\mathbf{u}))$$

for each \mathbf{u} in \mathbf{U} .

Example 1 □ Consider the linear transformations $S : \mathbf{R}_{2 \times 2} \rightarrow \mathbf{R}^3$ and $T : \mathbf{R}^3 \rightarrow \mathcal{P}_1$ over \mathbf{R} defined by

$$S \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = (a + 2b, b - 3c, c + d)$$

and

$$T(a_1, a_2, a_3) = (a_1 - 2a_2 - 6a_3) + (a_2 + 3a_3)x.$$

Computing the product TS , we have

$$\begin{aligned} TS \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) &= T \left(S \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) \right) \\ &= T(a + 2b, b - 3c, c + d) \\ &= (a + 2b - 2(b - 3c) - 6(c + d) + (b - 3c + 3(c + d))x \\ &= (a - 6d) + (b + 3d)x. \end{aligned}$$
■

One of the exercises for this section asks for verification that the product in Definition 5.21 is associative but not commutative. The most important property, as far as our study is concerned, is stated in Theorem 5.22.

For the remainder of this section, \mathbf{U} , \mathbf{V} , and \mathbf{W} will denote vector spaces over the same field \mathcal{F} . Also, S and T will denote linear transformations of \mathbf{U} into \mathbf{V} and \mathbf{V} into \mathbf{W} , respectively.

Theorem 5.22 *The product of two linear transformations is a linear transformation.*

Proof. It is clear from Definition 5.21 that TS is a mapping of \mathbf{U} into \mathbf{W} . To show that TS is a linear transformation, let $\mathbf{u}_1, \mathbf{u}_2 \in \mathbf{U}$ and $a, b \in \mathcal{F}$. Then

$$\begin{aligned} TS(a\mathbf{u}_1 + b\mathbf{u}_2) &= T(S(a\mathbf{u}_1 + b\mathbf{u}_2)) \\ &= T(aS(\mathbf{u}_1) + bS(\mathbf{u}_2)) \\ &= aT(S(\mathbf{u}_1)) + bT(S(\mathbf{u}_2)) \\ &= aTS(\mathbf{u}_1) + bTS(\mathbf{u}_2), \end{aligned}$$

and TS is indeed a linear transformation of \mathbf{U} into \mathbf{W} . ■

The relation between multiplication of linear transformation and multiplication of matrices is given in the next theorem.

Theorem 5.23 Suppose that \mathbf{U}, \mathbf{V} , and \mathbf{W} are finite-dimensional vector spaces with bases \mathcal{A} , \mathcal{B} , and \mathcal{C} , respectively. If S has matrix A relative to \mathcal{A} and \mathcal{B} and T has matrix B relative to \mathcal{B} and \mathcal{C} , then TS has matrix BA relative to \mathcal{A} and \mathcal{C} .

Proof. Assume that the hypotheses are satisfied with $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$, $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p\}$, $A = [a_{ij}]_{m \times n}$, and $B = [b_{ij}]_{p \times m}$. Let $C = [c_{ij}]_{p \times n}$ be the matrix of TS relative to \mathcal{A} and \mathcal{C} . For $j = 1, 2, \dots, n$ we have

$$\begin{aligned} \sum_{i=1}^p c_{ij} \mathbf{w}_i &= TS(\mathbf{u}_j) \\ &= T\left(\sum_{k=1}^m a_{kj} \mathbf{v}_k\right) \\ &= \sum_{k=1}^m a_{kj} T(\mathbf{v}_k) \\ &= \sum_{k=1}^m a_{kj} \left(\sum_{i=1}^p b_{ik} \mathbf{w}_i\right) \\ &= \sum_{k=1}^m \sum_{i=1}^p (a_{kj} b_{ik} \mathbf{w}_i) \\ &= \sum_{i=1}^p \sum_{k=1}^m (b_{ik} a_{kj} \mathbf{w}_i) \\ &= \sum_{i=1}^p \left(\sum_{k=1}^m b_{ik} a_{kj}\right) \mathbf{w}_i, \end{aligned}$$

and consequently $c_{ij} = \sum_{k=1}^m b_{ik} a_{kj}$ for all values of i and j . Therefore $C = BA$ and the theorem is proved. ■

Example 2 □ Let the bases \mathcal{A} of $\mathbf{R}_{2 \times 2}$, \mathcal{B} of \mathbf{R}^3 , and \mathcal{C} of \mathcal{P}_1 be given by

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

$\mathcal{B} = \mathcal{E}_3$, and $\mathcal{C} = \{1, x\}$. With S and T as in Example 1, we shall find the matrices of S , T , and TS relative to these bases and verify that the matrix of TS is the product of the matrix of T times the matrix of S .

Since

$$S \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) = (1, 0, 0), \quad S \left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right) = (2, 1, 0),$$

$$S \left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right) = (0, -3, 1), \quad S \left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = (0, 0, 1),$$

S has matrix

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

relative to \mathcal{A} and \mathcal{B} . Since

$$\begin{aligned} T(1, 0, 0) &= (1)(1) + (0)(x) \\ T(0, 1, 0) &= (-2)(1) + (1)(x) \\ T(0, 0, 1) &= (-6)(1) + (3)(x), \end{aligned}$$

the matrix B of T relative to \mathcal{B} and \mathcal{C} is

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

For the product TS , we have

$$TS \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) = (1)(1) + (0)(x), \quad TS \left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right) = (0)(1) + (1)(x),$$

$$TS \left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right) = (0)(1) + (0)(x), \quad TS \left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = (-6)(1) + (3)(x),$$

so the matrix of TS relative to \mathcal{A} and \mathcal{C} is

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

Routine arithmetic verifies that

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

We note that the product AB is not defined, and this is consistent with the fact that ST is not defined. ■

The operations of addition and multiplication of matrices are connected by the distributive property. The statement of this fact is in Theorem 5.24. Proofs are requested in the exercises.

Theorem 5.24 *Let $A = [a_{ij}]_{m \times n}$, $B = [b_{ij}]_{n \times p}$, and $C = [c_{ij}]_{n \times p}$ over \mathcal{F} . Then*

$$A(B + C) = AB + AC.$$

There is the possibility in Definition 5.10 that the vector spaces \mathbf{U} and \mathbf{V} may be identical, and yet the bases \mathcal{A} and \mathcal{B} may be different. In many instances, there is no condition present that requires that \mathcal{A} and \mathcal{B} be different. In these instances, it is convenient and conventional to choose $\mathcal{A} = \mathcal{B}$. If it is our intention to use just one base \mathcal{A} of \mathbf{V} , we simply use the phrase “matrix of T relative to \mathcal{A} ” rather than “matrix of T relative to \mathcal{A} and \mathcal{B} , where \mathcal{A} and \mathcal{B} are equal.” Similarly, “ A represents T relative to \mathcal{A} ” means that A represents T relative to \mathcal{A} and \mathcal{B} , where \mathcal{A} and \mathcal{B} are equal.

Let us consider the case where $\mathbf{U} = \mathbf{V} = \mathbf{W}$ in Definition 5.21. This allows us to define positive integral powers of T inductively by $T^{k+1} = T^k \circ T$ for each positive integer k . We define T^0 to be the identity transformation of \mathbf{V} . In combination with Definition 5.2, this determines the value of each polynomial $a_r T^r + a_{r-1} T^{r-1} + \cdots + a_1 T + a_0 T^0$ in T with coefficients in \mathcal{F} , and such a polynomial is always a linear transformation of \mathbf{V} into \mathbf{V} . In such a polynomial we shall write a_0 in place of $a_0 T^0$.

If T has matrix A relative to the basis \mathcal{A} , then Theorems 5.20 and 5.23 show that $a_r T^r + a_{r-1} T^{r-1} + \cdots + a_1 T + a_0$ has matrix $a_r A^r + a_{r-1} A^{r-1} + \cdots + a_1 A + a_0 I$ relative to \mathcal{A} . Consequently, $\sum_{i=0}^r a_i T^i$ is the zero linear transformation if and only if $\sum_{i=0}^r a_i A^i$ is the zero matrix. This means that T and A satisfy the same polynomial equations.

A linear transformation T of \mathbf{U} into \mathbf{V} is called **invertible** or **nonsingular** if there is a mapping S of \mathbf{V} into \mathbf{U} such that $ST(\mathbf{u}) = \mathbf{u}$ for all $\mathbf{u} \in \mathbf{U}$ and $TS(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in \mathbf{V}$. Whenever such a mapping S exists, it is denoted by $S = T^{-1}$ and is called the **inverse** of T . It is left as an exercise to prove that the inverse of T is a linear transformation of \mathbf{V} into \mathbf{U} . It follows from Theorem 5.23 that if T is invertible and has matrix A relative to the bases \mathcal{A} of \mathbf{U} and \mathcal{B} of \mathbf{V} , then T^{-1} has matrix A^{-1} relative to \mathcal{B} and \mathcal{A} .

The next example shows how polynomials in matrices and linear transformations can sometimes give interesting and surprising results.

Example 3 □ Let T be the linear operator on \mathbf{R}^3 defined by

$$T(x_1, x_2, x_3) = (2x_1 + x_2, 2x_2, 2x_1 + 3x_2 + x_3).$$

Then T has the matrix

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

relative to the standard basis \mathcal{E}_3 . Straightforward computations show that A is a zero of the polynomial $x^3 - 5x^2 + 8x - 4$. That is,

$$\begin{aligned} A^3 - 5A^2 + 8A - 4I &= \begin{bmatrix} 8 & 0 & 14 \\ 12 & 8 & 31 \\ 0 & 0 & 1 \end{bmatrix} - 5 \begin{bmatrix} 4 & 0 & 6 \\ 4 & 4 & 11 \\ 0 & 0 & 1 \end{bmatrix} + 8 \begin{bmatrix} 2 & 0 & 2 \\ 1 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix} - 4 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

This equation implies that

$$A^3 - 5A^2 + 8A = 4I$$

and therefore

$$A \left(\frac{1}{4}A^2 - \frac{5}{4}A + 2I \right) = I.$$

By Theorem 3.14, A is invertible and

$$A^{-1} = \frac{1}{4}A^2 - \frac{5}{4}A + 2I.$$

It follows from this that T^{-1} exists and T^{-1} can be expressed as a polynomial in T :

$$T^{-1} = \frac{1}{4}T^2 - \frac{5}{4}T + 2.$$

The equation $A^3 - 5A^2 + 8A - 4I = \mathbf{0}$ also has some implications concerning positive integral powers of A . For instance,

$$A^3 = 5A^2 - 8A + 4I$$

and this implies that

$$A^4 = 5A^3 - 8A^2 + 4A.$$

Substituting for A^3 , we have

$$\begin{aligned} A^4 &= 5(5A^2 - 8A + 4I) - 8A^2 + 4A \\ &= 17A^2 - 36A + 20I. \end{aligned}$$

This substitution procedure can be repeated so as to express any higher integral power of A as a quadratic polynomial in A , and the corresponding powers of T can be expressed as quadratic polynomials in T . ■

Exercises 5.5

1. Let S and T be the linear transformations of \mathbf{R}^4 into \mathbf{R}^3 given by

$$S(x_1, x_2, x_3, x_4) = (5x_1 - 3x_3, x_2 + 6x_3 - x_4, 2x_1 - 9x_2 + 5x_3 + 2x_4)$$

and

$$T(x_1, x_2, x_3, x_4) = (4x_1 + 10x_2, 6x_1 - x_2 + 7x_4, -3x_1 + 8x_2 - 5x_4).$$

- (a) Find the matrices of S , T , and $S + T$ relative to \mathcal{E}_4 and \mathcal{E}_3 .
 - (b) Find the matrix of $2S - 3T$ relative to \mathcal{E}_4 and \mathcal{E}_3 .
2. Let S and T be as given in Problem 1. Consider the bases \mathcal{A} and \mathcal{B} of \mathbf{R}^4 and \mathbf{R}^3 , respectively, where $\mathcal{A} = \{(1, 1, 1, 1), (0, 1, 1, 1), (0, 0, 1, 1), (0, 0, 0, 1)\}$, and $\mathcal{B} = \{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$.
- (a) Find the matrices of S , T , and $S + T$ relative to \mathcal{A} and \mathcal{B} .
 - (b) Find the matrix of $2S - 3T$ relative to \mathcal{A} and \mathcal{B} .
3. Suppose that S and T are linear operators on \mathbf{R}^2 such that S has matrix $\begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$ and T has matrix $\begin{bmatrix} 3 & 0 \\ -2 & 1 \end{bmatrix}$ relative to the standard basis \mathcal{E}_2 . Find the matrix that represents TS relative to \mathcal{E}_2 .
4. Let $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ and $S : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be given by
- $$T(x, y) = (x + y, 2x), \quad S(x, y) = (3y, x - y).$$
- (a) Find a formula for $ST(x, y)$.
 - (b) Find matrix representations for each of T , S , and ST .
5. Suppose the linear operator $S : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ has the matrix $\begin{bmatrix} 5 & -2 \\ 1 & -3 \end{bmatrix}$ relative to the basis $\{(1, 2), (0, 1)\}$ and the linear transformation $T : \mathbf{R}^2 \rightarrow \mathbf{R}^3$ has the matrix $\begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$ relative to the bases $\{(1, 2), (0, 1)\}$ of \mathbf{R}^2 and \mathcal{E}_3 of \mathbf{R}^3 . Find a matrix representation for TS .
6. Find the matrix representation of S^{-1} in Problem 5 with respect to the basis $\{(1, 2), (0, 1)\}$.

7. Assume that $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a basis for \mathbf{R}^3 , and let

$$\mathbf{v}_1 = \mathbf{u}_2 + 2\mathbf{u}_3$$

$$\mathbf{v}_2 = \mathbf{u}_3$$

$$\mathbf{v}_3 = \mathbf{u}_1 + 2\mathbf{u}_2 + 5\mathbf{u}_3.$$

- (a) Prove that $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis, and express each \mathbf{u}_i as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$, and \mathbf{v}_3 .
- (b) Let $T : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ be defined by $T(\mathbf{u}_i) = \mathbf{v}_i$ for $i = 1, 2, 3$. Find the matrix that represents T relative to \mathcal{A} .
8. Let \mathcal{A}, \mathcal{B} and T be as defined in Problem 7. Define $S : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ by $S(\mathbf{v}_i) = \mathbf{u}_i$ for $i = 1, 2, 3$.
- (a) Find the matrix that represents S relative to \mathcal{B} .
- (b) Find the matrix that represents TS relative to \mathcal{A} .
9. Given that T is a linear operator on \mathbf{R}^2 with matrix $\begin{bmatrix} 3 & 2 \\ -4 & -2 \end{bmatrix}$ relative to \mathcal{E}_2 , find the matrix of $2T^3 + T^2 - 3T + 7$ relative to \mathcal{E}_2 .
10. Let T be the linear operator on \mathbf{R}^2 that has the matrix $\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ relative to \mathcal{E}_2 .
- (a) Find the matrix of $T^2 + 3T - 6$ relative to \mathcal{E}_2 .
- (b) Given that $T^2 + T - 6 = \mathbf{0}$, write T^{-1} as a polynomial in T .
- (c) Write T^3 as a first-degree polynomial in T .
11. Given that T is a linear operator satisfying the polynomial equation
- $$T^3 - 3T^2 + 4T + 6 = \mathbf{0},$$
- write each of T^4 and T^5 as polynomials in T with degree less than 3.
12. Suppose A is any matrix such that $A^2 - 5A + 12I = \mathbf{0}$.
- (a) Prove that A is invertible and that A^{-1} is a polynomial in A .
- (b) Given that $A = \begin{bmatrix} 1 & 2 \\ -4 & 4 \end{bmatrix}$ is such a matrix, find A^{-1} by writing A^{-1} as a polynomial in A .
13. Prove the associative property for multiplication of linear transformations.

14. Give an example which shows that it may happen that $ST \neq TS$, even when both ST and TS are defined.
15. Let T_1 and T_2 be linear transformations of \mathbf{U} into \mathbf{V} , and let S be a linear transformation of \mathbf{V} into \mathbf{W} . Suppose that S , T_1 , and T_2 have matrices A , B , and C , respectively, relative to certain bases of \mathbf{U} , \mathbf{V} , and \mathbf{W} . Prove that $S(T_1 + T_2)$ has matrix $AB + AC$ relative to these same bases.
16. Use Problem 15 to prove Theorem 5.24.
17. Use Theorem 3.6 and the definition of addition of matrices to prove Theorem 5.24.
18. Prove that if T is an invertible linear transformation of \mathbf{U} into \mathbf{V} , then $T^{-1}(\mathbf{u}) = \mathbf{v}$ if and only if $T(\mathbf{v}) = \mathbf{u}$.
19. Let T be an invertible linear transformation of \mathbf{U} into \mathbf{V} . Prove that T^{-1} is a linear transformation of \mathbf{V} into \mathbf{U} .
20. Let T be a linear operator on \mathbf{V} with matrix A relative to the basis \mathcal{A} of \mathbf{V} . Prove that T is invertible if and only if A is invertible.
21. Let S and T be invertible linear operators on \mathbf{V} . Prove that ST is invertible, and that $(ST)^{-1} = T^{-1}S^{-1}$.
22. Let T be the linear operator in Problem 11. Write T^{-1} as a polynomial in T .

Chapter 6

Determinants

6.1 Introduction

In this chapter, the fundamentals of the theory of determinants are developed. A knowledge of this material is necessary in the study of eigenvalues and eigenvectors of linear transformations, and many of the applications of linear algebra involve a use of eigenvalues and eigenvectors. These topics will be studied in Chapter 7.

6.2 Permutations and Indices

The definition of a determinant presented in Section 6.3 depends on the concept of a permutation of a set of integers and on certain properties of these permutations. A **permutation** of a set $\{x_1, x_2, \dots, x_n\}$ is simply an arrangement of the elements of the set into a particular sequence or order. For example, the arrangements 1, 4, 5, 3, 2, 6 and 6, 3, 2, 1, 4, 5 are permutations of the set $\{1, 2, 3, 4, 5, 6\}$. If an element x_i appears to the left of an element x_j in a permutation of $\{x_1, x_2, \dots, x_n\}$, then we say that x_i **precedes** x_j .

Our interest in permutations is limited to the permutations of the first n positive integers. The permutation $1, 2, 3, \dots, n$ is referred to as the **natural ordering**.

Definition 6.1 *The index of an integer j_k in a permutation j_1, j_2, \dots, j_n of $\{1, 2, \dots, n\}$ is the number of integers greater than j_k that precede j_k in the permutation j_1, j_2, \dots, j_n . The index of j_k is denoted by $\mathcal{I}(j_k)$.*

Example 1 \square In the permutation 1,4,5,3,2,6, the index of 3 is given by $\mathcal{I}(3) = 2$ since 4 and 5 are greater than 3 and precede 3 in the permutation. No integer greater than 4 precedes 4 in this permutation, so $\mathcal{I}(4) = 0$. Since the three integers 4, 5, and 3 are greater than 2 and precede 2 in the permutation, $\mathcal{I}(2) = 3$. ■

Definition 6.2 *The index of a permutation j_1, j_2, \dots, j_n of $\{1, 2, \dots, n\}$ is the integer \mathcal{I} given by*

$$\mathcal{I} = \sum_{k=1}^n \mathcal{I}(j_k),$$

where $\mathcal{I}(j_k)$ is the index of j_k in the permutation j_1, j_2, \dots, j_n .

That is, the index of a given permutation is the sum of the indices of all of the elements in that permutation.

Example 2 □ The index of the permutation 1,4,5,3,2,6 in Example 1 is

$$\begin{aligned}\mathcal{I} &= \mathcal{I}(1) + \mathcal{I}(4) + \mathcal{I}(5) + \mathcal{I}(3) + \mathcal{I}(2) + \mathcal{I}(6) \\ &= 0 + 0 + 0 + 2 + 3 + 0 \\ &= 5.\end{aligned}\quad \blacksquare$$

Theorem 6.3 *Any single interchange of adjacent elements in a permutation j_1, j_2, \dots, j_n of $\{1, 2, \dots, n\}$ changes the index of the permutation by 1.*

Proof. Let $\mathcal{I} = \sum_{k=1}^n \mathcal{I}(j_k)$ be the index of the given permutation

$$j_1, j_2, \dots, j_m, j_{m+1}, \dots, j_n$$

and consider the index \mathcal{I}' of the permutation

$$j_1, j_2, \dots, j_{m+1}, j_m, \dots, j_n$$

that results from interchanging j_m and j_{m+1} in the original permutation. Now

$$\mathcal{I}' = \sum_{k=1}^{m-1} \mathcal{I}'(j_k) + \mathcal{I}'(j_{m+1}) + \mathcal{I}'(j_m) + \sum_{k=m+2}^n \mathcal{I}'(j_k),$$

where $\mathcal{I}'(j_k)$ denotes the index of j_k in the new permutation. It is clear that $\mathcal{I}'(j_k) = \mathcal{I}(j_k)$ if k is different from m and $m+1$. If $j_m > j_{m+1}$, then $\mathcal{I}'(j_m) = \mathcal{I}(j_m)$, $\mathcal{I}'(j_{m+1}) = \mathcal{I}(j_{m+1}) - 1$, and consequently $\mathcal{I}' = \mathcal{I} - 1$. On the other hand, if $j_m < j_{m+1}$, then $\mathcal{I}'(j_m) = \mathcal{I}(j_m) + 1$ and $\mathcal{I}'(j_{m+1}) = \mathcal{I}(j_{m+1})$, so that $\mathcal{I}' = \mathcal{I} + 1$. In either case, the index is changed by 1, and the theorem is proven. ■

Theorem 6.3 paves the way for obtaining the corresponding result concerning the interchange of any two (not necessarily adjacent) elements in a permutation.

Theorem 6.4 *Any interchange of two elements in a permutation j_1, j_2, \dots, j_n of the set $\{1, 2, \dots, n\}$ changes the index by an odd integer.*

Proof. Let

$$j_1, j_2, \dots, j_r, \dots, j_s, \dots, j_n$$

be the given permutation, and consider the interchange of j_r and j_s . Let m be the number of elements between j_r and j_s .

Now the permutation that results from the interchange of j_r and j_s can be accomplished by using only interchanges of adjacent elements. The element j_r can be moved to the position initially occupied by j_s by $m + 1$ interchanges with the adjacent element on the right. Then j_s can be moved to the position that j_r initially occupied by m interchanges with the adjacent element on the left. Thus, the interchange of j_r and j_s can be accomplished by $2m + 1$ interchanges of adjacent elements. These $2m + 1$ interchanges cause $2m + 1$ changes of 1 in the index of the ordering, and consequently the index has changed by an odd number. ■

Example 3 □ As an illustration of the proof of Theorem 6.4, we shall accomplish the interchange of $j_r = 4$ and $j_s = 2$ in the permutation 1,4,5,3,2,6 by using only interchanges of adjacent elements. We note that there are $m = 2$ elements between 4 and 2.

$$\begin{aligned} 1, 4, 5, 3, 2, 6 &\rightarrow 1, 5, 4, 3, 2, 6 \\ &\rightarrow 1, 5, 3, 4, 2, 6 \\ &\rightarrow 1, 5, 3, 2, 4, 6 \\ &\rightarrow 1, 5, 2, 3, 4, 6 \\ &\rightarrow 1, 2, 5, 3, 4, 6 \end{aligned}$$

We first moved 4 to the position initially occupied by 2 using 3 interchanges with the adjacent element on the right, and then we moved 2 to the position that 4 initially occupied by using 2 interchanges with the adjacent element on the left. The total number of interchanges of adjacent elements was $3 + 2 = 5$, an odd integer. ■

The main objective of this section is to establish Theorem 6.7 for use in Section 6.3. The following two lemmas are basic to our proof of Theorem 6.7.

Lemma 6.5 *If a given permutation of $\{1, 2, \dots, n\}$ is carried into another permutation by an odd number of interchanges of elements, then the index of the given permutation differs from the index of the final permutation by an odd number.*

Proof. Suppose that a given permutation of $\{1, 2, \dots, n\}$ is carried into another permutation by an odd number of interchanges of elements. According to Theorem 6.4, each of the interchanges of elements changes the index by an odd number. Thus, the index of the original permutation differs from the index of the final permutation by an odd number, since the sum of an odd number of odd integers is an odd integer. ■

Lemma 6.6 *If a given permutation of $\{1, 2, \dots, n\}$ is carried into another permutation by an even number of interchanges of elements, then the index of the given permutation differs from the index of the final permutation by an even number.*

Proof. The proof is an exact parallel to that of Lemma 6.5, except that the sum of an even number of odd integers is an even integer. ■

Theorem 6.7 *The number of interchanges used to carry a permutation j_1, j_2, \dots, j_n of $\{1, 2, \dots, n\}$ into the natural ordering is either always odd or always even.*

Proof. Since the index of the natural ordering is zero, the difference in the indices of j_1, j_2, \dots, j_n and the natural ordering is the same as the index \mathcal{I} of j_1, j_2, \dots, j_n . If j_1, j_2, \dots, j_n can be carried into $1, 2, \dots, n$ by an odd number of interchanges, then \mathcal{I} must be odd by Lemma 6.5. If j_1, j_2, \dots, j_n can be carried into $1, 2, \dots, n$ by an even number of interchanges, then \mathcal{I} must be even by Lemma 6.6. Thus the number of interchanges used to carry j_1, j_2, \dots, j_n into $1, 2, \dots, n$ must always be odd if \mathcal{I} is odd and must always be even if \mathcal{I} is even. ■

Example 4 □ We shall determine whether the index of the permutation

$$6, 4, 1, 7, 5, 2, 3$$

is odd or even by counting the number of interchanges used to carry this permutation into the natural ordering.

$$\begin{aligned} 6, 4, 1, 7, 5, 2, 3 &\rightarrow 1, 4, 6, 7, 5, 2, 3 \\ &\rightarrow 1, 2, 6, 7, 5, 4, 3 \\ &\rightarrow 1, 2, 3, 7, 5, 4, 6 \\ &\rightarrow 1, 2, 3, 4, 5, 7, 6 \\ &\rightarrow 1, 2, 3, 4, 5, 6, 7 \end{aligned}$$

Because we have used an odd number of interchanges to carry the original permutation into the natural ordering, the index \mathcal{I} of

$$6, 4, 1, 7, 5, 2, 3$$

must be odd by Lemma 6.5, and the number of interchanges used to carry the original permutation into the natural ordering would always be odd by Theorem 6.7. Although always odd, this number of interchanges may be different from the index of the permutation. In this example, we used 5 interchanges of elements and the index of

$$6, 4, 1, 7, 5, 2, 3$$

is

$$\begin{aligned} \mathcal{I} &= \mathcal{I}(1) + \mathcal{I}(2) + \mathcal{I}(3) + \mathcal{I}(4) + \mathcal{I}(5) + \mathcal{I}(6) + \mathcal{I}(7) \\ &= 2 + 4 + 4 + 1 + 2 + 0 + 0 \\ &= 13. \end{aligned}$$
■

Exercises 6.2

1. Find the index of the following permutations.
 - (a) 5, 3, 1, 4, 2
 - (b) 2, 4, 5, 3, 1
 - (c) 6, 2, 4, 5, 3, 1
 - (d) 5, 3, 4, 2, 1, 6
 - (e) 3, 4, 6, 1, 2, 7, 5
 - (f) 2, 5, 1, 4, 3, 7, 6
2. Determine whether the index of the given permutation is odd or even by counting the number of interchanges used to carry the given permutation into the natural ordering.
 - (a) 3, 4, 2, 5, 1
 - (b) 5, 1, 3, 2, 4
 - (c) 6, 2, 4, 5, 3, 1
 - (d) 5, 2, 4, 6, 1, 3
 - (e) 4, 1, 7, 5, 2, 6, 3
 - (f) 3, 7, 5, 4, 6, 2, 1
3. Write out a sequence of interchanges of *adjacent* elements that will accomplish the interchange of the given pair of elements in the permutation 6,2,4,5,3,1.
 - (a) 2 and 3
 - (b) 6 and 3
 - (c) 2 and 1
 - (d) 4 and 1
4. Prove that the number of interchanges used to carry a permutation j_1, j_2, \dots, j_n of $\{1, 2, \dots, n\}$ into itself (i.e., into the same permutation) must always be even.
5. If $n \geq 2$, what is the index of the permutation $n, n - 1, \dots, 3, 2, 1$? Justify your answer.

6.3 The Definition of a Determinant

With the results of Section 6.2 at our disposal, we can proceed to our definition of a determinant. Only square matrices are considered in this chapter since the determinant is defined only for this type of matrix. We shall simplify our notation from $A = [a_{ij}]_{n \times n}$ to $A = [a_{ij}]_n$ or even $A = [a_{ij}]$ if the order of A is not significant.

Definition 6.8 *The determinant of the square matrix $A = [a_{ij}]_n$ over \mathcal{F} is the scalar $\det(A)$ defined by*

$$\det(A) = \sum_{(j)} (-1)^t a_{1j_1} a_{2j_2} \cdots a_{nj_n},$$

where $\sum_{(j)}$ denotes the sum of all terms of the form $(-1)^t a_{1j_1} a_{2j_2} \cdots a_{nj_n}$ as j_1, j_2, \dots, j_n assumes all possible permutations of the numbers $1, 2, \dots, n$, and the exponent t is the number of interchanges used to carry j_1, j_2, \dots, j_n into the natural ordering $1, 2, \dots, n$.

The notations $\det A$ and $|A|$ are used interchangeably with $\det(A)$. When the elements of $A = [a_{ij}]_n$ are written as a rectangular array, we would have

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Although the number of interchanges used to carry j_1, j_2, \dots, j_n into $1, 2, \dots, n$ is not always the same, Theorem 6.7 assures us that this number is either always even or always odd. Hence the sign $(-1)^t$ of each term is well-defined, and $\det(A)$ is uniquely determined by A .

We observe that there are $n!$ terms in the sum $\det(A)$ since there are $n!$ possible orderings of $1, 2, \dots, n$. The determinant of an $n \times n$ matrix is referred to as an $n \times n$ determinant, or a **determinant of order n** .

Example 1 □ Consider a 3×3 matrix

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

By the definition,

$$\begin{aligned} \det(A) = & (-1)^{t_1} a_{11}a_{22}a_{33} + (-1)^{t_2} a_{11}a_{23}a_{32} + (-1)^{t_3} a_{12}a_{23}a_{31} \\ & + (-1)^{t_4} a_{12}a_{21}a_{33} + (-1)^{t_5} a_{13}a_{22}a_{31} + (-1)^{t_6} a_{13}a_{21}a_{32}. \end{aligned}$$

Since $1, 2, 3$ is the natural ordering, we may take $t_1 = 0$. Since $1, 3, 2$ can be carried into $1, 2, 3$ by the single interchange of 2 and 3, we may take $t_2 = 1$. The ordering $2, 3, 1$ can be carried into $1, 2, 3$ by an interchange of 2 and 1, followed by an interchange of 2 and 3. Thus we may take $t_3 = 2$. By the same method, we find $t_4 = 1, t_5 = 1$, and $t_6 = 2$. Hence

$$\begin{aligned} \det(A) = & a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} \\ & - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} + a_{13}a_{21}a_{32}. \quad \blacksquare \end{aligned}$$

It is worth noting that the value of $\det(A)$ obtained in Example 1 agrees with the value yielded by evaluation routines that are taught in high school algebra. One of the most popular of these routines evaluates a 3×3 determinant $|A|$ by reproducing the first two columns and forming signed products according to the following diagram.

$$\begin{aligned} \det(A) = & \begin{array}{c} + + + \\ \diagup \diagdown \diagup \diagdown \diagup \diagdown \\ a_{11} a_{12} a_{13} \quad a_{21} a_{22} a_{23} \quad a_{31} a_{32} a_{33} \\ | | | \\ a_{11} a_{12} a_{13} \quad a_{21} a_{22} a_{23} \quad a_{31} a_{32} a_{33} \\ \diagdown \diagup \diagdown \diagup \diagdown \diagup \\ - - - \end{array} \\ = & a_{11} a_{22} a_{33} + a_{12} a_{23} a_{31} + a_{13} a_{21} a_{32} \\ & - a_{13} a_{22} a_{31} - a_{11} a_{23} a_{32} - a_{12} a_{21} a_{33} \end{aligned}$$

A similar diagram for a 2×2 determinant given by

$$\det(A) = \begin{vmatrix} + & a_{11} & a_{12} \\ - & a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21}$$

gives a correct value, but it is important to know that **there is no similar scheme that works for determinants of order 4 or any order greater than 3.**

Definition 6.8 is frequently referred to as the “row” definition of a determinant, since the row subscripts on the factors a_{ij_i} are held fixed in the natural ordering.

The next theorem presents an alternate formulation (the “column” definition) in which the column subscripts are held fixed in the natural ordering.

Theorem 6.9 *For any matrix $A = [a_{ij}]_n$, $\det(A)$ is given by*

$$\det(A) = \sum_{(i)} (-1)^s a_{i_1 1} a_{i_2 2} \cdots a_{i_n n},$$

where $\sum_{(i)}$ denotes the sum over all possible permutations i_1, i_2, \dots, i_n of $1, 2, \dots, n$, and s is the number of interchanges used to carry i_1, i_2, \dots, i_n into the natural ordering.

Proof. Let $S = \sum_{(i)} (-1)^s a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}$. Now both S and $\det(A)$ have $n!$ terms. Except possibly for sign, each term of S is a term of $\det(A)$, and each term of $\det(A)$ is a term of S . Thus, S and $\det(A)$ consist of the same terms, with a possible difference in sign.

Consider a certain term $(-1)^s a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}$ and let $(-1)^t a_{1j_1} a_{2j_2} \cdots a_{nj_n}$ be the corresponding term in $\det(A)$. Then $a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}$ can be carried into $a_{1j_1} a_{2j_2} \cdots a_{nj_n}$ by s interchanges of factors since the permutation i_1, i_2, \dots, i_n can be changed into the natural ordering $1, 2, \dots, n$ by s interchanges of elements. This means that the natural ordering $1, 2, \dots, n$ can be changed into the permutation j_1, j_2, \dots, j_n by s interchanges since the column subscripts have been interchanged each time the factors were interchanged. But j_1, j_2, \dots, j_n can be carried into $1, 2, \dots, n$ by t interchanges, by the definition of $\det(A)$. Thus $1, 2, \dots, n$ can be carried into j_1, j_2, \dots, j_n and then back into itself by $s+t$ interchanges. Since $1, 2, \dots, n$ can be carried into itself by an even number (zero) of interchanges, $s+t$ is even by Theorem 6.7. Therefore $(-1)^{s+t} = 1$ and $(-1)^s = (-1)^t$. Now we have the corresponding terms in $\det(A)$ and S with the same sign, and therefore $\det(A) = S$. ■

We recall that the transpose of an $m \times n$ matrix $A = [a_{ij}]$ is the $n \times m$ matrix A^T with a_{ji} as the element in the i^{th} row and j^{th} column.

Theorem 6.10 *If $A = [a_{ij}]_n$, then $\det(A^T) = \det(A)$.*

Proof. Let $B = A^T$, so that $b_{ij} = a_{ji}$ for all pairs i, j . Thus

$$\det(B) = \sum_{(j)} (-1)^t b_{1j_1} b_{2j_2} \cdots b_{nj_n}$$

by the definition of $\det(B)$, and

$$\det(B) = \sum_{(j)} (-1)^t a_{j_1 1} a_{j_2 2} \cdots a_{j_n n}.$$

Therefore $\det(B) = \det(A)$, by Theorem 6.9. ■

Exercises 6.3

1. Determine whether t is even or odd in the given term of $\det(A)$, where $A = [a_{ij}]_n$.
 - (a) $(-1)^t a_{13} a_{21} a_{34} a_{42}$
 - (b) $(-1)^t a_{14} a_{21} a_{33} a_{42}$
 - (c) $(-1)^t a_{14} a_{23} a_{32} a_{41}$
 - (d) $(-1)^t a_{12} a_{24} a_{31} a_{43}$
 - (e) $(-1)^t a_{24} a_{43} a_{12} a_{31}$
 - (f) $(-1)^t a_{21} a_{33} a_{14} a_{42}$
 - (g) $(-1)^t a_{34} a_{21} a_{42} a_{13}$
 - (h) $(-1)^t a_{34} a_{41} a_{12} a_{23}$
2. Write out the complete expression for $\det(A)$ if $A = [a_{ij}]_4$.
3. If A is a 3×3 matrix, find the value of $\det(A + A)$ in terms of $\det(A)$.
4. Work Problem 3 for an $n \times n$ matrix A .
5. If $A = [a_{ij}]_n$ is a diagonal matrix, what is the value of $\det(A)$?
6. Evaluate $\det(A)$ if $A = [\delta_{ij}]_n$, where δ_{ij} is the Kronecker delta.
7. If $A = [a_{ij}]_n$ and c is any scalar, express the value of $\det(cA)$ in terms of $\det(A)$.
8. Find the value of the given determinant by use of Definition 6.8.

| | | | |
|---|---|--|---|
| (a) $\begin{vmatrix} 4 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 6 \end{vmatrix}$ | (b) $\begin{vmatrix} 4 & -5 & 7 \\ 0 & -3 & -8 \\ 0 & 0 & 6 \end{vmatrix}$ | (c) $\begin{vmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ | (d) $\begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix}$ |
|---|---|--|---|

9. Prove or disprove each of the given statements.

- (a) $|A + B| = |A| + |B|$.
- (b) $|A + A^T| = |A| + |A^T|$.

6.4 Cofactor Expansions

A comparison of Definition 6.8 and Theorem 6.9 shows that there is a certain duality in the determinant of a matrix, since emphasis can be placed on either the row subscripts or the column subscripts. This duality is complete in that the entire theory can be developed with either point of emphasis. For each property formulated in terms of rows, there is a dual formulation in terms of columns, and vice versa. With a few adjustments, the derivation of a property from one point of view can be changed into a derivation from the other point of view. It is for this reason that we include derivations from only one point of view. The approach that is adopted here is to use a formulation in terms of rows. Each result stated in terms of rows has a dual result stated in terms of columns. In some instances it is necessary to state the dual result as a separate theorem. In others, the dual result is indicated by the insertion of the word "column" in parentheses at appropriate places. In these instances, the dual statement is obtained simply by replacing the word "row" with the word "column."

In the course of our development, it will become apparent that the elementary operations play a fundamental role in the theory. In addition, they furnish a valuable tool for use in the evaluation of determinants.

The following theorem describes the effect on the determinant if an elementary row or column operation of type III is applied to a matrix. This result is invaluable in the derivation of the cofactor expansion in Theorem 6.14.

Theorem 6.11 *If the square matrix B is obtained from the matrix A by an elementary row (column) operation of type III, then $\det(B) = -\det(A)$.*

Proof. Let B be obtained from A by interchanging rows u and v of A , so that $b_{uj} = a_{vj}$ and $b_{vj} = a_{uj}$ for all j . Then

$$\begin{aligned}\det(B) &= \sum_{(j)} (-1)^{t_1} b_{1j_1} \cdots b_{uj_u} \cdots b_{vj_v} \cdots b_{nj_n} \\ &= \sum_{(j)} (-1)^{t_1} a_{1j_1} \cdots a_{vj_u} \cdots a_{uj_v} \cdots a_{nj_n} \\ &= \sum_{(j)} (-1)^{t_1} a_{1j_1} \cdots a_{uj_v} \cdots a_{vj_u} \cdots a_{nj_n},\end{aligned}$$

where t_1 is the number of interchanges used to carry

$$j_1, \dots, j_u, \dots, j_v, \dots, j_n$$

into $1, 2, \dots, n$. Now

$$\det(A) = \sum_{(j)} (-1)^{t_2} a_{1j_1} \cdots a_{uj_v} \cdots a_{vj_u} \cdots a_{nj_n},$$

where t_2 is the number of interchanges used to carry $j_1, \dots, j_v, \dots, j_u, \dots, j_n$ into $1, 2, \dots, n$. Except for the exponents t_1 and t_2 , $\det(B)$ would be equal to $\det(A)$. Since only one

interchange is necessary to obtain $j_1, \dots, j_u, \dots, j_v, \dots, j_n$ from $j_1, \dots, j_v, \dots, j_u, \dots, j_n$, t_1 and t_2 must differ by an odd number, and $(-1)^{t_1} = (-1)(-1)^{t_2}$. Therefore $\det(B) = -\det(A)$, and the proof is complete. ■

The main purpose of this section is to establish an expression for the value of a determinant that is known as “an expansion by cofactors.” Some new notation and terminology are needed to state this expansion by cofactors.

Definition 6.12 *The minor of the element a_{ij} in $A = [a_{ij}]_n$ is the determinant M_{ij} of the $(n - 1) \times (n - 1)$ submatrix of A obtained by deleting row i and column j of A .*

Example 1 □ To obtain the minor of a_{12} in $A = [a_{ij}]_3$, we first delete row 1 and column 2 of A as indicated below.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

We then evaluate the determinant of the submatrix that remains. The minor M_{12} of a_{12} is given by

$$M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = a_{21}a_{33} - a_{23}a_{31}. \blacksquare$$

Definition 6.13 *The cofactor of a_{ij} in $A = [a_{ij}]_n$ is the product of $(-1)^{i+j}$ and the minor of a_{ij} . The cofactor of a_{ij} is denoted by A_{ij} .*

Example 2 □ The minor of $a_{32} = 2$ in

$$A = \left[a_{ij} \right]_3 = \begin{bmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix}$$

is

$$M_{32} = \begin{vmatrix} 7 & 9 \\ 4 & 6 \end{vmatrix} = 42 - 36 = 6,$$

and the cofactor of 2 is

$$A_{32} = (-1)^{3+2} M_{32} = -M_{32} = -6. \blacksquare$$

Our next theorem shows that the evaluation of an $n \times n$ determinant can be reduced to the evaluation of n determinants of order $n - 1$. This is of little practical use except when used in combination with elementary operations. Aside from this fact, however, the theorem has substantial theoretical value. The expression given in the theorem is referred to as “an expansion by cofactors” or more precisely as “the expansion about the i^{th} row.” This expansion is the main result of this section.

Theorem 6.14 *If $A = [a_{ij}]_n$, then*

$$\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}.$$

Proof. For a fixed integer i , we collect all of the terms in the sum $\det(A) = \sum_{(j)} (-1)^t a_{1j_1}a_{2j_2} \cdots a_{nj_n}$ that contain a_{i1} as a factor in one group, all of the terms that contain a_{i2} as a factor in another group, and so on for each column number. This separates the terms in $\det(A)$ into n groups with no overlapping since each term contains exactly one factor from row i . In each of the terms containing a_{i1} , we factor out a_{i1} and let F_{i1} denote the remaining factor. Repeating this process for each of $a_{i2}, a_{i3}, \dots, a_{in}$ in turn, we obtain

$$\det(A) = a_{i1}F_{i1} + a_{i2}F_{i2} + \cdots + a_{in}F_{in}.$$

To finish the proof, we need only show that $F_{ij} = A_{ij} = (-1)^{i+j} M_{ij}$, where M_{ij} is the minor of a_{ij} .

Consider first the case where $i = 1$ and $j = 1$. We shall show that $a_{11}F_{11} = a_{11}M_{11}$. Each term in F_{11} was obtained by factoring a_{11} from a term $(-1)^{t_1} a_{11}a_{2j_2} \cdots a_{nj_n}$ in the expansion of $\det(A)$. Thus each term in F_{11} has the form $(-1)^{t_1} a_{2j_2}a_{3j_3} \cdots a_{nj_n}$ where t_2 is the number of interchanges used to carry j_2, j_3, \dots, j_n into $2, 3, \dots, n$. Letting j_2, j_3, \dots, j_n range over all permutations of $2, 3, \dots, n$, we see that each of F_{11} and M_{11} has $(n - 1)!$ terms. Now $1, j_2, \dots, j_n$ can be carried into the natural ordering by the same interchanges used to carry j_2, \dots, j_n into $2, \dots, n$. That is, we may take $t_1 = t_2$. This means that F_{11} and M_{11} have exactly the same terms, yielding $F_{11} = M_{11}$ and $a_{11}F_{11} = a_{11}M_{11}$.

Consider now an arbitrary a_{ij} . By $i - 1$ interchanges of the original row i with the adjacent row above and then $j - 1$ interchanges of column j with the adjacent column on the left, we obtain a matrix B that has a_{ij} in the first row, first column position. Since the order of the remaining rows and columns of A was not changed, the minor of a_{ij} in B is the same M_{ij} as it is in A . By Theorem 6.11,

$$\det(B) = (-1)^{i-1+j-1} \det(A) = (-1)^{i+j} \det(A).$$

This gives $\det(A) = (-1)^{i+j} \det(B)$.

The sum of all the terms in $\det(B)$ that contain a_{ij} as a factor is $a_{ij}M_{ij}$, from our first case. Since $\det(A) = (-1)^{i+j} \det(B)$, the sum of all the terms in $\det(A)$ that contain a_{ij} as a factor is $(-1)^{i+j} a_{ij}M_{ij}$. Thus $a_{ij}F_{ij} = (-1)^{i+j} a_{ij}M_{ij} = a_{ij}A_{ij}$, and the theorem is proved. ■

The dual statement for Theorem 6.14 is as follows.

Theorem 6.15 *If $A = [a_{ij}]_n$, then*

$$\det(A) = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj}.$$

Example 3 □ If $A = [a_{ij}]_3$, the expansion of $\det(A)$ about the 2nd row is given by

$$\begin{aligned} \left| \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right| &= a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23} \\ &= a_{21}(-1)^{2+1} \left| \begin{array}{cc} a_{12} & a_{13} \\ a_{32} & a_{33} \end{array} \right| + a_{22}(-1)^{2+2} \left| \begin{array}{cc} a_{11} & a_{13} \\ a_{31} & a_{33} \end{array} \right| \\ &\quad + a_{23}(-1)^{2+3} \left| \begin{array}{cc} a_{11} & a_{12} \\ a_{31} & a_{32} \end{array} \right| \\ &= -a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{22}(a_{11}a_{33} - a_{13}a_{31}) \\ &\quad - a_{23}|a_{11}a_{32} - a_{12}a_{31}|. \end{aligned}$$

This value agrees with that given in Example 1 of Section 6.3. ■

Theorem 6.14 proves to be extremely useful. For instance, it provides the key step in establishing our next theorem.

Theorem 6.16 *The expression $c_1A_{i1} + c_2A_{i2} + \cdots + c_nA_{in}$ is equal to the determinant of a matrix which is the same as $A = [a_{ij}]_n$ except that the elements a_{ij} of the i^{th} row have been replaced by the scalars c_j .*

Proof. By Theorem 6.14,

$$\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}.$$

Let B be obtained from A by replacing a_{i1} by c_1 , a_{i2} by c_2 , a_{in} by c_n . Then $\det(B)$ can be found from the above expansion by replacing each a_{ik} by c_k for $k = 1, 2, \dots, n$. In evaluating A_{ik} , the i^{th} row and k^{th} column are deleted from A . Thus, the values of the A_{ik} do not depend on any of the elements $a_{i1}, a_{i2}, \dots, a_{in}$. Therefore, the values of the A_{ik} do not change when each a_{ik} is replaced by c_k , and

$$\det(B) = c_1A_{i1} + c_2A_{i2} + \cdots + c_nA_{in}. \blacksquare$$

The result parallel to Theorem 6.16 has the following formulation in terms of columns.

Theorem 6.17 *The expression $c_1A_{1j} + c_2A_{2j} + \cdots + c_nA_{nj}$ is equal to the determinant of a matrix which is the same as $A = [a_{ij}]_n$ except that the elements a_{ij} of the j^{th} column have been replaced by the scalars c_i .*

Theorem 6.18 *The determinant of a matrix $A = [a_{ij}]_n$ that has two identical rows (columns) is zero.*

Proof. In contrast to the development thus far in this chapter, we must take into account the field \mathcal{F} that contains the elements a_{ij} of A .

Suppose first that $1 + 1 \neq 0$ in \mathcal{F} .¹ If the u^{th} and v^{th} rows of A are identical, let B be the matrix formed from A by the interchange of the u^{th} and v^{th} rows. Then $B = A$, but $\det(B) = -\det(A)$ by Theorem 6.11. Thus we have $\det(A) = -\det(A)$, and $(1 + 1)\det(A) = 0$. Since $1 + 1 \neq 0$, $\det(A) = 0$.

Consider now the case where $1 + 1 = 0$ in \mathcal{F} . According to Theorem 6.11, an interchange of two rows in A produces a matrix whose determinant has the value $-\det(A)$. But $\det(A) = -\det(A)$ since $\det(A)$ is in \mathcal{F} and $c + c = (1 + 1)c = 0$ for all c in \mathcal{F} . Thus, an interchange of rows does not change the value of the determinant, and there is no loss of generality if we assume that the first two rows are equal. That is, $a_{1j} = a_{2j}$ for all j . Since $-1 = 1$ in \mathcal{F} , $\det(A) = \sum_{(j)} a_{1j_1}a_{2j_2}\cdots a_{nj_n}$. For each term $a_{1j_1}a_{2j_2}a_{3j_3}\cdots a_{nj_n}$ in $\det(A)$, there is a corresponding term $a_{2j_1}a_{1j_2}a_{3j_3}\cdots a_{nj_n}$. If we group these terms in pairs and sum over only those orderings with $j_1 < j_2$, we have

$$\begin{aligned}\det(A) &= \sum_{j_1 < j_2} (a_{1j_1}a_{2j_2}a_{3j_3}\cdots a_{nj_n} + a_{1j_2}a_{2j_1}a_{3j_3}\cdots a_{nj_n}) \\ &= \sum_{j_1 < j_2} (a_{1j_1}a_{2j_2}a_{3j_3}\cdots a_{nj_n} + a_{2j_2}a_{1j_1}a_{3j_3}\cdots a_{nj_n}) \\ &= \sum_{j_1 < j_2} (1 + 1)a_{1j_1}a_{2j_2}a_{3j_3}\cdots a_{nj_n} \\ &= 0.\end{aligned}$$

This completes the proof. ■

Theorem 6.19 *The sum of the products of the elements of a row of $A = [a_{ij}]_n$ by the cofactors of the corresponding elements of a different row of A is zero. Hence*

$$a_{i1}A_{k1} + a_{i2}A_{k2} + \cdots + a_{in}A_{kn} = \delta_{ik} \det(A),$$

where δ_{ik} is the Kronecker delta.

Proof. If $i = k$, the equality follows from Theorem 6.14.

Suppose that $i \neq k$. According to Theorem 6.16,

$$a_{i1}A_{k1} + a_{i2}A_{k2} + \cdots + a_{in}A_{kn}$$

is the determinant of a matrix B that is the same as A except that the elements of row k have been replaced by $a_{i1}, a_{i2}, \dots, a_{in}$. Thus the matrix B has the i^{th} and k^{th} rows identical, and $\det(B) = 0$ by Theorem 6.18. ■

The dual statement of Theorem 6.19 is given next.

¹A field in which $1 + 1 = 0$ is a field of characteristic 2. For a discussion of characteristics of a field, the reader may consult any standard text in abstract algebra.

Theorem 6.20 *The sum of the products of the elements of a column of $A = [a_{ij}]_n$ by the cofactors of the corresponding elements of a different column is zero. Hence*

$$a_{1j}A_{1k} + a_{2j}A_{2k} + \cdots + a_{nj}A_{nk} = \delta_{jk} \det(A).$$

Exercises 6.4

1. Prove that if $A = [a_{ij}]_n$ has a row with all elements zero, then $\det(A) = 0$.
2. Compute the cofactor of the indicated element a_{ij} in $A = [a_{ij}]$, where

$$A = \begin{bmatrix} 2 & -2 & 2 & 1 \\ 2 & -1 & -2 & -1 \\ 0 & 2 & -4 & -6 \\ 2 & -3 & 10 & 4 \end{bmatrix}.$$

- (a) a_{23} (b) a_{34} (c) a_{42} (d) a_{31} (e) a_{14} (f) a_{32}

3. Evaluate $\det(A)$ with A as in Problem 2.

4. Let $A = [a_{ij}]$ be given by

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

- (a) Find the matrix $B = [A_{ij}]^T$, where A_{ij} is the cofactor of a_{ij} .
- (b) Compute AB , where A and B are as in part (a).
- (c) Use the results of parts (a) and (b) to find A^{-1} .
5. For an arbitrary matrix $A = [a_{ij}]_3$, let $B = [A_{ij}]^T$.
 - (a) Evaluate AB .
 - (b) From the product AB in part (a), deduce a formula for A^{-1} whenever $\det(A) \neq 0$ and $A = [a_{ij}]_3$.
6. Evaluate the given determinant by a cofactor expansion.

$$(a) \begin{vmatrix} 2 & -1 & 2 & 1 \\ 0 & 1 & -4 & -2 \\ 0 & 0 & 4 & -2 \\ 0 & 0 & 4 & 1 \end{vmatrix} \quad (b) \begin{vmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{vmatrix}$$

$$(c) \begin{vmatrix} 2 & 3 & 0 & -1 \\ -4 & -6 & 0 & 3 \\ 2 & 1 & -1 & 2 \\ 0 & -2 & -1 & 3 \end{vmatrix} \quad (d) \begin{vmatrix} 1 & 2 & 1 & 2 \\ 3 & 4 & -1 & 5 \\ -2 & 2 & -1 & -1 \\ 1 & -3 & -2 & -1 \end{vmatrix}$$

7. Find the values of x for which

$$\begin{vmatrix} 1-x & -1 & -1 \\ -1 & 1-x & -1 \\ -1 & -1 & 1-x \end{vmatrix} = 0.$$

8. Prove Theorem 6.11 for an elementary column operation of type III.
9. Prove Theorem 6.15.
10. Prove Theorem 6.17.
11. Prove the dual statement of Theorem 6.18.
12. Prove Theorem 6.20.

6.5 Elementary Operations and Cramer's Rule

In Theorem 6.11 of the preceding section, we have seen that the application of a single operation of type III to a square matrix has the effect of changing the sign of the determinant. We propose now to investigate the effects of the other types of elementary operations, and to indicate the usefulness of these operations in combination with the cofactor expansions.

For an elementary operation of type I, we have the following result.

Theorem 6.21 *If the square matrix B is obtained from the matrix A by an elementary row (column) operation of type I that multiplies every element of a row (column) by $c \neq 0$, then $\det(B) = c \det(A)$.*

Proof. Suppose that B is obtained from $A = [a_{ij}]_n$ by multiplying each entry of the k^{th} row by $c \neq 0$.

By Definition 6.8,

$$\det(B) = \sum_{(j)} (-1)^t b_{1j_1} b_{2j_2} \cdots b_{kj_k} \cdots b_{nj_n},$$

and since $b_{ij} = a_{ij}$ if $i \neq k$ and $b_{kj_k} = ca_{kj_k}$, we have

$$\begin{aligned} \det(B) &= c \sum_{(j)} (-1)^t a_{1j_1} a_{2j_2} \cdots a_{kj_k} \cdots a_{nj_n} \\ &= c \det(A). \end{aligned} \quad \blacksquare$$

For elementary operations of type II, we have the following description.

Theorem 6.22 *If B is obtained from the square matrix A by an elementary row (column) operation of type II, then $\det(B) = \det(A)$.*

Proof. Let $A = [a_{ij}]_n$ and suppose that $B = [b_{ij}]_n$ is formed by adding to each element a_{uj} of the u^{th} row of A the product of the scalar c and the corresponding element a_{vj} of the v^{th} row of A ($u \neq v$). Then B and A are the same except in the u^{th} rows, and the cofactor of b_{uj} in the u^{th} row of B is the same as the cofactor A_{uj} of the corresponding element a_{uj} in A . When $\det(B)$ is expanded about the u^{th} row, we find

$$\begin{aligned}\det(B) &= b_{u1}A_{u1} + b_{u2}A_{u2} + \cdots + b_{un}A_{un} \\ &= (a_{u1} + ca_{v1})A_{u1} + (a_{u2} + ca_{v2})A_{u2} + \cdots + (a_{un} + ca_{vn})A_{un} \\ &= a_{u1}A_{u1} + a_{u2}A_{u2} + \cdots + a_{un}A_{un} \\ &\quad + c(a_{v1}A_{u1} + a_{v2}A_{u2} + \cdots + a_{vn}A_{un}).\end{aligned}$$

By Theorem 6.14 and 6.19,

$$\det(B) = \det(A) + c(0) = \det(A). \blacksquare$$

Example 1 \square As an illustration of the usefulness of the elementary operations, consider the evaluation of the determinant

$$\det(A) = \begin{vmatrix} 5 & 2 & 2 & 15 \\ 2 & 2 & -4 & 6 \\ 2 & -4 & 2 & 6 \\ 0 & 5 & 7 & 1 \end{vmatrix}.$$

By Theorem 6.21,

$$\det(A) = 2 \begin{vmatrix} 5 & 2 & 2 & 15 \\ 1 & 1 & -2 & 3 \\ 2 & -4 & 2 & 6 \\ 0 & 5 & 7 & 1 \end{vmatrix}.$$

According to Theorem 6.22, the value of the determinant is unchanged if we (a) add to row 1 the product of -5 and row 2 and (b) add to row 3 the product of -2 and row 2. Thus

$$\det(A) = 2 \begin{vmatrix} 0 & -3 & 12 & 0 \\ 1 & 1 & -2 & 3 \\ 0 & -6 & 6 & 0 \\ 0 & 5 & 7 & 1 \end{vmatrix}.$$

Expanding about the first column, we have

$$\det(A) = 2(-1)^{2+1} \begin{vmatrix} -3 & 12 & 0 \\ -6 & 6 & 0 \\ 5 & 7 & 1 \end{vmatrix} = (-2)(-3)(6) \begin{vmatrix} 1 & -4 & 0 \\ -1 & 1 & 0 \\ 5 & 7 & 1 \end{vmatrix}.$$

Expanding now about the third column, we have

$$\det(A) = (-2)(-3)(6)(-1)^{3+3} \begin{vmatrix} 1 & -4 \\ -1 & 1 \end{vmatrix} = -108. \blacksquare$$

Our final result of this section makes an important connection between determinants and the solution of certain types of systems of linear equations. This theorem presents a formula for the unknowns in terms of certain determinants. This formula is commonly known as **Cramer's rule**.

Theorem 6.23 Consider a system of linear equations $AX = B$, in which $A = [a_{ij}]_{n \times n}$, $X = [x_1, x_2, \dots, x_n]^T$, and $B = [b_1, b_2, \dots, b_n]^T$. If $\det(A) \neq 0$, the unique solution of the system is given by

$$x_j = \frac{\sum_{k=1}^n b_k A_{kj}}{\det(A)}, \quad j = 1, 2, \dots, n.$$

Proof. We first show that the given values are a solution of the system. Substitution of these values for x_j into the left member of the i^{th} equation of the system yields

$$\begin{aligned} a_{i1}x_1 + \cdots + a_{in}x_n &= \frac{1}{\det(A)} \left(a_{i1} \left(\sum_{k=1}^n b_k A_{k1} \right) + \cdots + a_{in} \left(\sum_{k=1}^n b_k A_{kn} \right) \right) \\ &= \frac{1}{\det(A)} \sum_{j=1}^n \left(\sum_{k=1}^n a_{ij} b_k A_{kj} \right) \\ &= \frac{1}{\det(A)} \sum_{k=1}^n \left(b_k \sum_{j=1}^n a_{ij} A_{kj} \right) \\ &= \frac{1}{\det(A)} \sum_{k=1}^n b_k (\delta_{ik} \det(A)) \\ &= \frac{1}{\det(A)} b_i (\delta_{ii} \det(A)) \\ &= b_i. \end{aligned}$$

Thus, the values

$$x_j = \frac{\sum_{k=1}^n b_k A_{kj}}{\det(A)}$$

furnish a solution of the system.

To prove uniqueness, suppose that $x_j = y_j, j = 1, 2, \dots, n$, represents any solution to the system. Then the i^{th} equation $\sum_{k=1}^n a_{ik} y_k = b_i$ is satisfied for $i = 1, 2, \dots, n$. If we

multiply both members of the i^{th} equation by A_{ij} (j fixed) and form the sum of these equations, we find that

$$\sum_{i=1}^n \left(\sum_{k=1}^n a_{ik} A_{ij} y_k \right) = \sum_{i=1}^n b_i A_{ij}$$

or

$$\sum_{k=1}^n \left(\sum_{i=1}^n a_{ik} A_{ij} \right) y_k = \sum_{i=1}^n b_i A_{ij}.$$

But, for each k , $\sum_{k=1}^n a_{ik} A_{ij} = \delta_{kj} \det(A)$. Thus

$$\sum_{k=1}^n \delta_{kj} \det(A) y_k = \sum_{i=1}^n b_i A_{ij},$$

and

$$y_j = \frac{\sum_{i=1}^n b_i A_{ij}}{\det(A)}.$$

Hence these y_j 's are the same as the solution given in the statement of the theorem. ■

We note that the sum $\sum_{k=1}^n b_k A_{kj}$ is the determinant of the matrix obtained by replacing the j^{th} column of A by the column of constants $B = [b_1, b_2, \dots, b_n]^T$.

Example 2 □ For a system in three unknowns with $\det(A) \neq 0$,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3,$$

the solutions stated in Theorem 6.23 can be written as

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{|A|}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{|A|}, \quad x_3 = \frac{\begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}}{|A|}. \quad \blacksquare$$

Exercises 6.5

1. Evaluate the given determinant by an appropriate combination of elementary operations and expansion by cofactors.

| | | | |
|--|--|---|--|
| (a) $\begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 3 & 4 & -2 \end{vmatrix}$ | (b) $\begin{vmatrix} -4 & -5 & 3 \\ 3 & 3 & -2 \\ -1 & -1 & 1 \end{vmatrix}$ | (c) $\begin{vmatrix} a & -b \\ b & a \end{vmatrix}$ | (d) $\begin{vmatrix} a-b & 2a-b \\ b-a & 2b-a \end{vmatrix}$ |
|--|--|---|--|

$$(e) \begin{vmatrix} 4 & -2 & -9 \\ 7 & 2 & 10 \\ 4 & 1 & -3 \end{vmatrix} \quad (f) \begin{vmatrix} 2 & 4 & -3 \\ 4 & 8 & 9 \\ 4 & 4 & -9 \end{vmatrix} \quad (g) \begin{vmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & -2 \\ 4 & 6 & 1 & 4 \\ 8 & 9 & 1 & -8 \end{vmatrix} \quad (h) \begin{vmatrix} -1 & 1 & 0 & 2 \\ 10 & 2 & 3 & -1 \\ 4 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

2. Use Cramer's Rule to evaluate two of the unknowns, and then find the remaining unknown by substitution of these values into one of the equations.

| | |
|------------------------------|------------------------------|
| (a) $2x_1 + 2x_3 = 2$ | (b) $4x_1 - 2x_2 + 9x_3 = 1$ |
| $x_2 - 3x_3 = -4$ | $7x_1 + 2x_2 - 10x_3 = 0$ |
| $2x_1 + x_2 + x_3 = 6$ | $4x_1 + x_2 + 3x_3 = 2$ |
| | |
| (c) $4x_1 + 2x_2 + 4x_3 = 2$ | (d) $3x_1 - 4x_2 + 6x_3 = 1$ |
| $6x_1 - 5x_2 + 8x_3 = -3$ | $9x_1 + 8x_2 - 12x_3 = 3$ |
| $3x_2 + 2x_3 = 5$ | $9x_1 - 4x_2 + 12x_3 = 4$ |
| | |
| (e) $2x_1 - x_2 + 3x_3 = 17$ | (f) $2x_1 + 4x_2 + x_3 = 17$ |
| $5x_1 - 2x_2 + 4x_3 = 28$ | $5x_1 + 3x_2 - x_3 = 4$ |
| $3x_1 + 3x_2 - x_3 = -1$ | $2x_1 - 7x_2 + 3x_3 = -36$ |

3. Explain how one can use elementary operations to conclude that

$$\text{if } \begin{vmatrix} 2 & 5 & -6 \\ -1 & 2 & 4 \\ 3 & 1 & 5 \end{vmatrix} = 139, \text{ then } \begin{vmatrix} 4 & 15 & -6 \\ -2 & 6 & 4 \\ 6 & 3 & 5 \end{vmatrix} = 834.$$

4. Solve for x :

$$\begin{vmatrix} x & 0 & 0 & -8 \\ -1 & x & 0 & -10 \\ 0 & -1 & x & -1 \\ 0 & 0 & -1 & 1 \end{vmatrix} = 0.$$

5. Show that

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (a-b)(b-c)(c-a).$$

6. Show that

$$\begin{vmatrix} a & a & a \\ a & b & b \\ a & b & c \end{vmatrix} = a(a-b)(b-c).$$

7. Evaluate the following determinant.

$$\begin{vmatrix} (a+b)^2 & c^2 & c^2 \\ a^2 & (b+c)^2 & a^2 \\ b^2 & b^2 & (a+c)^2 \end{vmatrix}$$

8. Show that

$$\begin{vmatrix} -x & 1 & 0 & 0 \\ 0 & -x & 1 & 0 \\ 0 & 0 & -x & 1 \\ -c_0 & -c_1 & -c_2 & -c_3 - x \end{vmatrix} = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + x^4.$$

9. Show that

$$\begin{vmatrix} a & b & c \\ d & e & f \\ h & i & j \end{vmatrix} + \begin{vmatrix} a & b & c \\ d & e & f \\ k & m & n \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ h+k & i+m & j+n \end{vmatrix}.$$

10. Prove that an equation of the straight line through two distinct points with rectangular coordinates (x_1, y_1) and (x_2, y_2) is given by

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0.$$

11. Suppose that f_{11}, f_{12}, f_{21} , and f_{22} are differentiable functions of x and that g is defined by

$$g(x) = \begin{vmatrix} f_{11}(x) & f_{12}(x) \\ f_{21}(x) & f_{22}(x) \end{vmatrix}.$$

Use calculus formulas and show that

$$g' = \begin{vmatrix} f'_{11} & f'_{12} \\ f_{21} & f_{22} \end{vmatrix} + \begin{vmatrix} f_{11} & f_{12} \\ f'_{21} & f'_{22} \end{vmatrix}.$$

12. Use Theorem 6.21 to express $|cA|$ in terms of $|A|$ for $A = [a_{ij}]_n$.
13. Prove the dual statement of Theorem 6.21.
14. Prove the dual statement of Theorem 6.22.

6.6 Determinants and Matrix Multiplication

We turn now to an investigation of the determinant of a product of two matrices. Our principal goal is to obtain the result that $\det(AB) = \det(A)\det(B)$. Although the statement of this result is simple enough, the derivation is a bit involved. We adopt an approach that utilizes elementary matrices in this derivation.

Theorem 6.24 *The determinant of any elementary matrix is not zero.*

Proof. Let M be an elementary matrix. If M is of type I, then $|M| = c|I_n| = c$, where $c \neq 0$, by Theorem 6.21. If M is of type II, then $|M| = |I_n| = 1$ by Theorem 6.22. If M is of type III, then $|M| = -|I_n| = -1$ by Theorem 6.11. ■

Theorem 6.25 *If A and M are $n \times n$ matrices and M is an elementary matrix, then*

$$\det(MA) = \det(M)\det(A).$$

Proof. We recall that multiplying A on the left by an elementary matrix M performs the same row operation on A as that used to obtain M from the identity matrix I_n (see page 92).

If M is of type I as in Theorem 6.21, then

$$\det(MA) = c\det(A) = \det(M)\det(A),$$

where $c \neq 0$. If M is of type II, then

$$\det(MA) = \det(A) = \det(M)\det(A),$$

by Theorem 6.22. If M is of type III as in Theorem 6.11, then

$$\det(MA) = -\det(A) = (-1)\det(A) = \det(M)\det(A).$$

Thus the theorem is valid in all cases. ■

The result of the theorem extends readily to the following corollary, the proof of which is left as an exercise (Problem 6).

Corollary 6.26 *If M_1, M_2, \dots, M_k are elementary $n \times n$ matrices and A is $n \times n$, then*

$$\det(M_1 M_2 \cdots M_k A) = \det(M_1) \det(M_2) \cdots \det(M_k) \det(A).$$

Corollary 6.27 *If M_1, M_2, \dots, M_k are elementary $n \times n$ matrices, then*

$$\det(M_1 M_2 \cdots M_k) = \det(M_1) \det(M_2) \cdots \det(M_k).$$

Proof. The asserted equality follows at once from Corollary 6.26 with $A = I_n$. ■

Corollary 6.27 enables us to extend the result of Theorem 6.24 to arbitrary invertible matrices, and we are also able to establish the converse.

Theorem 6.28 *An $n \times n$ matrix A is invertible if and only if $\det(A) \neq 0$.*

Proof. Assume first that A is invertible. Then A is a product of elementary matrices (Theorem 3.19), say $A = M_1 M_2 \cdots M_k$. From Corollary 6.27,

$$\det(A) = \det(M_1) \det(M_2) \cdots \det(M_k),$$

and this product is not zero by Theorem 6.24.

Assume now that $\det(A) \neq 0$. Let P be an invertible matrix such that PA is in reduced row-echelon form. Now P is a product of elementary matrices, $P = M_1 M_2 \cdots M_k$, so that

$$\det(PA) = \det(M_1 M_2 \cdots M_k A) = \det(M_1) \det(M_2) \cdots \det(M_k) \det(A).$$

Since $\det(A) \neq 0$ and each $\det(M_i) \neq 0$, $\det(PA) \neq 0$. This implies that PA does not have a row of zeros, and consequently is of rank n by Corollary 3.49. But A and PA have the same rank since they are equivalent, so this means that A has rank n . It follows from Corollary 3.50 that A is invertible. ■

When the theory of matrices and determinants was in its early stages of development, the word *nonsingular* was normally used instead of invertible, and the word *singular* was used to indicate that a square matrix did not have an inverse. The last theorem gives the basis for the early terminology: singular corresponds to zero determinant, and nonsingular corresponds to nonzero determinant.

We are now in a position to prove the main theorem in this section.

Theorem 6.29 *If A and B are $n \times n$ matrices, then*

$$\det(AB) = \det(A) \det(B).$$

Proof. If A is not invertible, then AB is not invertible (for AB invertible implies $AB(AB)^{-1} = I_n$, which in turn implies A is invertible by Theorem 3.14). By Theorem 6.28, $\det(A) = 0$ and $\det(AB) = 0$ so that

$$\det(AB) = \det(A) \det(B)$$

in this case.

Assume that A is invertible. Then A is a product of elementary matrices, $A = M_1 M_2 \cdots M_k$, and Corollary 6.26 implies that

$$\begin{aligned} \det(AB) &= \det(M_1 M_2 \cdots M_k B) \\ &= \det(M_1) \det(M_2) \cdots \det(M_k) \det(B). \end{aligned}$$

But

$$\det(A) = \det(M_1) \det(M_2) \cdots \det(M_k)$$

by Corollary 6.27, and we have

$$\det(AB) = \det(A) \det(B). \blacksquare$$

The next theorem is of great interest since it provides a formula for the computation of the inverse of a matrix. Unfortunately the use of this formula is not very practical for matrices of higher orders, due to the large number of computations that are necessary.

The formula is conveniently expressed in terms of the adjoint of a matrix, defined as follows.

Definition 6.30 If $A = [a_{ij}]_n$, then the **adjoint** of A , denoted by $\text{adj}(A)$, is the matrix given by $\text{adj}(A) = [A_{ij}]_n^T$, where A_{ij} is the cofactor of a_{ij} in A .

Example 1 □ Let

$$A = [a_{ij}]_3 = \begin{bmatrix} 11 & -6 & 2 \\ 3 & -2 & 1 \\ 2 & -2 & 2 \end{bmatrix}.$$

The cofactor of 11 in A is

$$A_{11} = (-1)^{1+1} \begin{vmatrix} -2 & 1 \\ -2 & 2 \end{vmatrix} = -4 + 2 = -2,$$

and the cofactor of -6 is

$$A_{12} = (-1)^{1+2} \begin{vmatrix} 3 & 1 \\ 2 & 2 \end{vmatrix} = (-1)(6 - 2) = -4.$$

Continuing in this fashion, we find that

$$\text{adj}(A) = \begin{bmatrix} -2 & -4 & -2 \\ 8 & 18 & 10 \\ -2 & -5 & -4 \end{bmatrix}^T = \begin{bmatrix} -2 & 8 & -2 \\ -4 & 18 & -5 \\ -2 & 10 & -4 \end{bmatrix}. \blacksquare$$

Theorem 6.31 If $A = [a_{ij}]_n$ is invertible, then $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$.

Proof. Let $A \text{adj}(A) = [c_{ij}]_n$. Now

$$A \text{adj}(A) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix},$$

and hence $c_{ij} = \sum_{k=1}^n a_{ik} A_{jk}$. By Theorem 6.19, $\sum_{k=1}^n a_{ik} A_{jk} = \delta_{ij} \det(A)$. Thus

$$A \text{adj}(A) = \det(A) I_n.$$

Since $\det(A) \neq 0$, this yields

$$A \left(\frac{1}{\det(A)} \text{adj}(A) \right) = I_n,$$

and it follows that

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A). \blacksquare$$

Corollary 6.32 For any $A = [a_{ij}]_n$, $A \text{adj}(A) = \det(A) I_n$.

Example 2 \square Using the matrices A and $\text{adj}(A)$ from Example 1, we have

$$\det(A) = \begin{vmatrix} 11 & -6 & 2 \\ 3 & -2 & 1 \\ 2 & -2 & 2 \end{vmatrix} = -2$$

and

$$A^{-1} = \frac{1}{-2} \begin{bmatrix} -2 & -4 & -2 \\ 8 & 18 & 10 \\ -2 & -5 & -4 \end{bmatrix}^T = \begin{bmatrix} 1 & -4 & 1 \\ 2 & -9 & \frac{5}{2} \\ 1 & -5 & 2 \end{bmatrix}. \blacksquare$$

Example 3 \square The formula in Theorem 6.31 has its greatest usefulness with 2×2 matrices. If

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

and $ad - bc \neq 0$, then

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \blacksquare$$

Exercises 6.6

1. Use the formula of Theorem 6.31 to find the inverse of the given matrix.

$$\begin{array}{lll} \text{(a)} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} & \text{(b)} \begin{bmatrix} -4 & -5 & 3 \\ 3 & 3 & -2 \\ -1 & -1 & 1 \end{bmatrix} & \text{(c)} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 3 & 4 & -2 \end{bmatrix} \\ \text{(d)} \begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{bmatrix} & \text{(e)} \begin{bmatrix} 4 & 2 & 4 \\ 6 & -5 & 8 \\ 3 & 3 & 2 \end{bmatrix} & \text{(f)} \begin{bmatrix} 3 & 3 & 2 \\ 2 & 4 & 2 \\ 7 & 10 & 6 \end{bmatrix} \end{array}$$

2. Use Theorem 6.29 to express $\det(A^m)$ in terms of $\det(A)$ for an arbitrary integer m ($A^0 = I$). In particular, obtain the result $\det(A^{-1}) = \frac{1}{\det(A)}$.
3. Given that A and B are square matrices of the same order, $\det(A) = 2$, and $\det(B) = 6$. Find the value of $\det(A^{-1}B)$.
4. If A and B are invertible and of the same order, express $\det((AB)^{-1})$ in terms of $\det(A)$ and $\det(B)$.
5. Express $\det(\text{adj}(A))$ in terms of $\det(A)$, where A is an $n \times n$ matrix.
6. Prove Corollary 6.26.
7. A matrix A is called **skew-symmetric** if $A^T = -A$. Prove that any skew-symmetric matrix of odd order is singular if $1 + 1 \neq 0$ in \mathcal{F} .
8. Suppose that the matrices $A = [a_{ij}]_n$, $B = [b_{ij}]_n$, and $C = [c_{ij}]_n$ are such that $a_{ij} = b_{ij} = c_{ij}$ whenever $i \neq k$, and $c_{kj} = a_{kj} + b_{kj}$ for each j . Prove that $\det(C) = \det(A) + \det(B)$.
9. A **submatrix** of an $m \times n$ matrix A is a matrix obtained by deleting certain rows and/or columns of A . For an arbitrary matrix A (not necessarily square), the number $\rho(A)$ is defined as follows:
- (i) if $A = \mathbf{0}$, then $\rho(A) = 0$,
 - (ii) if $A \neq \mathbf{0}$, then $\rho(A)$ is the largest possible order for a square submatrix of A that has a nonzero determinant.
- (a) Prove that if the matrix B is obtained from A by an elementary row operation, then $\rho(B) \leq \rho(A)$. Hence conclude that $\rho(B) = \rho(A)$.
- (b) Prove that if the matrix B is row-equivalent to A , then $\rho(B) = \rho(A)$.
- (c) Prove that $\rho(A)$ is the rank of A . (Hint: See Corollary 3.49.)
10. Suppose that A is a nonzero square matrix of order 3 that is singular. What is the most precise statement that can be made about the rank of A ?

11. Find all values of x and y in the field of complex numbers for which the matrix

$$\begin{bmatrix} x & y & 8 \\ 1 & x & y \end{bmatrix} \text{ has rank 1.}$$

Chapter 7

Eigenvalues and Eigenvectors

7.1 Introduction

In Chapter 5, we studied linear transformations of a vector space \mathbf{U} into a vector space \mathbf{V} , where \mathbf{U} and \mathbf{V} were vector spaces over the same field \mathcal{F} . We turn our attention now to the special case in which $\mathbf{U} = \mathbf{V}$. More precisely, we shall be concerned here with linear operators on a finite-dimensional vector space \mathbf{V} . Throughout this chapter, \mathcal{F} will denote a field, \mathbf{V} will denote a finite-dimensional vector space over \mathcal{F} , and T will denote a linear operator on \mathbf{V} .

7.2 Eigenvalues and Eigenvectors

In many applications of linear algebra, particularly in engineering, those vectors in a vector space that are mapped onto multiples of themselves by a certain linear operator are of key importance. This type of vector also arises with the use of quadratic forms in statistics and other areas of mathematics.

Definition 7.1 *An eigenvector of T is a nonzero vector \mathbf{v} such that $T(\mathbf{v}) = \lambda\mathbf{v}$ for some scalar λ . The scalar λ is called an eigenvalue of T .*

We say that \mathbf{V} and λ are **associated** with each other, or that they **correspond** to each other. Thus a scalar λ is an eigenvalue of T if and only if there exists a nonzero vector \mathbf{v} such that $T(\mathbf{v}) = \lambda\mathbf{v}$. The set of all eigenvalues of T is called the **spectrum** of T .

The term “eigenvalue” in the definition is not completely standardized. Other terms used interchangeably are **characteristic value**, **characteristic root**, **proper value**, and **proper number**. The German word “eigen” translates into English as “characteristic,” but the hybrid word eigenvalue seems to be more widely used than any of the other terms. Similarly, other terms for eigenvector are **characteristic vector** and **proper vector**.

We have already studied the intimate relations between a linear transformation and the various matrices that represent it relative to different bases. We shall see presently that the matrices that represent a linear operator are also useful tools in investigating the eigenvalues of the operator. The principal connection between a linear operator T and a matrix A that represents it is provided here by the characteristic matrix of A .

Definition 7.2 Let A be an $n \times n$ matrix over \mathcal{F} , and let x represent an indeterminate scalar. Then the matrix $A - xI$ is called the **characteristic matrix** of A .

Example 1 □ The characteristic matrix of

$$A = \begin{bmatrix} 8 & 5 & 6 & 0 \\ 0 & -2 & 0 & 0 \\ -10 & -5 & -8 & 0 \\ 2 & 1 & 1 & 2 \end{bmatrix}$$

is the matrix given by

$$A - xI = \begin{bmatrix} 8 - x & 5 & 6 & 0 \\ 0 & -2 - x & 0 & 0 \\ -10 & -5 & -8 - x & 0 \\ 2 & 1 & 1 & 2 - x \end{bmatrix}. \blacksquare$$

If we concentrate on the *degrees* of the terms in the definition of

$$\det(A - xI) = \begin{vmatrix} a_{11} - x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - x \end{vmatrix}$$

it is clear that the term with highest degree is the product of the diagonal elements

$$(a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x).$$

Hence $\det(A - xI)$ is a polynomial in x of degree n with lead coefficient $(-1)^n$, say

$$\det(A - xI) = (-1)^n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0.$$

Upon setting $x = 0$, we find that $c_0 = \det(A)$.

Definition 7.3 For any square matrix A over \mathcal{F} , the polynomial $\det(A - xI)$ is the **characteristic polynomial** of A . The equation $\det(A - xI) = 0$ is the **characteristic equation** of A , and the solutions of $\det(A - xI) = 0$ are called the **eigenvalues** of A . The set of all eigenvalues of A is called the **spectrum** of A .

Example 2 □ With A as in Example 1,

$$\begin{aligned}\det(A - xI) &= \begin{vmatrix} 8-x & 5 & 6 & 0 \\ 0 & -2-x & 0 & 0 \\ -10 & -5 & -8-x & 0 \\ 2 & 1 & 1 & 2-x \end{vmatrix} \\ &= x^4 - 8x^2 + 16\end{aligned}$$

is the *characteristic polynomial* of A . The equation

$$x^4 - 8x^2 + 16 = 0$$

is the *characteristic equation* of A . Since

$$x^4 - 8x^2 + 16 = (x+2)^2(x-2)^2,$$

the *eigenvalues* of A are given by $\lambda_1 = \lambda_2 = -2, \lambda_3 = \lambda_4 = 2$. The *spectrum* of A is $\{-2, 2\}$. ■

The important connection between the eigenvalues of linear operators and those of matrices is given in our next theorem.

Theorem 7.4 *Let A be any matrix that represents the linear operator T . Then T and A have the same spectrum.*

Proof. Suppose that T is represented by the matrix A relative to the basis \mathcal{A} of \mathbf{V} .

Let λ be an eigenvalue of T , and let $X = [x_1, x_2, \dots, x_n]^T$ be the coordinate matrix relative to \mathcal{A} of a corresponding eigenvector \mathbf{v} . Then $T(\mathbf{v}) = \lambda\mathbf{v}$ and therefore $AX = \lambda X$ since A represents T . This gives $AX - \lambda X = \mathbf{0}$, and $(A - \lambda I)X = \mathbf{0}$. Since X is a nonzero matrix, the system of equations $(A - \lambda I)X = \mathbf{0}$ has a nontrivial solution given by the coordinates recorded in X . By Theorems 6.28 and 4.18, this implies that $\det(A - \lambda I) = 0$, and λ is an eigenvalue of A .

On the other hand, if $\det(A - \lambda I) = 0$, then the system $(A - \lambda I)X = \mathbf{0}$ has a nontrivial solution by Theorems 6.28 and 4.18. This solution provides the coordinate matrix X relative to \mathcal{A} of a nonzero vector \mathbf{v} . Since $AX = \lambda X, T(\mathbf{v}) = \lambda\mathbf{v}$, and λ is an eigenvalue of T . ■

The proof of the preceding theorem rests primarily on the equivalence of the equations

$$T(\mathbf{v}) = \lambda\mathbf{v} \text{ and } AX = \lambda X.$$

We see that λ is an eigenvalue of A if and only if there exists a nonzero $n \times 1$ matrix X such that $AX = \lambda X$. This motivates the following definition of eigenvectors of matrices.

Definition 7.5 *Let λ be an eigenvalue of the $n \times n$ matrix A . Then an **eigenvector** of A associated with λ is a nonzero $n \times 1$ matrix X such that $AX = \lambda X$.*

Thus the eigenvectors of matrices and linear operators are related in this way: If the $n \times n$ matrix A represents T relative to the basis \mathcal{A} of \mathbf{V} , then X is an eigenvector of A corresponding to λ if and only if X is the coordinate matrix relative to \mathcal{A} of an eigenvector of T corresponding to the same eigenvalue.

The method of proof of Theorem 7.4 provides a systematic method for determining the eigenvalues of a given linear operator. Any convenient choice of a basis \mathcal{A} of \mathbf{V} will determine the matrix A that represents T relative to \mathcal{A} , and the eigenvalues of T are precisely the solutions of the characteristic equation $\det(A - xI) = 0$. The eigenvectors \mathbf{v} corresponding to a particular eigenvalue λ are just those nonzero vectors \mathbf{v} in \mathbf{V} with coordinates X relative to \mathcal{A} that satisfy $(A - \lambda I)X = \mathbf{0}$. An illustration of this procedure is given in the next example.

Example 3 □ Consider the linear operator on \mathbf{R}^4 defined by

$$\begin{aligned} T(x_1, x_2, x_3, x_4) \\ = (8x_1 + 5x_2 + 6x_3, -2x_2, -10x_1 - 5x_2 - 8x_3, 2x_1 + x_2 + x_3 + 2x_4). \end{aligned}$$

The matrix of T relative to the standard basis of \mathbf{R}^4 is the matrix

$$A = \begin{bmatrix} 8 & 5 & 6 & 0 \\ 0 & -2 & 0 & 0 \\ -10 & -5 & -8 & 0 \\ 2 & 1 & 1 & 2 \end{bmatrix}$$

considered in Examples 1 and 2. From Example 2, we know that the characteristic polynomial of A is

$$\det(A - xI) = x^4 - 8x^2 + 16 = (x + 2)^2(x - 2)^2,$$

and the eigenvalues of T are $\lambda_1 = \lambda_2 = -2$ and $\lambda_3 = \lambda_4 = 2$.

To determine the eigenvectors corresponding to $\lambda_1 = -2$, we consider the system of equations $(A + 2I)X = \mathbf{0}$:

$$\begin{bmatrix} 10 & 5 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ -10 & -5 & -6 & 0 \\ 2 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Solving this system, we have

$$\begin{bmatrix} 10 & 5 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -10 & -5 & -6 & 0 & 0 \\ 2 & 1 & 1 & 4 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 1 & 1 & 4 & 0 \\ 0 & 0 & 1 & -20 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{2} & 0 & 12 & 0 \\ 0 & 0 & 1 & -20 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The solutions to this system are given by

$$\begin{aligned}x_1 &= -\frac{1}{2}x_2 - 12x_4 \\x_2 &= x_2 \\x_3 &= 20x_4 \\x_4 &= x_4,\end{aligned}$$

where x_2 and x_4 are arbitrary. The eigenvectors of T corresponding to the eigenvalue -2 are those vectors of the form

$$\begin{aligned}(x_1, x_2, x_3, x_4) &= \left(-\frac{1}{2}x_2 - 12x_4, x_2, 20x_4, x_4\right) \\&= x_2\left(-\frac{1}{2}, 1, 0, 0\right) + x_4(-12, 0, 20, 1).\end{aligned}$$

For the eigenvalue $\lambda_3 = \lambda_4 = 2$, the system of equations $(A - 2I)X = \mathbf{0}$ is given by

$$\begin{bmatrix} 6 & 5 & 6 & 0 \\ 0 & -4 & 0 & 0 \\ -10 & -5 & -10 & 0 \\ 2 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The reduced row-echelon form for the augmented matrix here is

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

so the solutions are given by $x_1 = 0, x_2 = 0, x_3 = 0, x_4$ arbitrary. The eigenvectors of T corresponding to the eigenvalue 2 are those vectors of the form

$$(x_1, x_2, x_3, x_4) = x_4(0, 0, 0, 1). \blacksquare$$

The procedure described just before Example 3 is not always as simple and effective as it might seem, for there are several complications that may arise. There is usually no difficulty in determining the matrix A . (In most cases, the matrix A is already known.) But one may encounter trouble in the solution of the resulting characteristic equation $\det(A - xI) = 0$.

First of all, some or all of the solutions to $\det(A - xI) = 0$ may not lie in the field \mathcal{F} . If an eigenvalue λ is not in \mathcal{F} , then the nonzero coordinates x_i in a solution of $AX = \lambda X$ are not in \mathcal{F} . For the nonzero elements of AX are in \mathcal{F} whenever those

of X are, whereas those of λX are not. Thus there are no eigenvectors of T in \mathbf{V} corresponding to λ whenever $\lambda \notin \mathcal{F}$.

Even when the eigenvalues are all in \mathcal{F} , difficulties may be encountered in the solution of the polynomial equation $\det(A - xI) = 0$. This is the typical situation in applications, and a large number of numerical methods have been devised to obtain approximate solutions to such problems.

Problems that call for the determination of eigenvalues or eigenvectors of a certain linear transformation or matrix are called **eigenvalue problems**. They are one of the most common types of problems encountered in the applications of linear algebra. Unfortunately, most of them are quite complicated in their formulation, and their solution frequently involves a knowledge of several areas of mathematics. One of the simplest types that occurs in physical situations is illustrated in our next example.

Example 4 □ Before stating the physical problem to be solved, we recall that the force required to stretch or compress a spring by an amount x is directly proportional to x , and that the constant of proportionality is called the *spring constant*. Thus $F = cx$, where F is the force on the spring and x is the change in length caused by F .

Consider now the mechanical system shown in Figure 7.1.

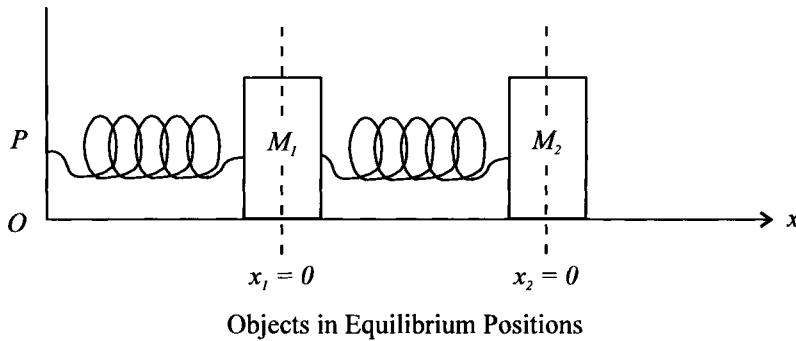


Figure 7.1

On the horizontal plane containing the line OX , an object M_1 of mass 1 unit is connected to the fixed point P by a first spring with spring constant $c_1 = 3$. A second object M_2 of mass 1 unit is then connected to M_1 by a second spring with spring constant $c_2 = 2$. The centers of gravity of M_1 and M_2 lie on a horizontal line through P . The object M_1 is displaced 1 unit toward P from its equilibrium position, and the object M_2 is displaced from its equilibrium position 2 units away from P . The two objects are released at time $t = 0$, and it is desired to find the positions of the objects at any subsequent time t . The masses of the springs and frictional forces are to be neglected, and no external forces act on the system.

Let x_1 and x_2 denote the displacement from the equilibrium positions of M_1 and M_2 , respectively. Each displacement x_i is measured with the positive direction to the right as shown in Figure 7.2.

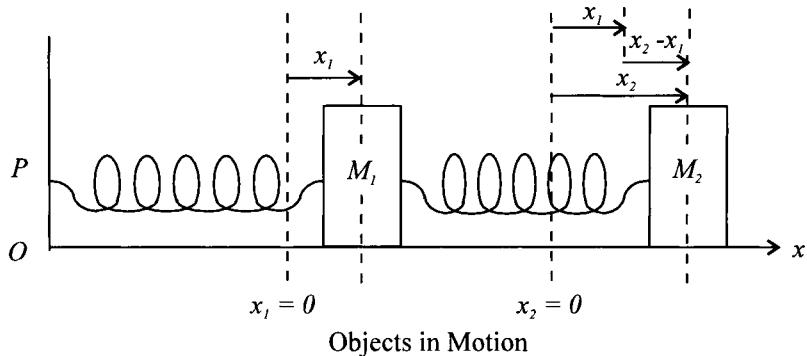


Figure 7.2

There are two forces acting on M_1 at any time $t > 0$, one from each spring. The first spring exerts a force F_1 given by $F_1 = -3x_1$ since F_1 acts in the direction opposite to the displacement. The second spring exerts a force F_2 given by $F_2 = 2(x_2 - x_1)$ since $x_2 - x_1$ is the (directed) change in the distance from the center of gravity in M_1 to that in M_2 . According to Newton's second law, the sum of the forces acting on M_1 is equal to the product of its mass and its acceleration. Since M_1 has unit mass, this requires that

$$\frac{d^2x_1}{dt^2} = -3x_1 + 2(x_2 - x_1) = -5x_1 + 2x_2.$$

The only force acting on the object M_2 is the force $-2(x_2 - x_1)$ due to the second spring, and Newton's second law yields

$$\frac{d^2x_2}{dt^2} = 2x_1 - 2x_2.$$

Thus the original problem has been reduced to that of solving the system of differential equations:

$$\begin{aligned}\frac{d^2x_1}{dt^2} &= -5x_1 + 2x_2 \\ \frac{d^2x_2}{dt^2} &= 2x_1 - 2x_2.\end{aligned}$$

We assume the existence of a solution of the form

$$x_1 = a_1 \cos \omega_1 t + a_2 \cos \omega_2 t,$$

$$x_2 = b_1 \cos \omega_1 t + b_2 \cos \omega_2 t.$$

(A justification for this assumption would be quite a digression, and we shall see momentarily that it is a valid one, at any rate.)

Substitution for $x_1, x_2, d^2x_1/dt^2$, and d^2x_2/dt^2 in the system yields

$$\begin{aligned} & -a_1\omega_1^2 \cos \omega_1 t - a_2\omega_2^2 \cos \omega_2 t \\ & = -5a_1 \cos \omega_1 t - 5a_2 \cos \omega_2 t + 2b_1 \cos \omega_1 t + 2b_2 \cos \omega_2 t, \\ & -b_1\omega_1^2 \cos \omega_1 t - b_2\omega_2^2 \cos \omega_2 t \\ & = 2a_1 \cos \omega_1 t + 2a_2 \cos \omega_2 t - 2b_1 \cos \omega_1 t - 2b_2 \cos \omega_2 t. \end{aligned}$$

In matrix form, we have

$$\begin{aligned} & \omega_1^2 \cos \omega_1 t \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + \omega_2^2 \cos \omega_2 t \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \\ & = \cos \omega_1 t \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + \cos \omega_2 t \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix}. \end{aligned}$$

Since this equation is an identity in t , we must have

$$\omega_1^2 \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$$

and

$$\omega_2^2 \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix}.$$

Thus, ω_1^2 and ω_2^2 must be eigenvalues of the matrix $A = \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}$, and $\begin{bmatrix} a_i \\ b_i \end{bmatrix}$ must be an eigenvector corresponding to ω_i^2 .

The eigenvalues of A are found to be $\lambda_1 = 1, \lambda_2 = 6$. With $\omega_1 = 1$ and $\omega_2 = \sqrt{6}$, the eigenvectors are given by

$$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = b_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

When the initial conditions $x_1 = -1, dx_1/dt = 0, x_2 = 2, dx_2/dt = 0$ are imposed, we find that $a_1 = \frac{3}{5}, b_2 = \frac{4}{5}$. The desired solution is given by

$$\begin{aligned} x_1 &= \frac{3}{5} \cos t - \frac{8}{5} \cos \sqrt{6}t, \\ x_2 &= \frac{6}{5} \cos t + \frac{4}{5} \cos \sqrt{6}t. \blacksquare \end{aligned}$$

Exercises 7.2

Find the eigenvalues of the given matrix A .

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

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

$$8. A = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & -1 \\ -1 & 2 & 0 \end{bmatrix} \quad 9. A = \begin{bmatrix} 3 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad 10. A = \begin{bmatrix} -1 & 2 & 2 \\ -2 & 3 & 2 \\ -1 & 1 & 2 \end{bmatrix}$$

$$11. A = \begin{bmatrix} 3 & 0 & 4 & 4 \\ 0 & -1 & 0 & 0 \\ 0 & -4 & -1 & -4 \\ 0 & 4 & 0 & 3 \end{bmatrix} \quad 12. A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 2 \end{bmatrix}$$

In Problems 13–16, find the spectrum of the given linear operator on \mathbf{R}^n .

$$13. T(x_1, x_2) = (x_1 + 2x_2, 2x_1 - 2x_2)$$

$$14. T(x_1, x_2) = (2x_1 + x_2, x_1 - x_2)$$

$$15. T(x_1, x_2, x_3) = (x_1 + x_2 - x_3, -x_1 + 3x_2 - x_3, -x_1 + 2x_2)$$

$$16. T(x_1, x_2, x_3, x_4) = (x_1 + x_3 + x_4, 2x_1 + x_2 + 3x_3 + x_4, -x_1 - x_3 - x_4, 3x_1 + 2x_2 + 5x_3 + x_4)$$

17. The linear operator T on \mathbf{R}^2 has the matrix $\begin{bmatrix} 4 & -5 \\ -4 & 3 \end{bmatrix}$ relative to the basis $\{(1, 2), (0, 1)\}$. Find the eigenvalues of T , and obtain an eigenvector corresponding to each eigenvalue.

18. Let T be the linear operator on \mathcal{P}_2 over \mathbf{R} that has the matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 2 & 3 & 1 \end{bmatrix}$ with respect to the basis $\{x^2, x - 2, x + 1\}$. Find the eigenvalues of T , and find an eigenvector of T corresponding to each eigenvalue.

In Problems 19 and 20, find the spectrum of the given linear operator T on \mathcal{P}_2 over \mathbf{R} and find an eigenvector of T corresponding to each eigenvalue.

19. $T(a_0 + a_1x + a_2x^2) = (2a_0 + a_1) + a_1x + (2a_0 + 2a_1 + a_2)x^2$
20. $T(a_0 + a_1x + a_2x^2) = (3a_0 - 2a_1 + a_2) + 2a_1x + (a_0 - 2a_1 + 3a_2)x^2$
21. Given that T is a linear operator on \mathbf{R}^2 with $T(2, 1) = (5, 2)$ and $T(1, 2) = (7, 10)$, determine the eigenvalues of T and a corresponding eigenvector for each eigenvalue.
22. Prove that a square matrix A is not invertible if and only if 0 is an eigenvalue of A .
23. Let A be an invertible matrix. Prove that if λ is an eigenvalue of A , then λ^{-1} is an eigenvalue of A^{-1} .
24. Assume that T is a linear operator on \mathbf{R}^3 that has eigenvalues 1, 2, and 3, with associated eigenvectors $(2, 1, 3)$, $(1, 4, 0)$, and $(1, 0, 0)$, respectively. Find the eigenvalues of T^{-1} , and give an eigenvector associated with each eigenvalue.
25. If 2 is an eigenvalue of T and $\mathbf{v} = (2, -1, 3)$ is a corresponding eigenvector, find an eigenvalue of the linear transformation $S = T^2 - 2T + 1$. ($1 = T^0$ as in Section 5.5.)
26. Prove that if λ is an eigenvalue of T and if S is a polynomial in T given by $S = p(T) = \sum_{i=0}^r a_i T^i$, then $p(\lambda) = \sum_{i=0}^r a_i \lambda^i$ is an eigenvalue of S . Find an eigenvector of S corresponding to the eigenvalue $p(\lambda)$.
27. Translate the results of Problem 26 into statements concerning the eigenvalues and eigenvectors of a polynomial $p(A)$ in a square matrix A .

7.3 Eigenspaces and Similarity

If \mathbf{v} is an eigenvector associated with the eigenvalue λ of T , then $T(c\mathbf{v}) = cT(\mathbf{v}) = c\lambda\mathbf{v} = \lambda(c\mathbf{v})$ for every choice of the scalar c . Thus, there are many eigenvectors associated with the same eigenvalue. The next theorem gives a great deal of insight into this collection of eigenvectors.

Theorem 7.6 *For each eigenvalue λ of T in \mathcal{F} , let \mathbf{V}_λ be the set consisting of the zero vector together with all eigenvectors of T in \mathbf{V} that are associated with λ . Then each \mathbf{V}_λ is a subspace of \mathbf{V} .*

Proof. The zero vector is in \mathbf{V}_λ , so \mathbf{V}_λ is nonempty. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}_\lambda$ and $a, b \in \mathcal{F}$. Then

$$T(a\mathbf{v}_1 + b\mathbf{v}_2) = aT(\mathbf{v}_1) + bT(\mathbf{v}_2) = a\lambda\mathbf{v}_1 + b\lambda\mathbf{v}_2 = \lambda(a\mathbf{v}_1 + b\mathbf{v}_2),$$

so $a\mathbf{v}_1 + b\mathbf{v}_2 \in \mathbf{V}_\lambda$, and \mathbf{V}_λ is a subspace of \mathbf{V} . ■

Definition 7.7 The subspace \mathbf{V}_λ of Theorem 7.6 is called the **eigenspace** of T that is associated with the eigenvalue λ . The dimension of \mathbf{V}_λ is called the **geometric multiplicity** of λ .

There is another approach that we could have taken in proving the last theorem. For $T(\mathbf{v}) = \lambda\mathbf{v}$ if and only if $(T - \lambda)\mathbf{v} = \mathbf{0}$, i.e., if and only if \mathbf{v} is in the kernel of $T - \lambda$. Thus \mathbf{V}_λ is nothing more than the kernel of $T - \lambda$, and the geometric multiplicity of λ is precisely the dimension of the kernel of $T - \lambda$. This translates the problem of finding an eigenspace into the familiar problem of finding a kernel. In Example 3 of Section 7.2,

$$\mathbf{V}_{-2} = (T + 2)^{-1}(\mathbf{0}) = \left\langle \left(-\frac{1}{2}, 1, 0, 0\right), (-12, 0, 20, 1) \right\rangle$$

and

$$\mathbf{V}_2 = (T - 2)^{-1}(\mathbf{0}) = \left\langle (0, 0, 0, 1) \right\rangle.$$

Concerning those eigenvectors of T that are associated with different eigenvalues, we have the following result.

Theorem 7.8 Let $\{\lambda_1, \lambda_2, \dots, \lambda_r\}$ be a set of distinct eigenvalues of the linear operator T . For each i , $1 \leq i \leq r$, let \mathbf{v}_i be an eigenvector of T corresponding to λ_i . Then $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is a linearly independent set.

Proof. The proof is by induction on r . The theorem is true for $r = 1$ since any eigenvector is nonzero.

Assume that the theorem is true for any set of k distinct eigenvalues. Let the set $\{\lambda_1, \dots, \lambda_k, \lambda_{k+1}\}$ be a set of $k + 1$ distinct eigenvalues of T , with \mathbf{v}_i an eigenvector corresponding to λ_i . Suppose that c_1, \dots, c_k, c_{k+1} are scalars such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k + c_{k+1}\mathbf{v}_{k+1} = \mathbf{0}. \quad (7.1)$$

Applying T to each member and using $T(\mathbf{v}_i) = \lambda_i\mathbf{v}_i$, we have

$$c_1\lambda_1\mathbf{v}_1 + \cdots + c_k\lambda_k\mathbf{v}_k + c_{k+1}\lambda_{k+1}\mathbf{v}_{k+1} = \mathbf{0}. \quad (7.2)$$

Multiplying (7.1) by λ_{k+1} , we have

$$c_1\lambda_{k+1}\mathbf{v}_1 + \cdots + c_k\lambda_{k+1}\mathbf{v}_k + c_{k+1}\lambda_{k+1}\mathbf{v}_{k+1} = \mathbf{0}. \quad (7.3)$$

Subtracting (7.3) from (7.2), we obtain

$$c_1(\lambda_1 - \lambda_{k+1})\mathbf{v}_1 + \cdots + c_k(\lambda_k - \lambda_{k+1})\mathbf{v}_k = \mathbf{0}.$$

From the induction hypothesis, $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is linearly independent so that

$$c_i(\lambda_i - \lambda_{k+1}) = 0$$

for $i = 1, 2, \dots, k$. Since $\lambda_i - \lambda_{k+1} \neq 0$ for $i = 1, 2, \dots, k$, we conclude that

$$c_1 = c_2 = \dots = c_k = 0.$$

Thus in (7.1) we have $c_{k+1}\mathbf{v}_{k+1} = \mathbf{0}$, and hence $c_{k+1} = 0$. This shows that the set of $k+1$ eigenvectors is linearly independent, and it follows that the theorem is true for all positive integers r . ■

Corollary 7.9 *If $\mathbf{V}_{\lambda_1}, \mathbf{V}_{\lambda_2}, \dots, \mathbf{V}_{\lambda_r}$ are distinct eigenspaces of T and $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is a set of eigenvectors such that $\mathbf{v}_i \in \mathbf{V}_{\lambda_i}$ for $i = 1, 2, \dots, r$, then $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is linearly independent.*

Proof. The eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$ must be distinct in order for the eigenspaces $\mathbf{V}_{\lambda_1}, \mathbf{V}_{\lambda_2}, \dots, \mathbf{V}_{\lambda_r}$ to be distinct. Since \mathbf{v}_i is an eigenvector associated with λ_i , then the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is linearly independent by the theorem. ■

We recall that a sum $\mathbf{W}_1 + \mathbf{W}_2 + \dots + \mathbf{W}_r$ of subspaces \mathbf{W}_i of \mathbf{V} is called *direct* if

$$\mathbf{W}_i \cap \sum_{\substack{j=1 \\ j \neq i}}^r \mathbf{W}_j = \{\mathbf{0}\}$$

for $i = 1, 2, \dots, r$.

Corollary 7.10 *If $\mathbf{V}_{\lambda_1}, \mathbf{V}_{\lambda_2}, \dots, \mathbf{V}_{\lambda_r}$ are distinct eigenspaces of T , the sum $\mathbf{V}_{\lambda_1} + \mathbf{V}_{\lambda_2} + \dots + \mathbf{V}_{\lambda_r}$ is direct.*

Proof. Suppose that the sum is not direct, and let \mathbf{v}_k be a nonzero vector in \mathbf{V}_{λ_k} that is also contained in $\sum_{\substack{j=1 \\ j \neq k}}^r \mathbf{V}_{\lambda_j}$. Then there are vectors \mathbf{v}_j in \mathbf{V}_{λ_j} and scalars a_j such that

$$\mathbf{v}_k = \sum_{\substack{j=1 \\ j \neq k}}^r a_j \mathbf{v}_j.$$

That is, \mathbf{v}_k is linearly dependent on the set

$$\mathcal{A} = \{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_r\}.$$

Since $\mathbf{v}_k \neq \mathbf{0}$, there are nonzero vectors in \mathcal{A} . Let $\mathcal{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_t\}$ be the nonempty set obtained by deleting all zero vectors from \mathcal{A} . Then \mathbf{v}_k is dependent on \mathcal{A}' so that the set $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_t, \mathbf{v}_k\}$ is linearly dependent. But $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_t, \mathbf{v}_k\}$ is a set of eigenvectors that satisfies the hypothesis of Corollary 7.9. Thus we have a contradiction, and it follows that the sum $\sum_{i=1}^r \mathbf{V}_{\lambda_i}$ is direct. ■

Suppose that T has matrix A relative to the basis \mathcal{A} of \mathbf{V} , and that T has matrix B relative to the basis \mathcal{A}' of \mathbf{V} . If P is the (invertible) matrix of transition from \mathcal{A} to \mathcal{A}' , then Theorem 5.15 asserts that $B = P^{-1}AP$. This leads to the following definition.

Definition 7.11 *Let A and B be $n \times n$ matrices over \mathcal{F} . Then B is similar to A over \mathcal{F} if there is an invertible matrix P with elements in \mathcal{F} such that $B = P^{-1}AP$.*

It is left as an exercise (Problem 16) to show that this relation of similarity is a true equivalence relation on the set of $n \times n$ matrices over \mathcal{F} . This relation proves to be a useful tool in the investigation of the eigenvalues of matrices.

The remarks just before Definition 7.11 show that two $n \times n$ matrices over \mathcal{F} are similar over \mathcal{F} if and only if they represent the same linear operator on an n -dimensional vector space \mathbf{V} over \mathcal{F} .

The strong connection between the relation of similarity and the eigenvalues of matrices becomes apparent in our next theorem.

Theorem 7.12 *Similar matrices have the same characteristic polynomial.*

Proof. If B is similar to A over \mathcal{F} , then there is an invertible matrix P such that $B = P^{-1}AP$. Thus

$$\begin{aligned}\det(B - xI) &= \det(P^{-1}AP - xI) \\ &= \det(P^{-1}AP - xP^{-1}P) \\ &= \det(P^{-1}(A - xI)P) \\ &= \det(P^{-1}) \cdot \det(A - xI) \cdot \det(P) \\ &= \det(P^{-1}P) \cdot \det(A - xI) \\ &= \det(A - xI),\end{aligned}$$

so that B and A have the same characteristic polynomial. ■

Corollary 7.13 *Similar matrices have the same spectrum.*

Near the beginning of this section, we defined the geometric multiplicity of an eigenvalue of a linear transformation. There is a second type of multiplicity for eigenvalues, the *algebraic multiplicity*.

Definition 7.14 *Let λ be an eigenvalue of T , and let A be any matrix that represents T . The algebraic multiplicity of λ is the multiplicity of λ as a root of $\det(A - xI) = 0$.*

From Theorem 7.12 and our discussion concerning Definition 7.11, it is clear that the algebraic multiplicity of an eigenvalue is well-defined. That is, the algebraic multiplicity is independent of the choice of the matrix A .

Example 1 \square In Example 3 of Section 7.2, the linear operator T had the two distinct eigenvalues $\lambda_1 = -2$ and $\lambda_3 = 2$. Examining the characteristic polynomial of A , we see that the algebraic multiplicity of each eigenvalue is 2. Upon comparing the algebraic and geometric multiplicities, we find that the two are equal for λ_1 , but that the algebraic multiplicity of λ_3 exceeds the geometric multiplicity, which is 1. Our next theorem shows that the situation in this example illustrates the only possibilities. ■

Theorem 7.15 *The geometric multiplicity of an eigenvalue does not exceed its algebraic multiplicity.*

Proof. Suppose that the geometric multiplicity of an eigenvalue λ of T is r , and let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ be a basis of \mathbf{V}_λ . This basis of \mathbf{V}_λ can be extended to a basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \dots, \mathbf{v}_n\}$ of \mathbf{V} . The matrix of T relative to this basis is

$$A = \begin{bmatrix} \lambda & 0 & \cdots & 0 & a_{1,r+1} & \cdots & a_{1n} \\ 0 & \lambda & \cdots & 0 & a_{2,r+1} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & a_{r,r+1} & \cdots & a_{rn} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & a_{n,r+1} & \cdots & a_{nn} \end{bmatrix},$$

and

$$\det(A - xI) = (\lambda - x)^r \begin{vmatrix} a_{r+1,r+1} - x & \cdots & a_{r+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,r+1} & \cdots & a_{nn} - x \end{vmatrix}.$$

Thus the algebraic multiplicity of λ is at least r . That is, the geometric multiplicity does not exceed the algebraic multiplicity. ■

Example 2 \square Let T be the linear operator on \mathcal{P}_2 over \mathbf{R} defined by

$$T(a_0 + a_1x + a_2x^2) = (-4a_0 - a_1 - a_2) + (4a_0 + 2a_2)x + (2a_0 + a_1 - a_2)x^2.$$

We shall find the algebraic and geometric multiplicity of each eigenvalue of T and obtain a basis for each eigenspace. The set $\{1, x, x^2\}$ is a more convenient choice of basis, but we shall use the basis

$$\mathcal{A} = \{1, 1+x, 1+x+x^2\}$$

to emphasize the difference between the eigenvectors of T and the eigenvectors of the matrix A that represents T relative to \mathcal{A} .

Since

$$\begin{aligned} T(1) &= -4 + 4x + 2x^2 \\ &= (-8)(1) + 2(1+x) + 2(1+x+x^2), \end{aligned}$$

$$\begin{aligned} T(1+x) &= -5 + 4x + 3x^2 \\ &= (-9)(1) + 1(1+x) + 3(1+x+x^2), \end{aligned}$$

$$\begin{aligned} T(1+x+x^2) &= -6 + 6x + 2x^2 \\ &= (-12)(1) + 4(1+x) + 2(1+x+x^2), \end{aligned}$$

the matrix of T relative to \mathcal{A} is

$$A = \begin{bmatrix} -8 & -9 & -12 \\ 2 & 1 & 4 \\ 2 & 3 & 2 \end{bmatrix}.$$

We find that the characteristic polynomial of A is

$$\begin{aligned} \det(A - xI) &= \begin{vmatrix} -8-x & -9 & -12 \\ 2 & 1-x & 4 \\ 2 & 3 & 2-x \end{vmatrix} \\ &= -(x^3 + 5x^2 + 8x + 4) \\ &= -(x+1)(x^2 + 4x + 4) \\ &= -(x+1)(x+2)^2. \end{aligned}$$

Thus the eigenvalues of T and A are -2 with algebraic multiplicity 2 and -1 with algebraic multiplicity 1.

For the eigenvalue $\lambda = -2$, the system $(A + 2I) = \mathbf{0}$ appears as

$$\begin{bmatrix} -6 & -9 & -12 \\ 2 & 3 & 4 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The reduced row-echelon form for the corresponding augmented matrix is

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

From this, we see that the solutions to $(A + 2I) = \mathbf{0}$ are given by $x_1 = -\frac{3}{2}x_2 - 2x_3$, with x_2 and x_3 arbitrary. That is, the coordinates of vectors in the eigenspace \mathbf{V}_{-2} are

given by the eigenvectors

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2}x_2 - 2x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -\frac{3}{2} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

of A . Thus \mathbf{V}_{-2} has dimension 2. We can find coordinates for a basis of \mathbf{V}_{-2} by first setting $x_2 = -2$ and $x_3 = 0$ to obtain

$$X = \begin{bmatrix} 3 \\ -2 \\ 0 \end{bmatrix}$$

and then setting $x_2 = 0$ and $x_3 = -1$ to obtain

$$X = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}.$$

(Any other linearly independent pair of coordinate matrices X would serve as well, of course.) Corresponding to these coordinates, we have the vectors $3(1) + (-2)(1+x) = 1 - 2x$ and $2(1) + (-1)(1+x+x^2) = 1 - x - x^2$ that form a basis of \mathbf{V}_{-2} .

For the eigenvalue $\lambda = -1$, $(A + I)X = \mathbf{0}$ is given by

$$\begin{bmatrix} -7 & -9 & -12 \\ 2 & 2 & 4 \\ 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The reduced row-echelon form this time is

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

and the coordinates of vectors in \mathbf{V}_{-1} are given by

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}.$$

Choosing $x_3 = -1$ yields $X = [3, -1, -1]^T$ and $\{1 - 2x - x^2\}$ as a basis of \mathbf{V}_{-1} .

Summarizing, the eigenvalue $\lambda = -2$ has algebraic multiplicity 2, geometric multiplicity 2, and $\{1 - 2x, 1 - x - x^2\}$ is a basis for the eigenspace \mathbf{V}_{-2} . The eigenvalue $\lambda = -1$ has algebraic multiplicity 1, geometric multiplicity 1, and $\{1 - 2x - x^2\}$ is a basis for \mathbf{V}_{-1} . ■

Exercises 7.3

In Problems 1-6, let T be the linear operator on \mathbf{R}^n that is represented by the given matrix A relative to the standard basis of \mathbf{R}^n . Find the algebraic multiplicity and the geometric multiplicity of each eigenvalue of T . The matrices here are taken from Problems 1-12 in Exercises 7.2.

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

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

7. For each matrix A below, let T be the linear operator on \mathbf{R}^3 that has matrix A relative to the basis $\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$. Find the algebraic and geometric multiplicities of each eigenvalue, and a basis for each eigenspace.

$$\text{(a) } A = \begin{bmatrix} 8 & 5 & -5 \\ 5 & 8 & -5 \\ 15 & 15 & -12 \end{bmatrix} \quad \text{(b) } A = \begin{bmatrix} -4 & -3 & -1 \\ -4 & 0 & -4 \\ 8 & 4 & 5 \end{bmatrix}$$

$$\text{(c) } A = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{bmatrix} \quad \text{(d) } A = \begin{bmatrix} 8 & 5 & 6 \\ 0 & -2 & 0 \\ -10 & -5 & -8 \end{bmatrix}$$

8. Let \mathcal{C} denote the field of complex numbers, and let \mathcal{C}^n be as defined in Example 1 of Section 4.2. Let \mathcal{E}_n be the basis $\mathcal{E}_n = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ of \mathcal{C}^n , where $\mathbf{e}_1 = (1, 0, 0, \dots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$, etc. If T is the linear operator that has the given matrix A relative to \mathcal{E}_n , find the algebraic and geometric multiplicities of each eigenvalue, and a basis for each eigenspace.

$$(a) A = \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix}$$

$$(b) A = \begin{bmatrix} 5 & i \\ -i & 2 \end{bmatrix}$$

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

$$(d) A = \begin{bmatrix} i & 0 & 0 \\ -2i & i & -2+i \\ 0 & 0 & -2 \end{bmatrix}$$

$$(e) A = \begin{bmatrix} 1+i & 0 & 0 \\ -2i & 1+i & 2i \\ i & 0 & 1 \end{bmatrix}$$

$$(f) A = \begin{bmatrix} -2+2i & 0 & -2+i \\ 0 & -i & 0 \\ 4-2i & 0 & 4-i \end{bmatrix}$$

In Problems 9-12, find the eigenvalues of the given linear operator T on \mathcal{P}_2 over \mathbf{R} . For each eigenvalue, (a) state the algebraic multiplicity, (b) state the geometric multiplicity, and (c) find a basis for each eigenspace.

$$9. T(a_0 + a_1x + a_2x^2) = (3a_0 + a_1 + a_2) + 2a_1x + 2a_2x^2$$

$$10. T(a_0 + a_1x + a_2x^2) = 2a_0 + (3a_0 + a_1 + 2a_2)x + (3a_0 - a_1 + 4a_2)x^2$$

$$11. T(a_0 + a_1x + a_2x^2) = (a_0 + 3a_1 + 2a_2) + 2a_1x + (a_1 + 2a_2)x^2$$

$$12. T(a_0 + a_1x + a_2x^2) = (2a_0 + a_1) + 2a_1x + (2a_0 + 3a_1 + a_2)x^2$$

13. Let T be the linear operator on \mathbf{R}^2 that has the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ relative to the standard basis of \mathbf{R}^2 .

(a) Find eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ of T such that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of \mathbf{R}^2 .

(b) Find the matrix of T relative to this basis.

14. The linear operator T on \mathbf{R}^2 has the matrix $\begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}$ relative to the basis $\mathcal{A} = \{(3, 3), (1, -1)\}$. Find the eigenvalues of T and obtain an eigenvector of T corresponding to each eigenvalue.

15. Suppose that the basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{V} consists entirely of eigenvectors of T . Determine the matrix of T relative to \mathcal{A} .

16. Prove that the relation of similarity over \mathcal{F} is an equivalence relation on the set of all $n \times n$ matrices over \mathcal{F} .

17. Which $n \times n$ matrices over \mathcal{F} are similar to I_n ?

18. Prove that if B is similar to A , then B^T is similar to A^T .
19. Prove that if A and B are $n \times n$ matrices over \mathcal{F} with A invertible, then BA is similar over \mathcal{F} to AB .
20. Prove that if B is similar to A over \mathcal{F} , then $p(B) = \sum_{i=0}^k a_i B^i$ is similar to $p(A) = \sum_{i=0}^k a_i A^i$ for any $a_0, a_1, \dots, a_k \in \mathcal{F}$.
21. Prove that similar matrices have the same rank.
22. Prove that if B is similar to A , then $\det(B) = \det(A)$.
23. Let $B = P^{-1}AP$ and suppose that X is an eigenvector of A corresponding to the eigenvalue λ . Show that λ is an eigenvalue of B , and find a corresponding eigenvector.
24. For any square matrix $A = [a_{ij}]_n$, the **trace** of A , $t(A)$, is defined by $t(A) = \sum_{i=1}^n a_{ii}$. That is, $t(A)$ is the sum of the diagonal elements of A . Prove that if B is similar to A , then $t(B) = t(A)$.

7.4 Representation by a Diagonal Matrix

The simplest form that the matrix of a linear operator can have is that of a scalar matrix. For if cI_n is the matrix of T relative to the basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{V} , then $T(\mathbf{v}_i) = c\mathbf{v}_i$ for each i , and this implies that for any $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i$ in \mathbf{V} ,

$$T(\mathbf{v}) = \sum_{i=1}^n a_i T(\mathbf{v}_i) = \sum_{i=1}^n a_i c \mathbf{v}_i = c \sum_{i=1}^n a_i \mathbf{v}_i = c\mathbf{v}.$$

That is, c is an eigenvalue of T and every nonzero vector in \mathbf{V} is an eigenvector corresponding to c . If $c > 0$, T can be described geometrically as an expansion about the origin if $c \geq 1$ and as a contraction about the origin if $c \leq 1$. If $c < 0$, T can be described as a reflection through the origin followed by an expansion or a contraction.

The next simplest form for a matrix is a diagonal matrix (of which the scalar matrix is a special case). If a linear operator has a diagonal matrix relative to a certain basis, this diagonal matrix displays at a glance the essential features of the transformation. With \mathbf{v} confined to an eigenspace \mathbf{V}_λ , $T(\mathbf{v}) = \lambda\mathbf{v}$ so that T maps \mathbf{V}_λ in the same fashion as a scalar linear operator maps the entire space \mathbf{V} .

Although the class of linear operators that can be represented by a diagonal matrix is quite large, not all linear operators can be represented in this way. Our primary objective in this section is to characterize those linear operators that can be represented by a diagonal matrix. This characterization is given in the next theorem.

Theorem 7.16 *The linear operator T on \mathbf{V} can be represented by a diagonal matrix if and only if there is a basis of \mathbf{V} that consists entirely of eigenvectors of T .*

Proof. Suppose that $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a basis of \mathbf{V} such that each \mathbf{v}_i is an eigenvector of T with λ_i as the corresponding eigenvalue. Then $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ and the matrix of T relative to \mathcal{A} is

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Thus T is represented by a diagonal matrix relative to \mathcal{A} .

On the other hand, if T has a diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix},$$

relative to the basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{V} , then $T(\mathbf{v}_i) = d_i \mathbf{v}_i$ so that each d_i is an eigenvalue of T with \mathbf{v}_i as an associated eigenvector. ■

There are several corollaries that are worthy of mention.

Corollary 7.17 *If T is represented by a diagonal matrix, the elements on the diagonal are the eigenvalues of T .*

Proof. This follows at once from the last part of the proof of the theorem. ■

Corollary 7.18 *If T has n distinct eigenvalues in \mathcal{F} , then T can be represented by a diagonal matrix.*

Proof. Suppose that T has n distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ in \mathcal{F} . Consider a set $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of n vectors in \mathbf{V} that contains exactly one eigenvector corresponding to each λ_i . The set \mathcal{A} is linearly independent by Theorem 7.8 and therefore forms a basis of the n -dimensional vector space \mathbf{V} . ■

Corollary 7.19 *If the $n \times n$ matrix A over \mathcal{F} has n distinct eigenvalues in \mathcal{F} , then A is similar over \mathcal{F} to a diagonal matrix.*

Proof. If A is an $n \times n$ matrix over \mathcal{F} with n distinct eigenvalues in \mathcal{F} , then any linear operator that A represents has n distinct eigenvalues, by Theorem 7.4. But such a linear operator can be represented by a diagonal matrix, and this diagonal matrix is similar to A (by Definition 7.11 and Theorem 5.15). ■

Theorem 7.20 *Suppose that all eigenvalues of T are in \mathcal{F} . Then T can be represented by a diagonal matrix if and only if the geometric multiplicity of each eigenvalue of T is equal to the algebraic multiplicity.*

Proof. Let T be a linear operator on the n -dimensional vector space \mathbf{V} . In view of Theorem 7.16, it is sufficient to show that there exists a basis of eigenvectors of T if and only if the geometric and algebraic multiplicities of each eigenvalue are equal. As stated in the theorem, all eigenvalues of T are assumed to be in \mathcal{F} .

Let $\lambda_1, \lambda_2, \dots, \lambda_r$ be the distinct eigenvalues of T , let $n_i = \dim(\mathbf{V}_{\lambda_i})$ be the geometric multiplicity of λ_i , and let m_i be the algebraic multiplicity of λ_i . Since m_i is the multiplicity of λ_i as a zero of the n^{th} -degree polynomial $\det(A - xI)$, $\sum_{i=1}^r m_i = n$. According to Theorem 7.15, $0 < n_i \leq m_i$ for each i . Hence $\sum_{i=1}^r n_i = n$ if and only if $n_i = m_i$ for each i .

Now let $\mathcal{B}_i = \{\mathbf{u}_{i1}, \mathbf{u}_{i2}, \dots, \mathbf{u}_{in_i}\}$ be a basis of \mathbf{V}_{λ_i} for each i , and put

$$\mathcal{B} = \{\mathbf{u}_{11}, \dots, \mathbf{u}_{1n_1}, \mathbf{u}_{21}, \dots, \mathbf{u}_{2n_2}, \dots, \mathbf{u}_{r1}, \dots, \mathbf{u}_{rn_r}\}.$$

The set \mathcal{B} contains $\sum_{i=1}^r n_i$ vectors and clearly spans

$$\mathbf{V}_{\lambda_1} + \mathbf{V}_{\lambda_2} + \cdots + \mathbf{V}_{\lambda_r}.$$

The sum $\sum_{i=1}^r \mathbf{V}_{\lambda_i}$ is direct by Corollary 7.10, and therefore

$$\dim \left(\sum_{i=1}^r \mathbf{V}_{\lambda_i} \right) = \sum_{i=1}^r \dim(\mathbf{V}_{\lambda_i}) = \sum_{i=1}^r n_i.$$

Hence \mathcal{B} is a basis of $\sum_{i=1}^r \mathbf{V}_{\lambda_i}$.

Assume that there exists a basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of eigenvectors of T . Each \mathbf{v}_j is in some \mathbf{V}_{λ_j} and therefore dependent on \mathcal{B} . This means that \mathcal{B} spans \mathbf{V} and consequently has n elements since it is linearly independent. Thus $\sum_{i=1}^r n_i = n$ and $n_i = m_i$ for $i = 1, 2, \dots, r$.

Assume now that $n_i = m_i$ for $i = 1, 2, \dots, r$. Then $\sum_{i=1}^r n_i = n$ so that \mathcal{B} has n vectors. Since \mathcal{B} is linearly independent, \mathcal{B} must be a basis of \mathbf{V} . And since \mathcal{B} is composed of eigenvectors of T , the proof is complete. ■

In the remainder of this chapter, the frequent references to diagonal matrices make it desirable to have a more compact notation for this type of matrix. This notational convenience is provided in the next definition.

Definition 7.21 The diagonal matrix $D = [d_{ij}]_n$ with $d_{ij} = 0$ for $i \neq j$ and $d_{ii} = \lambda_i$ will be denoted by $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$.

We have seen that the problem of finding a diagonal matrix and a basis such that a given linear operator is represented by the diagonal matrix is one type of eigenvalue problem. Since we have a systematic method for finding the eigenvalues and eigenvectors of a linear operator, we are already equipped to solve this type of problem. We also have available from Chapter 5 a method for finding an invertible matrix P such that $P^{-1}AP$ is diagonal. For, with any convenient choice of basis \mathcal{B} , P is the matrix of transition from \mathcal{A} to a basis \mathcal{A}' of eigenvectors of T .

In most eigenvalue problems, the linear operator T is not given explicitly. Instead, one encounters the matrix A and is confronted with the problem of finding an invertible P such that $P^{-1}AP$ is diagonal. In such a situation, the formulation of the problem in terms of linear operators, vectors, and bases is only an encumbrance. It is more efficient to proceed directly to the problem of finding the columns of P . For this procedure, it is desirable to formulate the problem $P^{-1}AP = D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ in the form $AP = PD$. With $A = [a_{ij}]_n$ and $P = [p_{ij}]_n$, the element in the i^{th} row and j^{th} column of AP is $\sum_{k=1}^n a_{ik}p_{kj}$, whereas the corresponding element in PD is $p_{ij}\lambda_j$. With j fixed, we have

$$\begin{bmatrix} \sum_{k=1}^n a_{1k}p_{kj} \\ \sum_{k=1}^n a_{2k}p_{kj} \\ \vdots \\ \sum_{k=1}^n a_{nk}p_{kj} \end{bmatrix} = \begin{bmatrix} p_{1j}\lambda_j \\ p_{2j}\lambda_j \\ \vdots \\ p_{nj}\lambda_j \end{bmatrix} = \lambda_j \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix}.$$

But the left-hand member of this equation is the same as the product

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix} = AP_j,$$

where P_j is the j^{th} column of P . Thus, we find that the equation

$$P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$$

is equivalent to the system of equations

$$AP_j = \lambda_j P_j, \quad j = 1, 2, \dots, n,$$

where P_j is the j^{th} column of P . But this system says precisely that the j^{th} column of P is an eigenvector of A corresponding to λ_j . The requirement that P be invertible is equivalent to the requirement that these eigenvectors of A be linearly independent. Eigenvectors that are associated with distinct eigenvalues automatically form a linearly independent set (as in Theorem 7.8), but care must be taken to ensure independence whenever the geometric multiplicity of an eigenvalue exceeds 1. The procedure is illustrated in our next example.

Example 1 \square Consider the problem of finding a real invertible matrix P such that $P^{-1}AP$ is diagonal, where

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

As the initial step, we find the characteristic equation of A , given by

$$(x - 3)^2(x - 1)(x + 1) = 0.$$

We consider first the repeated eigenvalue $\lambda_1 = \lambda_2 = 3$, for if the geometric multiplicity of this eigenvalue is less than 2, then no matrix P of the required type exists (Theorem 7.20). The equation $(A - 3I)X = \mathbf{0}$ is given by

$$\begin{bmatrix} 4 & 3 & 3 & 2 \\ 0 & -2 & 2 & -4 \\ -8 & -4 & -8 & 0 \\ 2 & 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Straightforward computations show that the solutions here are given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -\frac{3}{2} \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \end{bmatrix}.$$

Hence the eigenvalue 3 has geometric multiplicity 2, and

$$P_1 = \begin{bmatrix} 3 \\ -2 \\ -2 \\ 0 \end{bmatrix} \text{ and } P_2 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \end{bmatrix}.$$

provide two linearly independent columns of P . Repetition of the same procedure yields the solutions

$$P_3 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 0 \end{bmatrix} \text{ for } \lambda_3 = 1 \text{ and } P_4 = \begin{bmatrix} 1 \\ -6 \\ 4 \\ -1 \end{bmatrix} \text{ for } \lambda_4 = -1.$$

Thus the matrix

$$P = [P_1, P_2, P_3, P_4] = \begin{bmatrix} 3 & 1 & 1 & 1 \\ -2 & -2 & -2 & -6 \\ -2 & 0 & 0 & 4 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

is an invertible matrix such that $P^{-1}AP = \text{diag}\{3, 3, 1, -1\}$.

As a check on this solution, the student may verify that $AP = P \cdot \text{diag}\{3, 3, 1, -1\}$ or that

$$P^{-1} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} & 1 \\ -3 & -2 & -\frac{5}{2} & -1 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}$$

and $P^{-1}AP = \text{diag}\{3, 3, 1, -1\}$. ■

Exercises 7.4

In Problems 1-6, (a) determine whether the given matrix A is similar over \mathbf{R} to a diagonal matrix, and (b) whenever possible, find an invertible matrix P over \mathbf{R} such that $P^{-1}AP$ is a diagonal matrix. The matrices here are the same as in Problems 1-6 of Exercises 7.3.

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

In Problems 7-10, (a) determine whether the given linear operator T can be represented by a diagonal matrix, and (b) whenever possible, find a diagonal matrix and a basis such that T is represented by the diagonal matrix relative to the basis. The linear operators are the same as those in Problems 13-16 in Exercises 7.2.

7. $T(x_1, x_2) = (x_1 + 2x_2, 2x_1 - 2x_2)$ on \mathbf{R}^2
8. $T(x_1, x_2) = (2x_1 + x_2, x_1 - x_2)$ on \mathbf{R}^2
9. $T(x_1, x_2, x_3) = (x_1 + x_2 - x_3, -x_1 + 3x_2 - x_3, -x_1 + 2x_2)$ on \mathbf{R}^3
10. $T(x_1, x_2, x_3, x_4) = (x_1 + x_3 + x_4, 2x_1 + x_2 + 3x_3 + x_4, -x_1 - x_3 - x_4, 3x_1 + 2x_2 + 5x_3 + x_4)$ on \mathbf{R}^4

In Problems 11-14, let T be the linear operator on \mathbf{R}^3 that has the given matrix A relative to the basis $\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$. (a) Determine whether T can be represented by a diagonal matrix, and (b) whenever possible, find a diagonal matrix and a basis of \mathbf{R}^3 such that T is represented by the diagonal matrix relative to the basis. These linear operators are the same as those in Problem 7 of Exercises 7.3.

$$11. A = \begin{bmatrix} 8 & 5 & -5 \\ 5 & 8 & -5 \\ 15 & 15 & -12 \end{bmatrix} \quad 12. A = \begin{bmatrix} -4 & -3 & -1 \\ -4 & 0 & -4 \\ 8 & 4 & 5 \end{bmatrix}$$

$$13. A = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{bmatrix} \quad 14. A = \begin{bmatrix} 8 & 5 & 6 \\ 0 & -2 & 0 \\ -10 & -5 & -8 \end{bmatrix}$$

In Problems 15–20, (a) determine whether the given matrix A is similar over \mathbb{C} to a diagonal matrix, and (b) whenever possible, find an invertible matrix P over \mathbb{C} such that $P^{-1}AP$ is a diagonal matrix. The matrices here are the same as in Problem 8 of Exercises 7.3.

$$15. A = \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix} \quad 16. A = \begin{bmatrix} 5 & i \\ -i & 2 \end{bmatrix}$$

$$17. A = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix} \quad 18. A = \begin{bmatrix} i & 0 & 0 \\ -2i & i & -2+i \\ 0 & 0 & -2 \end{bmatrix}$$

$$19. A = \begin{bmatrix} 1+i & 0 & 0 \\ -2i & 1+i & 2i \\ i & 0 & 1 \end{bmatrix} \quad 20. A = \begin{bmatrix} -2+2i & 0 & -2+i \\ 0 & -i & 0 \\ 4-2i & 0 & 4-i \end{bmatrix}$$

21. Whenever possible, perform a check on the work in the indicated problem by verifying that $AP = PD$, where D is the diagonal matrix that is similar to A .
- (a) Problem 1 (b) Problem 2 (c) Problem 3 (d) Problem 4
 - (e) Problem 5 (f) Problem 6 (g) Problem 15 (h) Problem 16
 - (i) Problem 17 (j) Problem 18 (k) Problem 19 (l) Problem 20
22. Whenever possible, perform a check on the work in the indicated problem by computing P^{-1} and verifying that $P^{-1}AP$ is indeed a diagonal matrix.
- (a) Problem 1 (b) Problem 2 (c) Problem 3 (d) Problem 4
 - (e) Problem 5 (f) Problem 6 (g) Problem 15 (h) Problem 16
 - (i) Problem 17 (j) Problem 18 (k) Problem 19 (l) Problem 20
23. Give an example of a 2×2 matrix over \mathbf{R} that is not similar over \mathbf{R} to a diagonal matrix.

- 24.** Give an example of two 2×2 matrices that have the same characteristic equation but are not similar.
- 25.** Show that the characteristic polynomial of the matrix

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{n-1} \end{bmatrix}$$

is $p(x) = (-1)^n(x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0)$. The matrix C is called the **companion matrix** of the polynomial $p(x)$. (*Hint:* Expand $\det(C - xI)$ about the last row.)

- 26.** Use the result of Problem 25 to write down a matrix with the given polynomial $p(x)$ as its characteristic polynomial.
- (a) $p(x) = -x^3 + 5x^2 - 2$ (b) $p(x) = x^2 - 3x + 2$
 (c) $p(x) = x^4 + 5x^2 + 4$ (d) $p(x) = -x^5 + 1$
- 27.** Suppose that $\lambda_1, \dots, \lambda_r$ are the distinct eigenvalues of T , and that each λ_i is in \mathcal{F} . Prove that T can be represented by a diagonal matrix if and only if

$$\mathbf{V} = \mathbf{V}_{\lambda_1} \oplus \cdots \oplus \mathbf{V}_{\lambda_r}.$$

Chapter 8

Functions of Vectors

8.1 Introduction

There are several standard types of functions defined on a vector space that have found widespread application. The linear transformation, which we have already studied, is probably the most important of these, but there are others that are of great value. The linear functional is central to the study of linear programming. The quadratic form is frequently useful in statistics, engineering, and physics. We shall encounter each of these types of functions in this chapter.

8.2 Linear Functionals

We recall from Chapter 5 that whenever \mathbf{U} and \mathbf{V} are vector spaces over the same field \mathcal{F} , the set of all linear transformations of \mathbf{U} into \mathbf{V} is a vector space over \mathcal{F} (Theorem 5.3). Also, we have seen in Example 1 of Section 4.2 that a field \mathcal{F} may be regarded as a vector space over itself. Thus, for any vector space \mathbf{V} over \mathcal{F} , the set of all linear transformations of \mathbf{V} into \mathcal{F} is a vector space over \mathcal{F} . This type of linear transformation is of such importance that it has a special name.

Definition 8.1 *Let \mathbf{V} be a vector space over the field \mathcal{F} . A linear transformation of \mathbf{V} into \mathcal{F} is called a **linear functional** on \mathbf{V} . The set of all linear functionals on \mathbf{V} is denoted by \mathbf{V}^* .*

Thus, a linear functional on \mathbf{V} is a scalar-valued function f defined on \mathbf{V} that has the property

$$f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v})$$

for all $\mathbf{u}, \mathbf{v} \in \mathbf{V}$. As mentioned above, \mathbf{V}^* is a vector space over \mathcal{F} . There are several interesting relations between \mathbf{V} and \mathbf{V}^* that we will investigate later, but first let us consider some examples illustrating Definition 8.1.

Example 1 □ Let $\mathbf{V} = \mathbf{R}^n$, and let $\mathbf{c} = (c_1, c_2, \dots, c_n)$ be a fixed vector in \mathbf{V} . For each $\mathbf{v} = (x_1, x_2, \dots, x_n) \in \mathbf{V}$, define $f(\mathbf{v})$ to be

$$f(\mathbf{v}) = \mathbf{c} \cdot \mathbf{v} = c_1x_1 + c_2x_2 + \cdots + c_nx_n.$$

The function f so defined is scalar-valued, and

$$f(a\mathbf{u} + b\mathbf{v}) = \mathbf{c} \cdot (a\mathbf{u} + b\mathbf{v}) = a(\mathbf{c} \cdot \mathbf{u}) + b(\mathbf{c} \cdot \mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v}).$$

Thus f is a linear functional on \mathbf{R}^n . ■

Example 2 □ Let $\mathbf{V} = \mathbf{R}^2$, and define $f(\mathbf{v}) = \|\mathbf{v}\|$ for each $\mathbf{v} \in \mathbf{V}$. Then f is scalar-valued, but f is not a linear transformation. For example,

$$f(\mathbf{e}_1) + f(\mathbf{e}_2) = f(1, 0) + f(0, 1) = 2$$

and

$$f(\mathbf{e}_1 + \mathbf{e}_2) = f(1, 1) = \sqrt{2}.$$

Thus f is not a linear functional on \mathbf{R}^2 . ■

Example 3 □ For a fixed positive integer n , let \mathbf{V} be the vector space \mathcal{P}_n consisting of all polynomials in x with coefficients in the field \mathcal{F} and degree $\leq n$. For each polynomial $p(x) = \sum_{i=0}^n a_i x^i$, define $f(p(x)) = p(0)$. The mapping f is clearly scalar-valued. For any $p(x) = \sum_{i=0}^n a_i x^i$ and $q(x) = \sum_{i=0}^n b_i x^i$ in \mathcal{P}_n and any $a, b \in \mathcal{F}$,

$$\begin{aligned} f(ap(x) + bq(x)) &= f\left(\sum_{i=0}^n (aa_i + bb_i)x^i\right) \\ &= aa_0 + bb_0 \\ &= af(p(x)) + bf(q(x)). \end{aligned}$$

This shows that f is a linear functional on \mathcal{P}_n . ■

Example 4 □ Let \mathbf{V} be the vector space $\mathbf{R}_{n \times n}$ of all $n \times n$ matrices over \mathbf{R} as defined in Chapter 4. For each $A = [a_{ij}]_n \in \mathbf{V}$, the trace of A , denoted by $t(A)$, is given by $t(A) = \sum_{i=1}^n a_{ii}$. It is left as an exercise (Problem 2) to verify that t is a linear functional on $\mathbf{R}_{n \times n}$. ■

Example 5 □ The set of all convergent sequences of real numbers is a subspace \mathbf{W} of the vector space in Example 3 in Section 4.2. The function f defined by $f(\{a_n\}) = \lim_{n \rightarrow \infty} a_n$ is a linear functional on \mathbf{W} . ■

As mentioned earlier, it is already known that \mathbf{V}^* is a vector space over \mathcal{F} , just as \mathbf{V} is. The following theorem shows that \mathbf{V}^* has the same dimension as \mathbf{V} whenever \mathbf{V} is of finite dimension.

Theorem 8.2 *If \mathbf{V} is a vector space of finite dimension n over \mathcal{F} , then \mathbf{V}^* is also of dimension n over \mathcal{F} .*

Proof. Suppose that $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis of \mathbf{V} . For each j , ($j = 1, 2, \dots, n$), let \mathbf{p}_j be defined at $\mathbf{u} = \sum_{i=1}^n x_i \mathbf{u}_i$ in \mathbf{V} by

$$\mathbf{p}_j(\mathbf{u}) = \mathbf{p}_j \left(\sum_{i=1}^n x_i \mathbf{u}_i \right) = x_j.$$

Since each \mathbf{u} can be written uniquely as $\mathbf{u} = \sum_{i=1}^n x_i \mathbf{u}_i$, the value $\mathbf{p}_j(\mathbf{u})$ is well-defined, and \mathbf{p}_j is a mapping of \mathbf{V} into \mathcal{F} . For any a, b in \mathcal{F} and $\mathbf{u} = \sum_{i=1}^n x_i \mathbf{u}_i, \mathbf{v} = \sum_{i=1}^n y_i \mathbf{u}_i$ in \mathbf{V} ,

$$\begin{aligned} \mathbf{p}_j(a\mathbf{u} + b\mathbf{v}) &= \mathbf{p}_j \left(\sum_{i=1}^n (ax_i + by_i) \mathbf{u}_i \right) \\ &= ax_j + by_j \\ &= a\mathbf{p}_j(\mathbf{u}) + b\mathbf{p}_j(\mathbf{v}). \end{aligned}$$

Thus each \mathbf{p}_j is contained in \mathbf{V}^* .

The contention is that the set $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$ is a basis of \mathbf{V}^* . To show that \mathcal{A}^* spans \mathbf{V}^* , let f be any linear functional on \mathbf{V} . If we put $a_i = f(\mathbf{u}_i)$, then for any $\mathbf{u} = \sum_{i=1}^n x_i \mathbf{u}_i$ in \mathbf{V} ,

$$\begin{aligned} f(\mathbf{u}) &= \sum_{i=1}^n x_i f(\mathbf{u}_i) \\ &= \sum_{i=1}^n x_i a_i \\ &= \sum_{i=1}^n a_i \mathbf{p}_i(\mathbf{u}) \\ &= \left(\sum_{i=1}^n a_i \mathbf{p}_i \right) (\mathbf{u}). \end{aligned}$$

Hence $f = \sum_{i=1}^n a_i \mathbf{p}_i$, and \mathcal{A}^* spans \mathbf{V}^* . To see that \mathcal{A}^* is linearly independent, suppose that c_1, c_2, \dots, c_n is a set of scalars such that

$$c_1 \mathbf{p}_1 + c_2 \mathbf{p}_2 + \cdots + c_n \mathbf{p}_n = Z,$$

where Z is the zero linear functional. Then $(\sum_{i=1}^n c_i \mathbf{p}_i)(\mathbf{u}) = 0$ for each $\mathbf{u} \in \mathbf{V}$. In particular, for each \mathbf{u}_j ,

$$0 = \left(\sum_{i=1}^n c_i \mathbf{p}_i \right) (\mathbf{u}_j) = \sum_{i=1}^n c_i \mathbf{p}_i(\mathbf{u}_j) = \sum_{i=1}^n c_i \delta_{ij} = c_j.$$

That is, each $c_j = 0$, and this completes the proof. ■

Definition 8.3 The basis \mathcal{A}^* in the proof of Theorem 8.2 is called the **dual basis** of \mathcal{A} . The linear functionals \mathbf{p}_j in \mathcal{A}^* are called the **coordinate projections** relative to \mathcal{A} .

For later use we note that the defining property of the coordinate projections \mathbf{p}_j is that $\mathbf{p}_j(\mathbf{u}_i) = \delta_{ij}$, for each base vector \mathbf{u}_i .

Whenever \mathbf{V} is finite-dimensional, \mathbf{V}^* is called the **dual space** of \mathbf{V} . If \mathbf{V} is of infinite dimension, then \mathbf{V}^* is not necessarily isomorphic to \mathbf{V} , and the term “dual space” is not ordinarily used.

For the remainder of this section, \mathbf{V} will denote an n -dimensional vector space over a field \mathcal{F} . As a linear transformation of \mathbf{V} into \mathcal{F} , each linear functional has a $1 \times n$ matrix relative to each basis of \mathbf{V} . According to Definition 5.10, the matrix of f relative to the basis $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ of \mathbf{V} is $A = [a_1, a_2, \dots, a_n]$ where $a_j = f(\mathbf{u}_j)$. (We are using the basis $\{1\}$ of \mathcal{F} here, and will adhere to this choice consistently.) The matrix of f provides a convenient method of computing the values $f(\mathbf{u})$. For if \mathbf{u} has coordinate matrix $X = [x_1, x_2, \dots, x_n]^T$, then

$$\begin{aligned} f(\mathbf{u}) &= \sum_{i=1}^n x_i f(\mathbf{u}_i) \\ &= \sum_{i=1}^n x_i a_i \\ &= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= AX. \end{aligned}$$

Actually, this result is nothing new, but merely a special case of Theorem 5.12.

The set $\{1\}$ is clearly the simplest choice of basis for \mathcal{F} , so we do not propose to make any changes here. But there is no reason to restrict ourselves in the choice of basis in \mathbf{V} , and Theorem 5.15 describes completely the results of such a change. If f has matrix A relative to the basis \mathcal{A} of \mathbf{V} , and if Q is the matrix of transition from \mathcal{A} to the basis \mathcal{B} , then f has matrix $B = AQ$ relative to \mathcal{B} . (The space \mathbf{V} here is playing the role of \mathbf{U} in Theorem 5.15.)

Any basis of \mathbf{V} has a dual basis in \mathbf{V}^* , so a change of basis from \mathcal{A} to \mathcal{B} in \mathbf{V} induces a corresponding change of basis from \mathcal{A}^* to \mathcal{B}^* in \mathbf{V}^* . Our next theorem describes the relation between these changes of bases.

Theorem 8.4 *If Q is the transition matrix from the basis $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ to the basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{V} , then $(Q^T)^{-1}$ is the matrix of transition from \mathcal{A}^* to \mathcal{B}^* in \mathbf{V}^* .*

Proof. Rather than prove the statement in the conclusion, we shall prove the equivalent assertion that Q^T is the transition matrix from \mathcal{B}^* to \mathcal{A}^* .

Let $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$, and let $\mathcal{B}^* = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n\}$. Now $\mathbf{p}_j = \sum_{k=1}^n c_{kj} \mathbf{g}_k$,

where $C = [c_{ij}]_n$ is the matrix of transition from \mathcal{B}^* to \mathcal{A}^* . Then

$$\mathbf{p}_j(\mathbf{v}_i) = \sum_{k=1}^n c_{kj} \mathbf{g}_k(\mathbf{v}_i) = \sum_{k=1}^n c_{kj} \delta_{ki} = c_{ij}.$$

But $\mathbf{v}_i = \sum_{k=1}^n q_{ki} \mathbf{u}_k$ since $Q = [q_{ij}]_n$ is the transition matrix from \mathcal{A} to \mathcal{B} . Hence

$$\mathbf{p}_j(\mathbf{v}_i) = \sum_{k=1}^n q_{ki} \mathbf{p}_j(\mathbf{u}_k) = \sum_{k=1}^n q_{ki} \delta_{jk} = q_{ji},$$

so we have $c_{ij} = \mathbf{p}_j(\mathbf{v}_i) = q_{ji}$, and $C = Q^T$. ■

The ideas developed in this section are illustrated in the following example.

Example 6 □ Let \mathbf{V} be the vector space \mathcal{C}^3 of all ordered triples of complex numbers.¹ The mapping f given by $f(c_1, c_2, c_3) = ic_1 - ic_2 + c_3$ is a linear functional on \mathcal{C}^3 . Relative to the basis $\mathcal{A} = \{\mathbf{u}_1 = (1, 0, 0), \mathbf{u}_2 = (0, i, 0), \mathbf{u}_3 = (0, 1, i)\}$, f has matrix $A = [i, 1, 0]$. The coordinates x_i of $\mathbf{u} = (c_1, c_2, c_3)$ relative to \mathcal{A} are given by $x_1 = c_1, x_2 = -ic_2 + c_3, x_3 = -ic_3$, and

$$AX = \begin{bmatrix} i & 1 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ -ic_2 + c_3 \\ -ic_3 \end{bmatrix} = ic_1 - ic_2 + c_3 = f(\mathbf{u}).$$

The elements of the dual basis $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ are given by

$$\begin{aligned} \mathbf{p}_1(c_1, c_2, c_3) &= c_1, \\ \mathbf{p}_2(c_1, c_2, c_3) &= -ic_2 + c_3, \\ \mathbf{p}_3(c_1, c_2, c_3) &= -ic_3. \end{aligned}$$

The matrix

$$Q = \begin{bmatrix} 0 & -i & 0 \\ 1 & 0 & 1 \\ 0 & 0 & -i \end{bmatrix}$$

is the matrix of transition from \mathcal{A} to $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where $\mathbf{v}_1 = (0, i, 0), \mathbf{v}_2 = (-i, 0, 0), \mathbf{v}_3 = (0, 0, 1)$. The matrix of f relative to \mathcal{B} is

$$B = AQ = [1, 1, 1].$$

The coordinates y_i of $\mathbf{u} = (c_1, c_2, c_3)$ relative to \mathcal{B} are $y_1 = -ic_2, y_2 = ic_1, y_3 = c_3$, and

$$BY = -ic_2 + ic_1 + c_3 = f(\mathbf{u}).$$

¹The symbol i here denotes the square root of -1 .

The dual basis is $\mathcal{B}^* = \{\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3\}$, where

$$\begin{aligned}\mathbf{g}_1(c_1, c_2, c_3) &= -ic_2, \\ \mathbf{g}_2(c_1, c_2, c_3) &= ic_1, \\ \mathbf{g}_3(c_1, c_2, c_3) &= c_3.\end{aligned}$$

The matrix $(Q^T)^{-1}$ is given by

$$(Q^T)^{-1} = \begin{bmatrix} 0 & i & 0 \\ 1 & 0 & 0 \\ -i & 0 & i \end{bmatrix},$$

so that $\mathbf{g}_1 = \mathbf{p}_2 - i\mathbf{p}_3$, $\mathbf{g}_2 = i\mathbf{p}_1$, and $\mathbf{g}_3 = i\mathbf{p}_3$. The reader may verify that these last equalities are correct by evaluating both members for an arbitrary $\mathbf{u} \in \mathcal{C}^3$. ■

Exercises 8.2

- In Example 3 of this section, replace $f(p(x)) = p(0)$ by $f(p(x)) = p(c)$ for a constant c . Determine if the resulting function is a linear functional.
- Verify that $t(aA + bB) = at(A) + bt(B)$ in Example 4.
- Define f on the vector space $\mathbf{R}_{n \times n}$ in Example 5 of Section 4.2 by $f(A) = \det(A)$. Determine whether or not f is a linear functional.
- Let \mathbf{V} be the vector space of all real-valued continuous functions defined on the closed interval $[0, 1]$. For each $g \in \mathbf{V}$, put $h(g) = \int_0^1 g(t)dt$. Determine whether or not h is a linear functional on \mathbf{V} .
- If $f(x_1, x_2, x_3) = 2x_1 - x_2 + 4x_3$, find a vector $\mathbf{c} \in \mathbf{R}^3$ such that $f(\mathbf{u}) = \mathbf{c} \cdot \mathbf{u}$ for all $\mathbf{u} \in \mathbf{R}^3$.
- Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ where $\mathbf{u}_1 = (1, 2, 0)$, $\mathbf{u}_2 = (1, 0, 2)$, $\mathbf{u}_3 = (0, 1, 2)$, and let $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$. Find the three expressions that give the values of $\mathbf{p}_i(x_1, x_2, x_3)$ for $i = 1, 2, 3$.
- Let $\mathcal{A} = \{(1, 1, 0), (1, 0, 0), (1, 1, 1)\}$ and $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$, and let f be the linear functional that has coordinates $[1, 2, 3]^T$ relative to \mathcal{A}^* . Find the values of $f(5, 4, 3)$ and $f(x_1, x_2, x_3)$.
- Let \mathbf{u} be a fixed vector in the n -dimensional vector space \mathbf{V} .
 - Prove that there is a nonzero $f \in \mathbf{V}^*$ such that $f(\mathbf{u}) = 0$.
 - Let $\mathbf{u} = (1, 2, 3) \in \mathbf{R}^3$. Find a linear functional $f \neq Z$ such that $f(\mathbf{u}) = 0$.

9. For the given basis \mathcal{A} of \mathbf{R}^3 , find the dual basis \mathcal{A}^* .
- $\mathcal{A} = \{(0, 0, 1), (0, 1, 1), (1, 1, 1)\}$
 - $\mathcal{A} = \{(0, 1, 1), (1, 1, 0), (1, 0, 1)\}$
 - $\mathcal{A} = \{(2, 0, 2), (0, 1, 1), (2, -3, 1)\}$
 - $\mathcal{A} = \{(1, -1, 0), (2, -1, 1), (1, 1, 3)\}$
10. Let $\mathcal{E}_3^* = \{\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3\}$ denote the dual basis of $\mathcal{E}_3 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. In each part of Problem 9, find the coordinates of the elements \mathbf{p}_i of \mathcal{A}^* relative to \mathcal{E}_3^* .
11. Find the matrix of the given linear functional relative to the given basis of \mathbf{R}^n .
- $f(x_1, x_2, x_3) = 3x_1 - 2x_2 + 7x_3, \mathcal{A} = \{(1, 0, 1), (1, 1, 0), (0, 1, 1)\}$
 - $f(x_1, x_2, x_3) = 6x_1 + 5x_2 - 8x_3, \mathcal{A} = \{(4, -3, 1), (5, -3, 1), (3, -2, 1)\}$
 - $f(x_1, x_2, x_3, x_4) = 2x_1 + 4x_3 + 12x_4, \\ \mathcal{A} = \{(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)\}$
 - $f(x_1, x_2, x_3, x_4) = 9x_1 - 6x_2 + 3x_4, \\ \mathcal{A} = \{(1, 0, 1, 0), (1, 0, 0, 1), (0, 0, 1, 1), (0, 1, 1, 0)\}$
12. Use the matrices found in Problem 11 to compute $f(\mathbf{e}_i)$ for each \mathbf{e}_i in the standard basis \mathcal{E}_n .
13. Suppose that the linear functional f has matrix $A = [a_1, a_2, \dots, a_n]$ relative to the basis \mathcal{A} of \mathbf{V} . Prove that the coordinate matrix of f relative to \mathcal{A}^* is A^T .
14. Use the matrix of transition from \mathcal{E}_n^* to \mathcal{A}^* to find $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$ for the given basis \mathcal{A} of \mathbf{R}^n . Write each \mathbf{p}_i as a linear combination of the elements \mathbf{g}_i of \mathcal{E}_n^* .
- $\mathcal{A} = \{(4, -6, 4), (0, 3, 2), (6, -17, -13)\}$
 - $\mathcal{A} = \{(1, 1, 3), (2, -1, 1), (1, -1, 0)\}$
 - $\mathcal{A} = \{0, 1, 0, 0\}, (1, -2, 0, -1), (0, 1, 0, 1), (2, -7, 1, -2)\}$
 - $\mathcal{A} = \{(3, 6, 0, 2), (2, 4, 1, 1), (2, 5, -1, 1), (1, 3, 0, 1)\}$
15. Suppose that the linear functional f on \mathbf{V} has matrix $A = [a_1, a_2, \dots, a_n]$ relative to the basis \mathcal{A} of \mathbf{V} , and that Q is the matrix of transition from \mathcal{A} to \mathcal{B} . Without using Theorem 5.15, prove that the matrix of f relative to \mathcal{B} is AQ .
16. Let \mathbf{V} be a finite-dimensional vector space. For any nonempty subset \mathcal{M} of \mathbf{V} , the **annihilator** of \mathcal{M} is the set \mathcal{M}^0 of linear functionals given by
- $$\mathcal{M}^0 = \{f \in \mathbf{V}^* \mid f(\mathbf{u}) = 0 \text{ for all } \mathbf{u} \in \mathcal{M}\}.$$
- Prove that \mathcal{M}^0 is a subspace of \mathbf{V}^* .
 - Prove that if \mathbf{W} is a subspace of \mathbf{V} , then $\dim(\mathbf{W}) + \dim(\mathbf{W}^0) = \dim(\mathbf{V})$.
17. Find a basis for \mathcal{M}^0 .

- (a) $\mathcal{M} = \{(1, 2, -2, 4), (1, 1, 1, 6), (2, 3, -1, 10)\}$
- (b) $\mathcal{M} = \{(2, 0, -3, 6), (2, 1, 0, 4), (0, 1, 3, -2)\}$
- (c) $\mathcal{M} = \{(2, 2, 1, 0), (0, 4, 1, 0), (4, 8, 3, 0)\}$
- (d) $\mathcal{M} = \{(1, 0, 1, 0), (2, -3, -4, 1), (1, 1, -3, 0)\}$

18. For any nonempty subset \mathcal{T} of \mathbf{V}^* ,

$$\mathcal{T}^0 = \{\mathbf{u} \in \mathbf{V} \mid f(\mathbf{u}) = 0 \text{ for all } f \in \mathcal{T}\}.$$

- (a) Prove that \mathcal{T}^0 is a subspace of \mathbf{V} .
 - (b) Prove that if \mathcal{T} is a subspace of \mathbf{V}^* , that $\dim(\mathcal{T}) + \dim(\mathcal{T}^0) = \dim(\mathbf{V})$.
19. Prove that $(\mathbf{W}^0)^0 = \mathbf{W}$ for any subspace \mathbf{W} of \mathbf{V} , and $(\mathcal{T}^0)^0 = \mathcal{T}$ for any subspace \mathcal{T} of \mathbf{V}^* .
20. Prove that if $\mathcal{M}_1 \subseteq \mathcal{M}_2$, then $\mathcal{M}_1^0 \supseteq \mathcal{M}_2^0$.
21. Let \mathbf{W}_1 and \mathbf{W}_2 be subspaces of \mathbf{V} .
- (a) Prove that $(\mathbf{W}_1 + \mathbf{W}_2)^0 = \mathbf{W}_1^0 \cap \mathbf{W}_2^0$.
 - (b) Prove that $(\mathbf{W}_1 \cap \mathbf{W}_2)^0 = \mathbf{W}_1^0 + \mathbf{W}_2^0$.
22. Let \mathbf{V} be of dimension n over \mathcal{F} . It follows from Theorem 8.2 that $(\mathbf{V}^*)^* = \mathbf{V}^{**}$ is an n -dimensional vector space over \mathcal{F} .
- (a) For each $\mathbf{u} \in \mathbf{V}$, define the function $h_{\mathbf{u}}$ on \mathbf{V}^* by $h_{\mathbf{u}}(f) = f(\mathbf{u})$ for all f in \mathbf{V}^* . Prove that $h_{\mathbf{u}} \in \mathbf{V}^{**}$.
 - (b) Prove that the mapping from \mathbf{V} to \mathbf{V}^{**} defined by $\phi(\mathbf{u}) = h_{\mathbf{u}}$ is an isomorphism from \mathbf{V} to \mathbf{V}^{**} .

(Comment: Since \mathbf{V} and \mathbf{V}^{**} are n -dimensional spaces over \mathcal{F} , each is isomorphic to \mathcal{F}^n . The isomorphism ϕ is such a natural one, however, that it is ordinarily used to identify \mathbf{V} and \mathbf{V}^{**} as being the same space. That is, \mathbf{u} and $h_{\mathbf{u}}$ are regarded as the same entity. This point of view is advantageous in certain instances in linear programming.)

8.3 Real Quadratic Forms

In this section we turn our attention to a second type of vector function, the *quadratic form*. For the time being, we are concerned only with *real quadratic forms*. We consider quadratic forms in a more general setting in Section 8.9.

Definition 8.5 Let q be a mapping of \mathbf{R}^n into \mathbf{R} . Then q is a **real quadratic form** if there exist constants c_{ij} in \mathbf{R} such that

$$q(\mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_i x_j$$

for each $\mathbf{v} = (x_1, x_2, \dots, x_n)$ in \mathbf{R}^n .

For convenience, we shall refer to a real quadratic form in this section simply as a “quadratic form.” And since $q(\mathbf{v})$ is a polynomial in x_1, x_2, \dots, x_n , we refer to q as a quadratic form in the variables x_1, x_2, \dots, x_n . The use of parentheses in both $q(\mathbf{v})$ and $\mathbf{v} = (x_1, x_2, \dots, x_n)$ leads to the clumsy expression $q(\mathbf{v}) = q((x_1, x_2, \dots, x_n))$, so we shall drop one set of parentheses from this notation. Thus a quadratic form in x_1, x_2 is given by

$$q(x_1, x_2) = ax_1^2 + bx_2^2 + cx_1 x_2.$$

The student has no doubt encountered such expressions as

$$q(x, y) = ax^2 + by^2 + cxy$$

in analytic geometry or the calculus. For example, the left-hand member of

$$9x^2 + 16y^2 = 144$$

defines a quadratic form in x and y , as does the left-hand member of

$$xy = 1.$$

Typically, the set of all points (x, y) in \mathbf{R}^2 for which a given quadratic form has a constant value is a conic section with center at the origin.

The value of a real quadratic form $q(x, y, z)$ in x, y, z is a polynomial

$$q(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz,$$

where the coefficients are real numbers. Typically, the set of all points (x, y, z) in \mathbf{R}^3 for which a given quadratic form has a constant value is a quadric surface.

Example 1 □ Consider the real quadratic form q defined on \mathbf{R}^3 by

$$q(x_1, x_2, x_3) = x_1^2 - 2x_2^2 + 4x_3^2 + 4x_1 x_2 - 3x_2 x_3.$$

If we let $X = [x_1, x_2, x_3]^T$ and

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

a simple computation shows that X^TAX is a matrix with a single element $q(x_1, x_2, x_3)$:

$$X^TAX = [x_1^2 - 2x_2^2 + 4x_3^2 + 4x_1x_2 - 3x_2x_3].$$

When $q(x_1, x_2, x_3)$ and this matrix of order 1 are identified as being the same, we have

$$q(x_1, x_2, x_3) = X^TAX.$$

In choosing this matrix A , we simply entered the coefficient a_{ij} of $x_i x_j$ in the i^{th} row and j^{th} column of A . Since the cross-product terms can be written in many ways, the matrix A used in $q(x_1, x_2, x_3) = X^TAX$ is not unique. For instance, the term $4x_1x_2$ can be written as $4x_2x_1$ or as $3x_1x_2 + x_2x_1$. These variations would lead one to use

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

With the cross-product terms split into equal parts, the symmetric matrix

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

would be used. ■

For any $q(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n c_{ij}x_i x_j$, it is always possible to write

$$q(x_1, x_2, \dots, x_n) = X^TAX$$

with $X = [x_1, x_2, \dots, x_n]^T$ and with A a symmetric $n \times n$ matrix as in the foregoing example. For any $r \neq s$, the sum of the two terms $c_{rs}x_r x_s + c_{sr}x_s x_r$ can be split into two parts $a_{rs}x_r x_s$ and $a_{sr}x_s x_r$ with $a_{rs} = a_{sr} = \frac{c_{rs} + c_{sr}}{2}$. Together with $a_{rr} = c_{rr}$, this yields

$$q(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j = X^TAX,$$

where $A = [a_{ij}]_n$ is a symmetric matrix.

Definition 8.6 Whenever $q(x_1, x_2, \dots, x_n)$ is written as

$$q(x_1, x_2, \dots, x_n) = X^TAX$$

with

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

and with $A = [a_{ij}]_n$ a symmetric matrix, we say that A represents the quadratic form q , or that A is the **matrix of q** relative to x_1, x_2, \dots, x_n .

According to this definition, a matrix must be symmetric in order to be called the **matrix of a quadratic form**, or to say that the matrix *represents* the quadratic form.

Unless stated otherwise, it will be assumed from now on that each X^TAX is written with A symmetric. Under this restriction, the matrix that represents a certain quadratic form is unique in a given set of variables x_1, x_2, \dots, x_n . As a first step in establishing this result, we prove the following lemma.

Lemma 8.7 *If A is a real symmetric matrix of order n and $X^TAX = 0$ for all choices of $X = [x_1, x_2, \dots, x_n]^T$, then A is a zero matrix.*

Proof. Suppose that A is a matrix that satisfies the given conditions. Let $x_r = 1$ and all other $x_i = 0$. Then $0 = X^TAX = a_{rr}$. Now let $r \neq s$, and put $x_r = 1$ and $x_s = 1$, with all other $x_i = 0$. Then

$$\begin{aligned} 0 &= X^TAX \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j \\ &= a_{rs}x_r x_s + a_{sr}x_s x_r \\ &= a_{rs} + a_{sr}. \end{aligned}$$

But $a_{rs} + a_{sr} = 2a_{rs}$ since A is symmetric, so we have $a_{rs} = 0$. Since r and s were arbitrary, A is a zero matrix. ■

Theorem 8.8 *The matrix of the real quadratic form q relative to x_1, x_2, \dots, x_n is unique.*

Proof. Suppose that A and B both represent q relative to x_1, x_2, \dots, x_n . Then

$$\begin{aligned} X^TAX &= q(x_1, x_2, \dots, x_n) \\ &= X^TBX, \end{aligned}$$

and this implies that

$$X^T(A - B)X = 0$$

for all choices of $X = [x_1, x_2, \dots, x_n]^T$. Now $A - B$ is symmetric since A and B are symmetric (see Problem 6), and therefore $A = B$ by the lemma. ■

Let us now reexamine the definition of a real quadratic form. As stated in Definition 8.5, $q(\mathbf{v})$ is a scalar that is uniquely determined by the vector \mathbf{v} in \mathbf{R}^n . Although the discussion up to this point has been primarily in terms of the components x_i of \mathbf{v} , this should not obscure the fact that q is a function of the vector variable \mathbf{v} . Now the vector \mathbf{v} is uniquely determined by its coordinates relative to any given basis of \mathbf{R}^n , and so the

value $q(\mathbf{v})$ should also be uniquely determined by these coordinates. Our next theorem gives $q(\mathbf{v})$ explicitly in terms of these coordinates. Theorem 5.14 provides the key to the proof.

Theorem 8.9 Suppose that, for each $\mathbf{v} = (x_1, x_2, \dots, x_n)$ in \mathbf{R}^n ,

$$q(\mathbf{v}) = X^T AX,$$

where A is the matrix of q relative to x_1, x_2, \dots, x_n . Let P be the matrix of transition from \mathcal{E}_n to the basis \mathcal{A} and let $y = [y_1, y_2, \dots, y_n]^T$ be the coordinate matrix of \mathbf{v} relative to \mathcal{A} . Then

$$q(\mathbf{v}) = Y^T BY,$$

where $B = P^T AP$.

Proof. Since $X = [x_1, x_2, \dots, x_n]^T$ is the coordinate matrix of $\mathbf{v} = (x_1, x_2, \dots, x_n)$ relative to \mathcal{E}_n , $X = PY$ by Theorem 5.14. Hence

$$q(\mathbf{v}) = X^T AX = (PY)^T A(PY) = Y^T (P^T AP)Y. \blacksquare$$

Example 2 □ Let q be the quadratic form given by

$$q(x_1, x_2, x_3) = 13x_1^2 + 5x_2^2 + 26x_3^2 - 16x_1x_2 - 40x_1x_3 + 24x_2x_3,$$

and let $Y = [y_1, y_2, y_3]^T$ be the coordinate matrix of $\mathbf{v} = (x_1, x_2, x_3)$ relative to the basis $\mathcal{A} = \{(1, 2, 0), (0, -2, 1), (2, 5, -1)\}$. We shall find the expression for $q(\mathbf{v})$ in terms of y_1, y_2, y_3 .

Written in matrix form, $q(\mathbf{v})$ appears as

$$q(x_1, x_2, x_3) = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 13 & -8 & -20 \\ -8 & 5 & 12 \\ -20 & 12 & 26 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

The matrix of transition from \mathcal{E}_3 to \mathcal{A} is

$$P = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -2 & 5 \\ 0 & 1 & -1 \end{bmatrix},$$

so Theorem 8.9 assures us that $q(\mathbf{v}) = Y^T BY$, where

$$\begin{aligned} B &= P^T AP \\ &= \begin{bmatrix} 1 & 2 & 0 \\ 0 & -2 & 1 \\ 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} 13 & -8 & -20 \\ -8 & 5 & 12 \\ -20 & 12 & 26 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 2 & -2 & 5 \\ 0 & 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}. \end{aligned}$$

Thus the expression for $q(\mathbf{v})$ in terms of y_1, y_2, y_3 is given by the simple form

$$q(\mathbf{v}) = y_1^2 - 2y_2^2 + 3y_3^2. \blacksquare$$

Theorem 8.9 makes available many different expressions for the value $q(\mathbf{v})$ of a quadratic form. From one point of view, it describes the effect of a change of variables from the set x_1, x_2, \dots, x_n to the set y_1, y_2, \dots, y_n according to the rule that $X = PY$. Such a change of variables is called a **linear** change of variables. The terminology of Definition 8.6 can be applied to the variables y_i as well as the x_i .

Corollary 8.10 *Suppose that $\mathbf{v} = (x_1, x_2, \dots, x_n)$ and A represents the quadratic form q relative to x_1, x_2, \dots, x_n . If the variables y_1, y_2, \dots, y_n are related to the x_i by $X = PY$ with P invertible, then the matrix $B = P^T AP$ represents q relative to y_1, y_2, \dots, y_n .*

Proof. With $B = P^T AP$, we have $q(\mathbf{v}) = Y^T BY$ from the theorem. This expression is valid for any $\mathbf{v} = (x_1, x_2, \dots, x_n)$ since $Y = P^{-1}X$ can take on any prescribed value. Now B is symmetric, since

$$B^T = (P^T AP)^T = P^T A^T (P^T)^T = P^T AP = B.$$

Hence B represents q relative to y_1, y_2, \dots, y_n . \blacksquare

The results of the preceding theorem and corollary motivate the introduction of the relation of congruence on matrices.

Definition 8.11 *Let A and B be matrices over the field \mathcal{F} . Then B is **congruent** to A over \mathcal{F} if there is an invertible matrix P over \mathcal{F} such that $B = P^T AP$.*

It is left as an exercise (Problem 8) to prove that congruence over \mathcal{F} is an equivalence relation on the set of all square matrices over \mathcal{F} .

The principal objective of this and the next two sections is to show that the polynomial expression for the value of a quadratic form q takes on the particularly simple form of a “sum of squares”

$$q(\mathbf{v}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

whenever the basis \mathcal{A} in Theorem 8.9 is chosen appropriately. Example 2 illustrates this simple form, but we have no method for finding an appropriate basis \mathcal{A} at this time.

Exercises 8.3

1. Write $q(x_1, x_2, \dots, x_n)$ as $X^T AX$ with two different matrices A , one of which is symmetric.

- (a) $q(x_1, x_2) = 4x_1^2 + 6x_1x_2 - 2x_2^2$
- (b) $q(x_1, x_2) = 3x_1^2 + 5x_2^2$
- (c) $q(x_1, x_2, x_3) = -2x_1^2 - 3x_1x_3 + 8x_2x_3 - 10x_2x_1 - 7x_3^2$
- (d) $q(x_1, x_2, x_3) = 12x_1^2 - 8x_2x_1 + 4x_1x_3 - 5x_2^2$
- (e) $q(x_1, x_2, x_3) = 6x_2^2 - 9x_1x_2 - 7x_1^2$
- (f) $q(x_1, x_2, x_3, x_4) = -3x_2^2 + 4x_1^2 - 11x_1x_4 + 5x_2x_4 + 18x_1x_2 + 16x_4^2$
- (g) $q(x_1, x_2, x_3) = \sum_{i=1}^3 \sum_{j=1}^3 a_{ij}x_i x_j$ with $a_{ij} = i + j$
- (h) $q(x_1, x_2, x_3) = \sum_{i=1}^3 \sum_{j=1}^3 a_{ij}x_i x_j$ with $a_{ij} = i - j$

2. Suppose that for each $\mathbf{v} = (x_1, x_2, \dots, x_n)$ in \mathbf{R}^n , $q(\mathbf{v}) = X^T AX$ for the given matrix A . For the given basis \mathcal{A} of \mathbf{R}^n , find the expression for $q(\mathbf{v})$ in terms of the coordinates y_i of \mathbf{V} relative to \mathcal{A} .

- (a) $A = \begin{bmatrix} \frac{3}{2} & \sqrt{2} & -\frac{1}{2} \\ \sqrt{2} & 1 & -\sqrt{2} \\ -\frac{1}{2} & -\sqrt{2} & -\frac{5}{2} \end{bmatrix}$, $\mathcal{A} = \{(1, 0, 1), (3, \sqrt{2}, 1), (3\sqrt{2}, -4, \sqrt{2})\}$
- (b) $A = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 2 & -2 \\ 0 & -2 & 3 \end{bmatrix}$, $\mathcal{A} = \{(1, 0, 1), (0, 1, 1), (1, 1, 0)\}$
- (c) $A = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$, $\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$
- (d) $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 3 & \frac{1}{2} \\ 2 & \frac{1}{2} & 7 \end{bmatrix}$, $\mathcal{A} = \{(1, 0, 0), (1, -1, 0), (-11, 3, 4)\}$

3. Find the matrix that represents the given quadratic form relative to the variables y_i .

(a) $q(x_1, x_2, x_3) = 4x_1x_2 - 2x_1x_3 + x_2^2 + 2x_2x_3 - 2x_3^2,$

$$x_1 = y_1$$

$$x_2 = -y_1 + y_2 + 2y_3$$

$$x_3 = -y_1 + 2y_2 + 2y_3$$

(b) $q(x_1, x_2, x_3) = x_1^2 + 4x_3^2 - 2x_1x_2 + 4x_1x_3 - 6x_2x_3,$

$$x_1 = -4y_1 - 5y_2 + 3y_3$$

$$x_2 = 3y_1 + 3y_2 - 2y_3$$

$$x_3 = -y_1 - y_2 + y_3$$

(c) $q(x_1, x_2, x_3) = 4x_1x_2 + 4x_1x_3 + 4x_2x_3,$

$$x_1 = 2y_1 + y_2 - 2y_3$$

$$x_2 = 3y_1 + y_3$$

$$x_3 = 6y_1 - 3y_2 + 2y_3$$

(d) $q(x_1, x_2, x_3) = -x_1^2 + 2x_2^2 - 6x_3^2 + 4x_1x_2 - x_1x_3 - 4x_2x_3,$

$$x_1 = y_1 + 2y_2 + y_3$$

$$x_2 = -y_1 - y_2 + y_3$$

$$x_3 = y_2 + 3y_3$$

4. Verify that if $a_{rs} = a_{sr} = \frac{1}{2}(c_{rs} + c_{sr})$ for $r, s = 1, 2, \dots, n$, then $\sum_{i=1}^n \sum_{j=1}^n c_{ij}x_i x_j = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j$.
5. Recall that a matrix A is *skew-symmetric* if and only if $A^T = -A$. Prove that if A is skew-symmetric, then $X^T A X = 0$ for all $X = [x_1, x_2, \dots, x_n]^T$.
6. Prove that the sum and difference of two symmetric matrices of the same order are symmetric.
7. Prove that any square matrix can be written uniquely as the sum of a symmetric and a skew-symmetric matrix. (*Hint:* $A + A^T$ is symmetric.)
8. Prove that the relation of congruence over \mathcal{F} is an equivalence relation on the set of all square matrices over \mathcal{F} .
9. Prove that two symmetric matrices A and B over \mathbf{R} are congruent if and only if they represent the same real quadratic form relative to two sets of n variables that are related by an invertible linear change of variables.
10. Show that the set of symmetric matrices of order n over \mathbf{R} is not closed under multiplication.
11. Show that AA^T and $A^T A$ are symmetric for any matrix A .
12. Show that if A is an $n \times n$ matrix over \mathbf{R} such that $X^T A X = 0$ for all $X = [x_1, x_2, \dots, x_n]^T$, then A is skew-symmetric.

8.4 Orthogonal Matrices

Corollary 8.10 shows that the problem of reducing the expression for $q(\mathbf{v})$ by a linear change of variable to the form

$$q(\mathbf{v}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

is equivalent to that of finding an invertible matrix P such that

$$P^T AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}.$$

This problem has something of the same flavor as the diagonalization problem studied in Section 7.4. The difference is that $P^T AP$ has taken the place of $P^{-1}AP$. It is not unnatural, then, to consider the possibility of finding a matrix P over \mathbf{R} such that $P^T AP$ is diagonal and $P^T = P^{-1}$. These are clearly strong restrictions to be placed on P , but oddly enough it is true that such a P *always* exists (whenever A is real and symmetric). Those matrices P that have the property that $P^T = P^{-1}$ are quite useful in many instances, and have a special name.

Definition 8.12 A matrix P over \mathbf{R} is **orthogonal** if $P^T = P^{-1}$.

It would seem natural to expect a connection between the use of the word orthogonal to describe a matrix and the use of the same word to describe a set of vectors in Chapter 1. We recall that an orthogonal set of vectors is a set $\{\mathbf{u}_\lambda \mid \lambda \in \mathcal{L}\}$ such that $\mathbf{u}_{\lambda_1} \cdot \mathbf{u}_{\lambda_2} = 0$ whenever $\lambda_1 \neq \lambda_2$. An **orthogonal basis** of a subspace \mathbf{W} of \mathbf{R}^n , then, is a basis of \mathbf{W} that is an orthogonal set of vectors. The two uses of the word orthogonal are related, but not exactly in the way that one might expect. Definition 8.13 and Theorem 8.14 explain the relation completely.

Definition 8.13 A set of vectors $\{\mathbf{u}_\lambda \mid \lambda \in \mathcal{L}\}$ is **orthonormal** if the set is orthogonal and if each \mathbf{u}_λ has length 1.

The word “orthonormal” is a fusion of the words “orthogonal” and “normal.” A **normalized** set of vectors is a set in which each vector has unit length.

Theorem 8.14 Let P be an $r \times r$ matrix over \mathbf{R} . Then P is orthogonal if and only if P is the matrix of transition from one orthonormal set of r vectors in \mathbf{R}^n to another orthonormal set of r vectors in \mathbf{R}^n .

Proof. Let $P = [p_{ij}]_{r \times r}$ over \mathbf{R} , and let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ be an orthonormal set in \mathbf{R}^n . Then P is the matrix of transition from \mathcal{A} to a set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ of vectors in \mathbf{R}^n such that $\mathbf{v}_j = \sum_{k=1}^r p_{kj} \mathbf{u}_k$. Now \mathcal{B} is orthonormal if and only if $\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij}$ for

all pairs i, j . We have

$$\begin{aligned}\mathbf{v}_i \cdot \mathbf{v}_j &= \left(\sum_{k=1}^r p_{ki} \mathbf{u}_k \right) \left(\sum_{m=1}^r p_{mj} \mathbf{u}_m \right) \\ &= \sum_{k=1}^r \sum_{m=1}^r p_{ki} p_{mj} \mathbf{u}_k \cdot \mathbf{u}_m \\ &= \sum_{k=1}^r \sum_{m=1}^r p_{ki} p_{mj} \delta_{km} \\ &= \sum_{k=1}^r p_{ki} p_{kj},\end{aligned}$$

where $\mathbf{u}_k \cdot \mathbf{u}_m = \delta_{km}$ since \mathcal{A} is orthonormal. But this last sum is precisely the element in the i^{th} row and j^{th} column of $P^T P$. Thus $\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij}$ if and only if $P^T P = I_r$, and the proof is complete. ■

Thus a matrix P is orthogonal if and only if it preserves the property of being orthonormal from one basis to another. Whether the bases are orthonormal or not, we say that there is an **orthogonal change of basis** whenever the matrix of transition from one basis to the other is orthogonal. The associated linear change of variable $X = PY$ is called an **orthogonal change of variable**.

There is a question lurking in the background here that should be brought out into the open and answered. The question is this: Does every nonzero subspace of \mathbf{R}^n have an orthonormal basis? Theorem 8.15 shows that the answer is affirmative, and the proof describes a procedure for obtaining an orthonormal basis from any given basis. This procedure is known as the **Gram-Schmidt Orthogonalization Process**.

Theorem 8.15 *Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ be a basis of the subspace \mathbf{W} of \mathbf{R}^n . There exists an orthonormal basis $\mathcal{N} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ of \mathbf{W} such that each \mathbf{v}_i is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i$.*

Proof. The proof is by induction on the dimension r of \mathbf{W} . In order to describe the procedure clearly, the routine is presented in full for $r = 1, 2, 3$.

Since \mathcal{A} is linearly independent, each $\mathbf{u}_i \neq \mathbf{0}$ and $\|\mathbf{u}_i\| > 0$. Let $\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$. Then \mathbf{v}_1 is a unit vector, and the proof is complete if $r = 1$.

For $r > 1$, let

$$\mathbf{w}_2 = \mathbf{u}_2 - (\mathbf{v}_1 \cdot \mathbf{u}_2) \mathbf{v}_1.$$

Then $\mathbf{w}_2 \neq \mathbf{0}$ since $\{\mathbf{v}_1, \mathbf{u}_2\}$ is linearly independent. We have

$$\begin{aligned}\mathbf{v}_1 \cdot \mathbf{w}_2 &= \mathbf{v}_1 \cdot \mathbf{u}_2 - (\mathbf{v}_1 \cdot \mathbf{u}_2) \mathbf{v}_1 \cdot \mathbf{v}_1 \\ &= \mathbf{v}_1 \cdot \mathbf{u}_2 - \mathbf{v}_1 \cdot \mathbf{u}_2 \\ &= 0,\end{aligned}$$

so \mathbf{w}_2 is orthogonal to \mathbf{v}_1 . Put $\mathbf{v}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|}$. Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthonormal set and \mathbf{v}_i is a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_i$ for $i = 1, 2$.

If $r > 2$, let

$$\mathbf{w}_3 = \mathbf{u}_3 - (\mathbf{v}_1 \cdot \mathbf{u}_3)\mathbf{v}_1 - (\mathbf{v}_2 \cdot \mathbf{u}_3)\mathbf{v}_2.$$

Then $\mathbf{w}_3 \neq \mathbf{0}$, since $\mathbf{w}_3 = \mathbf{0}$ would require that \mathbf{u}_3 be dependent on $\{\mathbf{v}_1, \mathbf{v}_2\}$ and hence on $\{\mathbf{u}_1, \mathbf{u}_2\}$. Since

$$\begin{aligned}\mathbf{v}_1 \cdot \mathbf{w}_3 &= \mathbf{v}_1 \cdot \mathbf{u}_3 - (\mathbf{v}_1 \cdot \mathbf{u}_3)\mathbf{v}_1 - (\mathbf{v}_2 \cdot \mathbf{u}_3)\mathbf{v}_1 \cdot \mathbf{v}_2 \\ &= \mathbf{v}_1 \cdot \mathbf{u}_3 - \mathbf{v}_1 \cdot \mathbf{u}_3 \\ &= 0,\end{aligned}$$

\mathbf{w}_3 is orthogonal to \mathbf{v}_1 . Similarly, \mathbf{w}_3 is orthogonal to \mathbf{v}_2 . The vector $\mathbf{v}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|}$ is a unit vector and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal set such that \mathbf{v}_i is a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_i$ for $i = 1, 2, 3$.

Assume the theorem is true for all subspaces with $r = k$. Let $\mathcal{A} = \{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}\}$ be a basis of the $(k + 1)$ -dimensional subspace \mathbf{W} . By the induction hypothesis, the subspace $\langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \rangle$ has an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ such that \mathbf{v}_i is a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_i$ for each i . Let

$$\begin{aligned}\mathbf{w}_{k+1} &= \mathbf{u}_{k+1} - (\mathbf{v}_1 \cdot \mathbf{u}_{k+1})\mathbf{v}_1 - (\mathbf{v}_2 \cdot \mathbf{u}_{k+1})\mathbf{v}_2 - \cdots - (\mathbf{v}_k \cdot \mathbf{u}_{k+1})\mathbf{v}_k \\ &= \mathbf{u}_{k+1} - \sum_{j=1}^k (\mathbf{v}_j \cdot \mathbf{u}_{k+1})\mathbf{v}_j.\end{aligned}$$

Then $\mathbf{w}_{k+1} \neq \mathbf{0}$, since otherwise \mathbf{u}_{k+1} would be dependent on $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ and hence on $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$. Since

$$\begin{aligned}\mathbf{v}_i \cdot \mathbf{w}_{k+1} &= \mathbf{v}_i \cdot \mathbf{u}_{k+1} - \sum_{j=1}^k (\mathbf{v}_j \cdot \mathbf{u}_{k+1})\mathbf{v}_i \cdot \mathbf{v}_j \\ &= \mathbf{v}_i \cdot \mathbf{u}_{k+1} - \mathbf{v}_i \cdot \mathbf{u}_{k+1} \\ &= 0,\end{aligned}$$

\mathbf{w}_{k+1} is orthogonal to \mathbf{v}_i for $i = 1, 2, \dots, k$. Thus $\mathbf{v}_{k+1} = \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|}$ is a unit vector and $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}\}$ is an orthonormal set such that \mathbf{v}_i is a linear combination of $\{\mathbf{u}_1, \dots, \mathbf{u}_i\}$ for $i = 1, 2, \dots, k + 1$. Since $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}\}$ is a linearly independent set of $k + 1$ vectors in \mathbf{W} , it is a basis of \mathbf{W} by Theorem 1.33, and this completes the proof. ■

Example 1 □ We shall use the Gram-Schmidt Orthogonalization Process to obtain an orthonormal basis of $\langle \mathcal{A} \rangle$, where \mathcal{A} is the linearly independent set given by

$$\mathcal{A} = \{\mathbf{u}_1 = (2, 0, 2, 1), \mathbf{u}_2 = (0, 0, 4, 1), \mathbf{u}_3 = (8, 0, 3, 5)\}.$$

Using the same notation as in the proof of Theorem 8.15, we begin by writing

$$\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \frac{1}{3}(2, 0, 2, 1) = \left(\frac{2}{3}, 0, \frac{2}{3}, \frac{1}{3}\right).$$

Next we let

$$\begin{aligned}\mathbf{w}_2 &= \mathbf{u}_2 - (\mathbf{v}_1 \cdot \mathbf{u}_2)\mathbf{v}_1 \\ &= (0, 0, 4, 1) - 3\left(\frac{2}{3}, 0, \frac{2}{3}, \frac{1}{3}\right) \\ &= (-2, 0, 2, 0)\end{aligned}$$

and

$$\mathbf{v}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \frac{\sqrt{2}}{4}(-2, 0, 2, 0) = \left(-\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, 0\right).$$

To obtain the third vector in our orthonormal basis of $\langle \mathcal{A} \rangle$, we write

$$\begin{aligned}\mathbf{w}_3 &= \mathbf{u}_3 - (\mathbf{v}_1 \cdot \mathbf{u}_3)\mathbf{v}_1 - (\mathbf{v}_2 \cdot \mathbf{u}_3)\mathbf{v}_2 \\ &= (8, 0, 3, 5) - 9\left(\frac{2}{3}, 0, \frac{2}{3}, \frac{1}{3}\right) - \left(-\frac{5\sqrt{2}}{2}\right)\left(-\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, 0\right) \\ &= (8, 0, 3, 5) - (6, 0, 6, 3) + \left(-\frac{5}{2}, 0, \frac{5}{2}, 0\right) \\ &= \left(-\frac{1}{2}, 0, -\frac{1}{2}, 2\right).\end{aligned}$$

Finally, we let

$$\mathbf{v}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|} = \frac{\sqrt{2}}{6}(-1, 0, -1, 4).$$

Thus the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\langle \mathcal{A} \rangle$, where

$$\mathbf{v}_1 = \frac{1}{3}(2, 0, 2, 1), \quad \mathbf{v}_2 = \frac{\sqrt{2}}{2}(-1, 0, 1, 0), \quad \mathbf{v}_3 = \frac{\sqrt{2}}{6}(-1, 0, -1, 4). \blacksquare$$

Exercises 8.4

1. Given that $\{\mathbf{u}_1 = (1, -1, 0, 1), \mathbf{u}_2 = (4, 1, 3, 0), \mathbf{u}_3 = (2, -3, 9, 4)\}$ is linearly independent, find an orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ such that each \mathbf{v}_i is a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_i$.
2. Let $\mathcal{A} = \{(\frac{2}{3}, \frac{2}{3}, \frac{1}{3}), (-\frac{2}{3}, \frac{1}{3}, \frac{2}{3})\}$ and $P = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$.
 - (a) Verify that \mathcal{A} is an orthonormal set.
 - (b) Verify that P is an orthogonal matrix.
 - (c) Suppose that P is the transition matrix from \mathcal{A} to $\{\mathbf{v}_1, \mathbf{v}_2\}$. Find \mathbf{v}_1 and \mathbf{v}_2 , and verify that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is orthonormal.
3. Write \mathbf{v}_3 in Problem 1 as a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$.
4. Given that the set \mathcal{A} is linearly independent, use the Gram-Schmidt Orthogonalization Process to obtain an orthonormal basis of $\langle \mathcal{A} \rangle$.
 - (a) $\mathcal{A} = \{(1, 2, -2, 4), (1, 1, 1, 6), (5, 2, 2, 5)\}$

- (b) $\mathcal{A} = \{(2, 0, -3, 6), (2, 1, 0, 4), (1, 7, -1, 3)\}$
(c) $\mathcal{A} = \{(2, 2, 1), (0, 4, 1), (8, 3, 5)\}$
(d) $\mathcal{A} = \{(1, 0, 1, 0), (1, 1, -3, 0), (2, -3, -4, 1), (2, 3, -2, -3)\}$
5. Prove that the set of all orthogonal matrices of order n is closed under multiplication.
6. Prove that P^{-1} is orthogonal whenever P is orthogonal.
7. Prove that if P is orthogonal, then $\det(P) = \pm 1$.
8. Prove that, given any unit vector \mathbf{v}_1 in \mathbf{R}^n , there is an orthonormal basis \mathbf{R}^n that has \mathbf{v}_1 for its first element.
9. A linear operator T on \mathbf{R}^n is called **orthogonal** if and only if $\|T(\mathbf{u})\| = \|\mathbf{u}\|$ for every $\mathbf{u} \in \mathbf{R}^n$. Prove that T is orthogonal if and only if $T(\mathbf{u}) \cdot T(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$ for all $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$. (*Hint:* $\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 = 2\mathbf{u} \cdot \mathbf{v}$.)
10. (See Problem 9.) Prove that T is orthogonal if and only if T maps an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{R}^n onto an orthonormal basis $\{T(\mathbf{v}_1), T(\mathbf{v}_2), \dots, T(\mathbf{v}_n)\}$ of \mathbf{R}^n .
11. (See Problem 10.) Let $\mathcal{N} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be an arbitrary orthonormal basis of \mathbf{R}^n . Prove that T is orthogonal if and only if the matrix of T relative to \mathcal{N} is an orthogonal matrix.
12. (See Problem 11.) Show that the product of two orthogonal linear operators on \mathbf{R}^n is an orthogonal linear operator.
13. It follows from Problem 11 that any orthogonal linear operator is invertible. Prove that if T is orthogonal, then T^{-1} is orthogonal.

8.5 Reduction of Real Quadratic Forms

We return now to the problem posed at the beginning of Section 8.4, that of reducing a real quadratic form $q(\mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j$ to the form

$$q(\mathbf{v}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

by an *orthogonal* linear change of variables $X = PY$. This is equivalent to finding an orthonormal basis \mathcal{A} of \mathbf{R}^n such that the linear operator T with matrix $A = [a_{ij}]_n$ relative to \mathcal{E}_n is represented by $P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ relative to \mathcal{A} . This is the same situation as that considered in Section 7.4, except that now P must be orthogonal. If the eigenvalues λ_j are real, then the solutions P_j to the equation

$$(A - \lambda_j I)P_j = \mathbf{0}$$

can be taken to be real since A is real. We show in our next theorem that the λ_j 's are indeed real because A is symmetric. Before proceeding to this theorem, we introduce some needed terminology and notation.

Definition 8.16 Let $C = [c_{rs}]_{m \times n}$ be a matrix over the field \mathcal{C} of complex numbers. The conjugate of C is the matrix $\bar{C} = [\bar{c}_{rs}]_{m \times n}$, where $\bar{c}_{rs} = a_{rs} - ib_{rs}$ is the conjugate² of the complex number $c_{rs} = a_{rs} + ib_{rs}$.

If a and b are real numbers, the notation \bar{z} for the conjugate $a - bi$ of the complex number $z = a + bi$ is a standard one, and we have merely extended this notation to matrices. The basic properties $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$ and $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$ are valid for matrices: $\overline{AB} = \bar{A} \bar{B}$ and $\overline{A + B} = \bar{A} + \bar{B}$. (See Problem 3 of the exercises.)

Definition 8.17 If $C = [c_{rs}]_{m \times n}$ is a matrix over \mathcal{C} , then the conjugate transpose of C is the matrix $C^* = (\bar{C})^T$.

The proofs of the equalities $C^* = \overline{(C^T)}$, $(A_1 + A_2 + \dots + A_n)^* = A_1^* + A_2^* + \dots + A_n^*$, and $(A_1 A_2 \dots A_n)^* = A_n^* \dots A_2^* A_1^*$ are left as exercises.

Theorem 8.18 The spectrum of a real symmetric matrix is a set of real numbers. That is, all of the eigenvalues of a real symmetric matrix are real.

Proof. Let A be a real symmetric matrix, let λ be any eigenvalue³ of A in the field \mathcal{C} , and let X be an eigenvector over \mathcal{C} associated with λ . Then $AX = \lambda X$, and this implies that

$$X^* AX = \lambda X^* X.$$

We regard this equality of 1 by 1 matrices as an equality of numbers. Now

$$(X^* AX)^* = X^* A^* (X^*)^* = X^* AX$$

since $\overline{A} = A$ and $A^T = A$ imply that $A^* = A$. Hence $X^* AX$ is a real number. Also,

$$X^* X = \sum_{k=1}^n \bar{x}_k x_k = \sum_{k=1}^n |x_k|^2,$$

which is a positive real number since $X \neq \mathbf{0}$. This means that

$$\lambda = \frac{X^* AX}{X^* X},$$

a quotient of real numbers. ■

We proceed now to the main result of this section.

Theorem 8.19 If A is a real symmetric matrix of order n , there exists an orthogonal matrix P over \mathbf{R} such that $P^{-1}AP$ is a diagonal matrix.

²The symbol i here denotes the square root of -1 ; a_{rs} and b_{rs} are real numbers.

³We cannot prove it here, but every polynomial of degree n with coefficients in \mathcal{C} has exactly n zeros, all of which are in \mathcal{C} . In particular, the polynomial $\det(A - \lambda I)$ has n zeros in \mathcal{C} if A has order n .

Proof. The proof is by induction on the order n of A . The theorem is trivially true for $n = 1$.

Assume that the theorem is true for all $k \times k$ matrices, and let A be a real symmetric matrix of order $k+1$. Let T be the linear operator on \mathbf{R}^{k+1} that has matrix A relative to \mathcal{E}_{k+1} . Let λ_1 be any eigenvalue of T (and A). Then λ_1 is real and there is a corresponding eigenvector \mathbf{v}_1 in \mathbf{R}^{k+1} . Since any multiple of \mathbf{v}_1 is in the eigenspace \mathbf{V}_{λ_1} , we may assume that \mathbf{v}_1 is a unit vector. It follows from Theorem 8.15 that $\{\mathbf{v}_1\}$ can be extended to an orthonormal basis

$$\mathcal{N} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}\}$$

of \mathbf{R}^{k+1} (See Problem 8 of Exercises 8.4). The matrix of transition P_1 from \mathcal{E}_{k+1} to \mathcal{N} is orthogonal (Theorem 8.14), and T is represented by

$$A_1 = P_1^{-1}AP_1 = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & A_2 \end{bmatrix}$$

relative to \mathcal{N} . (The elements to the right of λ_1 in the first row must be zero because of symmetry.) The $k \times k$ matrix A_2 is real and symmetric, so it follows from our induction hypotheses that there is an orthogonal $k \times k$ matrix Q such that $Q^{-1}A_2Q = \text{diag}\{\lambda_2, \dots, \lambda_{k+1}\}$. It is readily verified that

$$P_2 = \begin{bmatrix} 1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & Q \end{bmatrix}$$

is orthogonal and that

$$P_2^{-1}A_1P_2 = P_2^{-1}P_1^{-1}AP_1P_2 = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_{k+1}\}.$$

Now P_1P_2 is orthogonal since

$$(P_1P_2)^T = P_2^TP_1^T = P_2^{-1}P_1^{-1} = (P_1P_2)^{-1},$$

and thus $P = P_1P_2$ is the desired matrix. ■

It is important to note that the diagonal elements of the matrix

$$P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$$

are the eigenvalues of A .

Corollary 8.20 Any real quadratic form q with $q(\mathbf{v}) = \mathbf{X}^TAX$ can be reduced to a diagonalized representation

$$q(\mathbf{v}) = Y^T(P^TAP)Y = \lambda_1y_1^2 + \lambda_2y_2^2 + \cdots + \lambda_ny_n^2$$

by an orthogonal change of variables $X = PY$. For each $\mathbf{v} = (x_1, x_2, \dots, x_n) \in \mathbf{R}^n$, $Y = [y_1, y_2, \dots, y_n]^T$ is the coordinate matrix of \mathbf{v} relative to \mathcal{A} , where P is the matrix of transition from \mathcal{E}_n to \mathcal{A} .

Proof. The proof is left as an exercise in Problem 13. ■

The proof of Theorem 8.19 is not exactly constructive, although it does suggest an iterative procedure to obtain P by beginning with the selection of λ_1, \mathbf{v}_1 and P_1 . It is not advantageous for us to pursue this lead, as the next theorem leads to a much more efficient procedure.

We observe that if \mathbf{u} and \mathbf{v} are vectors in \mathbf{R}^n with coordinate matrices $X = [x_1, x_2, \dots, x_n]^T$ and $y = [y_1, y_2, \dots, y_n]^T$ relative to \mathcal{E}_n , then $\mathbf{u} \cdot \mathbf{v} = \sum_{k=1}^n x_k y_k = X^T Y$. In particular, \mathbf{u} and \mathbf{v} are orthogonal if and only if $X^T Y = 0$.

Theorem 8.21 *Let A be a real symmetric matrix. If λ_r and λ_s are distinct eigenvalues of A with associated eigenvectors P_r and P_s , respectively, then $P_r^T P_s = 0$.*

Proof. Suppose that λ_r and λ_s are distinct eigenvalues of A with P_r and P_s as corresponding eigenvectors. Then $AP_r = \lambda_r P_r$ and $AP_s = \lambda_s P_s$. Now

$$P_r^T AP_s = P_r^T (\lambda_s P_s) = \lambda_s P_r^T P_s$$

and

$$(P_r^T AP_s)^T = P_s^T AP_r = P_s^T (\lambda_r P_r) = \lambda_r P_s^T P_r = \lambda_r (P_r^T P_s)^T.$$

But $(P_r^T AP_s)^T = P_r^T AP_s$ and $(P_r^T P_s)^T = P_r^T P_s$ since $P_r^T AP_s$ and $P_r^T P_s$ are matrices of order 1. Hence

$$\lambda_r P_r^T P_s = \lambda_s P_r^T P_s$$

and

$$(\lambda_r - \lambda_s) P_r^T P_s = 0.$$

Since $\lambda_r - \lambda_s \neq 0$, it must be that $P_r^T P_s = 0$. ■

In Section 7.4, we found that $P = [P_1, P_2, \dots, P_n]$ was an invertible matrix such that $P^{-1}AP$ was diagonal if and only if the columns P_j of P were the coordinate matrices relative to \mathcal{E}_n of a basis of eigenvectors \mathbf{v}_j of the associated linear transformation T . Since $P_r^T P_s$ is the element in row r and column s of $P^T P$, the requirement that $P^T = P^{-1}$ is satisfied if and only if $P_r^T P_s = \delta_{rs}$. Since $P_r^T P_s = \mathbf{v}_r \cdot \mathbf{v}_s$, this is equivalent to requiring that the basis of eigenvectors be orthonormal. Theorem 8.21 assures us that eigenvectors from distinct eigenspaces are automatically orthogonal. Thus, the only modification of the procedure in Section 7.4 that is necessary to make P orthogonal is to choose orthonormal bases of the eigenspaces \mathbf{V}_{λ_j} . This is illustrated in the next example.

Example 1 □ Consider the problem of finding an orthogonal matrix P such that $P^T AP$ is diagonal, where

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

As explained above, the basic problem is to find an orthonormal basis of eigenvectors of the linear transformation T that has matrix A relative to \mathcal{E}_3 . The characteristic equation of A is $-(x+1)(x-2)^2 = 0$. By solving the systems

$$(A - (-1)I)X = \mathbf{0} \text{ and } (A - 2I)X = \mathbf{0},$$

we find that $\{(1, 1, 1)\}$ is a basis of the eigenspace \mathbf{V}_{-1} and $\{(1, -1, 0), (1, 0, -1)\}$ is a basis of \mathbf{V}_2 . Applying the Gram-Schmidt process, we obtain

$$\left\{ \frac{1}{\sqrt{3}}(1, 1, 1) \right\}$$

and

$$\left\{ \frac{1}{\sqrt{2}}(1, -1, 0), \frac{1}{\sqrt{6}}(1, 1, -2) \right\}$$

as orthonormal bases of \mathbf{V}_{-1} and \mathbf{V}_2 , respectively. Hence

$$P = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{bmatrix}$$

is an orthogonal matrix such that $P^T AP = \text{diag}\{-1, 2, 2\}$. ■

Since the introduction of quadratic forms was partially motivated by references to the conic sections and quadric surfaces, it is of interest to relate our results here to these geometric quantities.

A conic section always has an equation of the form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

in rectangular coordinates x, y . According to Corollary 8.20, the quadratic form q with $q(x, y) = ax^2 + bxy + cy^2$ can be reduced by an orthogonal change of variables

$$\begin{aligned} x &= p_{11}x' + p_{12}y' \\ y &= p_{21}x' + p_{22}y' \end{aligned}$$

to a diagonalized form

$$\lambda_1(x')^2 + \lambda_2(y')^2.$$

It is shown in analytic geometry that such a change of variables corresponds to a rotation of the coordinate axes about the origin. The different possibilities for the signs of the eigenvalues correspond to different types of conic sections, with degenerate cases possible in each instance. If λ_1 and λ_2 are nonzero and of the same sign, the conic section is a circle or an ellipse. If λ_1 and λ_2 are nonzero and of opposite sign, the conic section is a hyperbola. If exactly one of λ_1, λ_2 is zero, the conic section is a parabola. If both λ_1 and λ_2 are zero, the graph is a straight line.

As mentioned in Section 8.3, the quadric surfaces can be related to quadratic forms in three variables. This relation can be analyzed in a manner analogous to that for the conic sections. However, this analysis becomes somewhat involved, and it is omitted here for this reason.

Exercises 8.5

1. Find an orthogonal matrix P such that $P^T AP$ is diagonal.

$$\begin{array}{ll} \text{(a)} A = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix} & \text{(b)} A = \begin{bmatrix} 1 & 2 & -4 \\ 2 & -2 & -2 \\ -4 & -2 & 1 \end{bmatrix} \\ \text{(c)} A = \begin{bmatrix} 17 & 2 & -2 \\ 2 & 14 & 4 \\ -2 & 4 & 14 \end{bmatrix} & \text{(d)} A = \begin{bmatrix} 4 & -1 & 0 & 1 \\ -1 & 5 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 1 & 0 & -1 & 5 \end{bmatrix} \end{array}$$

2. Let z_1, z_2, \dots, z_n and a_1, a_2, \dots, a_n be complex numbers.

- (a) Verify that $\overline{z_1 z_2} = \overline{z}_1 \overline{z}_2$ and $\overline{z_1 + z_2} = \overline{z}_1 + \overline{z}_2$.
- (b) By induction, extend the results of part (a) to $\overline{z_1 z_2 \cdots z_n} = \overline{z}_1 \overline{z}_2 \cdots \overline{z}_n$ and $\overline{z_1 + z_2 + \cdots + z_n} = \overline{z}_1 + \overline{z}_2 + \cdots + \overline{z}_n$.
- (c) Prove that $\overline{\sum_{k=1}^n a_k z_k} = \sum_{k=1}^n \overline{a}_k \overline{z}_k$.
- 3. Let $A = [a_{rs}]_{m \times n}$ and $B = [b_{rs}]_{n \times p}$ over \mathcal{C} .
 - (a) Prove that $\overline{AB} = \overline{A} \overline{B}$.
 - (b) Prove that $\overline{A+B} = \overline{A} + \overline{B}$ whenever $A+B$ is defined.
- 4. Prove that $C^* = \overline{(C^T)}$.
- 5. Prove that $(A_1 + A_2 + \cdots + A_n)^* = A_1^* + A_2^* + \cdots + A_n^*$.
- 6. Prove that $(A_1 A_2 \cdots A_n)^* = A_n^* \cdots A_2^* A_1^*$.
- 7. Prove that $(A^*)^{-1} = (A^{-1})^*$.
- 8. In the proof of Theorem 8.19, verify that the matrix P_2 is orthogonal and that $P_2^{-1} A_1 P_2 = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_{k+1}\}$.
- 9. Let \mathbf{u} and \mathbf{v} be vectors in \mathbf{R}^n with coordinate matrices $X = [x_1, x_2, \dots, x_n]^T$ and $Y = [y_1, y_2, \dots, y_n]^T$, respectively, relative to an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{R}^n . Prove that $\mathbf{u} \cdot \mathbf{v} = X^T Y$.

10. Let $\mathcal{N} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be an orthonormal basis of \mathbf{R}^n , and let $X = [x_1, x_2, \dots, x_n]^T$ be the coordinate matrix of \mathbf{u} relative to \mathcal{N} . Prove that $x_k = \mathbf{u} \cdot \mathbf{v}_k$ for $k = 1, 2, \dots, n$.
11. Prove that if $A = B^T B$ for some matrix B over \mathbf{R} , then $X^T A X \geq 0$ for all $X = [x_1, x_2, \dots, x_n]^T$ over \mathbf{R} .
12. Assume that the linear operator T of \mathbf{R}^n has a symmetric matrix relative to an orthonormal basis of \mathbf{R}^n . Prove that the eigenvectors of T which correspond to distinct eigenvalues are orthogonal.
13. Prove Corollary 8.20.

8.6 Classification of Real Quadratic Forms

We have seen in Section 8.5 that an arbitrary real quadratic form q with $q(\mathbf{v}) = X^T A X$ can be reduced by an invertible (even orthogonal) linear change of variables $X = PY$ to a diagonalized representation

$$q(\mathbf{v}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

where

$$P^T A P = D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}.$$

The matrix $Y = [y_1, y_2, \dots, y_n]^T$ is the coordinate matrix of $\mathbf{v} = (x_1, x_2, \dots, x_n)$ relative to the basis $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, where P is the transition matrix from \mathcal{E}_n to \mathcal{A} . An interchange of two vectors \mathbf{u}_i and \mathbf{u}_j in \mathcal{A} amounts to an interchange of the variables y_i and y_j , and such an interchange is reflected in the matrix D by an interchange of λ_i and λ_j . The diagonalized representation can thus be written as

$$q(\mathbf{v}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_r y_r^2,$$

where each of $\lambda_1, \lambda_2, \dots, \lambda_r$ is nonzero. The matrices A and D have the same rank since they are equivalent. Consequently, the number of nonzero λ_i 's is always the same as the rank r of A .

Definition 8.22 *The rank of the real quadratic form q with $q(\mathbf{v}) = X^T A X$ is the rank of the matrix A .*

The discussion above shows that the rank of q is the same as the number of variables having nonzero coefficients in a diagonalized representation of q . We shall examine these nonzero terms in more detail.

Theorem 8.23 *In any two diagonalized representations of a real quadratic form q , the number of variables with positive coefficients is the same and the number of variables with negative coefficients is the same.*

Proof. Suppose that q has rank r , and let

$$q(\mathbf{v}) = d_1 y_1^2 + d_2 y_2^2 + \cdots + d_r y_r^2$$

and

$$q(\mathbf{v}) = d'_1 z_1^2 + d'_2 z_2^2 + \cdots + d'_r z_r^2$$

be any two diagonalized representations of q .⁴ We may assume without loss of generality that the diagonal elements in both $D = \text{diag}\{d_1, d_2, \dots, d_r, 0, \dots, 0\}$ and $D' = \text{diag}\{d'_1, d'_2, \dots, d'_r, 0, \dots, 0\}$ are arranged so that the positive elements come first, followed by the negative elements. That is,

$$D = \text{diag}\{d_1, \dots, d_p, d_{p+1}, \dots, d_r, 0, \dots, 0\}$$

with the first p elements d_i positive, and

$$D' = \text{diag}\{d'_1, \dots, d'_k, d'_{k+1}, \dots, d'_r, 0, \dots, 0\}$$

with the first k elements d'_i positive.

Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be bases of \mathbf{R}^n chosen so that $[y_1, y_2, \dots, y_n]^T$ and $[z_1, z_2, \dots, z_n]^T$ are the coordinate matrices of $\mathbf{v} = (x_1, x_2, \dots, x_n)$ relative to \mathcal{A} and \mathcal{B} , respectively (see Corollary 8.20). Now $q(\mathbf{u}_i) = d_i > 0$ for $i = 1, 2, \dots, p$. Therefore, for any $\mathbf{w} = \sum_{i=1}^p y_i \mathbf{u}_i \in \mathbf{W}_1 = \langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p \rangle$,

$$q(\mathbf{w}) = \sum_{i=1}^p d_i y_i^2 \geq 0.$$

Also, $q(\mathbf{v}_i) = d'_i < 0$ for $i = k+1, k+2, \dots, n$, so that for each $\mathbf{w} = \sum_{i=k+1}^n z_i \mathbf{v}_i \in \mathbf{W}_2 = \langle \mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \dots, \mathbf{v}_n \rangle$,

$$q(\mathbf{w}) = \sum_{i=k+1}^n d'_i z_i^2 \leq 0.$$

Thus $\mathbf{W}_1 \cap \mathbf{W}_2 = \{\mathbf{0}\}$ and

$$\dim(\mathbf{W}_1 + \mathbf{W}_2) = \dim(\mathbf{W}_1) + \dim(\mathbf{W}_2).$$

But $\dim(\mathbf{W}_1 + \mathbf{W}_2) \leq n$ and

$$\dim(\mathbf{W}_1) + \dim(\mathbf{W}_2) = p + n - k.$$

Therefore, $p + n - k \leq n$ and $p \leq k$.

From the symmetry of the conditions on p and k , it follows that $k \leq p$ and $k = p$. The second part of the conclusion follows immediately from $r - k = r - p$. ■

Corollary 8.24 Let A be a real symmetric matrix. Any two diagonal matrices that are congruent to A over \mathbf{R} have the same number of positive elements and the same number of negative elements on the diagonal.

⁴These two diagonalized representations correspond to two diagonal matrices D and D' that are congruent to the original matrix A of q : $D = P_1^T A P_1$ and $D' = P_2^T A P_2$ for invertible P_1 and P_2 . That is, we are dealing here with congruence of matrices, not with orthogonal similarity.

Proof. See Problem 5. ■

Definition 8.25 The index of the quadratic form q is the number p of positive coefficients appearing in a diagonalized representation of q . The difference s between p and the number of negative coefficients in a diagonalized representation of q is the signature of q . That is, the signature of q is the number $s = p - (r - p) = 2p - r$, where r is the rank of q .

Theorem 8.23 shows that the index and signature of q are well-defined terms. The signature of q is a measure of the “positiveness” or “negativeness” of q .

Theorem 8.26 A quadratic form q on \mathbf{R}^n with rank r and index p can be represented as

$$q(\mathbf{v}) = z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_r^2$$

by a suitable invertible linear change of variables.

Proof. Let

$$q(\mathbf{v}) = d_1 y_1^2 + \cdots + d_p y_p^2 + d_{p+1} y_{p+1}^2 + \cdots + d_r y_r^2$$

be a diagonalized representation of q with d_1, \dots, d_p positive and d_{p+1}, \dots, d_r negative. For $i = 1, 2, \dots, p$, d_i has a positive real square root $\sqrt{d_i}$, and we put $z_i = \sqrt{d_i} y_i$. For $i = p+1, p+2, \dots, r$, $-d_i$ has a positive real square root $\sqrt{|d_i|}$, and we put $z_i = \sqrt{|d_i|} y_i$. For $i = r+1, \dots, n$, we put $z_i = y_i$. The linear change of variables

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \text{diag}\{\sqrt{d_1}, \dots, \sqrt{d_p}, \sqrt{|d_{p+1}|}, \dots, \sqrt{|d_r|}, 1, \dots, 1\} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

is clearly invertible, and

$$\begin{aligned} q(\mathbf{v}) &= d_1 y_1^2 + \cdots + d_p y_p^2 - (-d_{p+1} y_{p+1}^2) - \cdots - (-d_r y_r^2) \\ &= (\sqrt{d_1} y_1)^2 + \cdots + (\sqrt{d_p} y_p)^2 - (\sqrt{|d_{p+1}|} y_{p+1})^2 - \cdots - (\sqrt{|d_r|} y_r)^2 \\ &= z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_r^2. \end{aligned} \quad \blacksquare$$

The form $q(\mathbf{v}) = z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_r^2$ in Theorem 8.26 is called the **canonical form** for q . There are two corollaries to the theorem concerning symmetric matrices. Proofs are requested in the exercises.

Corollary 8.27 Any real symmetric matrix A is congruent over \mathbf{R} to a unique matrix of the form

$$C = \begin{bmatrix} I_p & | & \mathbf{0} & | & \mathbf{0} \\ \hline - & + & - & + & - \\ \mathbf{0} & | & -I_{r-p} & | & \mathbf{0} \\ \hline - & + & - & + & - \\ \mathbf{0} & | & \mathbf{0} & | & \mathbf{0} \end{bmatrix},$$

where r is the rank of A .

The number p of positive 1's in C is called the **index** of A . Since the linear change of variable in the proof of Theorem 8.26 is clearly not necessarily orthogonal, the matrix P used to obtain $P^T AP = C$ in Corollary 8.27 may not be orthogonal.

Corollary 8.28 Two symmetric $n \times n$ matrices over \mathbf{R} are congruent over \mathbf{R} if and only if they have the same rank and the same index.

Definition 8.29 Let q be a real quadratic form on \mathbf{R}^n with rank r and index p .

- (1) If $p = r = n$, q is called **positive definite**.
- (2) If $p = r$, q is called **positive semidefinite**.
- (3) If $p = 0$ and $r = n$, q is called **negative definite**.
- (4) If $p = 0$, q is called **negative semidefinite**.

Each of the conditions in Definition 8.29 can be formulated in terms of the range of values of q . These formulations are given in the next theorem, with proofs requested in Problem 8.

Theorem 8.30 Let q be a real quadratic form on \mathbf{R}^n .

- (1) q is positive definite if and only if $q(\mathbf{v}) > 0$ for all $\mathbf{v} \neq \mathbf{0}$ in \mathbf{R}^n .
- (2) q is positive semidefinite if and only if $q(\mathbf{v}) \geq 0$ for all $\mathbf{v} \in \mathbf{R}^n$.
- (3) q is negative definite if and only if $q(\mathbf{v}) < 0$ for all $\mathbf{v} \neq \mathbf{0}$ in \mathbf{R}^n .
- (4) q is negative semidefinite if and only if $q(\mathbf{v}) \leq 0$ for all $\mathbf{v} \in \mathbf{R}^n$.

Example 1 □ In Example 1 of Section 8.5, we obtained the orthogonal matrix

$$P_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{bmatrix}$$

such that $P_1^T AP_1 = \text{diag}\{-1, 2, 2\}$, where

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

Thus the change of variables

$$\begin{aligned}x_1 &= \frac{1}{\sqrt{6}} (\sqrt{2}y_1 + \sqrt{3}y_2 + y_3) \\x_2 &= \frac{1}{\sqrt{6}} (\sqrt{2}y_1 - \sqrt{3}y_2 + y_3) \\x_3 &= \frac{1}{\sqrt{6}} (\sqrt{2}y_1 - 2y_2)\end{aligned}$$

reduces the quadratic form

$$q(\mathbf{v}) = q(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3$$

to the form

$$q(\mathbf{v}) = -y_1^2 + 2y_2^2 + 2y_3^2.$$

In the terminology of the first paragraph of this section, an interchange of the variables y_1 and y_3 corresponds to an interchange of \mathbf{u}_1 and \mathbf{u}_3 , or an interchange of columns one and three in P_1 . Thus

$$P_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & \sqrt{3} & \sqrt{2} \\ 1 & -\sqrt{3} & \sqrt{2} \\ -2 & 0 & \sqrt{2} \end{bmatrix}$$

is an orthogonal matrix such that $P_2^T AP_2 = \text{diag}\{2, 2, -1\}$. This means that the orthogonal change of variable $[x_1, x_2, x_3]^T = P_2[y'_1, y'_2, y'_3]^T$ reduces $q(\mathbf{v})$ to the form

$$q(\mathbf{v}) = 2(y'_1)^2 + 2(y'_2)^2 - (y'_3)^2.$$

Following the method of the proof of Theorem 8.26, we write $y'_1 = \frac{1}{\sqrt{2}}z_1$, $y'_2 = \frac{1}{\sqrt{2}}z_2$, and $y'_3 = z_3$. This reduces q to the canonical form

$$q(\mathbf{v}) = z_1^2 + z_2^2 - z_3^2.$$

The matrix of transition P_3 corresponding to this last change of variables $[y'_1, y'_2, y'_3]^T = P_3[z_1, z_2, z_3]^T$ is

$$P_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Combining, we have the change of variable $[x_1, x_2, x_3]^T = P_2P_3[z_1, z_2, z_3]^T$, and

$$P = P_2P_3 = \frac{1}{2\sqrt{3}} \begin{bmatrix} 1 & \sqrt{3} & 2 \\ 1 & -\sqrt{3} & 2 \\ -2 & 0 & 2 \end{bmatrix}$$

is a matrix such that $P^T AP = \text{diag}\{1, 1, -1\}$. We note that P is not orthogonal. It is clear that q has rank 3, index 2, and signature 1. None of the terms in Definition 8.29 apply to q . ■

Exercises 8.6

1. For each $\mathbf{v} = (x_1, x_2, \dots, x_n)$ in \mathbf{R}^n , a quadratic form q is defined by $q(\mathbf{v}) = \mathbf{X}^T A \mathbf{X}$ for the given A . Find the rank, index, and signature of q .

$$(a) A = \begin{bmatrix} \frac{3}{2} & \sqrt{2} & -\frac{1}{2} \\ \sqrt{2} & 1 & -\sqrt{2} \\ -\frac{1}{2} & -\sqrt{2} & -\frac{5}{2} \end{bmatrix} \quad (b) A = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 2 & -2 \\ 0 & -2 & 3 \end{bmatrix}$$

$$(c) A = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \quad (d) A = \begin{bmatrix} -2 & 2 & 2 \\ 2 & 1 & 4 \\ 2 & 4 & 1 \end{bmatrix}$$

2. Find the canonical form for each of the quadratic forms referred to in Problem 1.
3. For each of the following matrices A , find an invertible real matrix P such that $P^T A P$ is of the form C given in Corollary 8.27.

$$(a) A = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix} \quad (b) A = \begin{bmatrix} 1 & 2 & -4 \\ 2 & -2 & -2 \\ -4 & -2 & 1 \end{bmatrix}$$

$$(c) A = \begin{bmatrix} 17 & 2 & -2 \\ 2 & 14 & 4 \\ -2 & 4 & 14 \end{bmatrix} \quad (d) A = \begin{bmatrix} 4 & -1 & 0 & 1 \\ -1 & 5 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 1 & 0 & -1 & 5 \end{bmatrix}$$

4. For each of the matrices A in Problem 3, let q be the quadratic form on \mathbf{R}^n with $q(\mathbf{v}) = \mathbf{X}^T A \mathbf{X}$. Find a basis \mathcal{B} of \mathbf{R}^n such that $q(\mathbf{v})$ has the form

$$q(\mathbf{v}) = z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_r^2$$

with $[z_1, z_2, \dots, z_n]^T$ as the coordinate matrix of \mathbf{v} relative to \mathcal{B} .

5. Prove Corollary 8.24.
6. Prove Corollary 8.27.
7. Prove Corollary 8.28.
8. Prove Theorem 8.30.
9. A real symmetric matrix A is defined to be **positive definite** if the quadratic form $q(x_1, x_2, \dots, x_n) = \mathbf{X}^T A \mathbf{X}$ is positive definite. The terms **positive semidefinite**, etc., are defined similarly.

- (a) Prove that A is positive definite if and only if $A = B^T B$ for some real invertible matrix B .
- (b) Prove that A is positive semidefinite if and only if there exists a (possibly singular) real matrix Q such that $A = Q^T Q$.
- (c) Prove that A is positive definite if and only if all of the eigenvalues of A are positive.

8.7 Bilinear Forms

The *Cartesian product* $\mathcal{S} \times \mathcal{T}$ of two sets \mathcal{S}, \mathcal{T} is the set of all ordered pairs of elements from \mathcal{S} and \mathcal{T} :

$$\mathcal{S} \times \mathcal{T} = \{(s, t) \mid s \in \mathcal{S} \text{ and } t \in \mathcal{T}\}.$$

Few concepts of such simplicity have been put to such a wide range of uses in mathematics. For this is the basic concept behind the coordinate systems, the definitions of binary relation and function, and the constructions of several number systems. Although not stated explicitly in our development, the idea is implicit in the familiar vector spaces \mathcal{F}^n . For the next three sections, we shall be concerned with Cartesian products of sets of vectors, or rather, with functions defined on such sets of vectors.

Definition 8.31 Let \mathbf{U} and \mathbf{V} be vector spaces over the same field \mathcal{F} . A **bilinear form** on \mathbf{U} and \mathbf{V} is a mapping f of pairs of vectors (\mathbf{u}, \mathbf{v}) in $\mathbf{U} \times \mathbf{V}$ onto scalars $f(\mathbf{u}, \mathbf{v})$ in \mathcal{F} that has the properties

$$(i) \quad f(a_1\mathbf{u}_1 + a_2\mathbf{u}_2, \mathbf{v}) = a_1f(\mathbf{u}_1, \mathbf{v}) + a_2f(\mathbf{u}_2, \mathbf{v})$$

and

$$(ii) \quad f(\mathbf{u}, b_1\mathbf{v}_1 + b_2\mathbf{v}_2) = b_1f(\mathbf{u}, \mathbf{v}_1) + b_2f(\mathbf{u}, \mathbf{v}_2).$$

The conditions (i) and (ii) are described by saying that f is a linear function of each of the vector variables. These conditions (i) and (ii) are together equivalent to the single requirement that

$$\begin{aligned} (iii) \quad & f(a_1\mathbf{u}_1 + a_2\mathbf{u}_2, b_1\mathbf{v}_1 + b_2\mathbf{v}_2) \\ &= a_1b_1f(\mathbf{u}_1, \mathbf{v}_1) + a_1b_2f(\mathbf{u}_1, \mathbf{v}_2) + a_2b_1f(\mathbf{u}_2, \mathbf{v}_1) + a_2b_2f(\mathbf{u}_2, \mathbf{v}_2). \end{aligned}$$

These linearity conditions extend readily to

$$(iv) \quad f\left(\sum_{i=1}^m a_i\mathbf{u}_i, \sum_{j=1}^n b_j\mathbf{v}_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j f(\mathbf{u}_i, \mathbf{v}_j)$$

for all positive integers m and n . Thus, a function f from $\mathbf{U} \times \mathbf{V}$ to \mathcal{F} is a bilinear form on \mathbf{U} and \mathbf{V} if and only if (iv) is satisfied as an identity.

We are primarily interested in the case where the vector spaces \mathbf{U} and \mathbf{V} in Definition 8.31 are finite-dimensional. For the remainder of this section, \mathbf{U} and \mathbf{V} will denote vector spaces over \mathcal{F} with dimensions m and n , respectively.

Definition 8.32 Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be bases of \mathbf{U} and \mathbf{V} , respectively, and let f be a bilinear form on \mathbf{U} and \mathbf{V} . The matrix of f relative to \mathcal{A} and \mathcal{B} is the matrix $A = [a_{ij}]_{m \times n}$, where

$$a_{ij} = f(\mathbf{u}_i, \mathbf{v}_j)$$

for $i = 1, 2, \dots, m; j = 1, 2, \dots, n$.

We say that the matrix A represents the bilinear form f . As with linear functionals, the function values $f(\mathbf{u}, \mathbf{v})$ can be expressed compactly by use of the matrix of f .

Theorem 8.33 Let $X = [x_1, x_2, \dots, x_m]^T$ denote the coordinate matrix of a vector $\mathbf{u} \in \mathbf{U}$ relative to the basis \mathcal{A} of \mathbf{U} , and let $Y = [y_1, y_2, \dots, y_n]^T$ be the coordinate matrix of a vector $\mathbf{v} \in \mathbf{V}$ relative to the basis \mathcal{B} of \mathbf{V} . If $A = [a_{ij}]_{m \times n}$, then A is the matrix of the bilinear form f relative to \mathcal{A} and \mathcal{B} if and only if the equation

$$f(\mathbf{u}, \mathbf{v}) = X^T A Y$$

is satisfied for all choices of $\mathbf{u} \in \mathbf{U}, \mathbf{v} \in \mathbf{V}$.

Proof. Assume first that $A = [a_{ij}]_{m \times n}$ is the matrix of f relative to \mathcal{A} and \mathcal{B} . This means that $a_{ij} = f(\mathbf{u}_i, \mathbf{v}_j)$ for $i = 1, 2, \dots, m; j = 1, 2, \dots, n$. For any $\mathbf{u} = \sum_{i=1}^m x_i \mathbf{u}_i$, and $\mathbf{v} = \sum_{j=1}^n y_j \mathbf{v}_j$, the condition (iv) stated previously implies that

$$\begin{aligned} f(\mathbf{u}, \mathbf{v}) &= \sum_{i=1}^m \sum_{j=1}^n x_i y_j f(\mathbf{u}_i, \mathbf{v}_j) \\ &= \sum_{i=1}^m \sum_{j=1}^n x_i a_{ij} y_j \\ &= \sum_{i=1}^m x_i \left(\sum_{j=1}^n a_{ij} y_j \right). \end{aligned}$$

Now $\sum_{j=1}^n a_{ij} y_j$ is the element in the i^{th} row of the $m \times 1$ matrix AY , and thus

$$f(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^m x_i \left(\sum_{j=1}^n a_{ij} y_j \right) = X^T A Y.$$

Assume, on the other hand, that $f(\mathbf{u}, \mathbf{v}) = X^T A Y$ is satisfied for all $\mathbf{u} \in \mathbf{U}, \mathbf{v} \in \mathbf{V}$. For the particular choices $\mathbf{u} = \mathbf{u}_i$ and $\mathbf{v} = \mathbf{v}_j$, we have $X = [\delta_{1i}, \delta_{2i}, \dots, \delta_{mi}]^T$ and

$Y = [\delta_{1j}, \delta_{2j}, \dots, \delta_{nj}]^T$. Hence

$$\begin{aligned}
f(\mathbf{u}_i, \mathbf{v}_j) &= X^T A Y \\
&= \begin{bmatrix} \delta_{1i} & \delta_{2i} & \cdots & \delta_{mi} \end{bmatrix} \begin{bmatrix} \sum_{k=1}^n a_{1k} \delta_{kj} \\ \sum_{k=1}^n a_{2k} \delta_{kj} \\ \vdots \\ \sum_{k=1}^n a_{mk} \delta_{kj} \end{bmatrix} \\
&= \begin{bmatrix} \delta_{1i} & \delta_{2i} & \cdots & \delta_{mi} \end{bmatrix} \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \\
&= \sum_{k=1}^m \delta_{ki} a_{kj} \\
&= a_{ij},
\end{aligned}$$

and $A = [a_{ij}]_{m \times n}$ is the matrix of f relative to \mathcal{A} and \mathcal{B} . ■

It follows from Definition 8.32 and Theorem 8.33 that the two bilinear forms f and g on \mathbf{U} and \mathbf{V} are equal if and only if they have the same matrix relative to fixed bases \mathcal{A} of \mathbf{U} and \mathcal{B} of \mathbf{V} .

As we have in similar situations previously, we ask about the effects of changes of bases in \mathbf{U} and \mathbf{V} . The answer is obtained quite easily.

Theorem 8.34 *Let A be the matrix of the bilinear form f relative to the bases \mathcal{A} of \mathbf{U} and \mathcal{B} of \mathbf{V} . If Q is the matrix of transition from \mathcal{A} to the basis \mathcal{A}' of \mathbf{U} and P is the matrix of transition from \mathcal{B} to the basis \mathcal{B}' of \mathbf{V} , then $Q^T AP$ is the matrix of f relative to \mathcal{A}' and \mathcal{B}' .*

Proof. Suppose that $\mathbf{u} \in \mathbf{U}$ has coordinate matrix X relative to \mathcal{A} and X' relative to \mathcal{A}' . Let $\mathbf{v} \in \mathbf{V}$ have coordinate matrix Y relative to \mathcal{B} and Y' relative to \mathcal{B}' . By Theorem 5.14, $X = QX'$ and $Y = PY'$. Combined with the “only if” part of Theorem 8.33, this yields

$$\begin{aligned}
f(\mathbf{u}, \mathbf{v}) &= X^T A Y \\
&= (QX')^T A (PY') \\
&= (X')^T (Q^T A P) Y'.
\end{aligned}$$

But, by the “if” part of Theorem 8.33, this means that $Q^T AP$ is the matrix of f relative to \mathcal{A}' and \mathcal{B}' . ■

Corollary 8.35 Two $m \times n$ matrices A and B over \mathcal{F} represent the same bilinear form on \mathbf{U} and \mathbf{V} relative to certain (not necessarily different) choices of bases of \mathbf{U} and \mathbf{V} if and only if A and B are equivalent over \mathcal{F} .

Proof. See Problem 5. ■

The **rank** of a bilinear form f is defined to be the rank r of any matrix that represents f .

Corollary 8.36 Let f be a bilinear form on \mathbf{U} and \mathbf{V} . With suitable choices of bases in \mathbf{U} and \mathbf{V} , f can be represented by a matrix D_r that has the first r diagonal elements equal to 1 and all other elements zero.

Proof. See Problem 6. ■

The results of this section are illustrated in our next example.

Example 1 □ Let f be the bilinear form on $\mathbf{U} = \mathbf{R}^3$ and $\mathbf{V} = \mathbf{R}^2$ that is defined by

$$f((x_1, x_2, x_3), (y_1, y_2)) = -7x_1y_1 - 10x_1y_2 - 2x_2y_1 - 3x_2y_2 + 12x_3y_1 + 17x_3y_2.$$

We consider the following problems.

- (a) Write the matrix A of f relative to $\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$ and $\mathcal{B} = \{(1, -1), (2, -1)\}$.
- (b) Use the matrix A to compute the value of $f((2, 3, 1), (0, -1))$.
- (c) Determine bases \mathcal{A}' of \mathbf{U} and \mathcal{B}' of \mathbf{V} such that the matrix of f relative to \mathcal{A}' and \mathcal{B}' is of the form D_r described in Corollary 8.36.

We obtain the matrix A by computing the values $a_{ij} = f(\mathbf{u}_i, \mathbf{v}_j)$. For example,

$$\begin{aligned} a_{11} &= f((1, 0, 0), (1, -1)) \\ &= -7(1)(1) - 10(1)(-1) - 2(0)(1) - 3(0)(-1) + 12(0)(1) + 17(0)(-1) \\ &= 3. \end{aligned}$$

Similarly,

$$\begin{aligned} a_{21} &= f((1, 1, 0), (1, -1)) = 4, \\ a_{31} &= f((1, 1, 1), (1, -1)) = -1, \end{aligned}$$

and so on. Thus

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

In order to use A to compute the value of $f((2, 3, 1), (0, -1))$, we first write

$$(2, 3, 1) = (-1)(1, 0, 0) + (2)(1, 1, 0) + (1)(1, 1, 1)$$

and

$$(0, -1) = 2(1, -1) + (-1)(2, -1).$$

Then the equation $f(\mathbf{u}, \mathbf{v}) = \mathbf{X}^T A \mathbf{Y}$ yields

$$f((2, 3, 1), (0, -1)) = [-1, 2, 1] \begin{bmatrix} 3 & -4 \\ 4 & -5 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 12.$$

By the method of Section 3.8, we find that

$$Q^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & -2 & 1 \end{bmatrix} \text{ and } P = \begin{bmatrix} -5 & 4 \\ -4 & 3 \end{bmatrix}$$

are invertible matrices such that

$$Q^T A P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Using Q as the matrix of transition from \mathcal{A} to \mathcal{A}' and P as the matrix of transition from \mathcal{B} to \mathcal{B}' , we obtain

$$\mathcal{A}' = \{(1, 0, 0), (1, 1, 0), (2, -1, 1)\} \text{ and } \mathcal{B}' = \{(-13, 9), (10, -7)\}. \blacksquare$$

Exercises 8.7

- Find the matrix of the given bilinear form f on \mathbf{U} and \mathbf{V} with respect to the given bases \mathcal{A} and \mathcal{B} .
 - $\mathbf{U} = \mathbf{R}^2, \mathbf{V} = \mathbf{R}^3, \mathcal{A} = \mathcal{E}_2, \mathcal{B} = \mathcal{E}_3,$
 $f((x_1, x_2), (y_1, y_2, y_3)) = 2x_1y_2 - x_1y_3 + 2x_2y_1 + x_2y_2 + x_2y_3$
 - $\mathbf{U} = \mathbf{V} = \mathbf{R}^3, \mathcal{A} = \mathcal{B} = \mathcal{E}_3,$
 $f((x_1, x_2, x_3), (y_1, y_2, y_3)) = x_1y_1 - x_1y_2 + 2x_2y_1 + x_2y_3 + 2x_3y_2 + x_3y_3$
 - $\mathbf{U} = \mathbf{V} = \mathbf{R}^3, \mathcal{A} = \mathcal{B} = \{(1, 1, 1), (1, 1, 0), (1, 0, 0)\},$
 $f((x_1, x_2, x_3), (y_1, y_2, y_3)) = 2x_1y_2 - x_1y_3 + 2x_2y_1 + x_2y_2 + x_2y_3 - x_3y_1 + x_3y_2 - 2x_3y_3$

- (d) $\mathbf{U} = \mathbf{R}^3, \mathbf{V} = \mathbf{R}^2, \mathcal{A} = \{(1, 1, 0), (1, -1, 1), (0, 1, 0)\}, \mathcal{B} = \{(1, 2), (2, 1)\},$
 $f((x_1, x_2, x_3), (y_1, y_2)) = 2x_1y_1 + 4x_1y_2 - 6x_2y_1 + 3x_2y_2 + x_3y_1$
- (e) $\mathbf{U} = \mathbf{C}^3, \mathbf{V} = \mathbf{C}^2, \mathcal{A} = \{(i, 0, 0), (1, i, 0), (0, 0, 2i)\}, \mathcal{B} = \{(1-i, i), (i, -i)\},$
 $f((x_1, x_2, x_3), (y_1, y_2)) = 5x_1y_1 + ix_1y_2 - ix_2y_1 + 2x_2y_2 + 2x_3y_1 - x_3y_2$
- (f) $\mathbf{U} = \mathbf{C}^3, \mathbf{V} = \mathbf{C}^2, \mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}, \mathcal{B} = \{(1, -1), (2, -1)\},$
 $f((x_1, x_2, x_3), (y_1, y_2)) = x_1y_2 + x_2y_1 + 2x_2y_2 - 2x_3y_1 + 2x_3y_2$

2. Use the matrix obtained in the corresponding part of Problem 1 to compute $f(\mathbf{u}, \mathbf{v})$ for the given vectors.

- (a) $\mathbf{u} = (3, -1), \mathbf{v} = (0, 4, -1)$ (b) $\mathbf{u} = (5, -6, 3), \mathbf{v} = (2, -1, 0)$
 (c) $\mathbf{u} = (1, -2, 2), \mathbf{v} = (0, 0, 1)$ (d) $\mathbf{u} = (0, 3, -1), \mathbf{v} = (8, 7)$
 (e) $\mathbf{u} = (i, 0, i), \mathbf{v} = (2, 0)$ (f) $\mathbf{u} = (1, 0, -1), \mathbf{v} = (1, 0)$

3. Each of parts (a)-(f) below relate to the corresponding part of Problem 1 above. In each case, a new pair of bases $\mathcal{A}', \mathcal{B}'$ is given for the vector spaces \mathbf{U}, \mathbf{V} . Find the matrix of f relative to \mathcal{A}' and \mathcal{B}' by use of Theorem 8.34.

- (a) $\mathcal{A}' = \{(4, -1), (-1, 5)\}, \mathcal{B}' = \{(1, 2, -4), (1, -1, -1), (-4, -2, 1)\}$
 (b) $\mathcal{A}' = \mathcal{B}' = \{(17, 2, -2), (1, 7, 2), (-1, 2, 7)\}$
 (c) $\mathcal{A}' = \mathcal{B}' = \{(5, 4, 2), (1, -1, -2), (1, -1, 1)\}$
 (d) $\mathcal{A}' = \{(3, 1, 2), (3, -1, 1), (0, 5, -2)\}, \mathcal{B}' = \{(1, -1), (0, 3)\}$
 (e) $\mathcal{A}' = \{(-1, 0, 0), (-1, -i, 0), (0, 0, 1)\}, \mathcal{B}' = \{(2, 1+i), (1, -1)\}$
 (f) $\mathcal{A}' = \{(1, 0, 0), (0, 1, 0), (-6, 2, 1)\}, \mathcal{B}' = \{(-2, 1), (1, 0)\}$

4. Let f be the bilinear form on \mathbf{R}^4 and \mathbf{R}^3 that has the given matrix A relative to \mathcal{E}_4 and \mathcal{E}_3 . Find bases \mathcal{A}' of \mathbf{R}^4 and \mathcal{B}' of \mathbf{R}^3 such that, relative to \mathcal{A}' and \mathcal{B}' , f has a matrix of the form D_r described in Corollary 8.36.

$$\begin{array}{ll} \text{(a)} A = \begin{bmatrix} 1 & 0 & 2 \\ 4 & 1 & 3 \\ 3 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix} & \text{(b)} A = \begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \\ 1 & 0 & -1 \\ -1 & 0 & 1 \end{bmatrix} \\ \\ \text{(c)} A = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 1 & 3 \\ 4 & 2 & 6 \\ 1 & 0 & 1 \end{bmatrix} & \text{(d)} A = \begin{bmatrix} 0 & 1 & 2 \\ 3 & -2 & -1 \end{bmatrix} \end{array}$$

5. Prove Corollary 8.35.

6. Prove Corollary 8.36.
7. Prove that the rank of a bilinear form as defined in this section is unique.
8. If f and g are two bilinear forms on \mathbf{U} and \mathbf{V} , then $f + g$ is defined by

$$(f + g)(\mathbf{u}, \mathbf{v}) = f(\mathbf{u}, \mathbf{v}) + g(\mathbf{u}, \mathbf{v})$$

for all $(\mathbf{u}, \mathbf{v}) \in \mathbf{U} \times \mathbf{V}$. Also, for any $a \in \mathcal{F}$, the product af is defined by

$$(af)(\mathbf{u}, \mathbf{v}) = a \cdot f(\mathbf{u}, \mathbf{v}).$$

Prove that the set of all bilinear forms on \mathbf{U} and \mathbf{V} is a vector space with respect to these operations of addition and scalar multiplication.

9. (See Problem 8.) Let \mathbf{W} be the vector space of all bilinear forms on \mathbf{R}^m and \mathbf{R}^n . Prove that the mapping $m(f) = A$ that sends a bilinear form onto its matrix A relative to \mathcal{E}_m and \mathcal{E}_n is an isomorphism of the vector space \mathbf{W} onto the vector space $\mathbf{R}_{m \times n}$ as defined in Section 4.2.

8.8 Symmetric Bilinear Forms

As in the case of linear transformations, there is a great deal of interest in those bilinear forms for the special case where $\mathbf{V} = \mathbf{U}$. In this case, we shorten the phrase “bilinear form on \mathbf{V} and \mathbf{V} ” to “bilinear form on \mathbf{V} .“ We also agree that, without a statement to the contrary, the bases \mathcal{A} and \mathcal{B} are the same whenever $\mathbf{V} = \mathbf{U}$. This means that $a_{ij} = f(\mathbf{u}_i, \mathbf{u}_j)$ in the matrix of f relative to \mathcal{A} . Throughout this section, we shall have $\mathbf{V} = \mathbf{U}$ and \mathbf{V} an n -dimensional vector space over \mathcal{F} .

The imposition of the conditions that $\mathbf{U} = \mathbf{V}$ and $\mathcal{A} = \mathcal{B}$ means that our bilinear forms are now represented by square matrices only. As a more interesting consequence of these conditions, we have the next theorem. This result should be compared with Corollary 8.35 in the preceding section.

Theorem 8.37 *Let A and B be square matrices of order n over \mathcal{F} . Then B is congruent to A over \mathcal{F} if and only if B and A represent the same bilinear form on \mathbf{V} .*

Proof. Let A and B be matrices of order n over \mathcal{F} , and suppose that A represents the bilinear form f on \mathbf{V} relative to the basis \mathcal{A} .

Assume first that B is congruent to A over \mathcal{F} . Then $B = P^T AP$ for some invertible P over \mathcal{F} . Since P is invertible with elements in \mathcal{F} , P is the transition matrix from \mathcal{A} to a basis \mathcal{A}' of \mathbf{V} . With $\mathbf{U} = \mathbf{V}$, $\mathcal{A} = \mathcal{B}$, $Q = P$, and $\mathcal{A}' = \mathcal{B}'$ in Theorem 8.34, we have $B = P^T AP$ is the matrix of f relative to \mathcal{A}' .

On the other hand, suppose that B represents f relative to some basis \mathcal{A}' of \mathbf{V} . If P is the matrix of transition from \mathcal{A} to \mathcal{A}' , then Theorem 8.34 asserts that $P^T AP$ is the matrix of f relative to \mathcal{A}' . But this matrix is unique, so it must be that $B = P^T AP$. Now P is invertible with elements in \mathcal{F} since it is the transition matrix from one basis of \mathbf{V} to another. Hence B is congruent to A over \mathcal{F} . ■

The connection established between bilinear forms and matrices in Theorem 8.33 suggests the possibility of describing properties of bilinear forms in terms of their matrices. Several interesting results along these lines can be obtained whenever the matrices are square.

Definition 8.38 A bilinear form f on \mathbf{V} is **symmetric** if $f(\mathbf{u}, \mathbf{v}) = f(\mathbf{v}, \mathbf{u})$ for all $\mathbf{u}, \mathbf{v} \in \mathbf{V}$.

Theorem 8.39 A bilinear form f on \mathbf{V} is symmetric if and only if every matrix that represents f is symmetric.

Proof. Suppose that f is symmetric, and let $A = [a_{ij}]_n$ be a matrix that represents f . Then $a_{ij} = f(\mathbf{u}_i, \mathbf{u}_j) = f(\mathbf{u}_j, \mathbf{u}_i) = a_{ji}$, and A is symmetric.

Suppose now that f is represented by a symmetric matrix $A = [a_{ij}]_n$ relative to the basis \mathcal{A} of \mathbf{V} . Let \mathbf{u} and \mathbf{v} be arbitrary vectors with coordinate matrices X and Y , respectively, relative to \mathcal{A} . Then $f(\mathbf{u}, \mathbf{v}) = X^T A Y$ and $f(\mathbf{v}, \mathbf{u}) = Y^T A X$. Since $Y^T A X$ is a 1 by 1 matrix, $Y^T A X = (X^T A Y)^T$. Thus

$$f(\mathbf{v}, \mathbf{u}) = (Y^T A X)^T = X^T A^T (Y^T)^T = X^T A Y = f(\mathbf{u}, \mathbf{v}),$$

where $A^T = A$, since A is symmetric. ■

Our next theorem has an immediate corollary concerning symmetric bilinear forms.

Theorem 8.40 If $1 + 1 \neq 0$ in \mathcal{F} , every symmetric matrix A of order n over \mathcal{F} is congruent over \mathcal{F} to a diagonal matrix.

Proof. The proof is by induction on the order n of A . The theorem is trivially true for $n = 1$.

Assume that the theorem is true for all symmetric matrices of order k over \mathcal{F} , and let A be a symmetric matrix of order $k + 1$ over \mathcal{F} . Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k+1}\}$ be a basis of \mathcal{F}^{k+1} , and let f be the symmetric bilinear form on \mathcal{F}^{k+1} that has matrix A relative to \mathcal{A} . If A is the zero matrix, there is nothing to prove. Assume, then, that $a_{rs} = f(\mathbf{u}_r, \mathbf{u}_s) \neq 0$ for the pair r, s . If $f(\mathbf{u}_r, \mathbf{u}_r) = 0$ and $f(\mathbf{u}_s, \mathbf{u}_s) = 0$, then

$$f(\mathbf{u}_r + \mathbf{u}_s, \mathbf{u}_r + \mathbf{u}_s) = 2f(\mathbf{u}_r, \mathbf{u}_s) \neq 0.$$

Thus there is a vector $\mathbf{v}_1 \in \mathcal{F}^{k+1}$ such that $d_1 = f(\mathbf{v}_1, \mathbf{v}_1) \neq 0$. The set $\{\mathbf{v}_1\}$ can be extended to a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}\}$ of \mathcal{F}^{k+1} . The set $\mathcal{A}' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_{k+1}\}$ is obtained from the basis \mathcal{B} as follows: $\mathbf{u}'_1 = \mathbf{v}_1$ and $\mathbf{u}'_j = \mathbf{v}_j - \frac{f(\mathbf{u}'_1, \mathbf{v}_j)}{d_1} \mathbf{u}'_1$ for $j = 2, \dots, k + 1$. That is,

$$\begin{aligned} \mathbf{u}'_1 &= \mathbf{v}_1 \\ \mathbf{u}'_2 &= \mathbf{v}_2 - \frac{f(\mathbf{u}'_1, \mathbf{v}_2)}{d_1} \mathbf{u}'_1 \\ &\vdots \\ \mathbf{u}'_{k+1} &= \mathbf{v}_{k+1} - \frac{f(\mathbf{u}'_1, \mathbf{v}_{k+1})}{d_1} \mathbf{u}'_1. \end{aligned}$$

It is clear, then, that \mathcal{A}' can be obtained from \mathcal{B} by a sequence of elementary operations. Therefore, \mathcal{A}' is a basis of \mathcal{F}^{k+1} . Let A_1 denote the matrix of f relative to \mathcal{A}' . The matrix A_1 is symmetric since f is symmetric. Now $f(\mathbf{u}'_1, \mathbf{u}'_1) = f(\mathbf{v}_1, \mathbf{v}_1) = d_1$, and if $j > 1$,

$$\begin{aligned} f(\mathbf{u}'_1, \mathbf{u}'_j) &= f\left(\mathbf{u}'_1, \mathbf{v}_j - \frac{f(\mathbf{u}'_1, \mathbf{v}_j)}{d_1}\mathbf{u}'_1\right) \\ &= f(\mathbf{u}'_1, \mathbf{v}_j) - \frac{f(\mathbf{u}'_1, \mathbf{v}_j)}{d_1}f(\mathbf{u}'_1, \mathbf{u}'_1) \\ &= 0. \end{aligned}$$

Thus A_1 is of the form

$$A_1 = \begin{bmatrix} d_1 & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \\ \hline \mathbf{0} & | & A_2 \end{bmatrix},$$

where A_2 is a symmetric $k \times k$ matrix. If Q_1 denotes the matrix of transition from \mathcal{A} to \mathcal{B} and Q_2 denotes the matrix of transition from \mathcal{B} to \mathcal{A}' , then $P_1 = Q_1 Q_2$ denotes the invertible matrix of transition from \mathcal{A} to \mathcal{A}' , and $A_1 = P_1^T A P_1$ by Theorem 8.34.

By the induction hypothesis, there is an invertible matrix Q of order k such that $Q^T A_2 Q = \text{diag}\{d_2, \dots, d_{k+1}\}$. Hence

$$P_1 = \begin{bmatrix} 1 & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \\ \hline \mathbf{0} & | & Q \end{bmatrix}$$

is an invertible matrix, and

$$\begin{aligned} (P_1 P_2)^T A (P_1 P_2) &= P_2^T P_1^T A P_1 P_2 \\ &= P_2^T A_1 P_2 \\ &= \begin{bmatrix} 1 & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \\ \hline \mathbf{0} & | & Q^T \end{bmatrix} \begin{bmatrix} d_1 & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \\ \hline \mathbf{0} & | & A_2 \end{bmatrix} \begin{bmatrix} 1 & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \\ \hline \mathbf{0} & | & Q \end{bmatrix} \\ &= \begin{bmatrix} d_1 & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{0} \\ \hline \mathbf{0} & | & Q^T A_2 Q \end{bmatrix} \\ &= \text{diag}\{d_1, d_2, \dots, d_{k+1}\}. \end{aligned}$$

Thus $P = P_1 P_2$ is an invertible matrix such that $P^T A P$ is diagonal. The theorem follows by induction. ■

Corollary 8.41 *If $1 + 1 \neq 0$ in \mathcal{F} , then every symmetric bilinear form on \mathbf{V} can be represented by a diagonal matrix.*

Proof. The proof is left as an exercise (Problem 3). ■

Corollary 8.42 *Every symmetric matrix over \mathcal{C} is congruent over \mathcal{C} to a diagonal matrix D_r with the first r diagonal elements equal to 1 and all other elements zero.*

Proof. See Problem 4. ■

Much the same as with Theorem 8.19, the proof of Theorem 8.40 provides a basis for an approach to the problem of finding an invertible matrix P over \mathcal{F} such that $P^T AP$ is diagonal, but this approach is not very efficient. We can proceed more directly in the following manner to obtain such a P . The first column of P is simply the coordinate matrix

$$P_1 = \begin{bmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{n1} \end{bmatrix}$$

relative to \mathcal{A} of a vector \mathbf{u}'_1 chosen such that

$$d_1 = f(\mathbf{u}'_1, \mathbf{u}'_1) \neq 0.$$

The vector \mathbf{u}'_2 must be chosen so that $\{\mathbf{u}'_1, \mathbf{u}'_2\}$ is linearly independent and $f(\mathbf{u}'_1, \mathbf{u}'_2) = 0$. This latter condition means that the coordinate matrix

$$P_2 = \begin{bmatrix} p_{12} \\ p_{22} \\ \vdots \\ p_{n2} \end{bmatrix}$$

of \mathbf{u}'_2 relative to \mathcal{A} must satisfy the equation

$$P_1^T A P_2 = 0.$$

That is, $x_1 = p_{12}, x_2 = p_{22}, \dots, x_n = p_{n2}$ must be a solution to the system $(P_1^T A)X = \mathbf{0}$. Similarly, the requirements that $f(\mathbf{u}'_1, \mathbf{u}'_3) = 0$ and $f(\mathbf{u}'_2, \mathbf{u}'_3) = 0$ are reflected by the conditions $(P_1^T A)P_3 = 0$ and $(P_2^T A)P_3 = 0$ on the coordinate matrix P_3 of \mathbf{u}'_3 relative to \mathcal{A} , and so on. The method is illustrated in the following example.

Example 1 □ The problem is to determine a real invertible matrix P such that P^TAP is diagonal, where

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

Let f be the symmetric bilinear form on \mathbf{R}^3 that has matrix A relative to \mathcal{E}_3 . Since $f(\mathbf{e}_i, \mathbf{e}_i) = 0$ for $i = 1, 2, 3$ and $f(\mathbf{e}_1, \mathbf{e}_2) \neq 0$, we choose $\mathbf{u}'_1 = \mathbf{e}_1 + \mathbf{e}_2$ such that $f(\mathbf{u}'_1, \mathbf{u}'_1) \neq 0$. Then $P_1 = [1, 1, 0]^T$ and

$$P_1^T A = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}.$$

The choice $P_2 = [1, 1, -1]^T$ satisfies $(P_1^T A)P_2 = 0$ and makes $\{\mathbf{u}'_1, \mathbf{u}'_2\}$ linearly independent. The matrix $P_2^T A$ is given by

$$P_2^T A = \begin{bmatrix} -1 & 1 & 2 \end{bmatrix}.$$

The choice $P_3 = [0, -2, 1]^T$ satisfies $(P_1^T A)P_3 = 0$ and $(P_2^T A)P_3 = 0$, and makes $\{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\}$ linearly independent. The matrix $P = [P_1, P_2, P_3]$ is given by

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

and $P^T AP = \text{diag}\{2, -2, 0\}$. ■

The following definition and theorem apply to infinite-dimensional vector spaces as well as those of finite dimension, although our main interest is in the latter case.

Definition 8.43 Let \mathbf{V} be a vector space over \mathcal{F} . A mapping q of \mathbf{V} into \mathcal{F} is a quadratic form on \mathbf{V} if and only if there is a symmetric bilinear form f on \mathbf{V} such that $q(\mathbf{u}) = f(\mathbf{u}, \mathbf{u})$ for all $\mathbf{u} \in \mathbf{V}$.

It should be observed that the definition of a real quadratic form that was given earlier is entirely consistent with this definition.

It is clear that each symmetric bilinear form f determines a unique associated quadratic form q by the rule that $q(\mathbf{u}) = f(\mathbf{u}, \mathbf{u})$. The following theorem shows that this correspondence between symmetric bilinear forms and quadratic forms is one-to-one if $1 + 1 \neq 0$ in \mathcal{F} .

Theorem 8.44 Let \mathbf{V} be a vector space over the field \mathcal{F} in which $1 + 1 \neq 0$. If the quadratic form q on \mathbf{V} is determined by the symmetric bilinear form f on \mathbf{V} , then

$$f(\mathbf{u}, \mathbf{v}) = \frac{1}{2}[q(\mathbf{u} + \mathbf{v}) - q(\mathbf{u}) - q(\mathbf{v})]$$

for all \mathbf{u}, \mathbf{v} in \mathbf{V} .

Proof. The theorem follows from the simplification

$$\begin{aligned}
 \frac{1}{2}[q(\mathbf{u} + \mathbf{v}) - q(\mathbf{u}) - q(\mathbf{v})] &= \frac{1}{2}[f(\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) - f(\mathbf{u}, \mathbf{u}) - f(\mathbf{v}, \mathbf{v})] \\
 &= \frac{1}{2}[f(\mathbf{u}, \mathbf{u}) + f(\mathbf{u}, \mathbf{v}) + f(\mathbf{v}, \mathbf{u}) + f(\mathbf{v}, \mathbf{v}) \\
 &\quad - f(\mathbf{u}, \mathbf{u}) - f(\mathbf{v}, \mathbf{v})] \\
 &= \frac{1}{2}[2f(\mathbf{u}, \mathbf{v})] \\
 &= f(\mathbf{u}, \mathbf{v}). \quad \blacksquare
 \end{aligned}$$

Definition 8.45 Let \mathbf{V} be a finite-dimensional vector space over the field \mathcal{F} in which $1 + 1 \neq 0$. The matrix of the quadratic form q relative to the basis \mathcal{A} of \mathbf{V} is the same as the matrix of the symmetric bilinear form f that determines q .

We note that the matrix of a quadratic form is required to be symmetric.

For the remainder of this chapter, we restrict our attention to those vector spaces that have a field of scalars \mathcal{F} in which $1 + 1 \neq 0$.

The intimate connection between quadratic forms and symmetric bilinear forms means that many of the results obtained for symmetric bilinear forms translate immediately into statements about quadratic forms. The most important of these are listed next.

1. If q has matrix A relative to the basis \mathcal{A} of \mathbf{V} and \mathbf{u} has coordinate matrix X relative to \mathcal{A} , then $q(\mathbf{u}) = X^T A X$ (Theorem 8.33).
2. If q has matrix A relative to the basis \mathcal{A} of \mathbf{V} and P is the matrix of transition from \mathcal{A} to \mathcal{A}' , then $P^T A P$ is the matrix of q relative to \mathcal{A}' (Theorem 8.34).
3. Every quadratic form on \mathbf{V} can be represented by a diagonal matrix (Corollary 8.41).
4. Every quadratic form on a vector space over \mathcal{C} can be represented by a matrix D_r with the first r diagonal elements equal to 1 and all other elements zero (Problem 6).

Exercises 8.8

1. For each matrix A , find an invertible matrix P such that $P^T A P$ is diagonal.

$$\begin{array}{lll}
 \text{(a)} \ A = \begin{bmatrix} 0 & 2 & -1 \\ 2 & 1 & 1 \\ -1 & 1 & -2 \end{bmatrix} & \text{(b)} \ A = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 0 \\ 1 & 0 & 0 \end{bmatrix} & \text{(c)} \ A = \begin{bmatrix} 0 & 2 & 3 \\ 2 & 0 & -2 \\ 3 & -2 & 0 \end{bmatrix} \\
 \text{(d)} \ A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & -2 & -3 \\ 0 & -3 & 4 \end{bmatrix} & \text{(e)} \ A = \begin{bmatrix} 0 & -1 & 2 \\ -1 & 0 & -3 \\ 2 & -3 & 0 \end{bmatrix} & \text{(f)} \ A = \begin{bmatrix} 0 & 2 & 3 \\ 2 & 1 & -1 \\ 3 & -1 & 1 \end{bmatrix}
 \end{array}$$

2. For each matrix A in Problem 1, let f be the bilinear form on \mathbf{R}^3 that has matrix A relative to the basis $\mathcal{A} = \{(1, 1, 1), (1, 0, 1), (0, 1, -1)\}$. Use the matrix P from Problem 1 to find a basis of \mathbf{R}^3 relative to which f is represented by a diagonal matrix.
3. Prove Corollary 8.41.
4. Prove Corollary 8.42.
5. Translate Corollary 8.42 into a statement concerning symmetric bilinear forms on vector spaces over \mathcal{C} .
6. Use Corollary 8.42 to prove (4) above in the list of properties of quadratic forms over \mathcal{C} .
7. Let q represent an arbitrary quadratic form.
 - (a) Prove or disprove: $q(a\mathbf{u} + b\mathbf{v}) = aq(\mathbf{u}) + bq(\mathbf{v})$.
 - (b) Prove that $q(\mathbf{u} + \mathbf{v}) + q(\mathbf{u} - \mathbf{v}) = 2[q(\mathbf{u}) + q(\mathbf{v})]$.
8. Let q be a quadratic form on \mathbf{R}^n with symmetric matrix A relative to $\mathcal{E}_n : q(x_1, x_2, \dots, x_n) = \mathbf{X}^T A \mathbf{X}$. If $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the (necessarily real) eigenvalues of A , prove that $\lambda_1 \geq q(\mathbf{v}) \geq \lambda_n$ for all $\mathbf{v} = (x_1, x_2, \dots, x_n)$ with length 1.
9. A bilinear form f on \mathbf{V} is by definition **skew-symmetric** if and only if

$$f(\mathbf{u}, \mathbf{v}) = -f(\mathbf{v}, \mathbf{u})$$
 for all $\mathbf{u}, \mathbf{v} \in \mathbf{V}$. Prove that a bilinear form on \mathbf{V} is skew-symmetric if and only if any matrix that represents f is skew-symmetric.
10. Let \mathbf{V} be a vector space over a field \mathcal{F} in which $1 + 1 \neq 0$, and let f be a bilinear form on \mathbf{V} . Prove that f is skew-symmetric if and only if $f(\mathbf{v}, \mathbf{v}) = 0$ for all $\mathbf{v} \in \mathbf{V}$.
11. Prove that if $1 + 1 \neq 0$ in \mathcal{F} , every bilinear form f on \mathbf{V} can be written uniquely as the sum of a symmetric and a skew-symmetric bilinear form. (See Problem 8 of Exercises 8.7.)
12. Assume that $1 + 1 \neq 0$ in \mathcal{F} . The definition of a quadratic form on \mathbf{V} can be “extended” so as to include those mappings q of \mathbf{V} into \mathcal{F} such that $q(\mathbf{v}) = f(\mathbf{v}, \mathbf{v})$ for some (not necessarily symmetric) bilinear form on \mathbf{V} .
 - (a) Prove that f is skew-symmetric if and only if f determines the zero quadratic form on \mathbf{V} .
 - (b) Prove that the two bilinear forms f_1 and f_2 on \mathbf{V} define the same quadratic form on \mathbf{V} if and only if $f_1 - f_2$ is skew-symmetric.
 - (c) Does this “extension” actually enlarge the set of quadratic forms on \mathbf{V} ? Why, or why not?

8.9 Hermitian Forms

We restrict our attention now to those vector spaces with a scalar field \mathcal{F} that is either \mathbf{R} or the field \mathcal{C} of complex numbers. Throughout this section, each statement involving \mathcal{F} may be read in only two ways: with $\mathcal{F} = \mathbf{R}$, and with $\mathcal{F} = \mathcal{C}$.

The development of Chapter 9 is based on the interesting and important concept of an inner product on a vector space. We shall see there that, so long as our work is with real vectors spaces, symmetric bilinear forms are sufficient for our needs. But with complex vector spaces, we shall have need for another type of function, the *hermitian form*. A hermitian form has somewhat the same relation to a complex bilinear form as a symmetric bilinear form has to a bilinear form.

Definition 8.46 *Let \mathbf{U} and \mathbf{V} be vector spaces over the field \mathcal{F} . A **complex bilinear form** on \mathbf{U} and \mathbf{V} is a mapping f of pairs of vectors $(\mathbf{u}, \mathbf{v}) \in \mathbf{U} \times \mathbf{V}$ onto scalars $f(\mathbf{u}, \mathbf{v}) \in \mathcal{F}$ that has the properties*

$$(i) \quad f(a_1\mathbf{u}_1 + a_2\mathbf{u}_2, \mathbf{v}) = \bar{a}_1 f(\mathbf{u}_1, \mathbf{v}) + \bar{a}_2 f(\mathbf{u}_2, \mathbf{v})$$

and

$$(ii) \quad f(\mathbf{u}, b_1\mathbf{v}_1 + b_2\mathbf{v}_2) = b_1 f(\mathbf{u}, \mathbf{v}_1) + b_2 f(\mathbf{u}, \mathbf{v}_2),$$

where \bar{a}_i denotes the complex conjugate of a_i .

We describe the conditions (i) and (ii) by saying that f is complex linear in the first variable and linear in the second variable. It is readily seen that a function f from $\mathbf{U} \times \mathbf{V}$ into \mathcal{F} is a complex bilinear form if and only if

$$f \left(\sum_{i=1}^r a_i \mathbf{u}_i, \sum_{j=1}^s b_j \mathbf{v}_j \right) = \sum_{i=1}^r \sum_{j=1}^s \bar{a}_i b_j f(\mathbf{u}_i, \mathbf{v}_j)$$

for all positive integers r, s .

Whenever the field of scalars is real, a complex bilinear form reduces to a bilinear form. In this sense, a complex bilinear form is a generalization of a bilinear form. We must keep in mind, however, that the two concepts are quite distinct whenever \mathcal{F} is the field \mathcal{C} .

Throughout the remainder of this section, \mathbf{U} and \mathbf{V} will denote vector spaces over \mathcal{F} with dimensions m and n , respectively, and f will denote a complex bilinear form on \mathbf{U} and \mathbf{V} .

Definition 8.47 *Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be bases of \mathbf{U} and \mathbf{V} , respectively. The **matrix of a complex bilinear form** f relative to \mathcal{A} and \mathcal{B} is the matrix $A = [a_{ij}]_{m \times n}$, where $a_{ij} = f(\mathbf{u}_i, \mathbf{v}_j)$ for $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$.*

At this point, a parallel can be perceived between the properties of bilinear forms and those of complex bilinear forms. Actually, the parallel is so strong that it is quite repetitious to develop the properties of complex bilinear forms in as much detail as

was done with bilinear forms. At the same time, the adjustments that are necessitated by complex linearity in the first variable are not altogether obvious. Consequently, we shall more or less outline the development here with statements of the major results as theorems, and leave the proofs of most of these theorems as exercises. In all cases, a proof can be obtained by a suitable modification of the proof of the corresponding result for bilinear forms.

We recall that the conjugate transpose $(\bar{A})^T$ of a matrix A is denoted by A^* .

Theorem 8.48 *Let $X = [x_1, x_2, \dots, x_m]^T$ denote the coordinate matrix of $\mathbf{u} \in \mathbf{U}$ relative to the basis \mathcal{A} of \mathbf{U} , and let $Y = [y_1, y_2, \dots, y_n]^T$ be the coordinate matrix of $\mathbf{v} \in \mathbf{V}$ relative to the basis \mathcal{B} of \mathbf{V} . If $A = [a_{ij}]_{m \times n}$, then A is the matrix of the complex bilinear form f on \mathbf{U} and \mathbf{V} if and only if the equation*

$$f(\mathbf{u}, \mathbf{v}) = X^* A Y$$

is satisfied for all choices of $\mathbf{u} \in \mathbf{U}, \mathbf{v} \in \mathbf{V}$.

Proof. See Problem 7. ■

Example 1 □ Consider the mapping f defined from $\mathcal{C}^2 \times \mathcal{C}^3$ into \mathcal{C} by

$$f(\mathbf{u}, \mathbf{v}) = 3i\bar{u}_1 v_2 + (-1 - 3i)\bar{u}_1 v_3 + i\bar{u}_2 v_1 + (4 - i)\bar{u}_2 v_2 + (-4 + 3i)\bar{u}_2 v_3,$$

where $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2, v_3)$. Guided by the equation $a_{ij} = f(\mathbf{e}_i, \mathbf{e}_j)$ in Definition 8.47, we see that the value of $f(\mathbf{u}, \mathbf{v})$ can be written in the form

$$\begin{aligned} f(\mathbf{u}, \mathbf{v}) &= f((u_1, u_2), (v_1, v_2, v_3)) \\ &= [\bar{u}_1, \bar{u}_2] \begin{bmatrix} 0 & 3i & -1 - 3i \\ i & 4 - i & -4 + 3i \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}. \end{aligned}$$

Hence f is the complex bilinear form on \mathcal{C}^2 and \mathcal{C}^3 that has the matrix

$$A = \begin{bmatrix} 0 & 3i & -1 - 3i \\ i & 4 - i & -4 + 3i \end{bmatrix}$$

relative to the standard bases \mathcal{E}_2 of \mathcal{C}^2 and \mathcal{E}_3 of \mathcal{C}^3 . ■

It follows from Definition 8.47 and Theorem 8.48 that two complex bilinear forms f and g on \mathbf{U} and \mathbf{V} are equal if and only if they have the same matrix relative to bases \mathcal{A} of \mathbf{U} and \mathcal{B} of \mathbf{V} .

Theorem 8.49 *Let A be the matrix of f relative to the bases \mathcal{A} of \mathbf{U} and \mathcal{B} of \mathbf{V} . If Q is the matrix of transition from \mathcal{A} to the basis \mathcal{A}' of \mathbf{U} and P is the matrix of transition from \mathcal{B} to the basis \mathcal{B}' of \mathbf{V} , then $Q^* AP$ is the matrix of f relative to \mathcal{A}' and \mathcal{B}' .*

Proof. See Problem 8. ■

Example 2 □ Let f be the same complex bilinear form as in Example 1, and consider the problem of finding the matrix of f relative to the bases $\mathcal{A}' = \{(1, i), (i, 0)\}$ of \mathcal{C}^2 and $\mathcal{B}' = \{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$ of \mathcal{C}^3 .

The matrix of transition from \mathcal{A} to \mathcal{A}' is

$$Q = \begin{bmatrix} 1 & i \\ i & 0 \end{bmatrix},$$

and the matrix of transition from \mathcal{B} to \mathcal{B}' is

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

By Theorem 8.49, the matrix of f relative to \mathcal{A}' and \mathcal{B}' is

$$\begin{aligned} Q^* AP &= \begin{bmatrix} 1 & -i \\ -i & 0 \end{bmatrix} \begin{bmatrix} 0 & 3i & -1 - 3i \\ i & 4 - i & -4 + 3i \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -i & 1 \\ i & 3 & 0 \end{bmatrix}. \end{aligned} \quad \blacksquare$$

Our main interest in complex bilinear forms is with the case where $\mathbf{V} = \mathbf{U}$. As with bilinear forms and linear transformations, we assume that the bases \mathcal{A} and \mathcal{B} of $\mathbf{V} = \mathbf{U}$ are the same unless it is stated otherwise. For the remainder of the section, we shall have $\mathbf{V} = \mathbf{U}$ and \mathbf{V} an n -dimensional vector space over \mathcal{F} .

The equivalence relation that we need to replace congruence of matrices over \mathcal{F} is given in our next definition.

Definition 8.50 Let A and B be matrices over the field \mathcal{F} . Then B is **conjunctive** (or **hermitian congruent**) to A over \mathcal{F} if there is an invertible matrix P over \mathcal{F} such that $B = P^* AP$.

Theorem 8.49 and Definition 8.50 lead to the following result.

Theorem 8.51 Let A and B be matrices of order n over \mathcal{F} . Then B is conjunctive to A over \mathcal{F} if and only if B and A represent the same complex bilinear form on \mathbf{V} .

The term “hermitian” applies to complex bilinear forms in about the same way as the term “symmetric” applies to bilinear forms. As a matter of fact, when \mathcal{F} is the field of real numbers, the two terms are coincident, just as “bilinear” and “complex bilinear” are. In this sense, a hermitian complex bilinear form is a generalization of a symmetric bilinear form.

Definition 8.52 A complex bilinear form f on \mathbf{V} is **hermitian** if and only if

$$f(\mathbf{u}, \mathbf{v}) = \overline{f(\mathbf{v}, \mathbf{u})}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbf{V}$.

Definition 8.53 A matrix H over the field \mathcal{F} is **hermitian** if and only if $H^* = H$.

Thus a real matrix A is hermitian if and only if it is symmetric. The relation between hermitian complex bilinear forms and hermitian matrices is exactly what one would expect.

Theorem 8.54 A complex bilinear form f on \mathbf{V} is hermitian if and only if every matrix that represents f is hermitian.

Proof. See Problem 10. ■

Our next theorem corresponds to the result in Theorem 8.40 for symmetric matrices. We include a proof for this theorem, since it is quite important and its proof is a bit more difficult than the others in this section.

Theorem 8.55 Every hermitian matrix H of order n over \mathcal{C} is conjunctive over \mathcal{C} to a diagonal matrix.

Proof. The proof is by induction on the order n of H . The theorem is trivially valid for $n = 1$.

Assume that the theorem is true for all hermitian matrices of order k over \mathcal{C} , and let H be a hermitian matrix of order $k + 1$ over \mathcal{C} . Let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k+1}\}$ be a basis of \mathcal{C}^{k+1} , and let h be the hermitian complex bilinear form on \mathcal{C}^{k+1} that has matrix H relative to \mathcal{A} . If $H = \mathbf{0}$, H is a diagonal matrix already. Assume now that $H \neq \mathbf{0}$ and $h(\mathbf{u}_r, \mathbf{u}_s) = a + bi \neq 0$ for the pair r, s . If $h(\mathbf{u}_r, \mathbf{u}_r) = 0$ and $h(\mathbf{u}_s, \mathbf{u}_s) = 0$, then

$$\begin{aligned} h(\mathbf{u}_r + i\mathbf{u}_s, \mathbf{u}_r + i\mathbf{u}_s) &= h(\mathbf{u}_r, \mathbf{u}_r) + ih(\mathbf{u}_r, \mathbf{u}_s) - ih(\mathbf{u}_s, \mathbf{u}_r) + h(\mathbf{u}_s, \mathbf{u}_s) \\ &= i[h(\mathbf{u}_r, \mathbf{u}_s) - h(\mathbf{u}_s, \mathbf{u}_r)] \\ &= i[h(\mathbf{u}_r, \mathbf{u}_s) - \overline{h(\mathbf{u}_r, \mathbf{u}_s)}] \\ &= -2b. \end{aligned}$$

A similar computation shows that

$$\begin{aligned} h(\mathbf{u}_r + \mathbf{u}_s, \mathbf{u}_r + \mathbf{u}_s) &= h(\mathbf{u}_r, \mathbf{u}_s) + \overline{h(\mathbf{u}_r, \mathbf{u}_s)} \\ &= 2a. \end{aligned}$$

Thus, there is a vector $\mathbf{v}_1 \in \mathcal{C}^{k+1}$ such that $d_1 = h(\mathbf{v}_1, \mathbf{v}_1) \neq 0$. The set $\{\mathbf{v}_1\}$ can be extended to a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}\}$ of \mathcal{C}^{k+1} . The set $\mathcal{A}' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_{k+1}\}$ is obtained from \mathcal{B} as follows:

$$\mathbf{u}'_1 = \mathbf{v}_1$$

and

$$\mathbf{u}'_j = \mathbf{v}_j - \frac{h(\mathbf{u}'_1, \mathbf{v}_j)}{d_1} \mathbf{u}'_1,$$

for $j = 2, \dots, k+1$. Since \mathcal{A}' is obtained from \mathcal{B} by elementary operations, \mathcal{A}' is a basis of \mathcal{C}^{k+1} . The matrix H_1 of h relative to \mathcal{A}' is hermitian, since h is hermitian. Now $h(\mathbf{u}'_1, \mathbf{u}'_1) = d_1$, and a simple computation shows that $h(\mathbf{u}'_1, \mathbf{u}'_j) = 0$ for $j = 2, \dots, k+1$. Hence H_1 is of the form

$$H_1 = \begin{bmatrix} d_1 & | & \mathbf{0} \\ \hline - & + & - \\ \mathbf{0} & | & H_2 \end{bmatrix},$$

where H_2 is a $k \times k$ hermitian matrix. The matrix of transition P_1 from \mathcal{A} to \mathcal{A}' is invertible over \mathcal{C} , and $H_1 = P_1^* H P_1$ by Theorem 8.49.

By the induction hypothesis, there is an invertible matrix Q of order k over \mathcal{C} such that $Q^* H_2 Q = \text{diag}\{d_2, \dots, d_{k+1}\}$. Now

$$P_2 = \begin{bmatrix} 1 & | & \mathbf{0} \\ \hline - & + & - \\ \mathbf{0} & | & Q \end{bmatrix}$$

is an invertible matrix over \mathcal{C} , and

$$\begin{aligned} (P_1 P_2)^* H (P_1 P_2) &= P_2^* P_1^* H P_1 P_2 \\ &= P_2^* H_1 P_2 \\ &= \begin{bmatrix} d_1 & | & \mathbf{0} \\ \hline - & + & - \\ \mathbf{0} & | & Q^* H_2 Q \end{bmatrix} \\ &= \text{diag}\{d_1, d_2, \dots, d_{k+1}\}. \end{aligned}$$

Thus, $P = P_1 P_2$ is an invertible matrix over \mathcal{C} such that $P^* H P$ is diagonal. The theorem follows from the principle of mathematical induction. ■

Definition 8.56 A hermitian complex bilinear form on \mathbf{V} is called a **hermitian form** on \mathbf{V} .

It follows from Theorems 8.55 and 8.51 that any hermitian form h on \mathbf{V} can be represented by a diagonal matrix. Now the condition $H^* = H$ requires that the diagonal elements of a hermitian matrix must be real. Hence the diagonal elements in a diagonalized representation of h are always real.

The proof of Theorem 8.23 can be modified so as to obtain a proof of the following theorem. With $q(\mathbf{v})$ replaced by $h(\mathbf{v}, \mathbf{v})$ and \mathbf{R}^n replaced by \mathcal{C}^n , the only other

changes necessary are in the expressions for $h(\mathbf{v}, \mathbf{v})$. For example, the two diagonalized representations would appear as $h(\mathbf{v}, \mathbf{v}) = \sum_{k=1}^r d_k |y_k|^2$ and $h(\mathbf{v}, \mathbf{v}) = \sum_{k=1}^r c_k |z_k|^2$.

Theorem 8.57 *In any two diagonalized representations of a hermitian form h , the number of positive terms is the same and the number of negative terms is the same.*

The definitions of *index*, *signature*, *positive definite*, etc., apply to hermitian forms as they are given in Definitions 8.25 and 8.29. The positive definite hermitian forms are those that are fundamental to Chapter 9. The index of a hermitian matrix is by definition the same as that of a hermitian form that it represents.

Corollary 8.58 *Any hermitian matrix H is conjunctive over \mathcal{C} to a unique matrix of the form*

$$C = \begin{bmatrix} I_p & | & \mathbf{0} & | & \mathbf{0} \\ - & + & - & + & - \\ \mathbf{0} & | & -I_{r-p} & | & \mathbf{0} \\ - & + & - & + & - \\ \mathbf{0} & | & \mathbf{0} & | & \mathbf{0} \end{bmatrix},$$

where r is the rank of H . Two $n \times n$ hermitian matrices are conjunctive over \mathcal{C} if and only if they have the same rank and index.

Proof. See Problem 12. ■

Corollary 8.59 *With a suitable choice of basis in \mathbf{V} , any hermitian form on \mathbf{V} can be represented by a matrix of the form C in Corollary 8.58.*

Proof. See Problem 13. ■

Corollary 8.60 *A hermitian form h is positive definite if and only if*

$$h(\mathbf{v}, \mathbf{v}) > 0$$

for all $\mathbf{v} \neq \mathbf{0}$.

Proof. See Problem 14. ■

Just after Corollary 8.42 in Section 8.8, we described a method for obtaining an invertible matrix P that would reduce a given symmetric matrix A to a diagonal matrix $P^T A P$, and we illustrated the method in Example 1 of that section. In order to obtain a method for finding an invertible matrix P that will reduce a hermitian matrix H to a diagonal matrix $P^* H P$, the only changes that are necessary in that description are that each $P_i^T A$ be replaced by $P_i^* H$. The method for obtaining P such that $P^* H P$ is of the form C in Corollary 8.58 is entirely analogous to the techniques of Section 8.8. Our next example gives a demonstration of this method.

Example 3 □ For the hermitian matrix

$$H = \begin{bmatrix} 5 & 4i & -4 \\ -4i & 3 & 5i \\ -4 & -5i & 3 \end{bmatrix},$$

we shall find an invertible matrix P such that

$$P^*HP = \begin{bmatrix} I_p & | & \mathbf{0} & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{---} & + & \mathbf{---} \\ \mathbf{0} & | & -I_{r-p} & | & \mathbf{0} \\ \hline \mathbf{0} & + & \mathbf{---} & + & \mathbf{---} \\ \mathbf{0} & | & \mathbf{0} & | & \mathbf{0} \end{bmatrix}.$$

We first find an invertible matrix P_1 such that $P_1^*HP_1$ is diagonal. Since $h_{11} = 5$ is not zero in $H = [h_{ij}]$, we can use

$$C_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

as the first column of P_1 . The second column C_2 of P_1 needs to make $\{C_1, C_2\}$ linearly independent and satisfy $(C_1^*H)C_2 = 0$, which appears as

$$\begin{bmatrix} 5 & 4i & -4 \end{bmatrix} C_2 = 0.$$

The choice

$$C_2 = \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix}$$

satisfies both conditions. The third column C_3 of P_1 needs to make $\{C_1, C_2, C_3\}$ linearly independent and satisfy $(C_1^*H)C_3 = 0$ and $(C_2^*H)C_3 = 0$. That is, we need a column C_3 that is not a linear combination of C_1 and C_2 and that satisfies both equations

$$\begin{bmatrix} 5 & 4i & -4 \end{bmatrix} C_3 = 0,$$

$$\begin{bmatrix} 0 & -9i & -1 \end{bmatrix} C_3 = 0.$$

The choice

$$C_3 = \begin{bmatrix} -8i \\ 1 \\ -9i \end{bmatrix}$$

satisfies all the conditions. Thus

$$P_1 = \begin{bmatrix} 1 & 4 & -8i \\ 0 & 0 & 1 \\ 0 & 5 & -9i \end{bmatrix}$$

is a matrix such that $P_1^*HP_1$ is diagonal. Performing the multiplication, we find that

$$P_1^*HP_1 = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 16 \end{bmatrix}.$$

Interchanging the second and third columns in P_1 yields

$$P_2 = \begin{bmatrix} 1 & -8i & 4 \\ 0 & 1 & 0 \\ 0 & -9i & 5 \end{bmatrix}$$

such that $P_2^*HP_2 = \text{diag}\{5, 16, -5\}$. If we now put

$$P_3 = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{\sqrt{5}} \end{bmatrix}$$

and $P = P_2P_3$, we get

$$\begin{aligned} P^*HP &= (P_2P_3)^*H(P_2P_3) \\ &= P_3^*(P_2^*HP_2)P_3 \\ &= \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & -5 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{\sqrt{5}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \end{aligned}$$

and this matrix has the form described in Corollary 8.58. ■

Exercises 8.9

1. Which of the following matrices are hermitian?

$$\begin{array}{llll}
 \text{(a)} \begin{bmatrix} 5 & i \\ -i & 2 \end{bmatrix} & \text{(b)} \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix} & \text{(c)} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} & \text{(d)} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \\
 \text{(e)} \begin{bmatrix} 5 & i & 2 \\ -i & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix} & \text{(f)} \begin{bmatrix} 0 & 2 & 1 \\ -2 & 1 & -1 \\ -1 & 1 & -2 \end{bmatrix} & \text{(g)} \begin{bmatrix} \frac{11}{9} & -\frac{2}{3}i & -\frac{1}{9} \\ \frac{2}{3}i & 2 & -\frac{1}{3}i \\ -\frac{1}{9} & \frac{1}{3}i & \frac{5}{9} \end{bmatrix} \\
 \text{(h)} \begin{bmatrix} 3i & 0 & 4i \\ 0 & 2i & -2 \\ -4i & 2 & i \end{bmatrix} & & &
 \end{array}$$

2. For each hermitian matrix H in Problem 1, answer the following questions.
- (a) Find an invertible matrix P such that P^*HP is diagonal.
 - (b) State the rank and index of H .
 - (c) Which of these matrices are conjuctive over \mathcal{C} ?
3. For each hermitian matrix H in Problem 1, find an invertible matrix P such that P^*HP is of the form C given in Corollary 8.58.
4. In each of parts (a)-(h) of Problem 1, let f be the complex bilinear form on \mathcal{C}^n that has the given matrix relative to \mathcal{E}_n .
- (a) Which of these functions f are hermitian forms?
 - (b) For each f that is a hermitian form, find a basis \mathcal{A}' of \mathcal{C}^n with respect to which f is represented by a diagonal matrix.
5. Prove the following corollary to Theorem 8.49. Two $m \times n$ matrices over \mathcal{F} represent the same complex bilinear form on \mathbf{U} and \mathbf{V} if and only if A and B are equivalent over \mathcal{F} .
6. Let f be a complex bilinear form on \mathbf{U} and \mathbf{V} over \mathcal{C} . Prove that, with suitable choices of bases in \mathbf{U} and \mathbf{V} , f can be represented by a matrix D_r that has the first r diagonal elements equal to 1 and all other elements zero.
7. Prove Theorem 8.48.
8. Prove Theorem 8.49.
9. Prove Theorem 8.51.

10. Prove Theorem 8.54.
11. Write out the details of the proof of Theorem 8.57.
12. Prove Corollary 8.58.
13. Prove Corollary 8.59.
14. Prove Corollary 8.60.
15. Prove that a hermitian matrix H is positive definite if and only if there exists an invertible matrix P such that $H = P^*P$.
16. Given that the matrix H in Problem 1(e) is a positive definite hermitian matrix, find an invertible matrix P such that $P^*P = H$.
17. Let A and B be hermitian matrices of the same dimensions.
 - (a) Prove that $A + B$ and $A - B$ are hermitian.
 - (b) Prove that if $AB = BA$, then AB is hermitian.
 - (c) Prove that A^T is hermitian.
 - (d) Prove that if A is invertible, then A^{-1} is hermitian.
18. A matrix A is called **skew-hermitian** if $A^* = -A$. Prove that any square matrix A can be written uniquely as $A = B + C$ with B hermitian and C skew-hermitian.

Chapter 9

Inner Product Spaces

9.1 Introduction

The vector spaces \mathbf{R}^n have been an intuitive guide in our development thus far, and we have extended most of the concepts introduced there to more general settings. The outstanding exceptions are the basic concepts of length and inner product in \mathbf{R}^n . In this chapter, we generalize these concepts and study those properties of vector spaces that are based on an inner product. It is possible to define an inner product for vector spaces over fields other than the field \mathbf{R} of real numbers or the field \mathcal{C} of complex numbers, but our interest is restricted to these cases. Accordingly, throughout this chapter, we shall always have either $\mathcal{F} = \mathbf{R}$ or $\mathcal{F} = \mathcal{C}$, and \mathbf{V} shall denote a finite-dimensional vector space over \mathcal{F} .

9.2 Inner Products

We begin with the definition of an inner product. As mentioned above, \mathcal{F} denotes a field that is either \mathbf{R} or \mathcal{C} , and \mathbf{V} is a finite-dimensional vector space over \mathcal{F} . The conjugate of the complex number z is denoted by \bar{z} .

Definition 9.1 *An inner product on \mathbf{V} is a mapping f of pairs of vectors (\mathbf{u}, \mathbf{v}) in $\mathbf{V} \times \mathbf{V}$ onto scalars $f(\mathbf{u}, \mathbf{v}) \in \mathcal{F}$ that has the following properties:*

- (i) $f(a_1\mathbf{u}_1 + a_2\mathbf{u}_2, \mathbf{v}) = \bar{a}_1f(\mathbf{u}_1, \mathbf{v}) + \bar{a}_2f(\mathbf{u}_2, \mathbf{v})$
- (ii) $f(\mathbf{u}, \mathbf{v}) = \overline{f(\mathbf{v}, \mathbf{u})}$
- (iii) $f(\mathbf{v}, \mathbf{v}) > 0$ if $\mathbf{v} \neq \mathbf{0}$ and $f(\mathbf{0}, \mathbf{0}) = 0$.

We note that property (ii) requires that $f(\mathbf{v}, \mathbf{v})$ be real and property (iii) requires that this real number be positive except when $\mathbf{v} = \mathbf{0}$.

If $\mathcal{F} = \mathbf{R}$ in Definition 9.1, \mathbf{V} is called a **real inner product space**, or a **Euclidean space**. If $\mathcal{F} = \mathcal{C}$, \mathbf{V} is called a **complex inner product space**, or a **unitary space**.

The term **inner product space** is used to refer collectively to Euclidean spaces and unitary spaces.

The properties listed in Definition 9.1 invite a comparison between inner products and hermitian forms. This comparison yields the following theorem, which is important even though the proof is trivial.

Theorem 9.2 *A mapping f of $\mathbf{V} \times \mathbf{V}$ into \mathcal{F} is an inner product on \mathbf{V} if and only if f is a positive definite hermitian form.*

Proof. Suppose first that f is a positive definite hermitian form. Then f is a complex bilinear form, so it follows from Definition 8.46 that f has property (i) of Definition 9.1. By Definition 8.52, the mapping f has property (ii) of Definition 9.1. Finally, f has property (iii) by Corollary 8.60. Hence f is an inner product on \mathbf{V} .

Assume, on the other hand, that f is an inner product on \mathbf{V} . The references cited in the preceding paragraph show that f satisfies all of the requirements of a positive definite hermitian form except possibly the condition that

$$f(\mathbf{u}, b_1\mathbf{v}_1 + b_2\mathbf{v}_2) = b_1 f(\mathbf{u}, \mathbf{v}_1) + b_2 f(\mathbf{u}, \mathbf{v}_2)$$

in Definition 8.46. But

$$\begin{aligned} f(\mathbf{u}, b_1\mathbf{v}_1 + b_2\mathbf{v}_2) &= \overline{f(b_1\mathbf{v}_1 + b_2\mathbf{v}_2, \mathbf{u})} \\ &= \overline{\bar{b}_1 f(\mathbf{v}_1, \mathbf{u}) + \bar{b}_2 f(\mathbf{v}_2, \mathbf{u})} \\ &= b_1 \overline{f(\mathbf{v}_1, \mathbf{u})} + b_2 \overline{f(\mathbf{v}_2, \mathbf{u})} \\ &= b_1 f(\mathbf{u}, \mathbf{v}_1) + b_2 f(\mathbf{u}, \mathbf{v}_2), \end{aligned}$$

so that f is indeed a positive definite hermitian form. ■

The next example presents the most commonly used inner products in \mathbf{R}^n and \mathcal{C}^n .

Example 1 □ For any two vectors $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ in \mathbf{R}^n , let the value $f(\mathbf{u}, \mathbf{v})$ be given by

$$f(\mathbf{u}, \mathbf{v}) = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{k=1}^n u_k v_k.$$

Then $f(\mathbf{u}, \mathbf{v})$ is the familiar scalar product $\mathbf{u} \cdot \mathbf{v}$ of Chapter 1. The properties required in Definition 9.1 follow easily from Theorem 1.20.

The corresponding inner product on \mathcal{C}^n is given by

$$g(\mathbf{u}, \mathbf{v}) = \bar{u}_1 v_1 + \bar{u}_2 v_2 + \cdots + \bar{u}_n v_n = \sum_{k=1}^n \bar{u}_k v_k,$$

where $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ in \mathcal{C}^n . The verification that this mapping is an inner product is left as an exercise (Problem 3). ■

The inner products in Example 1 will be referred to hereafter as the **standard inner products** on \mathbf{R}^n and \mathcal{C}^n , respectively. Whenever an inner product on \mathbf{R}^n or \mathcal{C}^n is not specified, it is understood to be the standard inner product.

Example 2 □ Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in \mathbf{R}^2 . We shall determine if the following rule determines an inner product on \mathbf{R}^2 :

$$f(\mathbf{u}, \mathbf{v}) = 2u_1v_1 + u_1v_2 + u_2v_1 + u_2v_2.$$

We see that the value of f can be written as

$$f((u_1, u_2), (v_1, v_2)) = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

and since we are dealing with a real vector space, f is a hermitian form. Thus f is an inner product if and only if f is positive definite; that is, if and only if the third property in Definition 9.1 is satisfied. Since

$$\begin{aligned} f(\mathbf{v}, \mathbf{v}) &= 2v_1^2 + 2v_1v_2 + v_2^2 \\ &= v_1^2 + (v_1 + v_2)^2, \end{aligned}$$

we see that $f(\mathbf{v}, \mathbf{v}) \geq 0$ and $f(\mathbf{v}, \mathbf{v}) = 0$ if and only if $v_1 = 0$ and $v_1 + v_2 = 0$, that is, if and only if $\mathbf{v} = \mathbf{0}$. Thus f is an inner product on \mathbf{R}^2 , one that looks quite different from the standard inner product. ■

The connection established in Theorem 9.2 makes available the results of Section 8.9 for use with inner products. In particular, an inner product f on \mathbf{V} has a unique matrix $A = [a_{ij}]_{n \times n}$ relative to each basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{V} . This matrix A is determined by the conditions $a_{ij} = f(\mathbf{v}_i, \mathbf{v}_j)$, and $f(\mathbf{u}, \mathbf{v}) = \mathbf{X}^* A \mathbf{Y}$ where \mathbf{u} and \mathbf{v} have coordinate matrices X and Y , respectively, relative to \mathcal{A} . If P is the matrix of transition form \mathcal{A} to \mathcal{A}' , then f has matrix $P^* A P$ relative to \mathcal{A}' , by Theorem 8.49. Since f is positive definite, Corollary 8.59 implies that there is a basis \mathcal{A}' of \mathbf{V} such that the inner product f has matrix I_n relative to \mathcal{A}' . Relative to this basis \mathcal{A}' , $f(\mathbf{u}, \mathbf{v})$ is given by $f(\mathbf{u}, \mathbf{v}) = \mathbf{X}^* Y = \sum_{k=1}^n \bar{x}_k y_k$. Thus the standard inner products in Example 1 furnish typical examples provided the choice of basis is appropriate.

The results obtained in this chapter are valid for all finite-dimensional inner-product spaces, in that they are not dependent on a particular choice of f . For this reason, it is customary to replace the notation $f(\mathbf{u}, \mathbf{v})$ by a more convenient one. We choose to drop the f from the notation, and simply write (\mathbf{u}, \mathbf{v}) instead of $f(\mathbf{u}, \mathbf{v})$. This change of notation has an additional advantage in that it reminds us that we are dealing with an inner product, and not just a complex bilinear form.

According to the result in Problem 15 of Exercises 8.9, a hermitian matrix H is positive definite if and only if $H = P^* P$ for some invertible matrix P . This result

provides an easy way to construct inner products on \mathcal{C}^n . For $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ in \mathcal{C}^n , let

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

We can choose an invertible P of order n , put $H = P^*P$, and then the rule

$$(\mathbf{u}, \mathbf{v}) = U^*HV$$

defines an inner product on \mathcal{C}^n . We say that the invertible matrix P **generates** this inner product.

Example 3 \square As a specific example of the preceding discussion, the matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ -i & 1 & i \\ 0 & i & 1 \end{bmatrix}$$

generates the inner product on \mathcal{C}^3 defined as follows. For $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ in \mathcal{C}^3 ,

$$\begin{aligned} (\mathbf{u}, \mathbf{v}) &= \begin{bmatrix} \bar{u}_1 & \bar{u}_2 & \bar{u}_3 \end{bmatrix} P^*P \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \\ &= \begin{bmatrix} \bar{u}_1 & \bar{u}_2 & \bar{u}_3 \end{bmatrix} \begin{bmatrix} 2 & i & -1 \\ -i & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \\ &= 2\bar{u}_1v_1 + i\bar{u}_1v_2 - \bar{u}_1v_3 - i\bar{u}_2v_1 + 2\bar{u}_2v_2 - \bar{u}_3v_1 + 2\bar{u}_3v_3. \blacksquare \end{aligned}$$

Exercises 9.2

1. Using the standard inner product in \mathcal{C}^3 , compute (\mathbf{u}, \mathbf{v}) and (\mathbf{v}, \mathbf{u}) for the given vectors.

(a) $\mathbf{u} = (2 - i, 1 - 5i, 1 + 2i)$, $\mathbf{v} = (2 + i, -3, -5 + i)$

(b) $\mathbf{u} = (11, -6i, -1)$, $\mathbf{v} = (2i, 6, -i)$

2. Repeat Problem 1 using the inner product in Example 3 of this section.
3. Let $f : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}$ and $g : \mathcal{C}^n \times \mathcal{C}^n \rightarrow \mathcal{C}$ be the standard inner products as defined in Example 1.
 - (a) Prove that f is an inner product on \mathbf{R}^n .
 - (b) Prove that g is an inner product on \mathcal{C}^n .
4. Let $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ in \mathbf{R}^n . Prove or disprove that the given rule defines an inner product on \mathbf{R}^n .
 - (a) $(\mathbf{u}, \mathbf{v}) = 2u_1v_1 + u_1v_2 + u_2v_1 + 2u_2v_2$ on \mathbf{R}^2
 - (b) $(\mathbf{u}, \mathbf{v}) = u_1v_2 + u_2v_1$ on \mathbf{R}^2
 - (c) $(\mathbf{u}, \mathbf{v}) = u_1v_1 + u_1v_2 + u_2v_1 + u_2v_2$ on \mathbf{R}^2
 - (d) $(\mathbf{u}, \mathbf{v}) = 2u_1v_1 + u_2v_2$ on \mathbf{R}^2
 - (e) $(\mathbf{u}, \mathbf{v}) = u_1v_1 + u_3v_3$ on \mathbf{R}^3
 - (f) $(\mathbf{u}, \mathbf{v}) = u_1^2v_1^2 + u_2^2v_2^2 + u_3^2v_3^2$ on \mathbf{R}^3
5. Assume that the matrix

$$H = \begin{bmatrix} 5 & i & 2 \\ -i & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix}$$
 is a positive definite hermitian matrix, and let (\mathbf{u}, \mathbf{v}) be the inner product on \mathcal{C}^3 that has matrix H relative to the basis

$$\mathcal{A} = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}.$$
 Write out the value of (\mathbf{u}, \mathbf{v}) for arbitrary vectors $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ in \mathcal{C}^3 .
6. Find a basis \mathcal{A}' of \mathcal{C}^3 for which the inner product in Problem 5 has the form $(\mathbf{u}, \mathbf{v}) = X^*Y$, where \mathbf{u} and \mathbf{v} have coordinate matrices X and Y , respectively, relative to \mathcal{A}' .
7. Recall that the *trace* of a matrix $A = [a_{ij}]_{n \times n}$ is the number $t(A) = \sum_{i=1}^n a_{ii}$. Prove that $(A, B) = t(A^T B)$ is an inner product on the vector space $\mathbf{R}_{n \times n}$ of all $n \times n$ matrices over \mathbf{R} .
8. Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a basis of the inner product space \mathbf{V} , and let \mathbf{u} and \mathbf{v} be arbitrary vectors with coordinate matrices X and Y , respectively, relative to \mathcal{A} . Prove that $(\mathbf{u}, \mathbf{v}) = X^*Y$ for all $\mathbf{u}, \mathbf{v} \in \mathbf{V}$ if and only if $(\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}$ for all pairs i, j .

9. Assume that the function f defined on $\mathcal{C}^2 \times \mathcal{C}^2$ by

$$f((a_1, a_2), (b_1, b_2)) = 5\bar{a}_1b_1 + i\bar{a}_1b_2 - i\bar{a}_2b_1 + 2\bar{a}_2b_2$$

is an inner product on \mathcal{C}^2 .

- (a) Find the matrix of f relative to the basis $\mathcal{E}_2 = \{(1, 0), (0, 1)\}$.
 - (b) Find a basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2\}$ of \mathcal{C}^2 such that $(\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}$ for all i, j .
10. Show that $(\mathbf{v}, \mathbf{0}) = 0$ for all $\mathbf{v} \in \mathbf{V}$.
11. Prove that if $(\mathbf{u}, \mathbf{v}) = 0$ for all $\mathbf{v} \in \mathbf{V}$, then $\mathbf{u} = \mathbf{0}$.
12. Prove that if \mathbf{u} and \mathbf{v} are two vectors in \mathbf{V} such that $(\mathbf{u}, \mathbf{w}) = (\mathbf{v}, \mathbf{w})$ for all \mathbf{w} in \mathbf{V} , then $\mathbf{u} = \mathbf{v}$.
13. In each of parts (a)-(h) of Problem 1 of Exercises 8.9, let f be the complex bilinear form on \mathcal{C}^n that has the given matrix relative to \mathcal{E}_n . Which of these functions are inner products on \mathcal{C}^n ?
14. The definition of an inner product in Definition 9.1 applies equally well to infinite-dimensional vector spaces \mathbf{V} . Assuming the usual properties of the definite integral, prove that $(f, g) = \int_0^1 f(x)g(x)dx$ defines an inner product on the vector space of all continuous real-valued functions on $[0, 1]$.

9.3 Norms and Distances

The concept of an inner product leads to the notion of the *length* or *norm* of a vector.

Definition 9.3 For a given inner product (\mathbf{u}, \mathbf{v}) on \mathbf{V} , the length or norm of a vector \mathbf{v} in \mathbf{V} is the real number $\|\mathbf{v}\|$ given by $\|\mathbf{v}\| = \sqrt{(\mathbf{v}, \mathbf{v})}$.

Example 1 □ We shall compute $\|\mathbf{v}\|$ for $\mathbf{v} = (2i, 6, -i)$ in \mathcal{C}^3 by using the standard inner product and also by using the inner product in Example 3 of Section 9.2.

Using the standard inner product, we have

$$(\mathbf{v}, \mathbf{v}) = (-2i)(2i) + (6)(6) + (i)(-i) = 41$$

and

$$\|\mathbf{v}\| = \sqrt{41}.$$

Using the inner product from Example 3 of Section 9.2, we have

$$(\mathbf{v}, \mathbf{v}) = \begin{bmatrix} -2i & 6 & i \end{bmatrix} \begin{bmatrix} 2 & i & -1 \\ -i & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2i \\ 6 \\ -i \end{bmatrix} = 110$$

and

$$\|\mathbf{v}\| = \sqrt{110}. \blacksquare$$

The norm of a vector as given in Definition 9.3 is readily seen to be a generalization of Definition 1.19. Our last example illustrates that the number obtained as the length of a vector depends on the particular inner product that is being used. However, we shall see in Theorem 9.5 that this length has the properties that one would naturally expect. In the derivation of these properties, the famous Cauchy-Schwarz inequality is of invaluable aid.

Theorem 9.4 (The Cauchy-Schwarz inequality) *For any vectors \mathbf{u}, \mathbf{v} in \mathbf{V} ,*

$$|(\mathbf{u}, \mathbf{v})| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|.$$

Proof. If $\mathbf{v} = \mathbf{0}$, the equality holds with both members zero. Consider the case where $\mathbf{v} \neq \mathbf{0}$. Since the inner product is a positive definite hermitian form, we have

$$\begin{aligned} 0 &\leq (\mathbf{u} - z\mathbf{v}, \mathbf{u} - z\mathbf{v}) \\ &= (\mathbf{u}, \mathbf{u}) - z(\mathbf{u}, \mathbf{v}) - \bar{z}(\mathbf{v}, \mathbf{u}) + z\bar{z}(\mathbf{v}, \mathbf{v}) \\ &= (\mathbf{u}, \mathbf{u}) - z(\mathbf{u}, \mathbf{v}) - \overline{z(\mathbf{u}, \mathbf{v})} + z\bar{z}(\mathbf{v}, \mathbf{v}) \end{aligned}$$

for any complex number z . Since $\mathbf{v} \neq \mathbf{0}$, we may let $z = \frac{\overline{(\mathbf{u}, \mathbf{v})}}{(\mathbf{v}, \mathbf{v})}$. This yields

$$\begin{aligned} 0 &\leq (\mathbf{u}, \mathbf{u}) - \frac{\overline{(\mathbf{u}, \mathbf{v})(\mathbf{u}, \mathbf{v})}}{(\mathbf{v}, \mathbf{v})} - \frac{(\mathbf{u}, \mathbf{v})\overline{(\mathbf{u}, \mathbf{v})}}{(\mathbf{v}, \mathbf{v})} + \frac{(\mathbf{u}, \mathbf{v})(\mathbf{u}, \mathbf{v})}{(\mathbf{v}, \mathbf{v})^2}(\mathbf{v}, \mathbf{v}) \\ &= (\mathbf{u}, \mathbf{u}) - \frac{(\mathbf{u}, \mathbf{v})(\mathbf{u}, \mathbf{v})}{(\mathbf{v}, \mathbf{v})}. \end{aligned}$$

Multiplication of both members by (\mathbf{v}, \mathbf{v}) yields to

$$(\mathbf{u}, \mathbf{v})\overline{(\mathbf{u}, \mathbf{v})} \leq (\mathbf{u}, \mathbf{u})(\mathbf{v}, \mathbf{v})$$

or

$$|(\mathbf{u}, \mathbf{v})|^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2.$$

The statement of the theorem follows from taking the positive square root of both members of the last inequality. ■

It is left as an exercise to prove that the equality holds in the Cauchy-Schwarz inequality if and only if $\{\mathbf{u}, \mathbf{v}\}$ is linearly dependent (Problem 7).

Example 2 □ We shall verify the Cauchy-Schwarz inequality using the inner product from Example 3 of the last section with

$$\mathbf{u} = (1+i, -3i, -2), \quad \mathbf{v} = (2i, 6, -i).$$

We have $\|\mathbf{v}\| = \sqrt{110}$ from our last example. Performing the other required computations, we get

$$(\mathbf{u}, \mathbf{v}) = \begin{bmatrix} 1-i & 3i & -2 \end{bmatrix} \begin{bmatrix} 2 & i & -1 \\ -i & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2i \\ 6 \\ -i \end{bmatrix} = 11 + 61i,$$

$$|(\mathbf{u}, \mathbf{v})| = \sqrt{121 + 3721} = \sqrt{3842},$$

$$(\mathbf{u}, \mathbf{u}) = \begin{bmatrix} 1-i & 3i & -2 \end{bmatrix} \begin{bmatrix} 2 & i & -1 \\ -i & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1+i \\ -3i \\ -2 \end{bmatrix} = 40,$$

$$\|\mathbf{u}\| = \sqrt{40}.$$

The inequality

$$\sqrt{3842} \leq \sqrt{40} \sqrt{110} = \sqrt{4400}$$

is a valid one, verifying the Cauchy-Schwarz inequality in this case. ■

We now derive the fundamental properties of the norm.

Theorem 9.5 *The norm of a vector has the following properties:*

- (i) $\|\mathbf{v}\| > 0$ if $\mathbf{v} \neq \mathbf{0}$, and $\|\mathbf{0}\| = 0$;
- (ii) $\|a\mathbf{v}\| = |a| \cdot \|\mathbf{v}\|$;
- (iii) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Proof. The first of these follows immediately from $\|\mathbf{v}\| = \sqrt{(\mathbf{v}, \mathbf{v})}$, since the inner product (\mathbf{v}, \mathbf{v}) is a positive definite hermitian form.

For any $a \in \mathcal{F}$ and any $\mathbf{v} \in \mathbf{V}$,

$$\|a\mathbf{v}\| = \sqrt{(a\mathbf{v}, a\mathbf{v})} = \sqrt{\overline{aa}(\mathbf{v}, \mathbf{v})} = |a| \sqrt{(\mathbf{v}, \mathbf{v})}.$$

This proves (ii).

For any complex number $z = a + bi$ with $a, b \in \mathbf{R}$, $z + \bar{z} = 2a \leq 2|z|$. Hence

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= (\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) \\ &= (\mathbf{u}, \mathbf{u}) + (\mathbf{u}, \mathbf{v}) + (\mathbf{v}, \mathbf{u}) + (\mathbf{v}, \mathbf{v}) \\ &\leq \|\mathbf{u}\|^2 + 2|(\mathbf{u}, \mathbf{v})| + \|\mathbf{v}\|^2. \end{aligned}$$

But $|(\mathbf{u}, \mathbf{v})| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|$ by the Cauchy-Schwarz inequality, so

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\| \cdot \|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2. \end{aligned}$$

This result is equivalent to property (iii), and the proof is complete. ■

There is one more basic concept to be introduced in an inner product space, that of a distance function.

Definition 9.6 For a given inner product (\mathbf{u}, \mathbf{v}) on \mathbf{V} , the distance $d(\mathbf{u}, \mathbf{v})$ between vectors \mathbf{u} and \mathbf{v} in \mathbf{V} is defined by $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

The distance thus defined has the properties listed in the next theorem. The proof is left as an exercise.

Theorem 9.7 The distance function d has the following properties:

- (i) $d(\mathbf{u}, \mathbf{v}) > 0$ if $\mathbf{u} \neq \mathbf{v}$, and $d(\mathbf{v}, \mathbf{v}) = 0$;
- (ii) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$;
- (iii) $d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}) \geq d(\mathbf{u}, \mathbf{w})$.

Proof. See Problem 8. ■

A set in which there is defined a distance function with the properties (i), (ii), (iii) of Theorem 9.7 is called a **metric space**, and the distance function is called a **metric** or **norm** for the space. The only norms that we are interested in are those connected with an inner product, but there are more general norms.

Exercises 9.3

1. Compute the norm of the given $\mathbf{v} \in \mathcal{C}^3$.
 - (a) $\mathbf{v} = (3, -i, 1)$
 - (b) $\mathbf{v} = (0, 2i, 1)$
 - (c) $\mathbf{v} = (2i, i, 0)$
 - (d) $\mathbf{v} = (5 + i, 5, 0)$
2. Using the inner product from Problem 5 of Exercises 9.2, compute the norm of each \mathbf{v} in Problem 1 above.
3. Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. The rule

$$(\mathbf{u}, \mathbf{v}) = 2u_1v_1 + u_1v_2 + u_2v_1 + 2u_2v_2$$

defines an inner product on \mathbf{R}^2 .

- (a) Use this inner product and verify the Cauchy-Schwarz inequality for $\mathbf{u} = (1, 2)$ and $\mathbf{v} = (3, -5)$.
- (b) Use this inner product and compute $\|\mathbf{w}\|$ for $\mathbf{w} = (-1, 3)$.
- (c) Using this inner product and the vectors from part (a), compute $d(\mathbf{u}, \mathbf{v})$.

4. Use the inner product on \mathbf{R}^2 generated by $\begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$ and verify the Cauchy-Schwarz inequality for $\mathbf{u} = (3, -1)$ and $\mathbf{v} = (2, 4)$.

5. Using the standard inner product on \mathcal{C}^2 , write out the value of $d(\mathbf{u}, \mathbf{v})$ for arbitrary vectors $\mathbf{u}, \mathbf{v} \in \mathcal{C}^2$.
6. Find the value $d(\mathbf{u}, \mathbf{v})$ for the distance function determined by the inner product in Problem 5 of Exercises 9.2.
7. Prove that the equality sign holds in Theorem 9.4 if and only if $\{\mathbf{u}, \mathbf{v}\}$ is linearly dependent. (*Hint:* Consider $((\mathbf{u}, \mathbf{u})\mathbf{v} - (\mathbf{u}, \mathbf{v})\mathbf{u}, (\mathbf{u}, \mathbf{u})\mathbf{v} - (\mathbf{u}, \mathbf{v})\mathbf{u})$ in case $\mathbf{u} \neq \mathbf{0}$.)
8. Prove Theorem 9.7.
9. Use the Cauchy-Schwarz inequality to prove the following statements.
 - (a) For any real numbers a_1, a_2, \dots, a_n and b_1, b_2, \dots, b_n ,

$$\left(\sum_{k=1}^n a_k b_k \right)^2 \leq \left(\sum_{k=1}^n a_k^2 \right) \left(\sum_{k=1}^n b_k^2 \right).$$

- (b) For any complex numbers a_1, a_2, \dots, a_n and b_1, b_2, \dots, b_n ,

$$\left| \sum_{k=1}^n \bar{a}_k b_k \right|^2 \leq \left(\sum_{k=1}^n |a_k|^2 \right) \left(\sum_{k=1}^n |b_k|^2 \right).$$

10. Let \mathbf{V} be a Euclidean space.

- (a) Show that the Cauchy-Schwarz inequality implies that

$$-1 \leq \frac{(\mathbf{u}, \mathbf{v})}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|} \leq 1$$

for any \mathbf{u}, \mathbf{v} in \mathbf{V} .

- (b) We define the angle θ between two nonzero vectors \mathbf{u} and \mathbf{v} in \mathbf{V} by

$$\cos \theta = \frac{(\mathbf{u}, \mathbf{v})}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|},$$

with $0 \leq \theta \leq \pi$. Prove the *law of cosines* in \mathbf{V} :

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2 \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cos \theta.$$

11. For $\mathbf{u} = (1, -4)$ and $\mathbf{v} = (2, -3)$, use the inner product from Problem 3 to compute $\cos \theta$, where θ is the angle between \mathbf{u} and \mathbf{v} .

12. Let $U = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix}$ and $V = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$. Given that the rule

$$(U, V) = u_1 v_1 + u_2 v_2 + u_3 v_3 + u_4 v_4$$

defines an inner product on $\mathbf{R}_{2 \times 2}$, use this inner product to compute $\cos \theta$, where θ is the angle between $A = \begin{bmatrix} 1 & 2 \\ -2 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 3 \\ 6 & 2 \end{bmatrix}$.

13. Assume that the trace function $(U, V) = t(U^T V)$ is an inner product on $\mathbf{R}_{2 \times 2}$, and find the norm of $\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$.

14. Prove that for any \mathbf{u}, \mathbf{v} in a Euclidean space \mathbf{V} ,

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2).$$

15. Prove that for any \mathbf{u}, \mathbf{v} in a Euclidean space \mathbf{V} ,

$$(\mathbf{u}, \mathbf{v}) = \frac{1}{4} \|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4} \|\mathbf{u} - \mathbf{v}\|^2.$$

16. Prove that for any \mathbf{u}, \mathbf{v} in a unitary space \mathbf{V} ,

$$(\mathbf{u}, \mathbf{v}) = \frac{1}{4} \|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4} \|\mathbf{u} - \mathbf{v}\|^2 + \frac{i}{4} \|\mathbf{u} + i\mathbf{v}\|^2 - \frac{i}{4} \|\mathbf{u} - i\mathbf{v}\|^2.$$

9.4 Orthonormal Bases

With the generalizations of the concepts of inner product and length of a vector, there are corresponding generalizations of many of the results in Chapter 8.

Definition 9.8 A set $\{\mathbf{v}_\lambda \mid \lambda \in \mathcal{L}\}$ of vectors in \mathbf{V} is **orthogonal** if $(\mathbf{v}_{\lambda_1}, \mathbf{v}_{\lambda_2}) = 0$ whenever $\lambda_1 \neq \lambda_2$. The set $\{\mathbf{v}_\lambda \mid \lambda \in \mathcal{L}\}$ is **orthonormal** if it is orthogonal and if each \mathbf{v}_λ has norm 1.

It is easily proved that an orthogonal set of nonzero vectors in \mathbf{V} is necessarily linearly independent (see Problem 9).

The definition of an orthogonal matrix over \mathbf{R} remains unchanged from Definition 8.12: P is orthogonal if and only if $P^T = P^{-1}$.

The proof of Theorem 8.14 modifies easily to yield the following result and Theorem 9.11 as well.

Theorem 9.9 Let P be an $r \times r$ matrix over \mathbf{R} , and let \mathbf{V} be a real inner product space. Then P is orthogonal if and only if P is the transition matrix from one orthonormal set of r vectors in \mathbf{V} to another orthonormal set of r vectors in \mathbf{V} .

Proof. See Problem 10. ■

Definition 9.10 A matrix U over \mathcal{C} is **unitary** if and only if $U^* = U^{-1}$.

As transition matrices, unitary matrices play the same role for complex inner product spaces as orthogonal matrices do for Euclidean spaces.

Theorem 9.11 Let U be an $r \times r$ matrix over \mathcal{C} , and let \mathbf{V} be a complex inner product space. Then U is unitary if and only if U is the transition matrix from one orthonormal set of r vectors in \mathbf{V} to another orthonormal set of r vectors in \mathbf{V} .

Proof. See Problem 12. ■

If $\mathbf{v}_j \cdot \mathbf{u}_{k+1}$ is replaced by $(\mathbf{v}_j, \mathbf{u}_{k+1})$ in the Gram-Schmidt orthogonalization process (Theorem 8.15), there results a proof of the following theorem. The verification of this is requested in Problem 13.

Theorem 9.12 Let \mathbf{V} be an inner product space, and let $\mathcal{A} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ be a basis of the subspace \mathbf{W} of \mathbf{V} . There exists an orthonormal basis $\mathcal{N} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ of \mathbf{W} such that each \mathbf{v}_i is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i$.

As an immediate corollary, every finite-dimensional inner product space has an orthonormal basis.

Corollary 9.13 Any orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ of vectors in \mathbf{V} can be extended to an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ of \mathbf{V} .

Proof. See Problem 14. ■

A basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{V} is orthonormal if and only if $(\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}$. This condition is equivalent to the requirement that $(\mathbf{u}, \mathbf{v}) = X^*Y$, where \mathbf{u} and \mathbf{v} have coordinate matrices X and Y , respectively, relative to \mathcal{A} (see Problem 8 of Exercises 9.2). Thus, the orthonormal bases of \mathbf{V} relative to a certain inner product are precisely those bases for which the inner product has I_n as its matrix (see Section 9.2).

Example 1 □ With the inner product on \mathcal{C}^3 that has the matrix

$$H = \begin{bmatrix} \frac{11}{9} & -\frac{2i}{3} & -\frac{1}{9} \\ \frac{2i}{3} & 2 & -\frac{i}{3} \\ -\frac{1}{9} & \frac{i}{3} & \frac{5}{9} \end{bmatrix}$$

relative to the standard basis \mathcal{E}_3 of \mathcal{C}^3 , the inner product of $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ in \mathcal{C}^3 is given by

$$(\mathbf{u}, \mathbf{v}) = \begin{bmatrix} \bar{u}_1 & \bar{u}_2 & \bar{u}_3 \end{bmatrix} H \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

If we let $\mathbf{v}_1 = \frac{1}{\sqrt{3}}(1, 0, 2)$, $\mathbf{v}_2 = \frac{1}{\sqrt{2}}(-1, i, 1)$, $\mathbf{v}_3 = \frac{1}{\sqrt{6}}(1, i, -1)$, it is readily verified that $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of \mathcal{C}^3 relative to this inner product. The matrix of transition from \mathcal{E}_3 to \mathcal{A} is given by

$$P = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{i}{\sqrt{2}} & \frac{i}{\sqrt{6}} \\ \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{bmatrix}$$

and $P^*HP = I_3$. We note that the matrix P is not unitary, but it should not be since \mathcal{E}_3 is not an orthonormal basis relative to this inner product. ■

Exercises 9.4

1. Show that, with the standard inner product in \mathcal{C}^3 , each of the following sets is orthogonal.
 - (a) $\left\{ \left(\frac{3}{2}, \sqrt{2}, -\frac{1}{2}\right), (\sqrt{2}, -2, -\sqrt{2}), (-3, \sqrt{2}, -5) \right\}$
 - (b) $\left\{ (3i, 2i\sqrt{2}, -i), (-i, 2i\sqrt{2}, 5i), (6i\sqrt{2}, -7i, 4i\sqrt{2}) \right\}$
2. As in \mathbf{R}^n , a set is called **normalized** if each vector in the set has norm 1. Normalize the sets in Problem 1.
3. Show that the set $\{(1, 0, 0), (1 + 2i, 1 + 2i, 1), (3 + 4i, 3 + 6i, 1 + 4i)\}$ is orthogonal with respect to the inner product of Problem 5 in Exercises 9.2.
4. With the standard inner product in \mathcal{C}^3 , use the Gram-Schmidt process to find an orthonormal basis of $\langle \mathbf{A} \rangle$.
 - (a) $\mathcal{A} = \{(1, -i, 1), (2, 0, 1 - i)\}$
 - (b) $\mathcal{A} = \{(0, -i, 1), (1 + i, 2, 1)\}$
5. Let the inner product (A, B) be defined on $\mathbf{R}_{2 \times 2}$ as in Problem 7 of Exercises 9.2: $(A, B) = t(A^T B)$. Obtain an orthonormal basis of $\mathbf{R}_{2 \times 2}$ by applying the Gram-Schmidt process to the basis

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\}.$$

6. Find an orthonormal basis of \mathbf{R}^3 that has $(\frac{1}{3}, -\frac{2}{3}, \frac{2}{3})$ as its first element.
7. Using the inner product as in Problem 3, find another orthogonal set of three nonzero vectors in \mathcal{C}^3 .
8. Given that $\mathcal{A} = \left\{ \left(\frac{1}{2}, -\frac{i}{\sqrt{2}}, \frac{1}{2}\right), \left(-\frac{1}{2}i, \frac{1}{\sqrt{2}}, -\frac{i}{2}\right), \left(\frac{1+i}{2}, 0, -\frac{1-i}{2}\right) \right\}$ is an orthonormal basis of \mathcal{C}^3 relative to the standard inner product, extend each of the orthonormal sets \mathcal{B} below to an orthonormal basis of \mathcal{C}^3 .
 - (a) $\mathcal{B} = \left\{ \left(\frac{i}{2}, \frac{1}{\sqrt{2}}, \frac{i}{2}\right) \right\}$
 - (b) $\mathcal{B} = \left\{ \left(\frac{1}{2}, \frac{i}{\sqrt{2}}, \frac{1}{2}\right), \left(\frac{2+i}{2\sqrt{2}}, -\frac{i}{2}, -\frac{i}{2\sqrt{2}}\right) \right\}$

9. Prove that every orthogonal set $\{v_\lambda \mid \lambda \in \mathcal{L}\}$ of nonzero vectors in \mathbf{V} is linearly independent.
10. Prove Theorem 9.9.
11. Prove that the set of all unitary matrices of order n is closed under multiplication.
12. Prove Theorem 9.11.
13. Write out a detailed proof of Theorem 9.12.
14. Prove Corollary 9.13.
15. Let \mathbf{V} be a Euclidean space, and let $\mathbf{u}, \mathbf{v} \in \mathbf{V}$.
 - (a) Prove the Pythagorean theorem in \mathbf{V} : $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$ if and only if \mathbf{u} is orthogonal to \mathbf{v} .
 - (b) Prove that $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ are orthogonal if and only if $\|\mathbf{u}\| = \|\mathbf{v}\|$.

9.5 Orthogonal Complements

The inner product in \mathbf{V} forms the basis for a special type of decomposition of \mathbf{V} as a direct sum with a given subspace as a summand. This type of decomposition involves the *orthogonal complement* of a subspace, a term that we shall define shortly.

Definition 9.14 *Let \mathcal{A} be a nonempty subset of \mathbf{V} . The set \mathcal{A}^\perp (read \mathcal{A} perp) is defined by*

$$\mathcal{A}^\perp = \{\mathbf{v} \in \mathbf{V} \mid (\mathbf{u}, \mathbf{v}) = 0 \text{ for all } \mathbf{u} \in \mathcal{A}\}.$$

That is, \mathcal{A}^\perp consists precisely of all those vectors in \mathbf{V} that are orthogonal to every vector in \mathcal{A} .

Example 1 \square Let $\mathbf{V} = \mathbf{R}^3$ with the standard inner product, and let us consider several possibilities for \mathcal{A} .

If $\mathcal{A} = \{(0, 0, 0)\}$, then $\mathcal{A}^\perp = \mathbf{R}^3$.

If \mathcal{A} consists of the single nonzero vector (a_1, a_2, a_3) , then \mathcal{A}^\perp is the set of all vectors (x_1, x_2, x_3) in the plane $a_1x_1 + a_2x_2 + a_3x_3 = 0$.

If \mathcal{A} consists of two linearly independent vectors, $\mathcal{A} = \{(a_1, a_2, a_3), (b_1, b_2, b_3)\}$, then \mathcal{A}^\perp is the line of intersection of the planes $a_1x_1 + a_2x_2 + a_3x_3 = 0$ and $b_1x_1 + b_2x_2 + b_3x_3 = 0$.

If \mathcal{A} is a basis of \mathbf{R}^3 , then \mathcal{A}^\perp is the zero subspace. ■

In each of the cases considered in the preceding example, \mathcal{A}^\perp was a subspace of \mathbf{R}^3 . It is not difficult to show that this is always the case.

Theorem 9.15 *For any nonempty subset \mathcal{A} of \mathbf{V} , \mathcal{A}^\perp is a subspace of \mathbf{V} .*

Proof. See Problem 1. ■

Theorem 9.16 *For any nonempty subset \mathcal{A} of \mathbf{V} , $\mathcal{A}^\perp = \langle \mathcal{A} \rangle^\perp$.*

Proof. If \mathbf{v} is a vector such that $(\mathbf{u}, \mathbf{v}) = 0$ for all $\mathbf{u} \in \langle \mathcal{A} \rangle$, then surely $(\mathbf{u}, \mathbf{v}) = 0$ for all $\mathbf{u} \in \mathcal{A}$ since $\mathcal{A} \subseteq \langle \mathcal{A} \rangle$. Thus $\langle \mathcal{A} \rangle^\perp \subseteq \mathcal{A}^\perp$.

Now let $\mathbf{v} \in \mathcal{A}^\perp$. Any vector $\mathbf{u} \in \langle \mathcal{A} \rangle$ can be written as $\mathbf{u} = \sum_{k=1}^r a_k \mathbf{u}_k$ with $\mathbf{u}_k \in \mathcal{A}$, and

$$(\mathbf{u}, \mathbf{v}) = \left(\sum_{k=1}^r a_k \mathbf{u}_k, \mathbf{v} \right) = \sum_{k=1}^r \bar{a}_k (\mathbf{u}_k, \mathbf{v}).$$

But $(\mathbf{u}_k, \mathbf{v}) = 0$ for each k since $\mathbf{v} \in \mathcal{A}^\perp$. Hence $(\mathbf{u}, \mathbf{v}) = 0$ and $\mathbf{v} \in \langle \mathcal{A} \rangle^\perp$ since \mathbf{u} was arbitrary in $\langle \mathcal{A} \rangle$. This gives $\mathcal{A}^\perp \subseteq \langle \mathcal{A} \rangle^\perp$, and the proof is complete. ■

Theorem 9.16 shows that there is no loss of generality if we restrict our attention to those \mathbf{W}^\perp where \mathbf{W} is a subspace of \mathbf{V} . In this case, \mathbf{W}^\perp is called the **orthogonal complement** of \mathbf{W} . The justification for this terminology is contained in the next theorem.

Theorem 9.17 *If \mathbf{W} is any subspace of \mathbf{V} , then the sum $\mathbf{W} + \mathbf{W}^\perp$ is direct and $\mathbf{W} \oplus \mathbf{W}^\perp = \mathbf{V}$.*

Proof. Now $\mathbf{v} \in \mathbf{W} \cap \mathbf{W}^\perp$ must satisfy $(\mathbf{v}, \mathbf{v}) = 0$, so $\mathbf{W} \cap \mathbf{W}^\perp = \{\mathbf{0}\}$, and the sum $\mathbf{W} + \mathbf{W}^\perp$ is direct.

If $\mathbf{W} = \{\mathbf{0}\}$ or $\mathbf{W} = \mathbf{V}$, the theorem is trivial. Suppose then that $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is an orthonormal basis of \mathbf{W} and $r < n$. By Corollary 9.13, the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ can be extended to an orthonormal basis

$$\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$$

of \mathbf{V} . We shall show that $\mathcal{B} = \{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is a basis of \mathbf{W}^\perp . As part of a basis, \mathcal{B} is linearly independent, so it is necessary only to show that \mathcal{B} spans \mathbf{W}^\perp . To do this, let $\mathbf{v} = \sum_{k=r+1}^n a_k \mathbf{v}_k$ be any vector in \mathbf{W}^\perp . Since \mathcal{A} is orthonormal, $(\mathbf{v}_i, \mathbf{v}_k) = \delta_{ik}$. Thus, for $i = 1, 2, \dots, r$, we have

$$\begin{aligned} 0 &= (\mathbf{v}_i, \mathbf{v}) \\ &= \left(\mathbf{v}_i, \sum_{k=1}^n a_k \mathbf{v}_k \right) \\ &= \sum_{k=1}^n a_k \delta_{ik} \\ &= a_i. \end{aligned}$$

This means that $\mathbf{v} = \sum_{k=r+1}^n a_k \mathbf{v}_k$, and therefore \mathcal{B} spans \mathbf{W}^\perp . ■

Corollary 9.18 *If \mathbf{W} has dimension r , then \mathbf{W}^\perp has dimension $n - r$.*

Proof. In the proof of the theorem, $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is a basis of \mathbf{W} and $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is a basis of \mathbf{W}^\perp . ■

Whenever $\mathbf{v} \in \mathbf{V}$ is written as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 \in \mathbf{W}$ and $\mathbf{v}_2 \in \mathbf{W}^\perp$, \mathbf{v}_1 is called the **orthogonal projection** of \mathbf{v} onto \mathbf{W} .

Since the orthogonal complement \mathbf{W}^\perp of a subspace \mathbf{W} is again a subspace, it in turn has an orthogonal complement $(\mathbf{W}^\perp)^\perp = \mathbf{W}^{\perp\perp}$. This procedure of forming orthogonal complements can be repeated endlessly, but the next theorem shows that such repetition leads only back and forth between the same two subspaces.

Theorem 9.19 *For any subspace \mathbf{W} of \mathbf{V} , $\mathbf{W}^{\perp\perp} = \mathbf{W}$.*

Proof. Since $\mathbf{W} \oplus \mathbf{W}^\perp = \mathbf{V}$, any $\mathbf{v} \in \mathbf{V}$ can be written uniquely as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 \in \mathbf{W}$ and $\mathbf{v}_2 \in \mathbf{W}^\perp$. To prove that $\mathbf{W}^{\perp\perp} = \mathbf{W}$, it is sufficient to show that $\mathbf{v} \in \mathbf{W}^{\perp\perp}$ if and only if $\mathbf{v}_2 = \mathbf{0}$.

Now $\mathbf{v} \in \mathbf{W}^{\perp\perp}$ if and only if

$$(\mathbf{v}_1, \mathbf{u}) + (\mathbf{v}_2, \mathbf{u}) = (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{u}) = (\mathbf{v}, \mathbf{u}) = 0$$

for all $\mathbf{u} \in \mathbf{W}^\perp$. Since $\mathbf{v}_1 \in \mathbf{W}$, $(\mathbf{v}_1, \mathbf{u}) = 0$ for all $\mathbf{u} \in \mathbf{W}^\perp$. This means that $\mathbf{v} \in \mathbf{W}^{\perp\perp}$ if and only if $(\mathbf{v}_2, \mathbf{u}) = 0$ for all $\mathbf{u} \in \mathbf{W}^\perp$. But $\mathbf{v}_2 \in \mathbf{W}^\perp$, so the last condition holds if and only if $(\mathbf{v}_2, \mathbf{v}_2) = 0$. Hence $\mathbf{v} \in \mathbf{W}^{\perp\perp}$ if and only if $\mathbf{v}_2 = \mathbf{0}$. ■

Exercises 9.5

1. Prove Theorem 9.15.
2. Prove that $(\mathbf{W}_1 + \mathbf{W}_2)^\perp = \mathbf{W}_1^\perp \cap \mathbf{W}_2^\perp$, for subspaces \mathbf{W}_1 and \mathbf{W}_2 of \mathbf{V} .
3. Prove that $(\mathbf{W}_1 \cap \mathbf{W}_2)^\perp = \mathbf{W}_1^\perp + \mathbf{W}_2^\perp$, for subspaces \mathbf{W}_1 and \mathbf{W}_2 of \mathbf{V} .
4. Let $\mathbf{V} = \mathbf{R}^4$, and let $\mathbf{W} = \langle(1, -2, 2, -7), (1, -2, 3, -9)\rangle$.
 - (a) Find an orthonormal basis of \mathbf{W}^\perp .
 - (b) Show how to write an arbitrary vector $\mathbf{v} \in \mathbf{V}$ as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, with $\mathbf{v}_1 \in \mathbf{W}$ and $\mathbf{v}_2 \in \mathbf{W}^\perp$.
5. Find an orthonormal basis of \mathbf{W}^\perp in \mathbf{C}^3 .
 - (a) $\mathbf{W} = \langle(1, i, 1-i), (i, -1, 0)\rangle$
 - (b) $\mathbf{W} = \langle(1, i, i-1), (1, 1+i, 2i-1)\rangle$
6. Let \mathbf{V} be Euclidean, and let \mathbf{W} be a subspace of \mathbf{V} . Prove that whenever $\mathbf{v} \in \mathbf{V}$ is written as $\mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 \in \mathbf{W}$ and $\mathbf{v}_2 \in \mathbf{W}^\perp$, then $\|\mathbf{v}\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2$.
7. Let f be a fixed linear functional on \mathbf{V} with matrix $A = [a_1, a_2, \dots, a_n]$ relative to the orthonormal basis \mathcal{A} of \mathbf{V} , and let $\phi(f)$ be the vector in \mathbf{V} that has coordinate matrix A^* relative to \mathcal{A} . Prove that $f(\mathbf{v}) = (\phi(f), \mathbf{v})$ for all $\mathbf{v} \in \mathbf{V}$.

8. Prove that the mapping ϕ that maps each $f \in \mathbf{V}^*$ onto $\phi(f)$ in \mathbf{V} is a bijective mapping from \mathbf{V}^* to \mathbf{V} , but that ϕ is not an isomorphism except when \mathbf{V} is Euclidean.
9. Prove that, for any subspace \mathbf{W} of \mathbf{V} , f is in the annihilator \mathbf{W}^0 if and only if $\phi(f) \in \mathbf{W}^\perp$.
10. With ϕ as in Problem 8, define (f, g) on \mathbf{V}^* by $(f, g) = (\phi(f), \phi(g))$. Prove that (f, g) defines an inner product on \mathbf{V}^* .

9.6 Isometries

For the remainder of the chapter, (\mathbf{u}, \mathbf{v}) will denote an arbitrarily chosen but fixed inner product on \mathbf{V} . With this fixed inner product, we direct our attention to those linear operators on \mathbf{V} that preserve distances, or leave the lengths of vectors unchanged. These linear operators are called *isometries*. More precisely, we have the following definition.

Definition 9.20 A linear operator T on \mathbf{V} is an **isometry** if $\|T(\mathbf{v})\| = \|\mathbf{v}\|$ for all $\mathbf{v} \in \mathbf{V}$. An isometry on a Euclidean space is called an **orthogonal operator**, and an isometry on a unitary space is called a **unitary operator**.

Since $\|\mathbf{v}\| = \sqrt{(\mathbf{v}, \mathbf{v})}$, one would expect that a linear operator would preserve the norm if and only if it preserved the inner product. Our next theorem shows that this is actually the case.

Theorem 9.21 A linear operator T on \mathbf{V} is an isometry if and only if

$$(\mathbf{u}, \mathbf{v}) = (T(\mathbf{u}), T(\mathbf{v}))$$

for all \mathbf{u}, \mathbf{v} in \mathbf{V} .

Proof. If $(\mathbf{u}, \mathbf{v}) = (T(\mathbf{u}), T(\mathbf{v}))$ for all \mathbf{u}, \mathbf{v} in \mathbf{V} , then

$$\|T(\mathbf{v})\| = \sqrt{(T(\mathbf{v}), T(\mathbf{v}))} = \sqrt{(\mathbf{v}, \mathbf{v})} = \|\mathbf{v}\|,$$

and T is an isometry.

Suppose conversely that T is an isometry. We first observe that

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + (\mathbf{u}, \mathbf{v}) + (\mathbf{v}, \mathbf{u}) + \|\mathbf{v}\|^2. \quad (9.1)$$

If \mathbf{V} is a Euclidean space, then $(\mathbf{v}, \mathbf{u}) = (\mathbf{u}, \mathbf{v})$, and

$$(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \left\{ \|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 \right\}.$$

Hence

$$\begin{aligned} (T(\mathbf{u}), T(\mathbf{v})) &= \frac{1}{2} \left\{ \|T(\mathbf{u}) + T(\mathbf{v})\|^2 - \|T(\mathbf{u})\|^2 - \|T(\mathbf{v})\|^2 \right\} \\ &= \frac{1}{2} \left\{ \|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 \right\} \\ &= (\mathbf{u}, \mathbf{v}), \end{aligned}$$

and the proof is complete for this case.

If \mathbf{V} is a unitary space, we have

$$(\mathbf{u}, \mathbf{v}) + \overline{(\mathbf{u}, \mathbf{v})} = \|\mathbf{u} + i\mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 \quad (9.2)$$

from equation (9.1), and another relation is required. We find that

$$\|\mathbf{u} + i\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + i(\mathbf{u}, \mathbf{v}) - i\overline{(\mathbf{u}, \mathbf{v})} + \|\mathbf{v}\|^2,$$

and therefore

$$(\mathbf{u}, \mathbf{v}) - \overline{(\mathbf{u}, \mathbf{v})} = -i \left\{ \|\mathbf{u} + i\mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 \right\}. \quad (9.3)$$

Equations (9.2) and (9.3) combine to yield

$$(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \left\{ \|\mathbf{u} + \mathbf{v}\|^2 - i\|\mathbf{u} + i\mathbf{v}\|^2 + (i-1) (\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2) \right\}. \quad (9.4)$$

It is left as an exercise (Problem 5) to show that this equation implies $(T(\mathbf{u}), T(\mathbf{v})) = (\mathbf{u}, \mathbf{v})$. ■

Another characterization of an isometry can be given in terms of the images of orthonormal bases.

Theorem 9.22 *A linear operator T on \mathbf{V} is an isometry if and only if T maps an orthonormal basis of \mathbf{V} onto an orthonormal basis of \mathbf{V} .*

Proof. Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be an orthonormal basis of \mathbf{V} , and consider the set $T(\mathcal{A}) = \{T(\mathbf{v}_1), T(\mathbf{v}_2), \dots, T(\mathbf{v}_n)\}$.

If T is an isometry, then by Theorem 9.21 we have $(T(\mathbf{v}_i), T(\mathbf{v}_j)) = (\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}$, and therefore $T(\mathcal{A})$ is an orthonormal basis of \mathbf{V} .

Assume now that $T(\mathcal{A})$ is an orthonormal basis of \mathbf{V} . For any $\mathbf{u} = \sum_{i=1}^n a_i \mathbf{v}_i$, and $\mathbf{v} = \sum_{j=1}^n b_j \mathbf{v}_j$ in \mathbf{V} ,

$$\begin{aligned} (T(\mathbf{u}), T(\mathbf{v})) &= \left(T\left(\sum_{i=1}^n a_i \mathbf{v}_i\right), T\left(\sum_{j=1}^n b_j \mathbf{v}_j\right) \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \bar{a}_i b_j (T(\mathbf{v}_i), T(\mathbf{v}_j)) \\ &= \sum_{i=1}^n \sum_{j=1}^n \bar{a}_i b_j \delta_{ij} \\ &= \sum_{i=1}^n \sum_{j=1}^n \bar{a}_i b_j (\mathbf{v}_i, \mathbf{v}_j) \\ &= \left(\sum_{i=1}^n a_i \mathbf{v}_i, \sum_{j=1}^n b_j \mathbf{v}_j \right) \\ &= (\mathbf{u}, \mathbf{v}). \end{aligned}$$

Therefore T is an isometry by Theorem 9.21. ■

Yet another characterization of an isometry can be made in terms of its matrix relative to an orthonormal basis. This time it is necessary to separate the real and complex cases.

Theorem 9.23 *Let T be a linear operator on the unitary space \mathbf{V} . Then T is unitary if and only if every matrix that represents T relative to an orthonormal basis of \mathbf{V} is a unitary matrix.*

Proof. Let $A = [a_{ij}]_{n \times n}$ be the matrix of T relative to an orthonormal basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of \mathbf{V} . Then $T(\mathbf{v}_j) = \sum_{i=1}^n a_{ij} \mathbf{v}_i$, so that A is the matrix of transition from \mathcal{A} to $T(\mathcal{A})$. By Theorem 9.22, T is unitary if and only if $T(\mathcal{A})$ is orthonormal. But $T(\mathcal{A})$ is orthonormal if and only if A is unitary by Theorem 9.11. ■

A similar application of Theorems 9.22 and 9.9 yields the corresponding result for orthogonal operators.

Theorem 9.24 *Let T be a linear operator on the Euclidean space \mathbf{V} . Then T is orthogonal if and only if every matrix that represents T relative to an orthonormal basis of \mathbf{V} is an orthogonal matrix.*

Example 1 □ Let T be the linear operator on \mathcal{C}^3 given by

$$T(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \left(\frac{1}{2}\mathbf{v}_1 - \frac{i}{2}\mathbf{v}_2 + \frac{1+i}{2}\mathbf{v}_3, -\frac{i}{\sqrt{2}}\mathbf{v}_1 + \frac{1}{\sqrt{2}}\mathbf{v}_2, \frac{1}{2}\mathbf{v}_1 - \frac{i}{2}\mathbf{v}_2 - \frac{1+i}{2}\mathbf{v}_3 \right).$$

The standard basis \mathcal{E}_3 is an orthonormal basis relative to the standard inner product on \mathcal{C}^3 . The matrix of T with respect to \mathcal{E}_3 is

$$A = \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} & \frac{1+i}{2} \\ -\frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{i}{2} & \frac{-1-i}{2} \end{bmatrix}.$$

According to Theorem 9.23, T is a unitary operator if and only if A is a unitary matrix. Checking to see $A^* = A^{-1}$, we find that

$$A^* A = \begin{bmatrix} \frac{1}{2} & \frac{i}{\sqrt{2}} & \frac{1}{2} \\ \frac{i}{2} & \frac{1}{\sqrt{2}} & \frac{i}{2} \\ \frac{1-i}{2} & 0 & \frac{-1+i}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} & \frac{1+i}{2} \\ -\frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{i}{2} & \frac{-1-i}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Therefore A is a unitary matrix, and T is a unitary operator. Illustrating Theorem 9.22, we see that T maps the orthonormal basis \mathcal{E}_3 onto the orthonormal basis

$$T(\mathcal{E}_3) = \left\{ \left(\frac{1}{2}, -\frac{i}{\sqrt{2}}, \frac{1}{2} \right), \left(-\frac{i}{2}, \frac{1}{\sqrt{2}}, -\frac{i}{2} \right), \left(\frac{1+i}{2}, 0, \frac{-1-i}{2} \right) \right\}. \blacksquare$$

Exercises 9.6

1. Determine whether or not each of the matrices below represents an orthogonal operator relative to an orthonormal basis of \mathbf{R}^3 .

(a) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \end{bmatrix}$

(c) $\frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 2 \end{bmatrix}$

2. Determine whether or not each of the given matrices represents a unitary operator relative to an orthonormal basis of \mathcal{C}^3 .

(a) $\frac{1}{6} \begin{bmatrix} 3-4i & 3+i & i \\ (i-2)\sqrt{2} & 2(i+1)\sqrt{2} & (2i-1)\sqrt{2} \\ -1 & 1+3i & 4-3i \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1-i & 0 \\ 1+i & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 & 0 \\ i & -i & i \\ -1 & 1 & 1 \end{bmatrix}$ (d) $\frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{bmatrix}$

3. Find a symmetric orthogonal matrix with $[\frac{1}{3}, -\frac{2}{3}, \frac{2}{3}]^T$ in the first column.
4. Given that $\left\{ \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right) \right\}$ is an orthonormal basis of \mathbf{R}^2 , find an isometry that maps $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$ onto $(0, -1)$.
5. Prove that equation (9.4) in the proof of Theorem 9.21 implies that $(T(\mathbf{u}), T(\mathbf{v})) = (\mathbf{u}, \mathbf{v})$.
6. Prove Theorem 9.24.
7. Prove that if A is either orthogonal or unitary, then $|\det(A)| = 1$.
8. Prove that the product of two isometries is an isometry.
9. Prove that T is an isometry if and only if the image of a unit vector under T is always a unit vector.

9.7 Normal Matrices

In Chapter 7, we formulated necessary and sufficient conditions in order for a linear operator on \mathbf{V} to be represented by a diagonal matrix. At that time, our investigation was concerned with vector spaces in general, whereas now we are concerned only with inner product spaces. This prompts us to pose a restricted form of the original question by asking which linear operators can be represented by a diagonal matrix relative to an orthonormal basis of \mathbf{V} .

A given linear operator T on \mathbf{V} will have a unique matrix A relative to a particular orthonormal basis of \mathbf{V} . If P is the transition matrix from this basis to a new basis of \mathbf{V} , the new matrix of T is $P^{-1}AP$. When \mathbf{V} is unitary, the new basis is orthonormal if and only if P is a unitary matrix. Thus, for a unitary space \mathbf{V} , T can be represented by a diagonal matrix relative to an orthonormal basis if and only if there exists a unitary matrix P such that $P^{-1}AP$ is diagonal. Similarly, for a Euclidean space \mathbf{V} , T can be represented by a diagonal matrix relative to an orthonormal basis if and only if there exists an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

Definition 9.25 *If A and B are square matrices over \mathcal{C} , then B is **unitarily similar** to A if and only if there exists a unitary matrix U such that $B = U^{-1}AU$. If B and A are square matrices over \mathbf{R} , then B is **orthogonally similar** to A if and only if $B = P^{-1}AP$ for some orthogonal matrix P .*

Unitary similarity is an equivalence relation on the square matrices over \mathcal{C} , and orthogonal similarity is an equivalence relation on the square matrices over \mathbf{R} (Problems 4 and 5).

Definition 9.26 *A matrix $B = [b_{ij}]_{m \times n}$ is **upper triangular** if $b_{ij} = 0$ for all $i > j$.*

That is, an upper triangular matrix is one that has only zero elements below the main diagonal. Similarly, a **lower triangular** matrix is one that has only zero elements above the main diagonal.

The next proof should be compared with that of Theorem 8.19, for the technique of proof is much the same.

Theorem 9.27 *Every $n \times n$ matrix A over \mathcal{C} is unitarily similar to an upper triangular matrix B , and the diagonal elements of B are the eigenvalues of A .*

Proof. The theorem is trivial for $n = 1$. We proceed by induction on n .

Assume that the theorem is true for all $k \times k$ matrices over \mathcal{C} , and let A be a matrix of order $k + 1$ over \mathcal{C} . Let \mathbf{V} be a unitary space of dimension $k + 1$, and let \mathcal{A} be an orthonormal basis of \mathbf{V} . The matrix A represents a unique linear operator T on \mathbf{V} relative to the basis \mathcal{A} . Let λ be an eigenvalue of T with \mathbf{v} a corresponding eigenvector of norm 1. By Corollary 9.13, there is an orthonormal basis

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$$

of \mathbf{V} with $\mathbf{v}_1 = \mathbf{v}$. The matrix of transition U_1 from \mathcal{A} to \mathcal{B} is unitary, and the matrix of T relative to \mathcal{B} is of the form

$$A_1 = U_1^{-1}AU_1 = \begin{bmatrix} \lambda & R_1 \\ \mathbf{0} & A_2 \end{bmatrix},$$

where R_1 is 1 by k and A_2 is of order k . By the induction hypothesis, there is a unitary matrix Q such that $Q^{-1}A_2Q$ is upper triangular. The matrix

$$U_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix}$$

is unitary, and

$$\begin{aligned} U_2^{-1}A_1U_2 &= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q^{-1} \end{bmatrix} \begin{bmatrix} \lambda & R_1 \\ \mathbf{0} & A_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix} \\ &= \begin{bmatrix} \lambda & R_1Q \\ \mathbf{0} & Q^{-1}A_2Q \end{bmatrix}. \end{aligned}$$

Thus, $U_2^{-1}A_1U_2$ is upper triangular since $Q^{-1}A_2Q$ is upper triangular. The matrix $U = U_1U_2$ is unitary since U_1 and U_2 are unitary, and

$$B = U^{-1}AU$$

is upper triangular since

$$B = U_2^{-1}U_1^{-1}AU_1U_2 = U_2^{-1}A_1U_2.$$

It is clear that the diagonal elements of B are the eigenvalues of $B = [b_{ij}]$ since

$$\det(B - xI) = (b_{11} - x)(b_{22} - x) \cdots (b_{nn} - x).$$

But B and A have the same eigenvalues since they are similar. ■

Corollary 9.28 *Every linear operator on a unitary space \mathbf{V} can be represented by an upper triangular matrix relative to an orthonormal basis of \mathbf{V} .*

On the surface, it would appear that there should be a result for real matrices that corresponds to Theorem 9.27. That is, it would seem likely that every $n \times n$ real matrix would be orthogonally similar to an upper triangular matrix over \mathbf{R} . But a closer examination shows that this is not the case at all. For the diagonal elements of a triangular matrix are the eigenvalues of that matrix, and the eigenvalues of a real matrix are not necessarily real (see Problem 1 for an example). However, for real matrices that have only real eigenvalues, the proof of Theorem 9.27 can be modified so as to prove the following theorem.

Theorem 9.29 *Let A be a real matrix of order n . Then A is orthogonally similar to an upper triangular matrix if and only if all the eigenvalues of A are real.*

Proof. See Problem 6. ■

Corollary 9.30 A linear operator T on a Euclidean space \mathbf{V} can be represented by an upper triangular matrix relative to an orthonormal basis of \mathbf{V} if and only if all the eigenvalues of T are real.

Now we are ready to establish the criterion for a matrix to be unitarily similar to a diagonal matrix.

Theorem 9.31 A square matrix A over \mathcal{C} is unitarily similar to a diagonal matrix if and only if $AA^* = A^*A$.

Proof. Assume that there exists a unitary matrix U such that

$$U^{-1}AU = D = \text{diag}\{d_1, d_2, \dots, d_n\}.$$

We first note that

$$DD^* = D^*D = \text{diag}\{|d_1|^2, |d_2|^2, \dots, |d_n|^2\}.$$

Since $A = UDU^{-1}$ and $U^{-1} = U^*$, we have

$$\begin{aligned} AA^* &= (UDU^*)(UD^*U^*) \\ &= UDD^*U^* \\ &= UD^*DU^* \\ &= UD^*U^*UDU^* \\ &= A^*A. \end{aligned}$$

Assume now that $AA^* = A^*A$. By Theorem 9.27, there is a unitary matrix U such that $B = U^{-1}AU = U^*AU$ is upper triangular. Now

$$BB^* = U^*AUU^*A^*U = U^*AA^*U,$$

and a similar calculation shows that

$$B^*B = U^*A^*AU.$$

Therefore, $BB^* = B^*B$, since $AA^* = A^*A$. Since $b_{rs} = 0$ in $B = [b_{ij}]$ whenever $r > s$, the element in the i^{th} row and j^{th} column of BB^* is $\sum_{k=1}^n b_{ik}\bar{b}_{jk} = \sum_{k=i}^n b_{ik}\bar{b}_{jk}$. Similarly, the element in the i^{th} row and j^{th} column of B^*B is $\sum_{k=1}^n \bar{b}_{ki}b_{kj} = \sum_{k=1}^i \bar{b}_{ki}b_{kj}$. Equating the diagonal elements of BB^* and B^*B , we have

$$\sum_{k=1}^i |b_{ki}|^2 = \sum_{k=i}^n |b_{ik}|^2. \quad (9.5)$$

For $i = 1$, this yields

$$|b_{11}|^2 = |b_{11}|^2 + |b_{12}|^2 + \cdots + |b_{1n}|^2.$$

Therefore $b_{1j} = 0$ for all $j > 1$, and all elements in the first row of B except b_{11} are zero. In particular, $b_{12} = 0$. With $i = 2$ in equation (9.5), we now have

$$|b_{22}|^2 = |b_{12}|^2 + |b_{22}|^2 = |b_{22}|^2 + |b_{23}|^2 + \cdots + |b_{2n}|^2.$$

Therefore $b_{2j} = 0$ for all $j > 2$, and all elements except b_{22} in the second row of B are zero. This procedure can be repeated with equation (9.5) to obtain

$$|b_{ii}|^2 = |b_{ii}|^2 + |b_{i,i+1}|^2 + \cdots + |b_{in}|^2$$

for $i = 1, 2, \dots, n$. Therefore, $b_{ij} = 0$ for all $j > i$, and B is a diagonal matrix. This completes the proof. ■

Definition 9.32 A square matrix A over a field $\mathcal{F} \subseteq \mathcal{C}$ is called **normal** if and only if $AA^* = A^*A$.

Corollary 9.33 A linear operator T on a unitary space \mathbf{V} can be represented by a diagonal matrix relative to an orthonormal basis of \mathbf{V} if and only if any matrix representing T relative to an orthonormal basis of \mathbf{V} is normal.

The last theorem says that those matrices over \mathcal{C} that are unitarily similar to a diagonal matrix are precisely the normal matrices. These include the symmetric real matrices, the hermitian matrices, the orthogonal matrices, and the unitary matrices. However, there are normal matrices that do not fall into any of these categories. Such an example is found in Problem 2.

Our next theorem gives a simple characterization of those real matrices of order n that are orthogonally similar to a diagonal matrix.

Theorem 9.34 A real matrix of order n is orthogonally similar to a diagonal matrix if and only if it is symmetric.

Proof. If A is real and symmetric, then A is orthogonally similar to a diagonal matrix by Theorem 8.19.

Assume that the real matrix A is orthogonally similar to a diagonal matrix D , and let P be an orthogonal matrix such that $P^{-1}AP = P^TAP = D$. Then $A = PDP^T$, so that

$$A^T = PD^TP^T = PDP^T = A,$$

and A is symmetric. ■

Corollary 9.35 A linear operator T on a Euclidean space \mathbf{V} can be represented by a diagonal matrix relative to an orthonormal basis of \mathbf{V} if and only if every matrix that represents T relative to an orthonormal basis of \mathbf{V} is symmetric.

The results established in this section give a practical way to determine whether or not a matrix is unitarily similar or orthogonally similar to a diagonal matrix, but a systematic procedure for finding a matrix that will accomplish the diagonalization is yet lacking. Such a procedure will be developed at the end of the next section.

Exercises 9.7

1. Determine which of the following real matrices are orthogonally similar to an upper triangular matrix.

$$(a) \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad (b) \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad (c) \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (d) \begin{bmatrix} 7 & -6 \\ -6 & -2 \end{bmatrix}$$

2. Determine which of the following matrices are orthogonal, which are unitary, and which are normal.

$$(a) \begin{bmatrix} -1 & -\frac{3}{\sqrt{2}} & \frac{3}{\sqrt{2}} \\ -\frac{3}{\sqrt{2}} & \frac{1}{2} & \frac{3}{2} \\ \frac{3}{\sqrt{2}} & \frac{3}{2} & \frac{1}{2} \end{bmatrix} \quad (b) \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad (c) \begin{bmatrix} i & i-1 \\ i+1 & 0 \end{bmatrix}$$

$$(d) \begin{bmatrix} 2 & i \\ i & 2 \end{bmatrix} \quad (e) \frac{1}{2} \begin{bmatrix} 1 & i & 1+i \\ -i & 1 & -1+i \\ -1+i & 1+i & 0 \end{bmatrix} \quad (f) \begin{bmatrix} 5 & i & 2 \\ -i & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix}$$

3. For what values of x, y, z is the matrix $A = \begin{bmatrix} a & x & y \\ 0 & b & z \\ 0 & 0 & c \end{bmatrix}$ a normal matrix?

4. Prove that unitary similarity is an equivalence relation on the square matrices over \mathcal{C} .
5. Prove that orthogonal similarity is an equivalence relation on the square matrices over \mathbf{R} .
6. Prove Theorem 9.29.
7. Prove that if a matrix U is unitary, then all eigenvalues of U have absolute value 1.
8. Prove that a square matrix A over \mathcal{C} is normal if and only if every matrix that is unitarily similar to A is normal.
9. Let $B = [b_{ij}]$ be an upper triangular matrix that is square. Prove that B^2 is an upper triangular matrix.

10. Let C be any square matrix over \mathcal{C} .

- (a) Prove that C can be written as $C = A + Bi$ with A and B hermitian.
- (b) Prove that C is normal if and only if $AB = BA$.

9.8 Normal Linear Operators

In the preceding section, we found that a linear operator T on \mathbf{V} can be represented by a diagonal matrix relative to an orthonormal basis if and only if every matrix that represents T relative to an orthonormal basis is normal. Our main purpose in this section is to interpret this requirement as a condition on the linear operator T . This turns naturally to an investigation of the linear operator that has matrix A^* whenever T has matrix A relative to an orthonormal basis \mathcal{A} .

Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be an orthonormal basis of \mathbf{V} . For an arbitrary vector $\mathbf{v} = \sum_{i=1}^n x_i \mathbf{v}_i$ in \mathbf{V} , we have $(\mathbf{v}_j, \mathbf{v}) = \sum_{i=1}^n x_i (\mathbf{v}_j, \mathbf{v}_i) = x_j$. That is, $\mathbf{v} = \sum_{i=1}^n (\mathbf{v}_i, \mathbf{v}) \mathbf{v}_i$. Thus a vector \mathbf{v} is determined uniquely whenever the inner product of \mathbf{v} with each base vector \mathbf{v}_j is specified.

Suppose now that T is a certain fixed linear operator on \mathbf{V} . For a fixed \mathbf{v} in \mathbf{V} , the set of equations

$$(\mathbf{v}_j, \mathbf{u}) = (T(\mathbf{v}_j), \mathbf{v}), \quad j = 1, 2, \dots, n, \quad (9.6)$$

determines a unique vector \mathbf{u} in \mathbf{V} since it specifies the inner product of \mathbf{u} with each \mathbf{v}_j in \mathcal{A} . Thus the rule $T^*(\mathbf{v}) = \mathbf{u}$ defines a mapping T^* of \mathbf{V} into \mathbf{V} . For an arbitrary $\mathbf{w} = \sum_{j=1}^n b_j \mathbf{v}_j$ in \mathbf{V} ,

$$(\mathbf{u}, \mathbf{w}) = \sum_{j=1}^n b_j (\mathbf{u}, \mathbf{v}_j)$$

and

$$(\mathbf{v}, T(\mathbf{w})) = \sum_{j=1}^n b_j (\mathbf{v}, T(\mathbf{v}_j)).$$

Thus the set of equations (9.6) is equivalent to the requirement that $(\mathbf{u}, \mathbf{w}) = (\mathbf{v}, T(\mathbf{w}))$ for all $\mathbf{w} \in \mathbf{V}$. In other words, for each $\mathbf{v} \in \mathbf{V}$, the value $T^*(\mathbf{v})$ is determined by the equation

$$(T^*(\mathbf{v}), \mathbf{w}) = (\mathbf{v}, T(\mathbf{w})) \text{ for all } \mathbf{w} \in \mathbf{V}. \quad (9.7)$$

This leads to the following definition.

Definition 9.36 For each linear operator T on \mathbf{V} , the **adjoint** of T is the mapping T^* of \mathbf{V} into \mathbf{V} that is defined by the equation

$$(T^*(\mathbf{v}), \mathbf{w}) = (\mathbf{v}, T(\mathbf{w}))$$

for all $\mathbf{v}, \mathbf{w} \in \mathbf{V}$.

Theorem 9.37 For any linear operator T on \mathbf{V} , the adjoint of T is a linear operator on \mathbf{V} .

Proof. See Problem 7. ■

Our next theorem provides a basis for the desired interpretation of the results in Section 9.7.

Theorem 9.38 *If the linear operator T on \mathbf{V} has matrix $A = [a_{ij}]$ relative to the orthonormal basis \mathcal{A} of \mathbf{V} , then T^* has matrix A^* relative to \mathcal{A} .*

Proof. Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$. Then $T(\mathbf{v}_i) = \sum_{k=1}^n a_{ki} \mathbf{v}_k$ since A is the matrix of T relative to \mathcal{A} . Suppose $B = [b_{ij}]_n$ is the matrix of T^* relative to \mathcal{A} , so that $T^*(\mathbf{v}_j) = \sum_{k=1}^n b_{kj} \mathbf{v}_k$. From the definition of T^* , we have $(T(\mathbf{v}_i), \mathbf{v}_j) = (\mathbf{v}_i, T^*(\mathbf{v}_j))$. But

$$\begin{aligned}(T(\mathbf{v}_i), \mathbf{v}_j) &= \left(\sum_{k=1}^n a_{ki} \mathbf{v}_k, \mathbf{v}_j \right) \\&= \sum_{k=1}^n \bar{a}_{ki} (\mathbf{v}_k, \mathbf{v}_j) \\&= \sum_{k=1}^n \bar{a}_{ki} \delta_{kj} \\&= \bar{a}_{ji}\end{aligned}$$

and

$$\begin{aligned}(\mathbf{v}_i, T^*(\mathbf{v}_j)) &= \left(\mathbf{v}_i, \sum_{k=1}^n b_{kj} \mathbf{v}_k \right) \\&= \sum_{k=1}^n b_{kj} (\mathbf{v}_i, \mathbf{v}_k) \\&= \sum_{k=1}^n b_{kj} \delta_{ik} \\&= b_{ij}.\end{aligned}$$

Thus $b_{ij} = \bar{a}_{ji}$ and $B = A^*$. ■

It is left as an exercise to prove that $(T^*)^* = T$ for any linear operator T on \mathbf{V} . A linear operator T is called **self-adjoint** if $T^* = T$. It follows from the preceding theorem that if \mathbf{V} is unitary, then T is self-adjoint if and only if every matrix that represents T relative to an orthonormal basis of \mathbf{V} is hermitian. If \mathbf{V} is Euclidean, then T is self-adjoint if and only if every matrix that represents T relative to an orthonormal basis is symmetric. A self-adjoint linear operator of a unitary space is called a **hermitian** operator, and a self-adjoint linear operator on a Euclidean space is a **symmetric** operator.

Definition 9.39 *A linear operator T on \mathbf{V} is **normal** if and only if $TT^* = T^*T$.*

If T has matrix A relative to an orthonormal basis \mathcal{A} of \mathbf{V} , then TT^* has matrix AA^* and T^*T has matrix A^*A relative to \mathcal{A} . Thus T is a normal linear transformation of \mathbf{V} if and only if every matrix of T relative to an orthonormal basis is a normal matrix.

Theorem 9.40 *A linear operator T on the unitary space \mathbf{V} can be represented by a diagonal matrix relative to an orthonormal basis of \mathbf{V} if and only if T is normal.*

Proof. This is a restatement of Corollary 9.33. ■

Theorem 9.41 *A linear operator T on the unitary space \mathbf{V} is normal if and only if there exists an orthonormal basis of \mathbf{V} that consists entirely of eigenvectors of T .*

Proof. By the preceding theorem, T is normal if and only if T can be represented by a diagonal matrix relative to an orthonormal basis. But, according to the proof of Theorem 7.16, $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a basis of eigenvectors of T if and only if the matrix of T relative to \mathcal{A} is diagonal. ■

As was promised in the last section, we proceed now to develop a systematic method for finding a unitary matrix that will accomplish a desired diagonalization.

Many of the results in Section 7.4 are helpful here, even though we are presently restricted to orthonormal bases. If T is represented by a diagonal matrix, the elements on the diagonal are the eigenvalues of T (Corollary 7.17). If T is normal, the geometric multiplicity of each eigenvalue is equal to the algebraic multiplicity (Theorem 7.20).

For orthogonal similarity with a real matrix A , Theorem 9.34 shows that the treatment in Section 8.5 is complete, and the methods developed there apply unchanged. For unitary similarity, some changes are necessary. We must first obtain the result for normal matrices that corresponds to Theorem 8.21 for real symmetric matrices. The first step in this direction is to relate the eigenvalues of T to those of T^* .

Theorem 9.42 *A number λ is an eigenvalue of the normal linear operator T with \mathbf{v} as an associated eigenvector if and only if $\bar{\lambda}$ is an eigenvalue of T^* with the same \mathbf{v} as an associated eigenvector.*

Proof. See Problem 12. ■

Theorem 9.43 *Let A be a normal matrix of order n . If λ_r and λ_s are distinct eigenvalues of A with associated eigenvectors U_r and U_s , then $U_r^* U_s = 0$.*

Proof. Let \mathbf{V} be an n -dimensional inner product space, and let \mathcal{A} be an orthonormal basis of \mathbf{V} . Let T be the linear operator on \mathbf{V} that has matrix A relative to \mathcal{A} . An eigenvector of A corresponding to a certain eigenvalue is the coordinate matrix of an eigenvector of T corresponding to the same eigenvalue. Hence U_r and U_s are the coordinate matrices of eigenvectors \mathbf{v}_r and \mathbf{v}_s that correspond to the eigenvalues λ_r and λ_s , respectively. By Problem 8 of Exercises 9.2, $U_r^* U_s = 0$ if and only if $(\mathbf{v}_r, \mathbf{v}_s) = 0$.

According to Theorem 9.42, $\bar{\lambda}_r$ is an eigenvalue of T^* and $T^*(\mathbf{v}_r) = \bar{\lambda}_r \mathbf{v}_r$. Hence

$$(T^*(\mathbf{v}_r), \mathbf{v}_s) = (\bar{\lambda}_r \mathbf{v}_r, \mathbf{v}_s) = \bar{\lambda}_r (\mathbf{v}_r, \mathbf{v}_s).$$

We also have

$$(\mathbf{v}_r, T(\mathbf{v}_s)) = (\mathbf{v}_r, \lambda_s \mathbf{v}_s) = \lambda_s (\mathbf{v}_r, \mathbf{v}_s).$$

But $(T^*(\mathbf{v}_r), \mathbf{v}_s) = (\mathbf{v}_r, T(\mathbf{v}_s))$, so this means that $\lambda_r (\mathbf{v}_r, \mathbf{v}_s) = \lambda_s (\mathbf{v}_r, \mathbf{v}_s)$ and

$$(\lambda_r - \lambda_s)(\mathbf{v}_r, \mathbf{v}_s) = 0.$$

Since $\lambda_r - \lambda_s \neq 0$, it must be that $(\mathbf{v}_r, \mathbf{v}_s) = 0$, and the proof is complete. ■

We can now formulate our method for finding a unitary matrix U that will diagonalize a given normal matrix $A = [a_{ij}]_n$. Let U_j denote the j^{th} column of $U = [u_{ij}]_n$, so that $U = [U_1, U_2, \dots, U_n]$. The requirement that $U^{-1}AU = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ is equivalent to the system of equations

$$AU_j = \lambda_j U_j, \quad j = 1, 2, \dots, n.$$

That is, each U_j must be an eigenvector of A corresponding to λ_j . Since $U_r^* U_s$ is the element in row r and column s of U^*U , the requirement that $U^* = U^{-1}$ is satisfied if and only if $U_r^* U_s = \delta_{rs}$. With the same notation as in the proof of Theorem 9.43, $U_r^* U_s = (\mathbf{v}_r, \mathbf{v}_s)$. Thus, the columns U_r of U must be the coordinates of an orthonormal basis of eigenvectors of T . Theorem 9.43 assures us that eigenvectors from different eigenspaces are automatically orthogonal. Thus the only modification of the procedure in Section 7.4 that is necessary to make U unitary is to choose orthonormal bases of the eigenspaces \mathbf{V}_{λ_j} . The Gram-Schmidt process can be used to obtain orthonormal bases of those \mathbf{V}_{λ_j} that have dimension greater than 1.

Example 1 □ Consider the problem of finding a unitary matrix U such that $U^{-1}AU$ is diagonal, where A is the normal matrix

$$A = \begin{bmatrix} \frac{2+i}{2} & 0 & \frac{-2+i}{2} \\ 0 & i & 0 \\ \frac{-2+i}{2} & 0 & \frac{2+i}{2} \end{bmatrix}.$$

The characteristic polynomial is found to be

$$\det(A - xI) = -(x - i)^2(x - 2).$$

The augmented matrix $[A - 2I, \mathbf{0}]$ reduces by row operations to

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

so the corresponding eigenvectors in $\mathbf{V}_2 \subseteq \mathcal{C}^3$ are of the form $x_3(-1, 0, 1)$. This means that \mathbf{V}_2 has an orthonormal basis consisting of a real unit vector. But in order to exhibit a solution with a unitary U that is not orthogonal, we choose $x_3 = -\sqrt{3} + i$ and obtain the orthonormal basis

$$\left\{ \left(\frac{\sqrt{3}-i}{2\sqrt{2}}, 0, \frac{-\sqrt{3}+i}{2\sqrt{2}} \right) \right\}$$

for \mathbf{V}_2 . The matrix $A - iI$ leads to solutions of the form $x_2(0, 1, 0) + x_3(1, 0, 1)$. With $x_2 = 1, x_3 = i$, we obtain $(i, 1, i)$, and with $x_2 = i, x_3 = 1$, we obtain $(1, i, 1)$. Application

of the Gram-Schmidt process to the basis $\{(i, 1, i), (1, i, 1)\}$ leads to the orthonormal basis

$$\left\{ \left(\frac{i}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{i}{\sqrt{3}} \right), \left(\frac{1}{\sqrt{6}}, \frac{2i}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right) \right\}$$

of \mathbf{V}_i . Thus

$$U = \begin{bmatrix} \frac{i}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{\sqrt{3}-i}{2\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{2i}{\sqrt{6}} & 0 \\ \frac{i}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-\sqrt{3}+i}{2\sqrt{2}} \end{bmatrix}$$

is a unitary matrix such that $U^{-1}AU = \text{diag}\{i, i, 2\}$. ■

Exercises 9.8

- For each of the following linear operators on \mathcal{C}^2 , write out the value of $T^*(a_1, a_2)$.
 - $T(a_1, a_2) = (a_1 + (1-i)a_2, (1+i)a_1 + 2a_2)$
 - $T(a_1, a_2) = (a_1 + ia_2, a_1 - a_2)$
 - $T(a_1, a_2) = (ia_1 - ia_2, a_1)$
 - $T(a_1, a_2) = (ia_1 + (i-1)a_2, (1+i)a_1)$
- For each linear operator T in Problem 1, find an orthonormal basis of eigenvectors whenever such a basis exists.
- Whenever possible, find a unitary matrix U such that $U^{-1}AU$ is diagonal. The matrices in parts (a)-(e) are from Problem 2 in Exercises 9.7.

$$\begin{array}{ll} (\text{a}) A = \begin{bmatrix} -1 & -\frac{3}{\sqrt{2}} & \frac{3}{\sqrt{2}} \\ -\frac{3}{\sqrt{2}} & \frac{1}{2} & \frac{3}{2} \\ \frac{3}{\sqrt{2}} & \frac{3}{2} & \frac{1}{2} \end{bmatrix} & (\text{b}) A = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \end{bmatrix} \\ (\text{c}) A = \begin{bmatrix} i & i-1 \\ i+1 & 0 \end{bmatrix} & (\text{d}) A = \begin{bmatrix} 2 & i \\ i & 2 \end{bmatrix} \\ (\text{e}) A = \frac{1}{2} \begin{bmatrix} 1 & i & 1+i \\ -i & 1 & -1+i \\ -1+i & 1+i & 0 \end{bmatrix} & (\text{f}) A = \begin{bmatrix} 3 & 1 & 0 \\ 1 & \frac{5}{2} & 1-i \\ 0 & 1+i & 3 \end{bmatrix} \end{array}$$

- Consider the following matrices.

$$\begin{aligned} A &= \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}, \\ D &= \begin{bmatrix} 2+i & -2+i \\ -2+i & 2+i \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \end{aligned}$$

- (a) Determine which of these matrices are orthogonal, which are unitary, and which are normal.
 - (b) Which of these matrices are unitarily similar to an upper triangular matrix?
 - (c) Which of these matrices are unitarily similar to a diagonal matrix?
 - (d) Which of these matrices are similar over \mathcal{C} to a diagonal matrix?
5. Let T be a linear operator on \mathbf{V} over \mathcal{F} .
- (a) Prove that $(T^*)^* = T$.
 - (b) Prove that $(aT)^* = \bar{a}T^*$, for all $a \in \mathcal{F}$.
6. Let S and T be two linear operators on \mathbf{V} .
- (a) Prove that $(S + T)^* = S^* + T^*$.
 - (b) Prove that $(ST)^* = T^*S^*$.
7. Prove Theorem 9.37.
8. Show that $T + T^*$ is self-adjoint.
9. Let T be a linear operator on the unitary space \mathbf{V} . Prove that T is hermitian if and only if $(T(\mathbf{v}), \mathbf{v})$ is real for all $\mathbf{v} \in \mathbf{V}$.
10. Show that T is an isometry if and only if $T^* = T^{-1}$.
11. Prove that a linear operator T on the unitary space \mathbf{V} is normal if and only if $\|T(\mathbf{v})\| = \|T^*(\mathbf{v})\|$ for all $\mathbf{v} \in \mathbf{V}$.
12. Prove Theorem 9.42.
13. Prove that a normal linear operator T is self-adjoint if and only if all the eigenvalues of T are real.
14. Prove that if T is normal, then T and T^* have the same kernel.
15. Prove that if T is an invertible linear operator on \mathbf{V} , then T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$.
16. A linear operator T on an inner product space \mathbf{V} is said to be **skew-adjoint** if $T^* = -T$. (For \mathbf{V} unitary, T is called **skew-hermitian**, and for \mathbf{V} Euclidean, T is called **skew-symmetric**.)
- (a) Prove that, for any linear operator T on \mathbf{V} , $T - T^*$ is skew-adjoint.
 - (b) Show that any linear operator T on \mathbf{V} can be written uniquely as the sum of a self-adjoint linear operator and a skew-adjoint linear operator.

Chapter 10

Spectral Decompositions

10.1 Introduction

In this chapter we consider once again the question of diagonalization of a linear operator on a finite-dimensional vector space \mathbf{V} . We have seen in Chapter 7 that a linear operator T on \mathbf{V} is diagonalizable (i.e., can be represented by a diagonal matrix) if and only if there exists a basis of \mathbf{V} that consists of eigenvectors of T . For a unitary vector space \mathbf{V} , those linear operators that can be represented by a diagonal matrix relative to an orthonormal basis of \mathbf{V} are the same as the normal linear operators on \mathbf{V} . One of our main objectives now is to describe the diagonalizable linear operators in terms that are free of any reference to an inner product. The characterization that we obtain is in terms of a spectral decomposition. A certain acquaintance with projections is essential to a formulation of the concept of a spectral decomposition.

For the entire chapter, \mathbf{V} shall denote an n -dimensional vector space over a field \mathcal{F} , and T shall denote a linear operator on \mathbf{V} .

10.2 Projections and Direct Sums

In Section 9.5, we have seen that a finite-dimensional inner product space \mathbf{V} can be decomposed as a direct sum $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^\perp$, where \mathbf{W} is an arbitrary subspace of \mathbf{V} and \mathbf{W}^\perp is the orthogonal complement of \mathbf{W} . Each vector $\mathbf{v} \in \mathbf{V}$ can be written uniquely as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 \in \mathbf{W}$ and $\mathbf{v}_2 \in \mathbf{W}^\perp$, and the linear operator P defined by $P(\mathbf{v}) = \mathbf{v}_1$ is called the **projection of \mathbf{V} onto \mathbf{W} along \mathbf{W}^\perp** (see Problem 1).

The situation in the preceding paragraph generalizes to any direct sum decomposition $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2$ of an arbitrary (not necessarily an inner product space) \mathbf{V} . For each \mathbf{v} in \mathbf{V} can be written uniquely as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_i \in \mathbf{W}_i$, and the mapping $P(\mathbf{v}) = \mathbf{v}_1$ is again a linear operator on \mathbf{V} . Corresponding to the situation above, P is called the **projection of \mathbf{V} onto \mathbf{W}_1 along \mathbf{W}_2** . (In order to make a distinction for the case where $\mathbf{W}_2 = \mathbf{W}_1^\perp$, P is frequently called the **orthogonal projection of \mathbf{V} onto \mathbf{W}_1 when $\mathbf{W}_2 = \mathbf{W}_1^\perp$** .) For any $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ in \mathbf{V} ,

$$P^2(\mathbf{v}_1 + \mathbf{v}_2) = P(\mathbf{v}_1) = \mathbf{v}_1 = P(\mathbf{v}_1 + \mathbf{v}_2),$$

so P has the property that $P^2 = P$.

Definition 10.1 A linear operator T on \mathbf{V} is called **idempotent** if $T^2 = T$.

Our discussion above shows that the projection P of \mathbf{V} onto \mathbf{W}_1 along \mathbf{W}_2 is idempotent, and that $\mathbf{W}_1 = P(\mathbf{V})$, $\mathbf{W}_2 = P^{-1}(\mathbf{0})$. The converse is also true: An idempotent linear operator T is always a projection of \mathbf{V} onto $T(\mathbf{V})$ along $T^{-1}(\mathbf{0})$.

Theorem 10.2 If T is an idempotent linear operator on \mathbf{V} , then T is the projection of \mathbf{V} onto $T(\mathbf{V})$ along $T^{-1}(\mathbf{0})$.

Proof. Assume that $T^2 = T$. For any $\mathbf{u} \in T(\mathbf{V})$, $\mathbf{u} = T(\mathbf{v})$ for some $\mathbf{v} \in \mathbf{V}$. Hence

$$T(\mathbf{u}) = T^2(\mathbf{v}) = T(\mathbf{v}) = \mathbf{u},$$

and T acts as the identity transformation on $T(\mathbf{V})$. Thus, for any \mathbf{v} in $T(\mathbf{V}) \cap T^{-1}(\mathbf{0})$, we have $\mathbf{v} = T(\mathbf{v}) = \mathbf{0}$, and the sum $T(\mathbf{V}) + T^{-1}(\mathbf{0})$ is direct. Let \mathbf{v} be an arbitrary vector, and let $\mathbf{v}_1 = T(\mathbf{v})$, $\mathbf{v}_2 = (1-T)(\mathbf{v})$, where 1 denotes the identity transformation. Now \mathbf{v}_1 is clearly in $T(\mathbf{V})$ and \mathbf{v}_2 is in $T^{-1}(\mathbf{0})$ since¹ $T(\mathbf{v}_2) = (T - T^2)(\mathbf{v}) = Z(\mathbf{v}) = \mathbf{0}$. Since

$$\mathbf{v} = T(\mathbf{v}) + (1 - T)(\mathbf{v}) = \mathbf{v}_1 + \mathbf{v}_2,$$

we have $\mathbf{V} = T(\mathbf{V}) \oplus T^{-1}(\mathbf{0})$, and T is the projection of \mathbf{V} onto $T(\mathbf{V})$ along $T^{-1}(\mathbf{0})$. ■

In view of Theorem 10.2, an idempotent linear operator is referred to simply as a projection: T is a **projection** if and only if $T^2 = T$.

The relation between projections and direct sums that was described in the first two paragraphs of this section can be extended to direct sums that involve more than two terms.

Let $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \cdots \oplus \mathbf{W}_r$, and let $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r$ denote the decomposition of each $\mathbf{v} \in \mathbf{V}$ with $\mathbf{v}_i \in \mathbf{W}_i$. For $i = 1, 2, \dots, r$, let P_i be defined by

$$P_i(\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r) = \mathbf{v}_i.$$

Then the set $\{P_1, P_2, \dots, P_r\}$ has the following properties:

- (a) each P_i is a projection;
- (b) $P_i P_j = Z$ whenever $i \neq j$;
- (c) $P_1 + P_2 + \cdots + P_r = 1$.

¹The symbol Z denotes the zero linear transformation.

Each P_i is a projection since $P_i^2(\mathbf{v}) = P_i(\mathbf{v}_i) = \mathbf{v}_i = P_i(\mathbf{v})$. If $i \neq j$, $P_i P_j(\mathbf{v}) = P_i(\mathbf{v}_j) = \mathbf{0} = Z(\mathbf{v})$ for all $\mathbf{v} \in \mathbf{V}$, so (b) holds. For any $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r$ in \mathbf{V} , we have

$$\begin{aligned}\mathbf{v} &= P_1(\mathbf{v}) + P_2(\mathbf{v}) + \cdots + P_r(\mathbf{v}) \\ &= (P_1 + P_2 + \cdots + P_r)(\mathbf{v}),\end{aligned}$$

and hence $P_1 + P_2 + \cdots + P_r = 1$.

The three properties of the set $\{P_1, P_2, \dots, P_r\}$ in the preceding paragraph motivate the following definition.

Definition 10.3 A set of projections $\{P_1, P_2, \dots, P_r\}$ is an **orthogonal set** if $P_i P_j = Z$ whenever $i \neq j$. A **complete set of projections** for \mathbf{V} is an orthogonal set $\{P_1, P_2, \dots, P_r\}$ of nonzero projections with the property that $P_1 + P_2 + \cdots + P_r = 1$.

Two projections P_1, P_2 are called **orthogonal** if $\{P_1, P_2\}$ forms an orthogonal set, that is, if $P_1 P_2 = Z = P_2 P_1$.

We have seen above that every direct sum decomposition $\mathbf{V} = \mathbf{W}_1 \oplus \cdots \oplus \mathbf{W}_r$ determines a complete set of projections $\{P_1, \dots, P_r\}$ for \mathbf{V} such that $P_j(\mathbf{v}) = \mathbf{v}_j$ whenever $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_r$ with $\mathbf{v}_i \in \mathbf{W}_i$. The converse is also true, in that if $\{P_1, \dots, P_r\}$ is a complete set of projections for \mathbf{V} , then $\mathbf{V} = P_1(\mathbf{V}) \oplus \cdots \oplus P_r(\mathbf{V})$ and $P_j(\mathbf{v}) = \mathbf{v}_j$ whenever $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_r$ with $\mathbf{v}_i \in P_i(\mathbf{V})$. Now $\mathbf{v} = P_1(\mathbf{v}) + \cdots + P_r(\mathbf{v})$ for each $\mathbf{v} \in \mathbf{V}$ since $P_1 + \cdots + P_r = 1$, and therefore $\mathbf{V} = P_1(\mathbf{V}) + \cdots + P_r(\mathbf{V})$. To see that the sum is direct, let

$$\mathbf{w} = P_i(\mathbf{u}_1) = \sum_{\substack{j=1 \\ j \neq i}}^r P_j(\mathbf{u}_2)$$

be in

$$P_i(\mathbf{V}) \cap \sum_{\substack{j=1 \\ j \neq i}}^r P_j(\mathbf{V}).$$

Then

$$\begin{aligned}\mathbf{w} &= P_i(\mathbf{u}_1) = P_i^2(\mathbf{u}_1) = P_i \left(\sum_{\substack{j=1 \\ j \neq i}}^r P_j(\mathbf{u}_2) \right) \\ &= \sum_{\substack{j=1 \\ j \neq i}}^r P_i P_j(\mathbf{u}_2) = \sum_{\substack{j=1 \\ j \neq i}}^r Z(\mathbf{u}_2) = \mathbf{0}.\end{aligned}$$

It is clear that $P_j(\mathbf{v}) = \mathbf{v}_j$ whenever $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_r$ with $\mathbf{v}_i \in P_i(\mathbf{V})$.

Definition 10.4 A complete set of projections $\{P_1, P_2, \dots, P_r\}$ for \mathbf{V} and a direct sum decomposition $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \cdots \oplus \mathbf{W}_r$ are said to **correspond** to each other, or to be **associated** with each other if $P_j(\mathbf{v}) = \mathbf{v}_j$ whenever $\mathbf{v} \in \mathbf{V}$ is written in the unique form $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r$ with $\mathbf{v}_i \in \mathbf{W}_i$.

Thus, $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \cdots \oplus \mathbf{W}_r$ and $\{P_1, P_2, \dots, P_r\}$ correspond if and only if $P_j(\mathbf{V}) = \mathbf{W}_j$ and P_j acts as the identity on \mathbf{W}_j .

Let us consider now some simple examples of complete sets of projections. The reader should note in each case the corresponding direct sum decomposition.

Example 1 □ Let $\mathbf{V} = \mathbf{R}^n$, and define P_i by

$$P_i(a_1, a_2, \dots, a_n) = a_i \mathbf{e}_i = (0, \dots, 0, a_i, 0, \dots, 0)$$

for $i = 1, 2, \dots, n$. Then $\{P_1, P_2, \dots, P_n\}$ is a complete set of projections for \mathbf{R}^n . This situation generalizes readily to an arbitrary n -dimensional vector space \mathbf{V} . For any given basis $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, let $P_i\left(\sum_{j=1}^n a_j \mathbf{v}_j\right) = a_i \mathbf{v}_i$. Then $\{P_1, P_2, \dots, P_n\}$ is a complete set of projections for \mathbf{V} . ■

Variations in the number of projections in a complete set can easily be made. For example, for each $\mathbf{v} = (a_1, a_2, a_3)$ in \mathbf{R}^3 , let the mappings T_1 and T_2 be defined by $T_1(\mathbf{v}) = (a_1, 0, 0)$ and $T_2(\mathbf{v}) = (0, a_2, a_3)$. Then $\{T_1, T_2\}$ is a complete set of projections for \mathbf{R}^3 .

Theorem 10.5 *Each eigenvalue of a projection is either 0 or 1.*

Proof. Suppose that λ is an eigenvalue of the projection P , and let \mathbf{v} be an associated eigenvector. Then $(P - \lambda)(\mathbf{v}) = \mathbf{0}$, and

$$\begin{aligned} (1 - \lambda)\lambda\mathbf{v} &= (1 - \lambda)P(\mathbf{v}) \\ &= (P - \lambda P)(\mathbf{v}) \\ &= (P^2 - \lambda P)(\mathbf{v}) \\ &= P((P - \lambda)(\mathbf{v})) \\ &= P(\mathbf{0}) \\ &= \mathbf{0}. \end{aligned}$$

Since $\mathbf{v} \neq \mathbf{0}$, it must be that $(1 - \lambda)\lambda = 0$, and λ is either 0 or 1. ■

The fact that a projection P has the property $P^2 = P$ and also acts as the identity transformation on its range might lead one to expect the matrix of a projection to look somehow like the identity matrix. This is not necessarily the case, however. For example, the matrix

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

is such that $A^2 = A$, and hence represents a projection. But with an appropriate choice of basis, the matrix of a projection P can be made to take on a form very much like I_n . Now $P(\mathbf{V})$ is the same as the eigenspace of P corresponding to the eigenvalue 1, and $\mathbf{V} = P(\mathbf{V}) \oplus P^{-1}(\mathbf{0})$. Thus, if a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ of $P(\mathbf{V})$ is extended to a basis $\mathcal{A} = \{\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ of \mathbf{V} with $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ a basis of $P^{-1}(\mathbf{0})$, then the matrix of P relative to \mathcal{A} is

$$D_r = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

An interchange of the \mathbf{v}_i in \mathcal{A} produces an interchange of the elements on the diagonal of D_r , so the 1's on the diagonal can be placed in any desired diagonal positions.

Exercises 10.2

1. If $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2$, prove that the projection of \mathbf{V} onto \mathbf{W}_1 along \mathbf{W}_2 is a linear operator on \mathbf{V} .
2. Let T be a linear operator on \mathbf{V} . Prove that T is a projection if and only if $1 - T$ is a projection.
3. Prove that if P is the projection of \mathbf{V} onto \mathbf{W}_1 along \mathbf{W}_2 , then $1 - P$ is the projection of \mathbf{V} onto \mathbf{W}_2 along \mathbf{W}_1 .
4. Let P_1, P_2, P_3 be nonzero projections of \mathbf{V} such that $P_1 + P_2 + P_3 = 1$, and assume that $1 + 1 \neq 0$ in \mathcal{F} . Prove that $\{P_1, P_2, P_3\}$ is a complete set of projections for \mathbf{V} .
5. Give an example of two projections P_1, P_2 such that $P_1 P_2 = Z$ but $P_2 P_1 \neq Z$.
6. Prove that if P is a projection on the inner product space \mathbf{V} , then P^* is a projection on \mathbf{V} .
7. Prove that any complete set of projections for \mathbf{V} is linearly independent.
8. Let $\mathbf{V} = \mathbf{W}_1 + \mathbf{W}_2 + \cdots + \mathbf{W}_r$, where each \mathbf{W}_i is a subspace of \mathbf{V} . Prove that $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \cdots \oplus \mathbf{W}_r$ if and only if $\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r = \mathbf{0}$ and $\mathbf{v}_i \in \mathbf{W}_i$ imply that each $\mathbf{v}_j = \mathbf{0}$.
9. Let P_1 be the projection of \mathbf{V} onto \mathbf{S}_1 along \mathbf{S}_2 , and let P_2 be the projection of \mathbf{V} onto \mathbf{W}_1 along \mathbf{W}_2 , and assume that $1 + 1 \neq 0$ in \mathcal{F} .
 - (a) Prove that $P_1 + P_2$ is a projection if and only if $\{P_1, P_2\}$ is orthogonal.
 - (b) Prove that if P_1 and P_2 are orthogonal projections, then $P_1 + P_2$ is the projection of \mathbf{V} onto $\mathbf{S}_1 \oplus \mathbf{W}_1$ along $\mathbf{S}_2 \cap \mathbf{W}_2$.
10. Let \mathbf{V} be an inner product space. Prove that if P is the projection of \mathbf{V} onto \mathbf{W}_1 along \mathbf{W}_2 , then P^* is the projection of \mathbf{V} onto \mathbf{W}_2^\perp along \mathbf{W}_1^\perp .

11. According to the definition in the second paragraph of this section, a projection P on an inner product space \mathbf{V} is called an orthogonal projection if and only if $P(\mathbf{V})$ and $P^{-1}(\mathbf{0})$ are orthogonal subspaces. Prove that a projection P is an orthogonal projection if and only if P is self-adjoint.
12. Let P be a projection on the inner product space \mathbf{V} . Prove that if $\|P(\mathbf{v})\| \leq \|\mathbf{v}\|$ for all $\mathbf{v} \in \mathbf{V}$, then P is an orthogonal projection.
13. Prove that a projection on an inner product space \mathbf{V} is self-adjoint if and only if it is normal.

10.3 Spectral Decompositions

With the contents of the preceding section at hand, we are prepared to formulate a new characterization of the diagonalizable linear operators on \mathbf{V} . It is clear that in order for T to be diagonalizable, all of its eigenvalues must be in \mathcal{F} . That is, the characteristic polynomial of T must factor into linear factors over \mathcal{F} . The problem, then is to determine what additional conditions are necessary in order that T be diagonalizable.

The characterization of diagonalizable operators that we shall obtain is phrased in terms of projections. The projections are a very special type of operator, so it is natural to ask how the set of projections fits into the vector space that consists of all linear operators on \mathbf{V} . The sum of two projections is usually *not* a projection (see Problem 1), so the projections do not form a subspace. However, our main interest is not with the set of all projections, but rather with those subspaces that are spanned by a complete set of projections. That is, we are interested primarily in those linear operators that are linear combinations of a complete set of projections. Our first result concerns the eigenvectors of this type of operator.

Theorem 10.6 *Let $\{P_1, P_2, \dots, P_r\}$ be a complete set of projections for \mathbf{V} , and suppose that $T = c_1P_1 + c_2P_2 + \dots + c_rP_r$ for some scalars c_i . Then each c_i is an eigenvalue of T , and each eigenvector \mathbf{v}_i associated with the eigenvalue 1 of P_i is an eigenvector of T associated with c_i .*

Proof. Let \mathbf{v}_i be an eigenvector associated with the eigenvalue 1 of P_i , so that $P_i(\mathbf{v}_i) = \mathbf{v}_i$. For any $j \neq i$, we have $P_j(\mathbf{v}_i) = P_j(P_i(\mathbf{v}_i)) = \mathbf{0}$. Therefore

$$\begin{aligned} T(\mathbf{v}_i) &= c_1P_1(\mathbf{v}_i) + c_2P_2(\mathbf{v}_i) + \dots + c_rP_r(\mathbf{v}_i) \\ &= c_iP_i(\mathbf{v}_i) \\ &= c_i\mathbf{v}_i, \end{aligned}$$

and c_i is an eigenvalue of T with \mathbf{v}_i as an associated eigenvector. ■

Theorem 10.7 *The eigenvectors of a complete set of projections span \mathbf{V} .*

Proof. Let $\{P_1, P_2, \dots, P_r\}$ be a complete set of projections for \mathbf{V} . For an arbitrary $\mathbf{v} \in \mathbf{V}$, consider the vector $P_i(\mathbf{v})$. Since

$$P_i(P_i(\mathbf{v})) = P_i^2(\mathbf{v}) = P_i(\mathbf{v}),$$

$P_i(\mathbf{v})$ is an eigenvector of P_i unless $P_i(\mathbf{v}) = \mathbf{0}$. Since $1 = P_1 + P_2 + \dots + P_r$,

$$\mathbf{v} = P_1(\mathbf{v}) + P_2(\mathbf{v}) + \dots + P_r(\mathbf{v}).$$

But $P_1(\mathbf{v}) + P_2(\mathbf{v}) + \dots + P_r(\mathbf{v})$ is a linear combination of eigenvectors of P_1, P_2, \dots, P_r , and the proof is complete. ■

We are now in a position to prove the main theorem of this section.

Theorem 10.8 *Let T be a linear operator on \mathbf{V} with distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$. Then T is diagonalizable if and only if*

$$T = \lambda_1 P_1 + \lambda_2 P_2 + \dots + \lambda_r P_r,$$

where $\{P_1, P_2, \dots, P_r\}$ is a complete set of projections for \mathbf{V} .

Proof. Suppose first that $T = \lambda_1 P_1 + \lambda_2 P_2 + \dots + \lambda_r P_r$, where $\{P_1, P_2, \dots, P_r\}$ is a complete set of projections for \mathbf{V} . Let $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a basis of \mathbf{V} , and consider the set of nr vectors

$$\mathcal{B} = \{P_1(\mathbf{v}_1), \dots, P_1(\mathbf{v}_n), P_2(\mathbf{v}_1), \dots, P_2(\mathbf{v}_n), \dots, P_r(\mathbf{v}_1), \dots, P_r(\mathbf{v}_n)\}.$$

For any $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i$, we have

$$\begin{aligned} \mathbf{v} &= (P_1 + P_2 + \dots + P_r)(\mathbf{v}) \\ &= (P_1 + P_2 + \dots + P_r) \left(\sum_{i=1}^n a_i \mathbf{v}_i \right) \\ &= \sum_{i=1}^n \sum_{j=1}^r a_i P_j(\mathbf{v}_i). \end{aligned}$$

Thus, \mathcal{B} spans \mathbf{V} , and therefore contains a basis $\mathcal{C} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ of \mathbf{V} . Each \mathbf{u}_k is a nonzero vector of the form $\mathbf{u}_k = P_j(\mathbf{v}_i)$, and hence is an eigenvector of P_j associated with the eigenvalue 1. By Theorem 10.6, \mathbf{u}_k is an eigenvector of T associated with the eigenvalue λ_j . Thus \mathcal{C} is a basis of eigenvectors of T , and T is diagonalizable by Theorem 7.16.

Suppose now that T is diagonalizable, and let $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$ be a basis of \mathbf{V} such that T is represented by a diagonal matrix D relative to \mathcal{B} . As in the proof of Theorem 7.16, the \mathbf{w}_i are eigenvectors of T , and the diagonal elements of D are the eigenvalues λ_i of T . By Theorem 7.20, the geometric multiplicity of λ_j is the same as the algebraic multiplicity m_j . That is, \mathbf{V}_{λ_j} is of dimension m_j . By Corollary 7.10, the sum $\sum_{j=1}^r \mathbf{V}_{\lambda_j}$ is direct. Since $\sum_{j=1}^r m_j = n$, it must be that $\mathbf{V} = \mathbf{V}_{\lambda_1} \oplus \mathbf{V}_{\lambda_2} \oplus \dots \oplus \mathbf{V}_{\lambda_r}$.

With $\mathbf{W}_j = \mathbf{V}_{\lambda_j}$ in Definition 10.4, the set of projections P_j defined there is a complete set of projections for \mathbf{V} . For each $\mathbf{v} \in \mathbf{V}$, we have

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r$$

with $\mathbf{v}_j = P_j(\mathbf{v})$. And since $\mathbf{v}_j \in \mathbf{V}_{\lambda_j}$, we have $T(\mathbf{v}_j) = \lambda_j \mathbf{v}_j$. Thus

$$\begin{aligned} T(\mathbf{v}) &= T(\mathbf{v}_1) + T(\mathbf{v}_2) + \cdots + T(\mathbf{v}_r) \\ &= \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_r \mathbf{v}_r \\ &= \lambda_1 P_1(\mathbf{v}) + \lambda_2 P_2(\mathbf{v}) + \cdots + \lambda_r P_r(\mathbf{v}) \\ &= (\lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r)(\mathbf{v}), \end{aligned}$$

and $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$. ■

Definition 10.9 If the linear operator T on \mathbf{V} can be written in the form

$$T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r,$$

where $\{P_1, P_2, \dots, P_r\}$ is a complete set of projections for \mathbf{V} and $\lambda_1, \lambda_2, \dots, \lambda_r$ are the distinct eigenvalues of T , then the expression $\lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$ is called a **spectral decomposition** of T .

Theorem 10.8 asserts that T is diagonalizable if and only if T has a spectral decomposition.

Theorem 10.10 If T has a spectral decomposition $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$, then $f(T) = f(\lambda_1)P_1 + f(\lambda_2)P_2 + \cdots + f(\lambda_r)P_r$ for every polynomial $f(x)$.

Proof. We first show that $T^i = \sum_{j=1}^r \lambda_j^i P_j$. This is trivial for $i = 0$ and $i = 1$. Assuming that $T^k = \sum_{j=1}^r \lambda_j^k P_j$, we have

$$\begin{aligned} T^{k+1} &= \left(\sum_{j=1}^r \lambda_j^k P_j \right) \left(\sum_{m=1}^r \lambda_m P_m \right) \\ &= \sum_{j=1}^r \sum_{m=1}^r \lambda_j^k \lambda_m P_j P_m \\ &= \sum_{j=1}^r \lambda_j^{k+1} P_j. \end{aligned}$$

By induction, $T^i = \sum_{j=1}^r \lambda_j^i P_j$ for each nonnegative integer i . For any polynomial $f(x) = \sum_{i=0}^s c_i x^i$, we thus have

$$\begin{aligned} f(T) &= \sum_{i=0}^s c_i \left(\sum_{j=1}^r \lambda_j^i P_j \right) \\ &= \sum_{j=1}^r \left(\sum_{i=0}^s c_i \lambda_j^i \right) P_j \\ &= \sum_{j=1}^r f(\lambda_j) P_j. \quad \blacksquare \end{aligned}$$

Theorem 10.10 will prove to be useful in the next section.

Theorem 10.11 *If $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$ is a spectral decomposition of T , then $\mathbf{V} = \mathbf{V}_{\lambda_1} \oplus \mathbf{V}_{\lambda_2} \oplus \cdots \oplus \mathbf{V}_{\lambda_r}$, and P_j is the projection of \mathbf{V} onto \mathbf{V}_{λ_j} along $\sum_{\substack{i=1 \\ i \neq j}}^r \mathbf{V}_{\lambda_i}$.*

Proof. Let $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$ be a spectral decomposition of T . Then T is diagonalizable by Theorem 10.8, and $\mathbf{V} = \mathbf{V}_{\lambda_1} \oplus \mathbf{V}_{\lambda_2} \oplus \cdots \oplus \mathbf{V}_{\lambda_r}$ as in the proof of that theorem. Since $1 = P_1 + P_2 + \cdots + P_r$, we have $\mathbf{v} = P_1(\mathbf{v}) + P_2(\mathbf{v}) + \cdots + P_r(\mathbf{v})$ for any \mathbf{v} in \mathbf{V} . But $P_j(\mathbf{v})$ is in \mathbf{V}_{λ_j} since

$$T(P_j(\mathbf{v})) = \left(\sum_{i=1}^r \lambda_i P_i \right) (P_j(\mathbf{v})) = \lambda_j (P_j(\mathbf{v})).$$

Thus we have $P_j(\mathbf{V}) \subseteq \mathbf{V}_{\lambda_j}$ and hence

$$\dim(P_j(\mathbf{V})) \leq \dim(\mathbf{V}_{\lambda_j}).$$

This implies the inequalities

$$n = \sum_{j=1}^r \dim(P_j(\mathbf{V})) \leq \sum_{j=1}^r \dim(\mathbf{V}_{\lambda_j}) = n$$

and therefore

$$\dim(P_j(\mathbf{V})) = \dim(\mathbf{V}_{\lambda_j}),$$

for $j = 1, 2, \dots, r$. It follows that $P_j(\mathbf{V}) = \mathbf{V}_{\lambda_j}$ for $j = 1, 2, \dots, r$. Hence P_j is the projection of \mathbf{V} onto \mathbf{V}_{λ_j} along the sum of the remaining eigenspaces. ■

As might be expected, a restriction to orthonormal bases of an inner product space \mathbf{V} corresponds to a restriction on the projections in the spectral decomposition. The exact description is as follows.

Theorem 10.12 *A linear operator T on an n -dimensional inner product space \mathbf{V} can be represented by a diagonal matrix relative to an orthonormal basis of \mathbf{V} if and only if T has a spectral decomposition in which the projections are self-adjoint.*

Proof. With the proofs of Theorems 7.16 and 10.8 in mind, it is sufficient to prove that an orthonormal basis of eigenvectors of T exists if and only if T has a spectral decomposition in which each projection is self-adjoint.

Suppose first that $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$, with the P_i self-adjoint. Let $\mathcal{B}_j = \{\mathbf{u}_{j1}, \mathbf{u}_{j2}, \dots, \mathbf{u}_{jn_j}\}$ be an orthonormal basis of the eigenspace \mathbf{V}_{λ_j} for $j = 1, 2, \dots, r$. As in the proof of Theorem 7.20, the set

$$\mathcal{B} = \{\mathbf{u}_{11}, \dots, \mathbf{u}_{1n_1}, \mathbf{u}_{21}, \dots, \mathbf{u}_{2n_2}, \dots, \mathbf{u}_{r1}, \dots, \mathbf{u}_{rn_r}\}$$

is a basis of \mathbf{V} . This basis is orthonormal if and only if $(\mathbf{u}_{it}, \mathbf{u}_{js}) = 0$ whenever $i \neq j$. According to Theorem 10.11, P_j is the projection of \mathbf{V} onto \mathbf{V}_{λ_j} along $\sum_{i \neq j} \mathbf{V}_{\lambda_i}$. Hence $P_j(\mathbf{v}) = \mathbf{v}$ for each \mathbf{v} in \mathbf{V}_{λ_j} . In other words, each nonzero vector of \mathbf{V}_{λ_j} is an eigenvector of P_j associated with 1. Consequently, we have

$$\begin{aligned} (\mathbf{u}_{it}, \mathbf{u}_{js}) &= (P_i(\mathbf{u}_{it}), P_j(\mathbf{u}_{js})) \\ &= (\mathbf{u}_{it}, P_i^* P_j(\mathbf{u}_{js})) \\ &= (\mathbf{u}_{it}, P_i P_j(\mathbf{u}_{js})) \\ &= (\mathbf{u}_{it}, \mathbf{0}) \\ &= 0 \end{aligned}$$

whenever $i \neq j$.

Conversely, suppose that there exists an orthonormal basis of eigenvectors of T . With the same notation as used in the proof of Theorem 10.8, $\mathbf{V} = \mathbf{V}_{\lambda_1} \oplus \mathbf{V}_{\lambda_2} \oplus \cdots \oplus \mathbf{V}_{\lambda_r}$. The set of projections $\{P_1, P_2, \dots, P_r\}$ is a complete set for \mathbf{V} , and $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$. The proof will be complete if we show that each P_i is self-adjoint. Let $\mathbf{u} = P_1(\mathbf{u}) + P_2(\mathbf{u}) + \cdots + P_r(\mathbf{u})$ and $\mathbf{v} = P_1(\mathbf{v}) + P_2(\mathbf{v}) + \cdots + P_r(\mathbf{v})$ be any two vectors in \mathbf{V} . For $i \neq j$, $P_i(\mathbf{u})$ and $P_j(\mathbf{v})$ are in distinct eigenspaces \mathbf{V}_{λ_i} and \mathbf{V}_{λ_j} , and so are orthogonal. This implies that

$$\begin{aligned} (\mathbf{u}, P_j(\mathbf{v})) &= \left(\sum_{i=1}^r P_i(\mathbf{u}), P_j(\mathbf{v}) \right) \\ &= (P_j(\mathbf{u}), P_j(\mathbf{v})) \\ &= \left(P_j(\mathbf{u}), \sum_{i=1}^r P_i(\mathbf{v}) \right) \\ &= (P_j(\mathbf{u}), \mathbf{v}), \end{aligned}$$

and therefore $P_j^* = P_j$. ■

We shall devise a method for obtaining the projections involved in a spectral decomposition near the end of the next section.

Exercises 10.3

1. Verify that the mappings of \mathbf{R}^2 into \mathbf{R}^2 defined by $P_1(x_1, x_2) = (x_1 + x_2, 0)$ and $P_2(x_1, x_2) = (x_2, x_2)$ are projections, and show that $P_1 + P_2$ is not a projection.
2. Let P_1 and P_2 be the projections defined on \mathcal{C}^2 by $P_1(x_1, x_2) = (x_2, x_1)$ and $P_2(x_1, x_2) = (x_2, x_2)$.
 - (a) Let $T = 3P_1 + 4P_2$ and determine if T is diagonalizable.
 - (b) Determine whether or not $P_1 + P_2$ is a projection.

3. Prove that if T is invertible and has a spectral decomposition, then T^{-1} has a spectral decomposition with the same complete set of projections.
4. Let $\{P_1, P_2, \dots, P_r\}$ be a complete set of projections for \mathbf{V} , and suppose that $T = c_1P_1 + c_2P_2 + \dots + c_rP_r$.
 - (a) Prove that each eigenvalue of T is equal to at least one of the scalars c_j .
 - (b) Give an example which shows that there may be a vector \mathbf{v} such that \mathbf{v} is an eigenvector of T associated with c_j , but \mathbf{v} is not an eigenvector of any P_i .
5. Prove that each projection P_i in a spectral decomposition

$$T = \lambda_1P_1 + \lambda_2P_2 + \dots + \lambda_rP_r$$

of T has rank equal to the geometric multiplicity of the eigenvalue λ_i of T .

6. Prove that the spectral decomposition of T is unique if it exists.
7. Let T be a normal linear transformation of the inner product space \mathbf{V} . Prove that T^* has a spectral decomposition.
8. Let $T = \sum_{i=1}^r \lambda_i P_i$ be a spectral decomposition for the linear operator T on the inner product space \mathbf{V} . Suppose that $\{f_1(x), f_2(x), \dots, f_r(x)\}$ is a set of polynomials with real coefficients such that $f_i(\lambda_j) = \delta_{ij}$. Prove that $f_i(T) = P_i$ for $i = 1, 2, \dots, r$.
9. A linear operator T on an inner product space \mathbf{V} is called **nonnegative** if T is self-adjoint and $(T(\mathbf{v}), \mathbf{v}) \geq 0$ for every $\mathbf{v} \in \mathbf{V}$.
 - (a) Prove that any nonnegative linear operator T on \mathbf{V} has a spectral decomposition $T = \sum_{i=1}^r \lambda_i P_i$ with each $\lambda_i \geq 0$.
 - (b) Show that $S = \sqrt{\lambda_1}P_1 + \sqrt{\lambda_2}P_2 + \dots + \sqrt{\lambda_r}P_r$ is a nonnegative linear operator such that $S^2 = T$ (i.e., S is a nonnegative square root of T).
 - (c) Show that the nonnegative square root of S in (b) is unique.

10.4 Minimal Polynomials and Spectral Decompositions

In this section, we shall need to use certain properties of addition and multiplication of polynomials in x over a field \mathcal{F} (i.e., polynomials in x with coefficients in \mathcal{F}). The set of all polynomials in x over \mathcal{F} is denoted by $\mathcal{F}[x]$. A derivation of the information that we need about $\mathcal{F}[x]$ would constitute a major digression at this point. For this reason, we shall simply state without proofs the results that are needed.

If $p(x)$ and $m(x)$ are in $\mathcal{F}[x]$ with $p(x)$ nonzero, then there exist polynomials $q(x), r(x)$ in $\mathcal{F}[x]$ such that

$$m(x) = p(x)q(x) + r(x)$$

with $r(x)$ either the zero polynomial or a polynomial of degree less than that of $p(x)$. This statement is known as the **division algorithm** for elements of $\mathcal{F}[x]$. If $r(x)$ is the zero polynomial, we say that $p(x)$ **divides** $m(x)$, and that $p(x)$ is a **divisor** of $m(x)$.

A nonzero polynomial $p(x)$ in $\mathcal{F}[x]$ is called **monic** if the coefficient of the highest degree term in $p(x)$ is 1. That is, the highest power of x that appears in $p(x)$ has 1 as its coefficient.

A monic polynomial $d(x)$ in $\mathcal{F}[x]$ is called the **greatest common divisor** of a set of nonzero polynomials $q_1(x), q_2(x), \dots, q_r(x)$ in $\mathcal{F}[x]$ if

- i. $d(x)$ is a divisor of each of the polynomials $q_i(x)$, and
- ii. every polynomial $p(x)$ that divides each $q_i(x)$ is also a divisor of $d(x)$.

Every set of nonzero polynomials $q_1(x), q_2(x), \dots, q_r(x)$ in $\mathcal{F}[x]$ has a unique monic greatest common divisor $d(x)$ in $\mathcal{F}[x]$. Moreover, there exists a set of polynomials $g_1(x), g_2(x), \dots, g_r(x)$ in $\mathcal{F}[x]$ such that

$$d(x) = g_1(x)q_1(x) + g_2(x)q_2(x) + \cdots + g_r(x)q_r(x).$$

In some of the examples and exercises, we assume that the student is familiar with the partial fraction decomposition of a quotient of two nonzero polynomials.

We have already encountered polynomials $p(T) = \sum_{i=0}^s c_i T^i$ in a linear operator T on \mathbf{V} and the corresponding polynomials $p(A) = \sum_{i=0}^s c_i A^i$ in a matrix A that represents T . It has been noted that T and A satisfy the same polynomial equations.

From one point of view, a polynomial $p(A) = \sum_{i=0}^s c_i A^i$ in the square matrix A can be thought of as being obtained from a polynomial $p(x) = \sum_{i=0}^s c_i x^i$ by replacing the powers x^i of the indeterminate x by the corresponding powers A^i of the matrix A . This suggests the construction of other types of polynomials involving matrices by making other replacements in $p(x)$. One might treat x as an indeterminate scalar and replace the scalar coefficients c_i by matrix coefficients C_i , or one might replace both the c_i and the x^i by matrix quantities. The algebra connected with this last type of polynomials is quite involved, and we shall not be concerned with them here. Our interest is confined to only two types of polynomials:

1. Those obtained by leaving the scalar coefficients c_i unchanged and replacing x by a square matrix A .
2. Those obtained by treating x as an indeterminate scalar and replacing the scalar coefficients c_i by matrix coefficients C_i .

There is another point of view from which the polynomials of the second type may be regarded. Any polynomial $\sum_{i=0}^s C_i x^i$ of this type can be considered to be a matrix with elements in $\mathcal{F}[x]$. For example,

$$\begin{bmatrix} 2 & -1 \\ 0 & 4 \end{bmatrix} x^3 + \begin{bmatrix} -1 & 0 \\ 8 & 3 \end{bmatrix} x + \begin{bmatrix} 4 & 5 \\ -7 & 0 \end{bmatrix} = \begin{bmatrix} 2x^3 - x + 4 & -x^3 + 5 \\ 8x - 7 & 4x^3 + 3x \end{bmatrix}.$$

In either form, this type of matrix is called a **matrix polynomial**. In our development, we have considered only matrices with elements in a field, and $\mathcal{F}[x]$ is not a field. But $\mathcal{F}[x]$ is contained in the field $\mathcal{F}(x)$ of all rational functions in x over \mathcal{F} , and we can consider matrix polynomials in this context. For future use, we note that two matrix polynomials are equal if and only if they have equal coefficients of each power of x .

Since powers of A commute with scalars and with each other, multiplication is commutative for polynomials of the first type: $p(A)q(A) = q(A)p(A)$ for any $p(x), q(x)$ in $\mathcal{F}[x]$. In this case, factorizations in $\mathcal{F}[x]$ remain valid: If $p(x) = g(x)h(x)$ in $\mathcal{F}[x]$, then $p(A) = g(A)h(A)$. A certain amount of care must be exercised when working with polynomials of the second type, since multiplication is not commutative then. Factorizations in $\mathcal{F}[x]$ no longer remain valid here. For example, if $AB \neq BA$, then

$$\begin{aligned}(Ax - B)(Ax + B) &= A^2x^2 + (AB - BA)x - B^2 \\ &\neq A^2x^2 - B^2.\end{aligned}$$

We recall from Chapter 4 that the set $\mathcal{F}_{n \times n}$ of all $n \times n$ matrices over \mathcal{F} is a vector space over \mathcal{F} . If E_{ij} denotes the $n \times n$ matrix that has the element in row i , column j equal to 1 and all other elements 0, then an arbitrary $A = [a_{ij}]$ in $\mathcal{F}_{n \times n}$ can be written uniquely as

$$A = \sum_{i=1}^n \sum_{j=1}^n a_{ij} E_{ij},$$

and the set $\{E_{ij}\}$ is a basis of $\mathcal{F}_{n \times n}$. Hence $\mathcal{F}_{n \times n}$ has dimension n^2 , and the set

$$I = A^0, A, A^2, \dots, A^{n^2}$$

is linearly dependent for any $n \times n$ matrix A over \mathcal{F} . That is, there is a polynomial $p(x) = \sum_{i=0}^{n^2} c_i x^i$ such that $p(A) = \mathbf{0}$. This means that there exists a monic polynomial $m(x)$ of smallest degree such that $m(A) = \mathbf{0}$.

Theorem 10.13 *Let A be an $n \times n$ matrix, and let $m(x)$ be a monic polynomial over \mathcal{F} of smallest degree such that $m(A) = \mathbf{0}$. For any $p(x)$ in $\mathcal{F}[x]$, $p(A) = \mathbf{0}$ if and only if $m(x)$ is a factor of $p(x)$.*

Proof. If $p(x) = m(x)h(x)$, then $p(A) = m(A)h(A) = \mathbf{0}$. Suppose conversely that $p(A) = \mathbf{0}$. Let $q(x)$ and $r(x)$ be the quotient and remainder upon division of $p(x)$ by $m(x)$:

$$p(x) = m(x)q(x) + r(x),$$

where either $r(x)$ is zero, or the degree of $r(x)$ is less than that of $m(x)$. Then

$$r(A) = m(A)q(A) + r(A) = p(A) = \mathbf{0}.$$

This means that $r(x)$ must be the zero polynomial, since otherwise $r(A) = \mathbf{0}$ would contradict the choice of $m(x)$ as having the smallest possible degree such that $m(A) = \mathbf{0}$. Hence $m(x)$ is a factor of $p(x)$. ■

This theorem makes it easy to prove that the polynomial $m(x)$ in the hypothesis is unique (see Problem 6).

Definition 10.14 Let A be an $n \times n$ matrix over \mathcal{F} . The unique monic polynomial $m(x)$ of smallest degree such that $m(A) = \mathbf{0}$ is called the **minimal polynomial** of A . The **minimal polynomial** of a linear operator T on \mathbf{V} is by definition the same as the minimal polynomial of any matrix that represents T .

Since a linear operator T and any matrix that represents it satisfy the same polynomial equations, the minimal polynomial of T is well-defined.

The discussion preceding Theorem 10.13 is an efficient argument for the existence of the minimal polynomial $m(x)$ of A , but it suggests no practical method for finding $m(x)$. The next theorem is a great help in this direction. It is known as the **Hamilton-Cayley theorem**.

Theorem 10.15 Let $A = [a_{ij}]$ be an arbitrary $n \times n$ matrix over \mathcal{F} , and let $f(x)$ be the characteristic polynomial of A . Then $f(A) = \mathbf{0}$.

Proof. Since the characteristic matrix

$$A - xI = [a_{ij} - x\delta_{ij}]$$

is a matrix with polynomials as elements, the minor of each $a_{ij} - x\delta_{ij}$ is the determinant of an $(n-1) \times (n-1)$ matrix with polynomial elements. Hence each element in $B = \text{adj}(A - xI)$ is a polynomial in x . Moreover,

$$\begin{aligned} B(A - xI) &= \text{adj}(A - xI) \cdot (A - xI) \\ &= \det(A - xI)I \\ &= f(x)I. \end{aligned}$$

If $f(x) = c_nx^n + \cdots + c_1x + c_0$, then

$$\begin{aligned} f(A) - f(x)I &= (c_nA^n + \cdots + c_1A + c_0I) - (c_nIx^n + \cdots + c_1Ix + c_0I) \\ &= c_n(A^n - Ix^n) + \cdots + c_2(A^2 - Ix^2) + c_1(A - Ix). \end{aligned}$$

The usual properties of matrix multiplication and matrix addition yield the factorization $A^k - Ix^k = P_k(A - Ix)$, where

$$P_k = Ix^{k-1} + Ax^{k-2} + \cdots + A^{k-2}x + A^{k-1}$$

is a matrix polynomial of degree $k-1$. Hence

$$f(A) - f(x)I = \sum_{k=1}^n c_k P_k(A - Ix),$$

and

$$\begin{aligned} f(A) &= f(x)I + \sum_{k=1}^n c_k P_k(A - Ix) \\ &= \text{adj}(A - Ix) \cdot (A - Ix) + \sum_{k=1}^n c_k P_k(A - Ix) \\ &= \left\{ \text{adj}(A - Ix) + \sum_{k=1}^n c_k P_k \right\} (A - Ix). \end{aligned}$$

The matrix inside the braces is a matrix polynomial, say $B_t x^t + \cdots + B_1 x + B_0$. This gives

$$\begin{aligned} f(A) &= (B_t x^t + \cdots + B_1 x + B_0)(A - Ix) \\ &= -B_t x^{t+1} + (B_t A - B_{t-1})x^t + \cdots + (B_1 A - B_0)x + B_0 A. \end{aligned}$$

Since two matrix polynomials are equal if and only if they have corresponding coefficients that are equal, this requires that $f(A) = B_0 A$ and the coefficients of each positive power of x on the right side be zero:

$$\begin{aligned} B_t &= \mathbf{0} \\ B_t A - B_{t-1} &= \mathbf{0} \\ &\vdots \\ B_2 A - B_1 &= \mathbf{0} \\ B_1 A - B_0 &= \mathbf{0}. \end{aligned}$$

Solving for $B_t, B_{t-1}, \dots, B_1, B_0$ in order we find

$$\mathbf{0} = B_t = B_{t-1} = \cdots = B_1 = B_0.$$

But this gives $f(A) = B_0 A = \mathbf{0} \cdot A = \mathbf{0}$, and the theorem is proved. ■

The minimal polynomial of a diagonalizable operator T can be used to obtain the projections for a spectral decomposition. Since the existence of a spectral decomposition requires that the eigenvalues of T be in \mathcal{F} , we assume for the remainder of this chapter that the characteristic polynomial $f(x)$ of T factors into linear factors over \mathcal{F} :

$$f(x) = (-1)^n (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_r)^{m_r}, \quad (10.1)$$

where the m_j are the algebraic multiplicities of the distinct eigenvalues λ_j . Since the minimal polynomial $m(x)$ divides $f(x)$, this means that

$$m(x) = (x - \lambda_1)^{t_1} (x - \lambda_2)^{t_2} \cdots (x - \lambda_r)^{t_r}. \quad (10.2)$$

We shall adopt the notation of this paragraph for the characteristic polynomial $f(x)$ and the minimal polynomial $m(x)$ of T throughout the remainder of the chapter.

There are some further notational conventions that are essential. For each factor $(x - \lambda_j)^{t_j}$ of $m(x)$, let $q_j(x)$ be the polynomial

$$\begin{aligned} q_j(x) &= \frac{m(x)}{(x - \lambda_j)^{t_j}} \\ &= (x - \lambda_1)^{t_1} \cdots (x - \lambda_{j-1})^{t_{j-1}} (x - \lambda_{j+1})^{t_{j+1}} \cdots (x - \lambda_r)^{t_r}. \end{aligned}$$

Any nonconstant common divisor of $q_1(x), q_2(x), \dots, q_r(x)$ would necessarily have a linear factor of the form $x - \lambda_i$ since the only linear factors of each $q_j(x)$ are of this type. But $x - \lambda_i$ is not a factor of $q_i(x)$, so there are no nonconstant common divisors of $q_1(x), q_2(x), \dots, q_r(x)$. That is, the greatest common divisor of $q_1(x), q_2(x), \dots, q_r(x)$ is 1. Hence there are polynomials $g_1(x), g_2(x), \dots, g_r(x)$ such that

$$1 = g_1(x)q_1(x) + g_2(x)q_2(x) + \cdots + g_r(x)q_r(x).$$

Since $q_j(x) = \frac{m(x)}{(x - \lambda_j)^{t_j}}$, this yields

$$\frac{1}{m(x)} = \frac{g_1(x)}{(x - \lambda_1)^{t_1}} + \frac{g_2(x)}{(x - \lambda_2)^{t_2}} + \cdots + \frac{g_r(x)}{(x - \lambda_r)^{t_r}}.$$

Let $p_j(x) = g_j(x)q_j(x)$ for $j = 1, 2, \dots, r$. The set of polynomials $p_1(x), p_2(x), \dots, p_r(x)$ has the following properties:

- (i) $p_i(x)p_j(x)$ has $m(x)$ as a factor if $i \neq j$, say $p_i(x)p_j(x) = h_{ij}(x)m(x)$
- (ii) $1 = p_1(x) + p_2(x) + \cdots + p_r(x)$.

Theorem 10.16 *Let $F_j = p_j(T)$, and let \mathcal{K}_j denote the kernel of $(T - \lambda_j)^{t_j}$ for $j = 1, 2, \dots, r$. Then $\{F_1, F_2, \dots, F_r\}$ is a complete set of projections for \mathbf{V} , and F_j is the projection of \mathbf{V} onto \mathcal{K}_j along $\sum_{i \neq j} \mathcal{K}_i$.*

Proof. Since $p_i(x)p_j(x) = h_{ij}(x)m(x)$ for $i \neq j$ and $m(T) = Z$, we have

$$F_i F_j = p_i(T)p_j(T) = h_{ij}(T)m(T) = Z$$

whenever $i \neq j$. It follows from (ii) above that $1 = F_1 + F_2 + \cdots + F_r$. Thus

$$F_j = F_j \left(\sum_{i=1}^r F_i \right) = \sum_{i=1}^r F_j F_i = F_j^2,$$

and $\{F_1, F_2, \dots, F_r\}$ is an orthogonal set of projections such that $1 = F_1 + F_2 + \cdots + F_r$. We shall show that F_j is the projection of \mathbf{V} onto \mathcal{K}_j along $\sum_{i \neq j} \mathcal{K}_i$, and it will follow that $F_j \neq Z$ since \mathcal{K}_j clearly contains the eigenspace \mathbf{V}_{λ_j} of T .

Now

$$\begin{aligned} (x - \lambda_j)^{t_j} p_j(x) &= (x - \lambda_j)^{t_j} g_j(x)q_j(x) \\ &= (x - \lambda_j)^{t_j} g_j(x) \frac{m(x)}{(x - \lambda_j)^{t_j}} \\ &= g_j(x)m(x), \end{aligned}$$

so $(T - \lambda_j)^{t_j} F_j = (T - \lambda_j)^{t_j} p_j(T) = g_j(T)m(T) = Z$. Hence $(T - \lambda_j)^{t_j} F_j(\mathbf{v}) = \mathbf{0}$ for all \mathbf{v} , and $F_j(\mathbf{V}) \subseteq \mathcal{K}_j$. Now let $\mathbf{v} \in \mathcal{K}_j$, so that $(T - \lambda_j)^{t_j}(\mathbf{v}) = \mathbf{0}$. For $i \neq j$, $(T - \lambda_j)^{t_j}$ is a factor of $F_i = p_i(T)$ since $(x - \lambda_j)^{t_j}$ is a factor of $p_i(x)$. Hence $F_i(\mathbf{v}) = \mathbf{0}$ for all $i \neq j$, and

$$\mathbf{v} = (F_1 + F_2 + \cdots + F_r)(\mathbf{v}) = F_j(\mathbf{v}).$$

Thus, $\mathbf{v} \in F_j(\mathbf{V})$, and $\mathcal{K}_j = F_j(\mathbf{V})$. This completes the proof of the theorem. ■

Definition 10.17 A subspace \mathbf{W} of \mathbf{V} is **invariant under T** , or **T -invariant**, if

$$T(\mathbf{W}) \subseteq \mathbf{W}.$$

That is, a subspace \mathbf{W} of \mathbf{V} is T -invariant if and only if $T(\mathbf{v}) \in \mathbf{W}$ for all $\mathbf{v} \in \mathbf{W}$. Every linear operator T on \mathbf{V} has invariant subspaces. The eigenspaces \mathbf{V}_λ of T are invariant under T , as are the zero subspace and \mathbf{V} .

If \mathbf{W} is a subspace of \mathbf{V} that is invariant under T , then T induces a linear transformation $T_{\mathbf{w}}$ of \mathbf{W} into \mathbf{W} defined by $T_{\mathbf{w}}(\mathbf{v}) = T(\mathbf{v})$ for all $\mathbf{v} \in \mathbf{W}$. That is, as long as $\mathbf{v} \in \mathbf{W}$, $T_{\mathbf{w}}$ and T map \mathbf{v} onto the same vector. The distinction between $T_{\mathbf{w}}$ and T is that $T_{\mathbf{w}}$ is defined only on \mathbf{W} . The transformation $T_{\mathbf{w}}$ is called the **restriction of T to \mathbf{W}** .

Theorem 10.18 Let \mathcal{K}_j be the kernel of $(T - \lambda_j)^{t_j}$ for $j = 1, 2, \dots, r$. Then

- (a) $\mathbf{V} = \mathcal{K}_1 \oplus \mathcal{K}_2 \oplus \cdots \oplus \mathcal{K}_r$;
- (b) each \mathcal{K}_j is invariant under T ;
- (c) if T_j denotes the restriction of T to \mathcal{K}_j , then the minimal polynomial of T_j is $(x - \lambda_j)^{t_j}$.

Proof. Consider the complete set of projections $\{F_1, F_2, \dots, F_r\}$ where $F_j = p_j(T)$. By Theorem 10.16, F_j is the projection of \mathbf{V} onto $\mathcal{K}_j = F_j(\mathbf{V})$ along $\sum_{i \neq j} \mathcal{K}_i$. Thus, $\mathbf{V} = \mathcal{K}_1 \oplus \mathcal{K}_2 \oplus \cdots \oplus \mathcal{K}_r$ is the direct sum decomposition corresponding to this complete set of projections.

To establish (b), let $\mathbf{v} \in \mathcal{K}_j$ and consider $T(\mathbf{v})$. Since $(T - \lambda_j)^{t_j}$ is a polynomial in T , it commutes with T to yield

$$(T - \lambda_j)^{t_j}(T(\mathbf{v})) = T((T - \lambda_j)^{t_j}(\mathbf{v})) = T(\mathbf{0}) = \mathbf{0}.$$

Hence $T(\mathbf{v})$ is in \mathcal{K}_j , and \mathcal{K}_j is invariant under T .

Now $(T_j - \lambda_j)^{t_j}$ is the zero transformation of \mathcal{K}_j since $(T - \lambda_j)^{t_j}(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in \mathcal{K}_j$. Therefore, the minimal polynomial of T_j divides $(x - \lambda_j)^{t_j}$. That is, the minimal polynomial of T_j is $(x - \lambda_j)^s$ for some $s \leq t_j$. Suppose that $s < t_j$. Now each $\mathbf{v} \in \mathbf{V}$ can be written uniquely as

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r$$

with $\mathbf{v}_i \in \mathcal{K}_i$. Since

$$(T - \lambda_i)^{t_i}(\mathbf{v}_i) = \mathbf{0}$$

for each i , then

$$q_j(T)(\mathbf{v}) = q_j(T)(\mathbf{v}_j).$$

And since

$$(T - \lambda_j)^s(\mathbf{v}_j) = (T_j - \lambda_j)^s(\mathbf{v}_j) = \mathbf{0}$$

for all $\mathbf{v}_j \in \mathcal{K}_j$, this means that $p(T) = (T - \lambda_j)^s q_j(T)$ is the zero transformation on \mathbf{V} . But this is a contradiction, since $p(x) = (x - \lambda_j)^s q_j(x)$ has degree less than that of $m(x) = (x - \lambda_j)^{t_j} q_j(x)$. Therefore, $s = t_j$. ■

Theorem 10.19 *The linear operator T on \mathbf{V} has a spectral decomposition if and only if the minimal polynomial $m(x)$ of T has the form*

$$m(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_r),$$

where $\lambda_1, \lambda_2, \dots, \lambda_r$ are the distinct eigenvalues of T .

Proof. If the minimal polynomial $m(x)$ had the given form, then each $t_i = 1$ in equation (10.2). Hence $\mathcal{K}_i = \mathbf{V}_{\lambda_i}$ for $i = 1, 2, \dots, r$, and

$$\mathbf{V} = \mathbf{V}_{\lambda_1} \oplus \mathbf{V}_{\lambda_2} \oplus \cdots \oplus \mathbf{V}_{\lambda_r}.$$

It follows from the proof of Theorem 7.20 that T is diagonalizable, and therefore T has a spectral decomposition by Theorem 10.8.

Conversely, suppose that T has a spectral decomposition

$$T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r.$$

For any polynomial $p(x)$,

$$p(T) = p(\lambda_1)P_1 + p(\lambda_2)P_2 + \cdots + p(\lambda_r)P_r$$

by Theorem 10.10. If $p(\lambda_i) \neq 0$ for some i , $p(T)$ has $p(\lambda_i)$ as a nonzero eigenvalue by Theorem 10.6. Hence $p(T) = Z$ if and only if $p(\lambda_i) = 0$ for $i = 1, 2, \dots, r$. This implies at once that

$$m(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_r). ■$$

The result that we have been working toward is now at hand.

Theorem 10.20 *If $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$ is a spectral decomposition, then $P_j = p_j(T)$.*

Proof. Let $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$ be a spectral decomposition of T . By Theorem 10.19,

$$m(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_r).$$

That is, each t_j in equation (10.2) is 1, and $\mathcal{K}_j = \mathbf{V}_{\lambda_j}$ for each j . With $F_j = p_j(T)$, then $\{F_1, F_2, \dots, F_r\}$ is a complete set of projections for \mathbf{V} , and F_j is the projection of \mathbf{V} onto \mathcal{K}_j along $\sum_{i \neq j} \mathcal{K}_i$ by Theorem 10.16. By Theorem 10.11, P_j is the projection

of \mathbf{V} onto \mathbf{V}_{λ_j} along $\sum_{i \neq j} \mathbf{V}_{\lambda_i}$. Since $\mathcal{K}_j = \mathbf{V}_{\lambda_j}$ for each j , $P_j = F_j = p_j(T)$ for each j , and the proof is complete. ■

The next corollary follows readily from the fact that each projection in a spectral decomposition of T is a polynomial in T .

Corollary 10.21 *If $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$ is a spectral decomposition of T , then a linear operator S on \mathbf{V} commutes with T if and only if S commutes with each P_j .*

Proof. See Problem 9. ■

One immediate and desirable consequence of Theorem 10.20 is that we now have a systematic procedure available for finding the projections in a spectral decomposition. This procedure, of course, is to determine the P_j by use of $P_j = p_j(T)$.

Example 1 □ Let T be the linear transformation of \mathbf{R}^3 that has the matrix A relative to the standard basis where

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

We shall determine whether or not T has a spectral decomposition, and shall obtain such a decomposition if there is one. The characteristic polynomial $f(x)$ is given by

$$f(x) = -(x+1)(x-2)^2.$$

It follows from Theorem 10.19 that T has a spectral decomposition if and only if

$$m(x) = (x+1)(x-2).$$

A simple calculation shows that

$$m(A) = A^2 - A - 2I = \mathbf{0},$$

so T does indeed have a spectral decomposition. With $\lambda_1 = -1, \lambda_2 = 2$, the polynomials $q_j(x)$ are given by $q_1(x) = x - 2, q_2(x) = x + 1$. The partial fraction decomposition

$$\frac{1}{m(x)} = \frac{g_1(x)}{x+1} + \frac{g_2(x)}{x-2}$$

leads to $g_1(x) = -\frac{1}{3}, g_2(x) = \frac{1}{3}$. Hence $p_1(x) = -\frac{1}{3}(x-2)$ and $p_2(x) = \frac{1}{3}(x+1)$. The projections P_1, P_2 thus have respective matrices E_1, E_2 relative to \mathcal{E}_3 given by

$$E_1 = -\frac{1}{3}(A - 2I) = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$E_2 = \frac{1}{3}(A + I) = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

It is easily checked that $E_1^2 = E_1$, $E_2^2 = E_2$, and $A = \lambda_1 E_1 + \lambda_2 E_2$. ■

In case T does not have a spectral decomposition, a somewhat similar decomposition can be obtained that in many cases is quite useful. We have seen that the mappings $P_j = F_j = p_j(T)$ form a complete set of projections $\{P_1, P_2, \dots, P_r\}$ for \mathbf{V} . Consider the linear operator

$$D = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r,$$

where $\lambda_1, \lambda_2, \dots, \lambda_r$ are the distinct eigenvalues of T . Since D has a spectral decomposition, it is diagonalizable. Let $N = T - D$. Using the representation

$$T = T(P_1 + P_2 + \cdots + P_r) = TP_1 + TP_2 + \cdots + TP_r$$

for T, N is given by

$$N = (T - \lambda_1)P_1 + (T - \lambda_2)P_2 + \cdots + (T - \lambda_r)P_r.$$

It is left as an exercise (Problem 10) to prove that

$$N^k = (T - \lambda_1)^k P_1 + (T - \lambda_2)^k P_2 + \cdots + (T - \lambda_r)^k P_r$$

for all positive integers k . Recall from the proof of Theorem 10.16 that $(T - \lambda_j)^{t_j} P_j = Z$, and thus $N^k = Z$ if $k \geq t_j$ for all j .

Definition 10.22 A linear operator T is called **nilpotent** if $T^k = Z$ for some positive integer k . The smallest positive integer t such that $T^t = Z$ is the **index** of nilpotency of T .

Theorem 10.23 If the minimal polynomial of T factors into a product of linear factors (not necessarily distinct) over \mathcal{F} , then T can be written as the sum $T = D + N$ of a diagonalizable transformation D and a nilpotent transformation N , where D and N are polynomials in T , and consequently commute.

Proof. Let D and N be as given in the paragraph preceding Definition 10.22. From the expressions there for D and N and the fact that $P_j = p_j(T)$, it is reasonably clear that D and N are polynomials in T and hence commute. Nevertheless, we furnish some additional details. The projections P_j commute with T since $P_j = p_j(T)$. Now

$$\begin{aligned} ND &= \left(\sum_{i=1}^r (T - \lambda_i)P_i \right) \left(\sum_{j=1}^r \lambda_j P_j \right) \\ &= \sum_{i=1}^r \sum_{j=1}^r (T - \lambda_i)\lambda_j P_i P_j, \end{aligned}$$

and

$$\begin{aligned} DN &= \left(\sum_{j=1}^r \lambda_j P_j \right) \left(\sum_{i=1}^r (T - \lambda_i) P_i \right) \\ &= \sum_{i=1}^r \sum_{j=1}^r \lambda_j P_j (T - \lambda_i) P_i. \end{aligned}$$

But

$$\lambda_j P_j (T - \lambda_i) P_i = (T - \lambda_i) \lambda_j P_i P_j,$$

so $DN = ND$. ■

In case \mathbf{V} is a unitary space, the hypothesis of Theorem 10.23 is automatically satisfied, and every linear operator on \mathbf{V} is the sum of a diagonalizable and a nilpotent transformation.

Spectral decompositions of linear operators have the usual parallel statements for matrices. If T has matrix A relative to a certain basis and the projections P_i in $T = \lambda_1 P_1 + \lambda_2 P_2 + \dots + \lambda_r P_r$ have matrices E_i for $i = 1, 2, \dots, r$, then

$$A = \lambda_1 E_1 + \lambda_2 E_2 + \dots + \lambda_r E_r$$

is called a **spectral decomposition** for A . The E_i are called **projection matrices** or **principal idempotents** for A . It is left as an exercise to prove that $AE_i = \lambda_i E_i$.

If T , D , and N in $T = D+N$ have matrices A , B , and C , respectively, then $A = B+C$. The matrix B is diagonalizable, and C is called a **nilpotent matrix**.

Exercises 10.4

1. Verify the Hamilton-Cayley theorem for each matrix A .

$$(a) A = \begin{bmatrix} -4 & -3 & -1 \\ -4 & 0 & -4 \\ 8 & 4 & 5 \end{bmatrix} \quad (b) A = \begin{bmatrix} 8 & 5 & -5 \\ 5 & 8 & -5 \\ 15 & 15 & -12 \end{bmatrix}$$

$$(c) A = \begin{bmatrix} 8 & 5 & 6 & 0 \\ 0 & -2 & 0 & 0 \\ -10 & -5 & -8 & 0 \\ 2 & 1 & 1 & 2 \end{bmatrix} \quad (d) A = \begin{bmatrix} 7 & 3 & 3 & 2 \\ 0 & 1 & 2 & -4 \\ -8 & -4 & -5 & 0 \\ 2 & 1 & 1 & 3 \end{bmatrix}$$

2. Find the minimal polynomial of each matrix A in Problem 1.
3. Determine all real 2×2 matrices A such that $A^2 = -I$.
4. Let T be the linear operator on \mathbf{R}^n that has the given matrix A relative to the standard basis \mathcal{E}_n . Find the spectral decomposition of T .

$$\begin{array}{ll}
 \text{(a)} \ A = \begin{bmatrix} 8 & 5 & -5 \\ 5 & 8 & -5 \\ 15 & 15 & -12 \end{bmatrix} & \text{(b)} \ A = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{bmatrix} \\
 \text{(c)} \ A = \begin{bmatrix} 7 & 3 & 3 & 2 \\ 0 & 1 & 2 & -4 \\ -8 & -4 & -5 & 0 \\ 2 & 1 & 2 & 3 \end{bmatrix} & \text{(d)} \ A = \begin{bmatrix} 4 & -1 & 0 & 1 \\ -1 & 5 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 1 & 0 & -1 & 5 \end{bmatrix}
 \end{array}$$

5. Write the given matrix A as the sum of a diagonalizable matrix and a nilpotent matrix.

$$\begin{array}{ll}
 \text{(a)} \ A = \begin{bmatrix} 7 & 3 & 3 & 2 \\ 0 & 1 & 2 & -4 \\ -8 & -4 & -5 & 0 \\ 2 & 1 & 1 & 3 \end{bmatrix} & \text{(b)} \ A = \begin{bmatrix} 8 & 5 & 6 & 0 \\ 0 & -2 & 0 & 0 \\ -10 & -5 & -8 & 0 \\ 2 & 1 & 1 & 2 \end{bmatrix} \\
 \text{(c)} \ A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ -1 & 0 & 2 & 1 \end{bmatrix} &
 \end{array}$$

6. Prove that the polynomial $m(x)$ in the hypothesis of Theorem 10.13 is unique.
7. Prove that, for an arbitrary square matrix A , A and A^T satisfy the same polynomial equations with scalar coefficients.
8. Prove that similar matrices have the same minimal polynomial.
9. Prove Corollary 10.21.
10. Let $N = (T - \lambda_1)P_1 + (T - \lambda_2)P_2 + \cdots + (T - \lambda_r)P_r$, where $\{P_1, P_2, \dots, P_r\}$ is a complete set of projections. Prove that

$$N^k = (T - \lambda_1)^k P_1 + (T - \lambda_2)^k P_2 + \cdots + (T - \lambda_r)^k P_r$$

for all positive integers k .

11. Let $\{P_1, P_2, \dots, P_r\}$ be the complete set of projections determined by $P_j = p_j(T)$. With $T = TP_1 + TP_2 + \cdots + TP_r$, show that $(TP_i) \cdot (TP_j) = Z$ if $i \neq j$, and that $p(T) = p(TP_1) + p(TP_2) + \cdots + p(TP_r)$ for any polynomial $p(x)$ with zero constant term.

12. Let $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r$ be a spectral decomposition for the linear operator T on \mathbf{V} . Prove that a linear operator S on \mathbf{V} commutes with T if and only if every \mathcal{K}_j is invariant under S .
13. Prove that if $A = \lambda_1 E_1 + \lambda_2 E_2 + \cdots + \lambda_r E_r$ is a spectral decomposition for A , then $AE_i = \lambda_i E_i$ for each i .

10.5 Nilpotent Transformations

We have seen in the preceding section that if the minimum polynomial of T factors into linear factors over \mathcal{F} , then T can be written as $T = D + N$, where D is diagonalizable and N is nilpotent. The degree of simplicity that can be obtained in a matrix representation of T thus depends on the type of representation that is possible for N . The representation of T that is accepted as being simplest, or nearest to a diagonal form, is the Jordan canonical form. The derivation of the Jordan canonical form is the principal objective of the remainder of this chapter. This derivation hinges on an investigation of the invariant subspaces of an operator.

Suppose that \mathbf{W} is an invariant subspace of \mathbf{V} under T . Let $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a basis of \mathbf{W} , and extend this set to a basis $\mathcal{A} = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$ of \mathbf{V} . Since \mathbf{W} is invariant under T , $T(\mathbf{v}_j) = \sum_{i=1}^k a_{ij} \mathbf{v}_i$ for $j = 1, 2, \dots, k$. Hence the matrix of T relative to \mathcal{A} is of the form

$$A = \begin{bmatrix} A_1 & A_3 \\ \mathbf{0} & A_2 \end{bmatrix},$$

where A_1 is the matrix of $T_{\mathbf{W}}$ relative to $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$.

If $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2$, where each of \mathbf{W}_1 and \mathbf{W}_2 is invariant under T , and if the basis $\mathcal{A} = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$ of \mathbf{V} is chosen so that $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a basis of \mathbf{W}_1 and $\{\mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$ is a basis of \mathbf{W}_2 , then we also have $T(\mathbf{v}_j) = \sum_{i=k+1}^n a_{ij} \mathbf{v}_i$ for $j = k+1, \dots, n$. Hence $A_3 = \mathbf{0}$, and the matrix of T relative to \mathcal{A} is of the form

$$A = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{bmatrix},$$

where A_i is the matrix of the restriction of T to \mathbf{W}_i .

The results of the preceding paragraph generalize readily to direct sums with more than two terms. If $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_r$ are T -invariant subspaces of \mathbf{V} such that \mathbf{V} is the direct sum $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \cdots \oplus \mathbf{W}_r$, and if the basis

$$\mathcal{B} = \{\mathbf{u}_{11}, \dots, \mathbf{u}_{1n_1}, \mathbf{u}_{21}, \dots, \mathbf{u}_{2n_2}, \dots, \mathbf{u}_{r1}, \dots, \mathbf{u}_{rn_r}\}$$

of \mathbf{V} is such that $\{\mathbf{u}_{i1}, \dots, \mathbf{u}_{in_i}\}$ is a basis of \mathbf{W}_i , then the matrix of T relative to \mathcal{B} is

of the form

$$A = \begin{bmatrix} A_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_r \end{bmatrix},$$

where A_i is the matrix of the restriction of T to \mathbf{W}_i . A matrix such as this A is called a **diagonal block matrix**.

The type of invariant subspace that we shall be mainly concerned with is the cyclic subspace, to be defined shortly.

Let \mathbf{v} be any nonzero vector in \mathbf{V} . Since a linearly independent subset of \mathbf{V} can have at most n elements, there is a unique positive integer k such that $\{\mathbf{v}, T(\mathbf{v}), \dots, T^{k-1}(\mathbf{v})\}$ is linearly independent and $T^k(\mathbf{v})$ is dependent on $\{\mathbf{v}, T(\mathbf{v}), \dots, T^{k-1}(\mathbf{v})\}$. Then there are scalars a_0, a_1, \dots, a_{k-1} such that $T^k(\mathbf{v}) = \sum_{i=0}^{k-1} a_i T^i(\mathbf{v})$. It follows from this that the subspace

$$\langle \mathbf{v}, T(\mathbf{v}), \dots, T^{k-1}(\mathbf{v}) \rangle$$

is T -invariant (Problem 3).

Definition 10.24 Let \mathbf{v} be a nonzero vector in \mathbf{V} , and let k be the unique positive integer such that $\{\mathbf{v}, T(\mathbf{v}), \dots, T^{k-1}(\mathbf{v})\}$ is linearly independent and $T^k(\mathbf{v})$ is dependent on $\{\mathbf{v}, T(\mathbf{v}), \dots, T^{k-1}(\mathbf{v})\}$. The subspace $\langle \mathbf{v}, T(\mathbf{v}), \dots, T^{k-1}(\mathbf{v}) \rangle$ is denoted by $\mathbf{C}(\mathbf{v}, T)$ and is called the **cyclic subspace of \mathbf{v} relative to T** , or the **cyclic subspace generated by \mathbf{v} under T** . The particular basis $\{T^{k-1}(\mathbf{v}), \dots, T(\mathbf{v}), \mathbf{v}\}$ of $\mathbf{C}(\mathbf{v}, T)$ is called the **cyclic basis generated by \mathbf{v} under T** and is denoted by $\mathcal{B}(\mathbf{v}, T)$.

The arrangement of the elements of $\mathcal{B}(\mathbf{v}, T)$ according to descending rather than ascending powers of T seems a bit unnatural, but this is done so that the restriction of T to $\mathbf{C}(\mathbf{v}, T)$ will have a special form of matrix relative to $\mathcal{B}(\mathbf{v}, T)$. Since

$$T(T^{k-1}(\mathbf{v})) = T^k(\mathbf{v}) = a_{k-1}T^{k-1}(\mathbf{v}) + \cdots + a_1T(\mathbf{v}) + a_0\mathbf{v}$$

and $T(T^j(\mathbf{v})) = T^{j+1}(\mathbf{v})$ for $j = 0, 1, \dots, k-2$, the restriction of T to $\mathbf{C}(\mathbf{v}, T)$ has the matrix

$$A = \begin{bmatrix} a_{k-1} & 1 & 0 & \cdots & 0 \\ a_{k-2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & 0 & 0 & \cdots & 1 \\ a_0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

relative to $\mathcal{B}(\mathbf{v}, T)$. The special form that this matrix A takes on for a nilpotent transformation T is especially useful.

Theorem 10.25 *If T is a nilpotent transformation of \mathbf{V} , then the restriction of T to $\mathbf{C}(\mathbf{v}, T)$ has the matrix*

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

relative to $\mathcal{B}(\mathbf{v}, T)$.

Proof. With the notation of the paragraph just before the theorem, we shall show that $T^k(\mathbf{v}) = \mathbf{0}$. Our proof of this fact is an “overkill”: We show that the restriction $T_{\mathbf{c}}$ of T to $\mathbf{C}(\mathbf{v}, T)$ has minimal polynomial x^k .

Since T is nilpotent on \mathbf{V} , $T_{\mathbf{c}}$ is nilpotent, say with index s . The minimal polynomial of $T_{\mathbf{c}}$ must therefore divide x^s . Let x^r denote this minimal polynomial. Since $T^{k-1}(\mathbf{v})$ is in the basis $\mathcal{B}(\mathbf{v}, T)$, then $r \geq k$.

Consider the polynomial $p(x) = a_0 + a_1x + \cdots + a_{k-1}x^{k-1} - x^k$. We shall show that

$$S = p(T_{\mathbf{c}}) = a_0 + a_1T_{\mathbf{c}} + \cdots + a_{k-1}T_{\mathbf{c}}^{k-1} - T_{\mathbf{c}}^k$$

is the zero transformation of $\mathbf{C}(\mathbf{v}, T)$. To do this, it is sufficient to show that S takes on the value $\mathbf{0}$ at each vector in $\mathcal{B}(\mathbf{v}, T)$. Since $T_{\mathbf{c}}(\mathbf{v}) = T(\mathbf{v})$,

$$S(\mathbf{v}) = a_0\mathbf{v} + a_1T(\mathbf{v}) + \cdots + a_{k-1}T^{k-1}(\mathbf{v}) - T^k(\mathbf{v}) = \mathbf{0}$$

by the choice of the scalars a_i . For $j = 1, \dots, k-1$, we have

$$\begin{aligned} S(T^j(\mathbf{v})) &= a_0T^j(\mathbf{v}) + \cdots + a_{k-1}T^{k-1}(T^j(\mathbf{v})) - T^k(T^j(\mathbf{v})) \\ &= T^j(a_0\mathbf{v} + \cdots + a_{k-1}T^{k-1}(\mathbf{v}) - T^k(\mathbf{v})) \\ &= T^j(\mathbf{0}) \\ &= \mathbf{0}. \end{aligned}$$

Hence $S = p(T_{\mathbf{c}})$ is the zero transformation, and x^r divides $p(x)$. This requires that $r \leq k$. Therefore $r = k$ and $p(x) = x^r = x^k$. ■

Our principal concern is with the connection between cyclic subspaces and the kernels of the powers of a nilpotent transformation. Before proceeding to our main result in this direction, some preliminary lemmas are in order. These lemmas by themselves are of little consequence, but they are essential to the proof of Theorem 10.28.

Lemma 10.26 *Suppose that T is a nilpotent transformation of index t on \mathbf{V} , and let \mathbf{W}_j be the kernel of T^j . Then $\{\mathbf{0}\} = \mathbf{W}_0 \subseteq \mathbf{W}_1 \subseteq \cdots \subseteq \mathbf{W}_t = \mathbf{V}$, and $\mathbf{W}_{j-1} \neq \mathbf{W}_j$ for $j = 1, 2, \dots, t$.*

Proof. The equality $\{\mathbf{0}\} = \mathbf{W}_0$ follows from the fact that T^0 is the identity transformation, and $\mathbf{W}_t = \mathbf{V}$ since T has index t on \mathbf{V} .

Since $T^{j-1}(\mathbf{v}) = \mathbf{0}$ implies $T^j(\mathbf{v}) = T(\mathbf{0}) = \mathbf{0}$, $\mathbf{W}_{j-1} \subseteq \mathbf{W}_j$ for $j = 1, 2, \dots, t$. Let \mathbf{u} be a vector in \mathbf{V} such that $T^{t-1}(\mathbf{u}) \neq \mathbf{0}$, and let $1 \leq j \leq t$. Then $T^{t-j}(\mathbf{u})$ is in \mathbf{W}_j since $T^j(T^{t-j}(\mathbf{u})) = \mathbf{0}$, and $T^{t-j}(\mathbf{u})$ is not in \mathbf{W}_{j-1} since $T^{j-1}(T^{t-j}(\mathbf{u})) = T^{t-1}(\mathbf{u}) \neq \mathbf{0}$. Thus $\mathbf{W}_{j-1} \neq \mathbf{W}_j$. ■

Lemma 10.27 *Let the subspaces \mathbf{W}_i be as in Lemma 10.26, with $j \geq 2$. Suppose that $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a basis of \mathbf{W}_{j-2} , and that $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ is a linearly independent set of vectors in \mathbf{W}_j such that*

$$\langle \mathbf{w}_1, \dots, \mathbf{w}_r \rangle \cap \mathbf{W}_{j-1} = \{\mathbf{0}\}.$$

Then $\{\mathbf{u}_1, \dots, \mathbf{u}_k, T(\mathbf{w}_1), \dots, T(\mathbf{w}_r)\}$ is a linearly independent subset of \mathbf{W}_{j-1} .

Proof. Since $T^{j-1}(T(\mathbf{w}_i)) = T^j(\mathbf{w}_i) = \mathbf{0}$ for each i , the set

$$\{\mathbf{u}_1, \dots, \mathbf{u}_k, T(\mathbf{w}_1), \dots, T(\mathbf{w}_r)\}$$

is contained in \mathbf{W}_{j-1} . Assume that

$$\sum_{i=1}^k b_i \mathbf{u}_i + \sum_{m=1}^r c_m T(\mathbf{w}_m) = \mathbf{0}.$$

Then

$$\sum_{m=1}^r c_m T(\mathbf{w}_m) = - \sum_{i=1}^k b_i \mathbf{u}_i \text{ is in } \mathbf{W}_{j-2},$$

so that

$$T^{j-2} \left(\sum_{m=1}^r c_m T(\mathbf{w}_m) \right) = \mathbf{0}.$$

This implies that $T^{j-1}(\sum_{m=1}^r c_m \mathbf{w}_m) = \mathbf{0}$, and therefore $\sum_{m=1}^r c_m \mathbf{w}_m$ is in \mathbf{W}_{j-1} . Since

$$\langle \mathbf{w}_1, \dots, \mathbf{w}_r \rangle \cap \mathbf{W}_{j-1} = \{\mathbf{0}\},$$

it must be that $\sum_{m=1}^r c_m \mathbf{w}_m = \mathbf{0}$. Hence each $c_m = 0$, and this implies $\sum_{i=1}^k b_i \mathbf{u}_i = \mathbf{0}$. This requires in turn that all $b_i = 0$ since $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is linearly independent. Therefore, $\{\mathbf{u}_1, \dots, \mathbf{u}_k, T(\mathbf{w}_1), \dots, T(\mathbf{w}_r)\}$ is linearly independent. ■

The **superdiagonal elements** in a matrix $A = [a_{ij}]_{m \times n}$ are those elements $a_{i,i+1}$ for $i = 1, 2, \dots, m-1$. That is, the superdiagonal elements are those immediately above the diagonal elements.

Theorem 10.28 Let T be a nilpotent operator of index t on \mathbf{V} , let \mathbf{W}_j be the kernel of T^j for $j = 1, 2, \dots, t$, and let $s = \text{nullity}(T)$. There exists a basis \mathcal{B} of \mathbf{V} such that the matrix of T relative to \mathcal{B} is given by

$$A = \begin{bmatrix} A_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_s \end{bmatrix},$$

where each A_i is a square matrix that has all superdiagonal elements 1 and all other elements 0, A_1 is of order t , and $\text{order}(A_i) \geq \text{order}(A_{i+1})$ for each i . The matrix A is uniquely determined by T .

Proof. By Lemma 10.27, there is a basis

$$\mathcal{A} = \{\mathbf{v}_{11}, \dots, \mathbf{v}_{1s_1}, \mathbf{v}_{21}, \dots, \mathbf{v}_{2s_2}, \dots, \mathbf{v}_{t1}, \dots, \mathbf{v}_{ts_t}\}$$

of \mathbf{V} such that $\{\mathbf{v}_{11}, \dots, \mathbf{v}_{js_j}\}$ is a basis of \mathbf{W}_j for $j = 1, \dots, t$. Such a basis can be obtained by extending a basis $\{\mathbf{v}_{11}, \dots, \mathbf{v}_{1s_1}\}$ of \mathbf{W}_1 to a basis

$$\{\mathbf{v}_{11}, \dots, \mathbf{v}_{1s_1}, \mathbf{v}_{21}, \dots, \mathbf{v}_{2s_2}\}$$

of \mathbf{W}_2 , then extending this basis to a basis of \mathbf{W}_3 , and so on. For each j , the elements of \mathcal{A} that are in \mathbf{W}_j but not in \mathbf{W}_{j-1} are precisely those in the segment $\mathbf{v}_{j1}, \dots, \mathbf{v}_{js_j}$. We shall replace each of these segments by a new set of vectors in order to obtain a basis \mathcal{A}' with certain properties. Roughly, the idea is this. The vectors in segment $j-1$ will be replaced by the images of the vectors in the next segment to the right, and then this set will be extended to a basis of \mathbf{W}_{j-1} .

No change is required in the last s_t vectors of \mathcal{A} . That is, we put

$$\mathbf{v}'_{t1} = \mathbf{v}_{t1}, \mathbf{v}'_{t2} = \mathbf{v}_{t2}, \dots, \mathbf{v}'_{ts_t} = \mathbf{v}_{ts_t}.$$

If $\sum_{m=1}^{s_t} c_m \mathbf{v}'_{tm}$ is in \mathbf{W}_{t-1} , then

$$\sum_{m=1}^{s_t} c_m \mathbf{v}'_{tm} = \sum_{i=1}^{t-1} \sum_{j=1}^{s_i} a_{ij} \mathbf{v}_{ij}.$$

The linear independence of \mathcal{A} requires that each $a_{ij} = 0$ and each $c_m = 0$. Hence $\langle \mathbf{v}'_{t1}, \dots, \mathbf{v}'_{ts_t} \rangle \cap \mathbf{W}_{t-1} = \{\mathbf{0}\}$. Let

$$\mathbf{v}'_{t-1,1} = T(\mathbf{v}_{t1}), \mathbf{v}'_{t-1,2} = T(\mathbf{v}_{t2}), \dots, \mathbf{v}'_{t-1,s_t} = T(\mathbf{v}_{ts_t}).$$

By Lemma 10.27, the set

$$\{\mathbf{v}_{11}, \dots, \mathbf{v}_{1s_1}, \dots, \mathbf{v}_{t-2,1}, \dots, \mathbf{v}_{t-2,s_{t-2}}, \mathbf{v}'_{t-1,1}, \dots, \mathbf{v}'_{t-1,s_t}\}$$

is linearly independent, and so can be extended to a basis

$$\{\mathbf{v}_{11}, \dots, \mathbf{v}_{1s_1}, \dots, \mathbf{v}_{t-2,1}, \dots, \mathbf{v}'_{t-1,1}, \dots, \mathbf{v}'_{t-1,s_t}, \dots, \mathbf{v}'_{t-1,s_{t-1}}\}$$

of \mathbf{W}_{t-1} . At this point, the last two segments in the basis

$$\{\mathbf{v}_{11}, \dots, \mathbf{v}_{1s_1}, \dots, \mathbf{v}_{t-2,s_{t-2}}, \mathbf{v}'_{t-1,1}, \dots, \mathbf{v}'_{t-1,s_{t-1}}, \mathbf{v}'_{t1}, \dots, \mathbf{v}'_{ts_t}\}$$

of \mathbf{V} have the desired form.

The procedure is repeated, replacing the segment $\mathbf{v}_{j-1,1}, \dots, \mathbf{v}_{j-1,s_{j-1}}$ by the images

$$\mathbf{v}'_{j-1,1} = T(\mathbf{v}'_{j1}), \mathbf{v}'_{j-1,2} = T(\mathbf{v}'_{j2}), \dots, \mathbf{v}'_{j-1,s_j} = T(\mathbf{v}'_{js_j}),$$

and then adjoining elements $\mathbf{v}'_{j-1,s_j+1}, \dots, \mathbf{v}'_{j-1,s_{j-1}}$ so as to form a basis of \mathbf{W}_{j-1} . This repetition leads finally to a basis

$$\mathcal{A}' = \{\mathbf{v}'_{11}, \dots, \mathbf{v}'_{1s_1}, \mathbf{v}'_{21}, \dots, \mathbf{v}'_{2s_2}, \dots, \mathbf{v}'_{t1}, \dots, \mathbf{v}'_{ts_t}\}$$

such that

$$\mathbf{v}'_{j-1,1} = T(\mathbf{v}'_{j1}), \dots, \mathbf{v}'_{j-1,s_j} = T(\mathbf{v}'_{js_j})$$

for each j .

The vectors in \mathcal{A}' are now rearranged so that the first vectors from each segment are written first, followed in order by the second vectors from each segment, and so on until the vectors in the last segment are exhausted. Whenever a segment of vectors is exhausted, we continue the same procedure with the remaining segments, if there are any. This process leads to the basis

$$\mathcal{B} = \{\mathbf{v}'_{11}, \mathbf{v}'_{21}, \dots, \mathbf{v}'_{t1}, \dots, \mathbf{v}'_{1s_t}, \dots, \mathbf{v}'_{ts_t}, \mathbf{v}'_{1,s_t+1}, \dots, \mathbf{v}'_{t-1,s_t+1}, \dots, \mathbf{v}'_{1s_1}, \dots, \mathbf{v}'_{ks_1}\}.$$

Since

$$\mathbf{v}'_{11} = T(\mathbf{v}'_{21}), \mathbf{v}'_{21} = T(\mathbf{v}'_{31}), \dots, \mathbf{v}'_{t-1,1} = T(\mathbf{v}'_{t1}),$$

we have

$$\mathbf{v}'_{11} = T^{t-1}(\mathbf{v}'_{t1}), \mathbf{v}'_{21} = T^{t-2}(\mathbf{v}'_{t1}), \dots, \mathbf{v}'_{t-1,1} = T(\mathbf{v}'_{t1}),$$

and $\{\mathbf{v}'_{11}, \mathbf{v}'_{21}, \dots, \mathbf{v}'_{t1}\}$ is the same as the cyclic basis $\mathcal{B}(\mathbf{v}'_{t1}, T)$ of $\mathbf{C}(\mathbf{v}'_{t1}, T)$. Similarly, those \mathbf{v}'_{ij} with the same second subscript (i.e., with the same position in the segments of \mathcal{A}') are the same as the vectors in the cyclic basis $\mathcal{B}(\mathbf{v}'_{mj}, T)$ of $\mathbf{C}(\mathbf{v}'_{mj}, T)$:

$$\mathcal{B}(\mathbf{v}'_{mj}, T) = \{\mathbf{v}'_{1j}, \mathbf{v}'_{2j}, \dots, \mathbf{v}'_{mj}\},$$

where m is the number of the last segment in \mathcal{A}' that has at least j elements. Thus

$$\mathcal{B} = \{\mathcal{B}(\mathbf{v}'_{t1}, T), \dots, \mathcal{B}(\mathbf{v}'_{ts_t}, T), \mathcal{B}(\mathbf{v}'_{t-1,s_t+1}, T), \dots, \mathcal{B}(\mathbf{v}'_{1,s_2+1}, T), \dots, \mathcal{B}(\mathbf{v}'_{1s_1}, T)\},$$

where $\mathcal{B}(\mathbf{v}'_{ij}, T)$ is the cyclic basis of $\mathbf{C}(\mathbf{v}'_{ij}, T)$. By Theorem 10.25, the matrix of the restriction of T to $\mathbf{C}(\mathbf{v}'_{ij}, T)$ is of the form

$$A_j = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Since each $\mathbf{C}(\mathbf{v}'_{ij}, T)$ is T -invariant and \mathbf{V} is the direct sum of the $\mathbf{C}(\mathbf{v}'_{ij}, T)$, the matrix of T with respect to \mathcal{B} is of the form A given in the statement of the theorem. There are s_1 submatrices A_j in A , one for each $\mathbf{C}(\mathbf{v}'_{ij}, T)$, and $s_1 = \dim(\mathbf{W}_1) = \text{nullity}(T)$. The matrix A_1 is of order t , since $\mathcal{B}(\mathbf{v}'_{t1}, T)$ has t elements. The inequality

$$\text{order}(A_i) \geq \text{order}(A_{i+1})$$

is clear from the construction of the basis \mathcal{B} .

Now T determines the subspaces \mathbf{W}_j , and the dimensions of the \mathbf{W}_j determine the matrices A_j . Hence A is uniquely determined by T . ■

Example 1 □ Let T be the linear transformation of \mathbf{R}^5 that has matrix

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

relative to the standard basis. It is easily verified that T is nilpotent of index 3. We follow the proof of Theorem 10.28 to obtain a basis \mathcal{B} of \mathbf{R}^5 such that T has a matrix A of the form described in the theorem.

The row-echelon form of M leads to the basis $\{(1, 0, 0, 1, 0), (1, 2, 2, 0, 1)\}$ of the kernel \mathbf{W}_1 of T . The row-echelon form for M^2 shows that the kernel \mathbf{W}_2 of T^2 is the set of all $(x_1, x_2, x_3, x_4, x_5)$ in \mathbf{R}^5 such that $x_2 = x_3$. The basis of \mathbf{W}_1 can then be extended to the basis

$$\{(1, 0, 0, 1, 0), (1, 2, 2, 0, 1); (0, 1, 1, 0, 0), (0, 1, 1, 0, 1)\}$$

of \mathbf{W}_2 . (The semicolon separates the segments in the basis.) This basis can in turn be extended to the basis

$$\mathcal{A} = \{(1, 0, 0, 1, 0), (1, 2, 2, 0, 1); (0, 1, 1, 0, 0), (0, 1, 1, 0, 1); (0, 0, 1, 0, 0)\}$$

of the desired type.

As the first step in modifying the basis, we find $T(0, 0, 1, 0, 0) = (-2, -1, -1, -1, 0)$ and replace the segment $(0, 1, 1, 0, 0), (0, 1, 1, 0, 1)$ in the basis of \mathbf{W}_2 by the vector $(-2, -1, -1, -1, 0)$. The vector $(0, 1, 1, 0, 0)$ can be used to extend this set to the basis

$$\{(1, 0, 0, 1, 0), (1, 2, 2, 0, 1); (-2, -1, -1, -1, 0), (0, 1, 1, 0, 0)\}$$

of \mathbf{W}_2 . [Incidentally, the vector $(0, 1, 1, 0, 1)$ cannot be used.] Next we replace the segment $(1, 0, 0, 1, 0), (1, 2, 2, 0, 1)$ by $T(-2, -1, -1, -1, 0), T(0, 1, 1, 0, 0)$ to obtain the basis

$$\begin{aligned} \mathcal{A}' = & \{(-1, -2, -2, 0, -1), (-2, -2, -2, -1, -1); \\ & (-2, -1, -1, -1, 0), (0, 1, 1, 0, 0); (0, 0, 1, 0, 0)\} \end{aligned}$$

as described in the proof of the theorem. By rearrangement of the vectors according to their positions in the segments, we obtain the basis

$$\begin{aligned} \mathcal{B} = & \{(-1, -2, -2, 0, -1), (-2, -1, -1, -1, 0), (0, 0, 1, 0, 0); \\ & (-2, -2, -2, -1, -1), (0, 1, 1, 0, 0)\}. \end{aligned}$$

It is easily verified that the matrix of T relative to \mathcal{B} is

$$A = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{bmatrix},$$

where

$$A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \blacksquare$$

Exercises 10.5

1. Let T be the nilpotent transformation of \mathbf{R}^n that has the given matrix relative to the standard basis. Find a basis \mathcal{B} of \mathbf{R}^n that satisfies the conditions of Theorem 10.28, and exhibit the matrix A described in that theorem.

$$\begin{array}{ll} (\text{a}) \begin{bmatrix} 4 & -1 & -19 & 3 \\ -3 & 0 & 12 & -2 \\ 1 & 0 & -4 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} & (\text{b}) \begin{bmatrix} 2 & 0 & -1 & -2 \\ 4 & -1 & -1 & -4 \\ 4 & -1 & -1 & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$(c) \begin{bmatrix} -1 & -1 & -2 & 0 & 1 \\ 0 & 1 & 3 & 0 & -2 \\ 2 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 2 & 0 & -1 \end{bmatrix} \quad (d) \begin{bmatrix} 3 & 0 & -2 & -2 & 1 \\ 4 & -1 & -1 & -2 & 0 \\ 4 & -1 & -1 & -3 & 0 \\ 1 & 0 & -1 & -1 & 1 \\ 2 & -1 & 0 & -2 & 0 \end{bmatrix}$$

2. In each part below, let T be the nilpotent transformation in the corresponding part of Problem 1. Find $\mathcal{B}(\mathbf{v}, T)$ for the given \mathbf{v} , and find the matrix relative to $\mathcal{B}(\mathbf{v}, T)$ of the restriction of T to $\mathbf{C}(\mathbf{v}, T)$.
- (a) $\mathbf{v} = (1, 1, 0, -1)$
 - (b) $\mathbf{v} = (1, 1, 1, 1)$
 - (c) $\mathbf{v} = (-1, -1, 1, 0, 1)$
 - (d) $\mathbf{v} = (1, 2, -1, 1, 2)$
3. Prove that the subspace $\mathbf{C}(\mathbf{v}, T)$ in Definition 10.24 is T -invariant.
4. Prove that T is nilpotent of index t if and only if the minimum polynomial of T is x^t .
5. Prove that T is nilpotent if and only if the characteristic polynomial of T is x^n .
6. A square matrix A is called **nilpotent** if $A^k = \mathbf{0}$ for some positive integer k . The **index** of nilpotency of A is the least positive integer t such that $A^t = \mathbf{0}$. Prove that if A is nilpotent of index t , then any matrix that is similar to A is also nilpotent of index t .
7. Show that the zero operator Z is the only normal linear operator on a unitary space \mathbf{V} that is nilpotent.
8. Prove that the cyclic subspace of \mathbf{v} relative to T is the set of all vectors of the form $g(T)(\mathbf{v})$, where $g(T)$ is a polynomial in T .
9. Prove that $\mathbf{C}(\mathbf{v}, T)$ is the intersection of all T -invariant subspaces that contain \mathbf{v} .

10.6 The Jordan Canonical Form

Throughout the development, a good portion of our effort has been connected in one way or another with representations of linear operators by diagonal matrices, and this is in keeping with the importance of the topic. A person's knowledge of this area might be regarded as a good rough measure of his knowledge of the fundamentals of linear algebra.

Not all linear operators can be represented by a diagonal matrix, but a standard form can be obtained that comes very close to being a diagonal matrix. This standard form is known as the *Jordan canonical form*. In a sense, it is the ultimate result concerning diagonalization, and this is the main reason for its inclusion here. There are situations in which it is of great value.

We begin with the definition of a Jordan matrix.

Definition 10.29 *The **Jordan matrix** of order k with eigenvalue λ is the $k \times k$ matrix*

$$\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}.$$

Theorem 10.30 *Let T have characteristic polynomial*

$$f(x) = (-1)^n(x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}$$

and minimal polynomial

$$m(x) = (x - \lambda_1)^{t_1} \cdots (x - \lambda_r)^{t_r}$$

over \mathcal{F} , and let n_i be the geometric multiplicity of λ_i . Then there exists a basis of \mathbf{V} such that the matrix of T relative to this basis has the following form:

$$J = \begin{bmatrix} J_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & J_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & J_r \end{bmatrix},$$

with J_i a square matrix of order m_i given by

$$J_i = \begin{bmatrix} J_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & J_{i2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & J_{in_i} \end{bmatrix},$$

*where each J_{ik} is a Jordan matrix with eigenvalue λ_i such that J_{i1} has order t_i and order(J_{ik}) \geq order($J_{i,k+1}$) for all k . For a prescribed ordering of the eigenvalues, the matrix J is uniquely determined by T and is called the **Jordan canonical matrix** for T .*

Proof. Let $T = D + N$ be the expression of T as the sum of a diagonalizable transformation D and a nilpotent transformation N as in the proof of Theorem 10.23.

With the same notation as used in Section 10.4, let \mathcal{K}_i denote the kernel of $(T - \lambda_i)^{t_i}$, and let $F_i = p_i(T)$. Then $\mathbf{V} = \mathcal{K}_1 \oplus \cdots \oplus \mathcal{K}_r$, and F_i is the projection of \mathbf{V} onto \mathcal{K}_i along $\sum_{j \neq i} \mathcal{K}_j$, by Theorem 10.16. Hence any nonzero vector in \mathcal{K}_i is an eigenvector of F_i corresponding to the eigenvalue 1, and the restriction of F_i to \mathcal{K}_j is the zero transformation if $i \neq j$. For the time being, let d_i denote the dimension of \mathcal{K}_i . With every choice of bases $\mathcal{B}_i = \{\mathbf{u}_{i1}, \dots, \mathbf{u}_{id_i}\}$ for the subspaces \mathcal{K}_i , the set

$$\mathcal{B} = \{\mathbf{u}_{11}, \dots, \mathbf{u}_{1d_1}, \mathbf{u}_{21}, \dots, \mathbf{u}_{2d_2}, \dots, \mathbf{u}_{r1}, \dots, \mathbf{u}_{rd_r}\}$$

is a basis of \mathbf{V} . Since $F_i(\mathcal{K}_i) = \mathcal{K}_i$ and $F_i(\mathcal{K}_j) = \{\mathbf{0}\}$ if $i \neq j$, each \mathcal{K}_i is invariant under $D = \lambda_1 F_1 + \cdots + \lambda_r F_r$, and the restriction of D to \mathcal{K}_i has matrix $\lambda_i I_{d_i}$ relative to \mathcal{B}_i . It follows from the discussion at the beginning of Section 10.5 that D has the matrix

$$K = \begin{bmatrix} \lambda_1 I_{d_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \lambda_2 I_{d_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \lambda_r I_{d_r} \end{bmatrix}$$

relative to \mathcal{B} .

Consider now the matrix of N relative to a basis of this type. Since

$$(T - \lambda_i)F_i(\mathcal{K}_i) \subseteq (T - \lambda_i)(\mathcal{K}_i) \subseteq \mathcal{K}_i$$

and $(T - \lambda_i)F(\mathcal{K}_j) = \{\mathbf{0}\}$ if $i \neq j$, each \mathcal{K}_i is invariant under

$$N = (T - \lambda_1)F_1 + \cdots + (T - \lambda_r)F_r.$$

Now the restriction N_i of N to \mathcal{K}_i is the same as the restriction of $T - \lambda_i$ to \mathcal{K}_i . By part (c) of Theorem 10.18, the minimal polynomial of the restriction of T to \mathcal{K}_i is $(x - \lambda_i)^{t_i}$. Therefore, the restriction of $T - \lambda_i$ to \mathcal{K}_i has minimal polynomial x^{t_i} . This means that the minimal polynomial of N_i is x^{t_i} , and that N_i is nilpotent of index t_i . According to Theorem 10.28, there exists a basis $\mathcal{B}'_i = \{\mathbf{u}'_{i1}, \dots, \mathbf{u}'_{id_i}\}$ of \mathcal{K}_i such that N_i has a $d_i \times d_i$ matrix of the form

$$L_i = \begin{bmatrix} L_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & L_{i2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & L_{in_i} \end{bmatrix},$$

where each L_{ij} has the form

$$L_{ij} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

L_{i1} has order t_i and order $(L_{ij}) \geq$ order $(L_{i,j+1})$ for all j . The number of diagonal blocks L_{ij} in L_i is the nullity of N_i . Since N_i is the same as the restriction of $T - \lambda_i$ to \mathcal{K}_i , and since the kernel of $T - \lambda_i$ is contained in \mathcal{K}_i , we have

$$\begin{aligned} \text{nullity}(N_i) &= \text{nullity}(T - \lambda_i) \\ &= \dim(T - \lambda_i)^{-1}(\mathbf{0}) \\ &= n_i. \end{aligned}$$

Relative to the basis

$$\mathcal{B}' = \{\mathbf{u}'_{11}, \dots, \mathbf{u}'_{1d_1}, \mathbf{u}'_{21}, \dots, \mathbf{u}'_{2d_2}, \dots, \mathbf{u}'_{r1}, \dots, \mathbf{u}'_{rd_r}\},$$

N has the matrix

$$L = \begin{bmatrix} L_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & L_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & L_r \end{bmatrix}$$

since each \mathcal{K}_i is invariant under N . Thus T has the matrix

$$J = K + L = \begin{bmatrix} \lambda_1 I_{d_1} + L_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \lambda_2 I_{d_2} + L_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \lambda_r I_{d_r} + L_r \end{bmatrix}$$

relative to \mathcal{B}' . Now J is an upper triangular matrix, so

$$\det(J - xI) = (-1)^n(x - \lambda_1)^{d_1} \cdots (x - \lambda_r)^{d_r}.$$

But $\det(J - xI) = f(x)$ since J represents T . Therefore $d_i = m_i$ for each i . Letting $J_i = \lambda_i I_{m_i} + L_i$, the matrix J is in the required form.

Now K is uniquely determined by $f(x)$ whenever the ordering of the eigenvalues is designated. By Theorem 10.28, N_i determines the matrix L_i , and a prescribed ordering of the eigenvalues then determines L . Hence $J = K + L$ is unique if the order of the eigenvalues is specified. ■

Theorem 10.31 *If an $n \times n$ matrix A over \mathcal{F} has characteristic polynomial $f(x)$ and minimal polynomial $m(x)$ that factor over \mathcal{F} as given in Theorem 10.30, then A is similar over \mathcal{F} to a unique matrix J as described in that theorem. The matrix J is called the **Jordan canonical form** for A .*

Proof. Let A be a matrix that satisfies the given conditions. With any given choices of an n -dimensional vector space \mathbf{V} over \mathcal{F} and a basis of \mathbf{V} , A determines a unique linear operator T on \mathbf{V} . Any such linear operator has $f(x)$ as its characteristic polynomial and $m(x)$ as its minimal polynomial. Since any n -dimensional vector space over \mathcal{F} is isomorphic to \mathcal{F}^n , we may assume without loss of generality that T is a linear operator on \mathcal{F}^n . The projections F_i , the subspaces \mathcal{K}_i , and the operators D and N are determined independently of the choice of basis in \mathcal{F}^n . Thus the matrix J is uniquely determined by A . ■

The proof of Theorem 10.30 furnishes a method for obtaining the basis and the matrix J described in that theorem, but several short-cuts can be made in the procedure. This is illustrated in the following example.

Example 1 □ Consider the linear operator T on \mathbf{R}^6 that has the matrix

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

relative to \mathcal{E}_6 . We shall (a) find a basis \mathcal{B}' of \mathbf{R}^6 such that T has the Jordan canonical matrix J relative to \mathcal{B}' , (b) determine the matrix J , and (c) find a matrix P such that $P^{-1}AP = J$.

The characteristic polynomial of A is $f(x) = (x+1)^3(x+2)^3$, so $\lambda_1 = -1$ and $\lambda_2 = -2$ are the distinct eigenvalues of T . It is actually not necessary to find the minimal polynomial.

Our first step is to find a basis \mathcal{B}'_1 for the kernel \mathcal{K}_1 of $(T+1)^{t_1}$ that is of the type in the proof of Theorem 10.30. We have seen that the restriction N_1 of N to \mathcal{K}_1 is the same as the restriction of $T - \lambda_1$ to \mathcal{K}_1 . It follows that $(N_1)^j$ is the restriction of $(T - \lambda_1)^j$ to \mathcal{K}_1 for each positive integer j . Since the kernel \mathbf{W}_j of $(T - \lambda_1)^j$ is contained in the kernel \mathbf{W}_{j+1} of $(T - \lambda_1)^{j+1}$, we begin by finding a basis of the kernel \mathbf{W}_1 of

$T - \lambda_1$, extending to a basis of \mathbf{W}_2 , and so on. Since the kernel of $T - \lambda_1$ is contained in \mathcal{K}_1 , the kernel of the restriction of $T - \lambda_1$ to \mathcal{K}_1 is the same as the kernel of $T - \lambda_1$.

The reduction of $A - \lambda_1 I = A + I$ to row-echelon form yields the basis $\{(1, 0, 1, 0, 1, 0)\}$ of the kernel \mathbf{W}_1 . By use of the row-echelon form of $(A + I)^2$, this extends to the basis $\{(1, 0, 1, 0, 1, 0); (1, 1, 1, 1, 0, 1)\}$ of \mathbf{W}_2 . Repetition of this procedure with $(A + I)^3$ produces the basis

$$\{(1, 0, 1, 0, 1, 0); (1, 1, 1, 1, 0, 1); (0, 0, 1, 0, 0, 0)\}$$

of \mathbf{W}_3 . Upon finding the dimension of \mathbf{W}_4 , we discover that $\mathbf{W}_3 = \mathbf{W}_4$. By Lemma 10.26, this indicates that the index of N_1 is $t_1 = 3$. Following the procedure of Theorem 10.28, we replace $(1, 1, 1, 1, 0, 1)$ by

$$N_1(0, 0, 1, 0, 0, 0) = (T - \lambda_1)(0, 0, 1, 0, 0, 0)$$

and

$$(1, 0, 1, 0, 1, 0) \text{ by } N_1^2(0, 0, 1, 0, 0, 0)$$

to obtain

$$\mathcal{B}'_1 = \{(1, 0, 1, 0, 1, 0), (0, 1, 0, 1, -1, 1), (0, 0, 1, 0, 0, 0)\}.$$

The matrix of N_1 relative to this basis is

$$L_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Following the same procedure with the eigenvalue $\lambda_2 = -2$, and letting $\mathbf{W}_j = \ker(N_2^j) = \ker(T + 2)^j$, we find

$$\begin{aligned} \mathbf{W}_1 &= \langle (0, 0, 0, 1, 1, 0), (0, 0, 0, 0, 0, 1) \rangle, \\ \mathbf{W}_2 &= \langle (0, 0, 0, 1, 1, 0), (0, 0, 0, 0, 0, 1); (0, 0, 0, 0, 1, 0) \rangle. \end{aligned}$$

We find that $\mathbf{W}_2 = \mathbf{W}_3$, and this indicates that $t_2 = 2$. We then replace the first vector in the basis of \mathbf{W}_2 by

$$N_2(0, 0, 0, 0, 1, 0) = (T + 2)(0, 0, 0, 0, 1, 0) = (0, 0, 0, 1, -1, 0)$$

and extend this to a basis of \mathbf{W}_1 to obtain the basis

$$\{(0, 0, 0, 1, -1, 0), (0, 0, 0, 0, 0, 1); (0, 0, 0, 0, 1, 0)\}$$

of \mathbf{W}_2 . Rearrangement of the vectors in this basis yields

$$\mathcal{B}'_2 = \{(0, 0, 0, 1, -1, 0), (0, 0, 0, 0, 1, 0), (0, 0, 0, 0, 0, 1)\}.$$

The matrix of N_2 relative to \mathcal{B}'_2 is

$$L_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The desired basis \mathcal{B}' is thus given by

$$\mathcal{B}' = \{(1, 0, 1, 0, 1, 0), (0, 1, 0, 1, -1, 1), (0, 0, 1, 0, 0, 0), (0, 0, 0, 1, -1, 0), (0, 0, 0, 0, 1, 0), (0, 0, 0, 0, 0, 1)\},$$

and the matrix of T relative to this basis is

$$J = \begin{bmatrix} \lambda_1 I_{m_1} + L_1 & \mathbf{0} \\ \mathbf{0} & \lambda_2 I_{m_2} + L_2 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 \end{bmatrix}.$$

The matrix of transition from \mathcal{E}_6 to \mathcal{B}' is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

It is easily verified that $P^{-1}AP = J$. ■

Exercises 10.6

1. Given that the minimum polynomial of

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

is $m(x) = (x - 2)^2(x - 1)$, find an invertible matrix P such that $P^{-1}AP$ is in Jordan canonical form.

2. Find the Jordan canonical form for each matrix A .

$$(a) A = \begin{bmatrix} 7 & 3 & 3 & 2 \\ 0 & 1 & 2 & -4 \\ -8 & -4 & -5 & 0 \\ 2 & 1 & 1 & 3 \end{bmatrix} \quad (b) A = \begin{bmatrix} 8 & 5 & 6 & 0 \\ 0 & -2 & 0 & 0 \\ -10 & -5 & -8 & 0 \\ 2 & 1 & 1 & 2 \end{bmatrix}$$

$$(c) A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ -1 & 0 & 2 & 1 \end{bmatrix} \quad (d) A = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -3 & 3 & 0 \\ 1 & 3 & -1 & 2 \end{bmatrix}$$

3. For each part of Problem 2, let T be the linear transformation of \mathbf{R}^4 that has the matrix A relative to \mathcal{E}_4 . Find a basis of \mathbf{R}^4 such that the matrix of T relative to this basis is the Jordan canonical matrix J for T , and write down a matrix P such that $P^{-1}AP = J$.
4. Use the results of Problems 2 and 3 to write each matrix A in Problem 2 as the sum of a diagonalizable matrix and a nilpotent matrix.
5. Write the transformation T in Example 1 of this section as $T = D + N$, where D is diagonalizable and N is nilpotent.
6. Suppose A is a matrix that has characteristic polynomial $(-1)^7(x - 2)^4(x - 7)^3$ and minimal polynomial $(x - 2)^2(x - 7)^2$. List all the possible Jordan canonical forms that A might have with $\lambda_1 = 2$ and $\lambda_2 = 7$.

Answers to Selected Exercises

Exercises 1.2, page 8

1. (a) Minimum of m and n (c) 12 3. (a) Yes (c) No (e) Yes
5. (a) Yes (c) No
6. (a) $(1, 1, 0) + (0, 1, 1) - (-1, 1, 1) - (2, 1, 0) = (0, 0, 0)$
 $(2, 1, 0) = (1, 1, 0) + (0, 1, 1) - (-1, 1, 1)$
7. $(a_1, a_2, a_3) = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3$
8. $(a_1, a_2, a_3) = (a_1 - a_2)(1, 0, 0) + (a_2 - a_3)(1, 1, 0) + a_3(1, 1, 1)$
9. $(1, 0, 1)$ is such a vector.

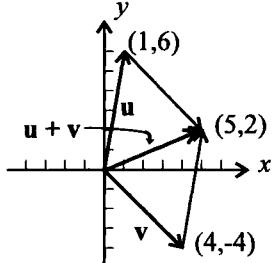
Exercises 1.3, page 17

2. The set of all vectors with components that satisfy a given equation in 5. is a subspace of the type in 4. The set of all vectors with components that satisfy the system of equations in 5. is the intersection of these m subspaces, and this intersection is a subspace by Theorem 1.11.
3. $\bigcup_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda = \{x \in \mathbf{R} \mid -1 < x < 1\}$, $\bigcap_{\lambda \in \mathcal{L}} \mathcal{M}_\lambda = \emptyset$
4. (a) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ (b) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$
5. **Definition 1.4** Let $\mathcal{A} = \{\mathbf{u}_\lambda \mid \lambda \in \mathcal{L}\}$ be a nonempty set of vectors in \mathbf{R}^n . A vector \mathbf{v} is *linearly dependent* on \mathcal{A} if there exist vectors $\mathbf{u}_{\lambda_1}, \mathbf{u}_{\lambda_2}, \dots, \mathbf{u}_{\lambda_k}$ in \mathcal{A} and a_1, a_2, \dots, a_k in \mathbf{R} such that $\mathbf{v} = a_1\mathbf{u}_{\lambda_1} + a_2\mathbf{u}_{\lambda_2} + \dots + a_k\mathbf{u}_{\lambda_k}$. A vector of the form $\sum_{i=1}^k a_i\mathbf{u}_{\lambda_i}$ is called a *linear combination* of the vectors in \mathcal{A} .
- Definition 1.5** Let $\mathcal{A} = \{\mathbf{u}_\lambda \mid \lambda \in \mathcal{L}\}$ be a nonempty set of vectors in \mathbf{R}^n . Then \mathcal{A} is *linearly dependent* if there exist vectors $\mathbf{u}_{\lambda_1}, \mathbf{u}_{\lambda_2}, \dots, \mathbf{u}_{\lambda_k}$ in \mathcal{A} and scalars a_1, a_2, \dots, a_k , not all zero, such that $a_1\mathbf{u}_{\lambda_1} + a_2\mathbf{u}_{\lambda_2} + \dots + a_k\mathbf{u}_{\lambda_k} = \mathbf{0}$. If \mathcal{A} is not linearly dependent, it is *linearly independent*.

6. (a) No (c) Yes (e) No 7. (a) Yes (c) No

Exercises 1.4, page 27

1.



3. $\bigcap_{\lambda \in \mathcal{L}} M_\lambda$ is the origin. $\bigcup_{\lambda \in \mathcal{L}} M_\lambda$ is the set of all points except those with coordinates $(0, y)$, where $y \neq 0$.
4. (a) $2x + 3y - z = 0$ 5. (a) 13 (c) 5 (e) $\sqrt{30}$ 7. (a) $\frac{1}{13}(3, -4, 12)$
8. (a) $-4\sqrt{21}/21$ 9. $2\sqrt{5}$
11. (a) $\{(1, 0, 2)\}$ (b) $\{(1, 0, 2), (2, -1, 1), (1, 1, -1)\}$
13. $\{\mathbf{u}_1, \mathbf{u}_2\}$, where δ_{ij} is the Kronecker delta and $\mathbf{u}_1 = (\delta_{11}, \delta_{12}, \dots, \delta_{1n})$, $\mathbf{u}_2 = (\delta_{21}, \delta_{22}, \dots, \delta_{2n})$

Exercises 1.5, page 37

1. (a) $\{(2, 6, -3), (5, 15, -8), (5, 3, -2)\}$ (c) $\{(1, 1, 0), (2, 4, 1), (1, 2, 1)\}$
2. (a) $\{(1, 0, 2, 3), (0, 1, -2, -3), (0, 1, 1, 0), (0, 0, 0, 1)\}$
 (c) $\{(1, 1, 0, 0), (0, 0, 1, 1), (1, 0, 1, 0), (0, 1, 0, -1)\}$
6. (a) Linearly independent (c) Linearly dependent
 (e) Linearly independent
7. (a) Does not span \mathbf{R}^3 (c) Does not span \mathbf{R}^3 (e) Spans \mathbf{R}^3
8. (a) Not a basis for \mathbf{R}^3 (c) Basis for \mathbf{R}^3
9. (a) Not a basis of \mathbf{R}^4 (c) Basis of \mathbf{R}^4
11. (a) $\{(1, 0, 1, -1), (3, -2, 2, 5)\}$ (c) $\{(2, -1, 0, 1), (1, 2, 1, 0), (5, 3, 2, 1)\}$
12. (a) 2 (c) 2

Exercises 2.2, page 45

1. $\{(1, 0, 0, 0, 0), (0, 1, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 1, 0), (0, 0, 0, 0, 1)\}$
3. Replace the second vector by the sum of the second vector and (-2) times the first vector.
5. Apply the sequence E_1, E_2, E_3, E_4 where
 E_1 : replace the second vector by the sum of the second and (-1) times the first,
 E_2 : interchange the two vectors,
 E_3 : multiply the second vector by 3,
 E_4 : replace the second vector by the sum of the second and the first.
7. E_1 : multiply the second vector by 2.
 E_2 : replace the second vector by the sum of the second vector and 3 times the third vector.
 E_3 : replace the third vector by the sum of the third vector and the first vector.
 E_4 : multiply the first vector by 2.
9. E_1 : replace the third vector by the sum of the third and (2) times the second.
 E_2 : replace the second vector by the sum of the second and (3) times the first.
 E_3 : replace the first vector by the sum of the first and (-1) times the fourth.
 E_4 : multiply the fourth vector by 2.
11. Use the inverses of the elementary operations in Problem 6 in reverse order.
13. The elementary operation of type II which replaces the first vector by the sum of the first vector and zero times the second vector is the identity operation.

Exercises 2.3, page 48

5. (a) Linearly independent (c) Linearly dependent (e) Linearly dependent

Exercises 2.4, page 56

1. $\mathcal{A}' = \{(1, 0, 5, 0, 2), (0, 1, 2, 0, 1), (0, 0, 0, 1, 1), (0, 0, 0, 0, 0)\}$
3. (a) 2 (c) 3
4. (a) $\{(1, 0, 0, 1), (0, 1, 0, -1), (0, 0, 1, 1)\}$ (c) $\{(1, 0, 1, 1), (0, 1, 1, -1)\}$
5. (a) 3, so the set is linearly dependent.
6. (a) $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ (c) $\langle \mathcal{A} \rangle \neq \langle \mathcal{B} \rangle$ (e) $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ 7. (a) 3

8. E_1 : interchange the first and third vectors.
 E_2 : multiply the first vector by $\frac{1}{2}$.
 E_3 : replace the second vector by the sum of the second and (-3) times the first.
 E_4 : replace the fourth vector by the sum of the fourth and the first.
 E_5 : multiply the second vector by $(-\frac{1}{2})$.
 E_6 : replace the third vector by the sum of the third and the second.
 E_7 : replace the fourth vector by the sum of the fourth and (-3) times the second.
 E_8 : replace the first vector by the sum of the first and (-1) times the second.
9. (a) E_1 : replace the second vector by the sum of the second vector and (-1) times the first.
 E_2 : replace the first vector by the sum of the first and second.
 E_3 : multiply the second vector by (-1) .
 E_4 : replace the first vector by the sum of the first and $(-\frac{1}{2})$ times the second.
 E_5 : multiply the first vector by 2.
- (c) E_1 : interchange the first and second vectors.
 E_2 : replace the second vector by the sum of the second vector and (-4) times the first vector.
 E_3 : replace the first vector by the sum of the first vector and (-1) times the second vector.
 E_4 : replace the first vector by the sum of the first vector and (-1) times the second vector.
 E_5 : multiply the second vector by 2.

Exercises 3.2, page 63

1. $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ 3. $x = 2y$

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

6. (a) $\{(4, 4), (-1, 1)\}$
(c) $\{(0, 0, 0, 0), (0, 1, 4, 4), (8, -2, 8, -16), (5, 3, 10, -4), (5, 2, 6, -8)\}$
(e) $\{(0, 1, -4), (-1, 2, 4), (1, -2, 0)\}$

7. (a) $\begin{bmatrix} 3 & -8 & -5 \\ 0 & 4 & 4 \end{bmatrix}$ is a matrix of transition from the first set to the second set.
(c) No matrix of transition exists.

8. (a) $(7, 14)$

(c) $(2, 4, -14, 1)$

9. (a) $\begin{bmatrix} 6 \\ 2 \end{bmatrix}$

(c) $\begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}$

11. $d_1 = x_3, d_2 = x_2 - x_3, d_3 = x_1 - x_2$ 13. $c_i = \sum_{j=1}^r p_{ij}d_j$ for $i = 1, 2, \dots, r$

Exercises 3.3, page 70

1. (a) $\begin{bmatrix} 29 & -5 \\ 6 & -6 \\ 32 & -5 \end{bmatrix}$

(c) $\begin{bmatrix} 9 & -18 \\ 29 & -2 \end{bmatrix}$

(e) $\begin{bmatrix} 11 & 12 \\ -75 & -7 \end{bmatrix}$

2. (a) BA does not exist.

(c) $\begin{bmatrix} 1 & -6 & -5 \\ 12 & -2 & -11 \\ 18 & 32 & 8 \end{bmatrix}$

(e) $\begin{bmatrix} 4 & -3 & -4 & 2 \\ -10 & 32 & -25 & 2 \\ -36 & 69 & -24 & -6 \\ -16 & 12 & 16 & -8 \end{bmatrix}$

3. (a) $n = r$

(c) $n = r$ and $m = t$

5. $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

7. $A = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}$

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

10. (a) $c_{ij} = 28 + 12i - 6j - 3ij$ (c) $c_{ij} = 6i - 3j + 2ij - 10$

11. $x_1 - 2x_2 + x_3 = 4$

$2x_1 + 3x_3 = 5$

$x_1 + 4x_2 - x_3 = 6$

Exercises 3.4, page 80

1. (a) Singular (c) Nonsingular

2. (a) Elementary (c) Not elementary (e) Elementary (g) Elementary

3. (a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix}$

4. (a) $\begin{bmatrix} 1 & -4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (c) Not elementary (e) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (g) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$

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

6. (a) $\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{3}{4} & 1 \end{bmatrix}$ (c) $\begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{4}{5} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$

7. (a) $\begin{bmatrix} -1 & 1 \\ \frac{3}{4} & -\frac{1}{2} \end{bmatrix}$ (c) $\begin{bmatrix} -\frac{1}{5} & \frac{4}{5} \\ \frac{2}{5} & -\frac{3}{5} \end{bmatrix}$ 8. (a) $\begin{bmatrix} -2 & -3 \\ 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} \frac{2}{5} & -\frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} \end{bmatrix}$

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

Exercises 3.5, page 89

1. (a) Interchange the first and second columns.
(c) Replace the second column by the sum of the second column and 2 times the third column.
2. (a) In reduced column-echelon form. (c) Not in reduced row-echelon form.
(e) Not in reduced column-echelon form.

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

5. Not always. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$ and $M = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Then A is the matrix of transition from \mathcal{E}_3 to $\mathcal{A} = \{(1, 0, 1), (0, 0, 1)\}$ and $MA = \begin{bmatrix} 1 & 0 \\ -2 & 0 \\ 1 & 1 \end{bmatrix}$ is the matrix of transition form \mathcal{E}_3 to $\mathcal{B} = \{(1, -2, 1), (0, 0, 1)\}$. Now $(1, -2, 1)$ is in \mathcal{B} but not in $\langle \mathcal{A} \rangle$ since it has a nonzero second component. Thus $\langle \mathcal{B} \rangle \neq \langle \mathcal{A} \rangle$ in this case.

6. (a) $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$

(g) $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$

7. (a) $\begin{bmatrix} -1 & 2 & -1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & -1 & -1 \\ -2 & 1 & -1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & -\frac{7}{8} & -\frac{3}{8} & \frac{41}{8} \\ 0 & \frac{1}{4} & \frac{1}{4} & -\frac{7}{4} \\ 0 & \frac{3}{8} & -\frac{1}{8} & -\frac{5}{8} \\ 0 & 0 & 0 & 1 \end{bmatrix}$

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

Exercises 3.6, page 97

1. (a) Multiply the first row by 4. (c) Interchange the first and third rows.

2. (a) Not in reduced row-echelon form. (c) Not in reduced row-echelon form.
 (e) Not in reduced row-echelon form.

3. (a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 & 0 & -\frac{41}{8} \\ 0 & 1 & 0 & \frac{7}{8} \\ 0 & 0 & 1 & \frac{5}{8} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

(g) $\begin{bmatrix} 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

4. (a) $\begin{bmatrix} 0 & -1 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & -2 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & -2 & 0 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 & -\frac{7}{8} & 0 & -\frac{3}{8} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{3}{8} & 0 & -\frac{1}{8} & 0 \\ -1 & 0 & -1 & 0 & -1 & 1 \\ -1 & 0 & -2 & 1 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ (g) $\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{3}{2} & -\frac{3}{2} & 0 & 1 \\ -2 & 0 & 1 & 0 \end{bmatrix}$

5. (a) $\begin{bmatrix} -\frac{1}{2} & \frac{3}{2} & 1 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}$ (c) $\begin{bmatrix} \frac{10}{9} & -\frac{4}{9} & \frac{1}{9} \\ -\frac{8}{9} & \frac{5}{9} & \frac{1}{9} \\ -\frac{1}{9} & \frac{4}{9} & -\frac{1}{9} \end{bmatrix}$

7. (a) $\mathcal{A} = \{(1, -1, 0), (0, 1, 1), (2, 0, -1)\}$ (b) $\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$

Exercises 3.7, page 104

1. (a) A and B are column-equivalent. (c) A and B are not column-equivalent.
 2. (a) A and B are not row-equivalent. (c) A and B are row-equivalent.

3.
$$\begin{bmatrix} 4 & 14 & -7 \\ 6 & 29 & -14 \\ 5 & 25 & -12 \end{bmatrix}$$

5. (a) $Q = \begin{bmatrix} 1 & 0 & \frac{9}{4} \\ 0 & -\frac{2}{3} & 0 \\ 1 & -2 & \frac{19}{8} \end{bmatrix}$ (c) A and B are not column-equivalent.

6. (a) A and B are not row-equivalent. (c) $P = \begin{bmatrix} 0 & 0 & 0 & -1 \\ \frac{1}{2} & 0 & 0 & -\frac{7}{2} \\ \frac{1}{2} & 0 & 1 & -\frac{3}{2} \\ \frac{5}{2} & 1 & 0 & -\frac{3}{2} \end{bmatrix}$

7. (a) A solution is given by

$$B = A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{8} \end{bmatrix} \begin{bmatrix} 1 & 0 & \frac{9}{4} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

- (c) A and B are not column-equivalent.
 9. (a) All nonsingular $n \times n$ matrices. (b) All nonsingular $n \times n$ matrices.

Exercises 3.8, page 111

1. (a) 2 (c) 2
 2. (a) $\{(1, 0, -1, 2), (0, 1, 2, -1)\}$ (c) $\{(1, 0, -1, -2), (0, 1, 1, 1)\}$
 3. In each part except (d), A and B are equivalent.
 5. Only A and B are equivalent. 6. (a) No (c) Yes

7. $B = \{(1, 2, 0), (0, 0, 1), (0, 1, 0)\}, P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$

8. (a) $\mathcal{A} = \{(1, 1, 0, 0), (0, 1, 1, 0), (0, 0, 1, 1), (1, 2, 2, 1)\}$

$$\mathcal{A}' = \{(1, 0, 0, 1), (0, 1, 0, -1), (0, 0, 1, 1), (0, 0, 0, 0)\}$$

$$\mathcal{B} = \{(1, 0, 0, 1), (0, 1, 0, -1), (0, 0, 1, 1), (0, 0, 0, 1)\}$$

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 1 & -1 & 1 \end{bmatrix}, Q = \begin{bmatrix} 1 & 0 & 0 & -1 \\ -1 & 1 & 0 & -1 \\ 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

9. (a) $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}, Q = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -2 & 0 & 0 \end{bmatrix}$

11. $P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} 0 & 0 & 0 & 1 \\ \frac{3}{2} & -2 & \frac{1}{2} & \frac{3}{2} \\ 0 & 1 & 0 & 0 \\ -5 & 10 & 0 & -5 \end{bmatrix}$

Exercises 4.2, page 120

2. $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$

3. mn

5. Conditions 3, 4, and 9 fail to hold.

7. V is a vector space with respect to the given operations.

9. Condition 10 fails to hold.

11. V is a vector space with respect to the given operations.

13. Conditions 8 and 9 fail to hold.

15. W is a vector space with respect to the given operations.

Exercises 4.3, page 125

1. (a) Linearly dependent (c) Linearly independent (e) Linearly dependent

2. (a) The set spans P_2 . (c) The set does not span P_2 .

3. (a) Linearly dependent (c) Linearly dependent
 4. (a) The set is not a basis for P_2 . (c) The set is a basis for P_2 .

5. (a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

6. (a) $\{p_1(x), p_2(x)\}$ or any subset consisting of two distinct vectors.
 (c) $\left\{ \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \right\}$ or any subset consisting of two distinct vectors.
 (e) $\{p_1(x), p_2(x), p_3(x), p_4(x)\}$
 7. (a) $\{(1, 0, 0, 0), (0, 0, 1, 0)\}$ is a basis for \mathbf{W} .
 9. \mathbf{W} is not a subspace of \mathbf{R}^3 . The vectors $\mathbf{u} = (1, 1, 0)$ and $\mathbf{v} = (2, 4, 6)$ are in \mathbf{W} , but $\mathbf{u} + \mathbf{v} = (3, 5, 6)$ is not in \mathbf{W} since $5 \neq 3^2$.
 11. \mathbf{W} is not a subspace of \mathbf{R}^3 . The vectors $\mathbf{u} = (2, 3, 6)$ and $\mathbf{v} = (4, 5, 20)$ are in \mathbf{W} , but $\mathbf{u} + \mathbf{v} = (6, 8, 26)$ is not in \mathbf{W} since $26 \neq (6)(8)$.
 13. \mathbf{W} is a subspace of P_2 over \mathbf{R} .

15. \mathbf{W} is not a subspace of $\mathbf{R}_{2 \times 2}$. The vectors $\mathbf{u} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 & 5 \\ 6 & 7 \end{bmatrix}$ are in \mathbf{W} , but $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 2 & 7 \\ 9 & 11 \end{bmatrix}$ is not in \mathbf{W} because its first row, first column element is not equal to 1.

17. \mathbf{W} is not a subspace of $\mathbf{R}_{2 \times 2}$. The vectors $\mathbf{u} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$ are in \mathbf{W} , but $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is not in \mathbf{W} since it is nonzero and not invertible.

19. \mathbf{W} is not a subspace of $\mathbf{R}_{2 \times 2}$. The vectors $\mathbf{u} = \begin{bmatrix} 2 & 3 \\ 4 & -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 & 5 \\ 6 & 2 \end{bmatrix}$ are in \mathbf{W} , but $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 4 & 8 \\ 10 & 0 \end{bmatrix}$ is not in \mathbf{W} since $4^2 \neq 0^2$.

Exercises 4.4, page 129

1. $f\left(\sum_{i=0}^n a_i x^i\right) = (a_0, a_1, \dots, a_n)$

3. (a) $r = 2, f(a_1(1, -1, 1) + a_2(0, 1, 0)) = (a_1, a_2)$

(c) $r = 3, f(a_1(2, 0, 4, -1) + a_2(5, -1, 11, 8) + a_3(0, 1, -7, 9)) = (a_1, a_2, a_3)$

(e) $r = 2, f(a_1 p_1(x) + a_2 p_2(x)) = (a_1, a_2)$

(g) $r = 2, f\left(a_1 \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} + a_2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = (a_1, a_2)$

5. $f(a_1 p_1(x) + a_2 p_2(x) + a_3 p_3(x)) = (a_1, a_2, a_3)$

Exercises 4.5, page 132

1. (a) $\{(3, 3, 2, 0), (1, 1, 1, 1)\}$ (c) $\{(3, 2, 2, 0), (1, 1, 0, 0), (1, 1, 1, 1)\}$

2. (a) $\{p_1(x), p_2(x), p_3(x)\}$ 3. (a) $\left\{ \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}, \begin{bmatrix} 3 & 3 \\ 2 & 2 \end{bmatrix} \right\}$

4. (a) $\langle A \rangle = \langle B \rangle$ (b) $\langle A \rangle \neq \langle B \rangle$ 5. (a) $\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & -3 \\ 2 & 0 & 2 \end{bmatrix}$

Exercises 4.7, page 142

1. $x_1 = -27, x_2 = 19, x_3 = -6$ 3. No solution 5. No solution

7. $x_1 = -\frac{9}{2}x_3, x_2 = -\frac{1}{4}x_3, x_4 = 0; x_3$ arbitrary

9. $x_1 = 1 + 2x_2 + x_4, x_3 = -x_4; x_2, x_4$ arbitrary

11. $x_1 = 5 - 3x_2 - 2x_4 - 2x_5, x_3 = 3 - 2x_4; x_2, x_4, x_5$ arbitrary

13. $x_1 = 1 - 2x_2, x_3 = 1, x_4 = 0, x_5 = 1; x_2$ arbitrary

15. (a) 2 (b) 3 (c) The system is inconsistent since $2 \neq 3$.

17. (a) $\{(1, 2, 0, -1, 3), (0, 0, 1, -2, 2)\}$

(b) $(3, 6, 1, -5, 11) = (3)(1, 2, 0, -1, 3) + (1)(0, 0, 1, -2, 2)$

$(2, 4, 1, -4, 8) = (2)(1, 2, 0, -1, 3) + (1)(0, 0, 1, -2, 2)$

$(-2, -4, 2, -2, -2) = (-2)(1, 2, 0, -1, 3) + (2)(0, 0, 1, -2, 2)$

$(7, 17, 3, -13, 27) = (7)(1, 2, 0, -1, 3) + (3)(0, 0, 1, -2, 2)$

19. (a) $a = 2$ (b) All real numbers except 2 and -2 (c) $a = -2$

21. (b) $n + 1$

- 22.** 1. \emptyset 3. $\{(1, -1, -2)\}$ 5. \emptyset 7. $\{(18, 1, -4, 0)\}$
 9. $\{(2, 1, 0, 0), (1, 0, -1, 1)\}$ 11. $\{(-3, 1, 0, 0, 0), (-2, 0, -2, 1, 0), (-2, 0, 0, 0, 1)\}$
 13. $\{(-2, 1, 0, 0, 0)\}$

Exercises 5.2, page 154

1. (b) $(S + T)(x_1, x_2) = (x_1, -x_2)$, $(2S - 3T)(x_1, x_2) = (2x_1 + 5x_2, 5x_1 - 2x_2)$
2. (a) T is a linear operator. (c) T is not a linear operator.
 (e) T is a linear operator.
3. (a) T is a linear transformation. (c) T is not a linear transformation.
4. (a) T is a linear operator. (c) T is not a linear operator.
5. (a) T is not a linear operator. (c) T is a linear operator.
6. (a) T is a linear transformation. (c) T is a linear transformation.
 (e) T is a linear transformation.
7. (a) $\{(3, 1, 2), (-2, 1, -3)\}$ (b) $\{(1, 1, 1, 0), (2, 1, 0, 1)\}$
9. (a) $\text{rank}(T) = 3, \text{nullity}(T) = 1$ 10. (a) $\{(1, 0, -\frac{1}{3}, \frac{2}{3})\}$
13. (a) $\{(1, 2, 3), (1, 1, 1)\}$

Exercises 5.3, page 165

1. $T(x, y) = (-y, x + 2y, x + y)$ 3. $T(a_0 + a_1x + a_2x^2) = (a_1, a_2 - a_1)$

5.
$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & 5 & -2 \\ 4 & 0 & 1 \\ 2 & 3 & 1 \end{bmatrix}$$
 7.
$$\begin{bmatrix} 2 & 4 & -2 \\ -1 & -2 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$
 9. $(3, 0, 2)$ 11. $\frac{3}{4} + \frac{1}{2}x - \frac{3}{4}x^2 - 2x^3$

13. $(3, 5)$ 15.
$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}$$
 17.
$$\begin{bmatrix} 4 & \frac{3}{2} \\ -1 & 0 \\ 3 & 1 \end{bmatrix}$$

19. (a) $\{(1, 0, -3), (0, 1, 2)\}$ (c) $\{(1, 0, 1, 0), (0, 1, 1, 0), (0, 0, 0, 1)\}$
20. (a) $\{(1, -1, 1, 0), (3, -2, 0, 1)\}$ (c) $\{(6, -\frac{5}{2}, 1, 0, 0), (-7, \frac{5}{2}, 0, 1, 0)\}$
21. (a) $\{(1, 0, 1), (0, 1, -\frac{1}{3})\}$

22. (a) $\{(-2, -2, -2, -2, -3), (-1, -1, -2, -2, -2), (0, 1, 1, 1, 2)\}$

23. $\{(0, 1, 0), (0, 0, 1)\}$

Exercises 5.4, page 174

1. $1 + 2x + 3x^2$

3. $\mathcal{B} = \{1 + x + x^2, 2 + x, 2 + x + x^2\}$

5. (a) $\begin{bmatrix} 4 \\ -4 \end{bmatrix}$

(b) $(-2, 1)$

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

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

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

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

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

17. $\mathcal{A}' = \{(1, 0, 0), (0, 1, 0), (1, -1, 1)\}, \mathcal{B}' = \{(2, 1), (3, 2)\}$

18. (a) $\mathcal{A}' = \{(-2, 0, 1, 0, -1), (-2, 0, 1, 0, 0), (7, 0, -3, 0, 1), (-2, 0, 0, 1, -2), (-3, 1, 0, 0, 0)\}$

$\mathcal{B}' = \mathcal{E}_3$

Exercises 5.5, page 183

1. (a) Matrix of S : $\begin{bmatrix} 5 & 0 & -3 & 0 \\ 0 & 1 & 6 & -1 \\ 2 & -9 & 5 & 2 \end{bmatrix}$, Matrix of T : $\begin{bmatrix} 4 & 10 & 0 & 0 \\ 6 & -1 & 0 & 7 \\ -3 & 8 & 0 & -5 \end{bmatrix}$

Matrix of $S + T$: $\begin{bmatrix} 9 & 10 & -3 & 0 \\ 6 & 0 & 6 & 6 \\ -1 & -1 & 5 & -3 \end{bmatrix}$

(b) Matrix of $2S - 3T$: $\begin{bmatrix} -2 & -30 & -6 & 0 \\ -18 & 5 & 12 & -23 \\ 13 & -42 & 10 & 19 \end{bmatrix}$

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

5. TS has matrix $\begin{bmatrix} 11 & -7 \\ 4 & 1 \\ 6 & -5 \end{bmatrix}$ relative to the bases $\{(1, 2), (0, 1)\}$ of \mathbf{R}^2 and \mathcal{E}_3 of \mathbf{R}^3 .

$$\begin{array}{lll} \mathbf{u}_1 = -2\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 & & \text{(b)} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 2 \\ 2 & 1 & 5 \end{bmatrix} \\ \text{7. (a)} \quad \mathbf{u}_2 = \mathbf{v}_1 - 2\mathbf{v}_2 & & \text{9.} \begin{bmatrix} -11 & -8 \\ 16 & 9 \end{bmatrix} \\ \mathbf{u}_3 = & \mathbf{v}_2 & \end{array}$$

11. $T^4 = 5T^2 - 18T - 18, T^5 = -3T^2 - 38T - 30$

Exercises 6.2, page 191

1. (a) 7 (c) 11 (e) 8 2. (a) Even (c) Odd (e) Even
3. (a) Apply the following interchanges in the given order: 5 and 3, 4 and 3, 2 and 3, 2 and 4, 2 and 5.
 (c) Apply the following interchanges in the given order: 3 and 1, 5 and 1, 4 and 1, 2 and 1, 2 and 4, 2 and 5, 2 and 3.

5. $\mathcal{I} = \sum_{k=1}^n \mathcal{I}(k) = (n-1) + (n-2) + \cdots + 2 + 1 = \frac{n(n-1)}{2}$

Exercises 6.3, page 194

1. (a) Odd (c) Even (e) Odd (g) Odd 3. $8 \det(A)$
5. $a_{11}a_{22} \cdots a_{nn}$ 7. $\det(cA) = c^n \det(A)$ 8. (a) -72 (c) $a_{11}a_{22}a_{33}$

Exercises 6.4, page 200

2. (a) 0 (c) 32 (e) -32 3. 24
5. (a) $AB = \det(A)I_3$ (b) $A^{-1} = \frac{1}{\det(A)} [A_{ij}]^T$ 6. (a) 24 (c) 0
7. 2, -1

Exercises 6.5, page 204

1. (a) -9 (c) $a^2 + b^2$ (e) -177 (g) 60
2. (a) $x_1 = -3, x_2 = 8, x_3 = 4$ (c) $x_1 = -1, x_2 = 1, x_3 = 1$
 (e) $x_1 = 2, x_2 = -1, x_3 = 4$
3. The second determinant is obtained from the first by multiplying the first column by 2 and the second column by 3. Therefore the second determinant has the value $(2)(3)(139) = 834$.

7. $2abc(a + b + c)^3$

Exercises 6.6, page 211

$$\text{1. (a) } \frac{1}{2} \begin{bmatrix} -1 & 3 & 2 \\ 1 & -1 & 0 \\ 1 & -1 & -2 \end{bmatrix} \quad \text{(c) } -\frac{1}{9} \begin{bmatrix} -10 & 4 & -1 \\ 8 & -5 & -1 \\ 1 & -4 & 1 \end{bmatrix} \quad \text{(e) } \frac{1}{20} \begin{bmatrix} -34 & 8 & 36 \\ 12 & -4 & -8 \\ 33 & -6 & -32 \end{bmatrix}$$

2. $\det(A^m) = [\det(A)]^m$ **3.** 3 **5.** $\det(\text{adj}(A)) = [\det(A)]^{n-1}$

11. $x = 2, y = 4; x = -1 + i\sqrt{3}, y = 2(-1 - i\sqrt{3}); x = -1 - i\sqrt{3}, y = 2(-1 + i\sqrt{3})$

Exercises 7.2, page 221

1. 1, 1 **3.** $\frac{5+\sqrt{33}}{2}, \frac{5-\sqrt{33}}{2}$ **5.** 1, -2, 3 **7.** 1, 1, 2 **9.** 1, 2, 2

11. -1, -1, 3, 3 **13.** {2, -3} **15.** {1, 2} **17.** 8, (-5, -6); -1, (1, 3)

19. {1, 2}; 1, 1 - x, x^2 ; 2, 1 + $2x^2$ **21.** 3, (3, 2); 4, (1, 1) **25.** 1

Exercises 7.3, page 229

For Problems 1, 3, and 5, each eigenvalue is followed by its algebraic multiplicity and geometric multiplicity, in that order.

1. 1, 2, 1 **3.** 2, 1, 1; 1, 2, 1 **5.** -1, 2, 2; 3, 2, 2

For Problems 7 and 8, each eigenvalue is followed by its algebraic multiplicity and a basis for the eigenspace.

7. (a) 3, 2, {(0, 1, 0), (2, 1, 1)}; -2, 1, {(5, 4, 3)}
 (c) 1, 1, {(0, 1, 1)}; 2, 1{(-1, 1, 0)}; 3, 1, {(0, 0, 1)}

8. (a) 1, 1, {(1 - i , -1)}; 4, 1, {(1, 1 + i)}
 (c) 1, 2, {(1, 1, -3), (0, 1, -2)}; 6, 1, {(2, 1, 1)}
 (e) 1 + i , 2, {(1, 0, 1), (0, 1, 0)}; 1, 1, {(0, 2, -1)}

9. For $\lambda = 3$: (a) 1, (b) 1, (c) {1};
 For $\lambda = 2$: (a) 2, (b) 2, (c) {1 - $x, 1 - x^2$ }

11. For $\lambda = 1$: (a) 1, (b) 1, (c) {1}; For $\lambda = 2$: (a) 2, (b) 1, (c) {2 + x^2 }

13. (a) {(1, -2), (2, 1)} (b) $\begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix}$

15. $\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$, where λ_i is the eigenvalue of T that corresponds to \mathbf{v}_i .

17. Only I_n 23. $P^{-1}X$ is a corresponding eigenvector.

Exercises 7.4, page 236

1. (a) A is not similar over \mathbf{R} to a diagonal matrix. (b) Not possible.

3. (a) A is not similar over \mathbf{R} to a diagonal matrix. (b) Not possible.

5. (a) A is similar over \mathbf{R} to a diagonal matrix. (b) $P = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & -1 \end{bmatrix}$

7. (a) T can be represented by a diagonal matrix. (b) $\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}, \{(2, 1), (1, -2)\}$

9. (a) T cannot be represented by a diagonal matrix. (b) Not possible.

11. (a) T can be represented by a diagonal matrix.

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

13. (a) T can be represented by a diagonal matrix.

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \{(0, 1, 1), (-1, 1, 0), (0, 0, 1)\}$

15. (a) A is similar over \mathbf{C} to a diagonal matrix. (b) $P = \begin{bmatrix} 1-i & 1 \\ -1 & 1+i \end{bmatrix}$

17. (a) A is similar over \mathbf{C} to a diagonal matrix. (b) $P = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1 \end{bmatrix}$

19. (a) A is similar over \mathcal{C} to a diagonal matrix.

$$(b) P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & -1 \end{bmatrix}$$

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

$$26. (a) \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 0 & 5 \end{bmatrix}$$

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

Exercises 8.2, page 244

1. f is a linear functional. 3. f is not a linear functional. 5. $\mathbf{c} = (2, -1, 4)$
7. $f(5, 43) = 12, f(x_1, x_2, x_3) = 2x_1 - x_2 + 2x_3$
9. (a) $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$, where $\mathbf{p}_1(\mathbf{u}) = x_3 - x_2, \mathbf{p}_2(\mathbf{u}) = x_2 - x_1, \mathbf{p}_3(\mathbf{u}) = x_1$ for $\mathbf{u} = (x_1, x_2, x_3)$.
- (c) $\mathcal{A}^* = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$, where $\mathbf{p}_1(\mathbf{u}) = \frac{1}{2}(2x_1 + x_2 - x_3), \mathbf{p}_2(\mathbf{u}) = \frac{1}{2}(-3x_1 - x_2 + 3x_3), \mathbf{p}_3(\mathbf{u}) = \frac{1}{2}(-x_1 - x_2 + x_3)$ for $\mathbf{u} = (x_1, x_2, x_3)$.
10. (a) $\begin{bmatrix} 0 & -1 & 1 \end{bmatrix}^T, \begin{bmatrix} -1 & 1 & 0 \end{bmatrix}^T, \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$
(c) $\frac{1}{2} \begin{bmatrix} 2 & 1 & -1 \end{bmatrix}^T, \frac{1}{2} \begin{bmatrix} -3 & -1 & 3 \end{bmatrix}^T, \frac{1}{2} \begin{bmatrix} -1 & -1 & 1 \end{bmatrix}^T$
11. (a) $\begin{bmatrix} 10 & 1 & 5 \end{bmatrix}$ (c) $\begin{bmatrix} 2 & 2 & 6 & 18 \end{bmatrix}$
12. (a) $f(\mathbf{e}_1) = 3, f(\mathbf{e}_2) = -2, f(\mathbf{e}_3) = 7$
(c) $f(\mathbf{e}_1) = 2, f(\mathbf{e}_2) = 0, f(\mathbf{e}_3) = 4, f(\mathbf{e}_4) = 12$
14. (a) $\mathbf{p}_1 = \frac{1}{164}(5\mathbf{g}_1 - 12\mathbf{g}_2 + 18\mathbf{g}_3), \mathbf{p}_2 = \frac{1}{82}(73\mathbf{g}_1 + 38\mathbf{g}_2 - 16\mathbf{g}_3), \mathbf{p}_3 = \frac{1}{41}(6\mathbf{g}_1 + 2\mathbf{g}_2 - 3\mathbf{g}_3)$
(c) $\mathbf{p}_1 = \mathbf{g}_1 + \mathbf{g}_2 + 3\mathbf{g}_3 - \mathbf{g}_4, \mathbf{p}_2 = \mathbf{g}_1 - 2\mathbf{g}_3, \mathbf{p}_3 = \mathbf{g}_1 + \mathbf{g}_4, \mathbf{p}_4 = \mathbf{g}_3$
17. (a) $\{4f_1 - 3f_2 - f_3, 8f_1 - 2f_2 - f_4\}$ (c) $\{f_1 + f_2 - 4f_3, f_4\}$

Exercises 8.3, page 252

1. Only the required symmetric matrix A is given.

$$(a) \begin{bmatrix} 4 & 3 \\ 3 & -2 \end{bmatrix}$$

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

$$(e) \begin{bmatrix} -7 & -\frac{9}{2} & 0 \\ -\frac{9}{2} & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(g) \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$$

2. (a) $q(\mathbf{v}) = -2y_1^2 + 18y_2^2$ (c) $q(\mathbf{v}) = y_1^2 - 2y_1y_3 - 4y_2y_3 - 3y_3^2$

3. (a) $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ (c) $\begin{bmatrix} 144 & -12 & 0 \\ -12 & -12 & 12 \\ 0 & 12 & -16 \end{bmatrix}$

Exercises 8.4, page 257

1. $\mathbf{v}_1 = \frac{1}{\sqrt{3}}(1, -1, 0, 1)$, $\mathbf{v}_2 = \frac{1}{\sqrt{23}}(3, 2, 3, -1)$, $\mathbf{v}_3 = \frac{1}{\sqrt{15}}(-2, -1, 3, 1)$

3. $\mathbf{v}_3 = \frac{-1}{2\sqrt{15}}(2\mathbf{u}_1 + \mathbf{u}_2 - \mathbf{u}_3)$

4. (a) $\left\{ \frac{1}{5}(1, 2, -2, 4), \frac{1}{\sqrt{14}}(0, -1, 3, 2), \frac{1}{\sqrt{19}}(4, 1, 1, -1) \right\}$

(c) $\left\{ \frac{1}{3}(2, 2, 1), \frac{1}{\sqrt{2}}(-1, 1, 0), \frac{1}{3\sqrt{2}}(-1, -1, 4) \right\}$

Exercises 8.5, page 263

1. (a) $P = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & 0 & 2 \\ \sqrt{2} & \sqrt{3} & -1 \\ \sqrt{2} & -\sqrt{3} & -1 \end{bmatrix}$ (c) $P = \frac{1}{3\sqrt{2}} \begin{bmatrix} \sqrt{2} & 0 & 4 \\ -2\sqrt{2} & 3 & 1 \\ 2\sqrt{2} & 3 & -1 \end{bmatrix}$

Exercises 8.6, page 269

1. (a) $r = 2, p = 1, s = 0$ (c) $r = 3, p = 2, s = 1$

2. (a) $q(\mathbf{v}) = z_1^2 - z_2^2$ (c) $q(\mathbf{v}) = z_1^2 + z_2^2 - z_3^2$

3. (a) $P = \frac{1}{2\sqrt{3}} \begin{bmatrix} 1 & 0 & -2 \\ 1 & -\sqrt{3} & 1 \\ 1 & \sqrt{3} & 1 \end{bmatrix}$ (c) $P = \frac{1}{18} \begin{bmatrix} 2 & 0 & 4 \\ -4 & 3 & 1 \\ 4 & 3 & -1 \end{bmatrix}$

4. (a) $\mathcal{B} = \left\{ \frac{1}{2\sqrt{3}}(1, 1, 1), \frac{1}{2}(0, -1, 1), \frac{1}{2\sqrt{3}}(-2, 1, 1) \right\}$

(c) $\mathcal{B} = \left\{ \frac{1}{9}(1, -2, 2), \frac{1}{6}(0, 1, 1), \frac{1}{18}(4, 1, -1) \right\}$

Exercises 8.7, page 274

1. (a) $\begin{bmatrix} 0 & 2 & -1 \\ 2 & 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 3 & 5 & 1 \\ 5 & 5 & 2 \\ 1 & 2 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 5+4i & -5+i \\ 3-6i & 3+6i \\ 6+4i & -6 \end{bmatrix}$

2. (a) 24 (c) -7 (e) $14i$

3. (a) $\begin{bmatrix} 32 & -4 & -11 \\ -8 & 1 & -40 \end{bmatrix}$ (c) $\begin{bmatrix} 84 & 0 & -15 \\ 0 & -3 & 3 \\ -15 & 3 & -9 \end{bmatrix}$ (e) $\begin{bmatrix} -9-i & -5+i \\ -9-3i & -6+3i \\ 3-i & 3 \end{bmatrix}$

4. (a) $\mathcal{A}' = \{(1, 0, 0, 0), (-4, 1, 0, 0), (1, -1, 1, 0), (2, -1, 0, 1)\},$
 $\mathcal{B}' = \{(1, 0, 0), (0, 1, 0), (-2, 5, 1)\}$
(c) $\mathcal{A}' = \{(0, 0, 0, 1), (0, 1, 0, -2), (0, -2, 1, 0), (1, 0, 0, 0)\},$
 $\mathcal{B}' = \{(1, 0, 0), (0, 1, 0), (-1, -1, 1)\}$

Exercises 8.8, page 281

1. (a) $P = \begin{bmatrix} 1 & 0 & -6 \\ 1 & 0 & 4 \\ 0 & 1 & 5 \end{bmatrix}$ (c) $P = \begin{bmatrix} 1 & 1 & -2 \\ 1 & -1 & 3 \\ 0 & 0 & -2 \end{bmatrix}$ (e) $P = \begin{bmatrix} 1 & 1 & -3 \\ 1 & -1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$
2. (a) $\{(2, 1, 2), (0, 1, -1), (-2, -1, -7)\}$ (c) $\{(2, 1, 2), (0, 1, 0), (1, -4, 3)\}$
(e) $\{(2, 1, 2), (0, 1, 0), (-1, -2, -2)\}$

Exercises 8.9, page 291

1. Those in parts (a), (b), (d), (e), (g).

2. (a) For part (a), $P = \begin{bmatrix} 1 & 1 \\ 0 & 5i \end{bmatrix}$; for part (c), the matrix is not hermitian;
for part (e), $P = \begin{bmatrix} 1 & 1 & -4-i \\ 0 & 5i & 5-2i \\ 0 & 0 & 9 \end{bmatrix}$; for part (g), $P = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 2i \\ 0 & 11 & 1 \end{bmatrix}$
(b) In part (a), $r = p = 2$; in part (c), the matrix is not hermitian; in part (e),
 $r = p = 3$; in part (g), $r = p = 3$
(c) Those in parts (a) and (b) are conjunctive, and those in parts (e) and (g) are
conjunctive.

3. (a) $\frac{1}{3\sqrt{5}} \begin{bmatrix} 3 & 0 \\ 0 & 5i \end{bmatrix}$ (c) The matrix in 1(c) is not hermitian.
(e) $\frac{1}{3\sqrt{5}} \begin{bmatrix} 3 & 1 & -4-i \\ 0 & 5i & 5-2i \\ 0 & 0 & 9 \end{bmatrix}$ (g) $\frac{1}{\sqrt{66}} \begin{bmatrix} 3\sqrt{6} & 1 & -\sqrt{11} \\ 0 & 0 & 2i\sqrt{11} \\ 0 & 11 & \sqrt{11} \end{bmatrix}$

4. (a) Those for the matrices in parts (a), (b), (d), (e), (g).
 (b) For f in part (a), $\mathcal{A}' = \{(1, 0), (1, 5i)\}$;
 for f in part (e), $\mathcal{A}' = \{(1, 0, 0), (1, 5i, 0), (4+i, -5+2i, -9)\}$;
 for f in part (g), $\mathcal{A}' = \{(1, 0, 0), (1, 0, 11), (1, -2i, -1)\}$

Exercises 9.2, page 296

1. (a) $(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u}) = -3$
 2. (a) $(\mathbf{u}, \mathbf{v}) = 15 + 3i, (\mathbf{v}, \mathbf{u}) = 15 - 3i$
 4. (a) The rule defines an inner product on \mathbf{R}^2 .
 (c) The rule does not define an inner product on \mathbf{R}^2 . If $\mathbf{v} = (1, -1)$, then $(\mathbf{v}, \mathbf{v}) = 0$, and this contradicts the requirement that $(\mathbf{v}, \mathbf{v}) > 0$ if $\mathbf{v} \neq \mathbf{0}$.
 (e) The rule does not define an inner product on \mathbf{R}^3 . If $\mathbf{v} = (0, 1, 0)$, then $(\mathbf{v}, \mathbf{v}) = 0$, contradicting the property that requires $(\mathbf{v}, \mathbf{v}) > 0$ if $\mathbf{v} \neq \mathbf{0}$.
 5. $(\mathbf{u}, \mathbf{v}) = 5\bar{u}_1v_1 + (-5+i)\bar{u}_1v_2 + (2-i)\bar{u}_1v_3 + (-5-i)\bar{u}_2v_1 + 7\bar{u}_2v_2 + (-5+i)\bar{u}_2v_3 + (2+i)\bar{u}_3v_1 + (-5-i)\bar{u}_3v_2 + 6\bar{u}_3v_3$
 6. $\mathcal{A}' = \left\{ \frac{1}{\sqrt{5}}(1, 0, 0), \frac{1}{\sqrt{30}}(-3, -5, -5), \frac{1}{\sqrt{30}}(9, 11+2i, 5+2i) \right\}$
 9. (a) $\begin{bmatrix} 5 & i \\ -i & 2 \end{bmatrix}$ (b) $\mathcal{A} = \left\{ \left(\frac{\sqrt{5}}{5}, 0\right), \left(\frac{\sqrt{5}}{15}, \frac{i\sqrt{5}}{3}\right) \right\}$
 13. Those in parts (a), (b), (e), (g).

Exercises 9.3, page 301

1. (a) $\sqrt{11}$ (c) $\sqrt{5}$ 2. (a) $\sqrt{74}$ (c) $\sqrt{7}$
 3. (a) We have $|(\mathbf{u}, \mathbf{v})| = |-13| = 13$ and $\|\mathbf{u}\| \cdot \|\mathbf{v}\| = \sqrt{14}\sqrt{38} = \sqrt{532} = 2\sqrt{133}$. Since $13 < 2\sqrt{133}$, the inequality is verified.
 (b) $\sqrt{14}$ (c) $\sqrt{78}$
 5. $d(\mathbf{u}, \mathbf{v}) = \sqrt{|u_1 - v_1|^2 + |u_2 - v_2|^2}$, where $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$
 11. $\cos \theta = \frac{17}{2\sqrt{91}}$ 13. 3

Exercises 9.4, page 305

2. (a) $\left\{ \frac{\sqrt{2}}{6}(3, 2\sqrt{2}, -1), \frac{1}{2}(1, -\sqrt{2}, -1), \frac{1}{6}(-3, \sqrt{2}, -5) \right\}$
 4. (a) $\left\{ \frac{1}{\sqrt{3}}(1, -i, 1), \frac{1}{2\sqrt{6}}(3+i, 1+3i, -2i) \right\}$

5. $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$

7. $\{(1, 1, 1), (0, -1, -1), (-9, -11 - 2i, -5 - 2i)\}$

8. (a) $\left\{ \left(\frac{i}{2}, \frac{1}{\sqrt{2}}, \frac{i}{2} \right), \left(-\frac{i}{2}, \frac{1}{\sqrt{2}}, -\frac{i}{2} \right), \left(\frac{1+i}{2}, 0, \frac{-1-i}{2} \right) \right\}$

Exercises 9.5, page 308

4. (a) $\left\{ \frac{1}{\sqrt{5}}(2, 1, 0, 0), \frac{1}{\sqrt{170}}(3, -6, 10, 5) \right\}$ 5. (a) $\left\{ \frac{1}{\sqrt{2}}(i, 1, 0) \right\}$

Exercises 9.6, page 312

1. (a) No (c) Yes 2. (a) Yes (c) No 3. $\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \end{bmatrix}$

Exercises 9.7, page 317

1. Those in parts (a), (c), (d).
2. The matrix in (b) is orthogonal. Those in parts (b) and (e) are unitary. All are normal.
3. Only $x = y = z = 0$

Exercises 9.8, page 322

1. (a) $T^*(a_1, a_2) = (a_1 + (1 - i)a_2, (1 + i)a_1 + 2a_2)$

(c) $T^*(a_1, a_2) = (-ia_1 + a_2, ia_1)$

2. (a) $\left\{ \frac{1}{\sqrt{3}}(-1 + i, 1), \frac{1}{\sqrt{3}}(1, 1 + i) \right\}$ (c) No such basis exists.

3. All of the matrices A are normal. A matrix U of the required type is given for parts (a), (c), and (e).

(a) $U = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$ (c) $U = \frac{1}{\sqrt{3}} \begin{bmatrix} i - 1 & -i \\ i & 1 + i \end{bmatrix}$

(e) $U = \frac{1}{2\sqrt{2}} \begin{bmatrix} \sqrt{2} & 1 - i & 2 \\ \sqrt{2}i & 1 + i & -2i \\ -\sqrt{2} - \sqrt{2}i & 2 & 0 \end{bmatrix}$

Exercises 10.2, page 329

5. Let P_1, P_2 be defined on \mathbf{R}^2 by $P_1(x_1, x_2) = (x_1, x_1)$, and $P_2(x_1, x_2) = (0, x_1 + x_2)$

Exercises 10.3, page 334

2. (a) T is diagonalizable. (b) $P_1 + P_2$ is not a projection.

Exercises 10.4, page 345

2. (a) $(x - 2)^2(x + 3)$ (c) $(x - 2)^2(x + 2)$

3. $A = \begin{bmatrix} \pm\sqrt{-1-bc} & b \\ c & \mp\sqrt{-1-bc} \end{bmatrix}$, with $-bc \geq 1$.

4. (a) $T = 3P_1 - 2P_2$, where P_1, P_2 have matrices E_1, E_2 relative to \mathcal{E}_3 given by

$$E_1 = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & -1 \\ 3 & 3 & -2 \end{bmatrix}, E_2 = \begin{bmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ -3 & -3 & 3 \end{bmatrix}.$$

- (c) $T = P_1 - P_2 + 3P_3$, where P_1, P_2, P_3 have matrices E_1, E_2, E_3 relative to

$$\mathcal{E}_4 \text{ given by } E_1 = -\frac{1}{2} \begin{bmatrix} 6 & 4 & 5 & 2 \\ -12 & -8 & -10 & -4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, E_2 = \frac{1}{4} \begin{bmatrix} 2 & 1 & 2 & 0 \\ -12 & -6 & -12 & 0 \\ 8 & 4 & 8 & 0 \\ -2 & -1 & -2 & 0 \end{bmatrix},$$

$$E_3 = \frac{1}{4} \begin{bmatrix} 14 & 7 & 8 & 4 \\ -12 & -6 & -8 & -8 \\ -8 & -4 & -4 & 0 \\ 2 & 1 & 2 & 4 \end{bmatrix}$$

5. (a) $A = D + N$, where the diagonalizable matrix D and the nilpotent matrix N

are given by $D = \begin{bmatrix} 5 & 2 & 2 & 2 \\ 4 & 3 & 4 & -4 \\ -8 & -4 & -5 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$, and $N = \begin{bmatrix} 2 & 1 & 1 & 0 \\ -4 & -2 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 \end{bmatrix}$.

(c) $A = D + N$, where the diagonalizable matrix D and the nilpotent matrix N

$$\text{are given by } D = \frac{1}{2} \begin{bmatrix} 2 & 3 & 0 & -1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ -1 & 0 & 3 & 2 \end{bmatrix}, \text{ and } N = \frac{1}{2} \begin{bmatrix} 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{bmatrix}.$$

Exercises 10.5, page 354

1. (a) $\mathcal{B} = \{(-4, 3, -1, 0), (-5, 3, -1, 0), (3, -2, 1, 0), (0, 0, 0, 1)\}$,

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

- (c) $\mathcal{B} = \{(-2, 4, 0, 0, 2), (-1, 0, 2, 1, 1), (1, 0, 0, 0, 0); (1, -2, 0, 1, -1), (0, 0, 0, 0, 1)\}$,

$$A = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{bmatrix} \text{ with } A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and } A_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

2. (a) $\mathcal{B}(\mathbf{v}, T) = \{(4, -3, 1, 0), (1, 0, 0, 0), (0, -1, 0, 0), (1, 1, 0, -1)\}$,

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

- (c) $\mathcal{B}(\mathbf{v}, T) = \{(2, -4, 0, 0, -2), (1, 0, -2, 0, -1), (-1, -1, 1, 0, 1)\}$,

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

Exercises 10.6, page 361

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

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

$$(c) \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

3. (a) $\mathcal{B}' = \{(1, -2, 0, 1), (0, 2, -1, -1), (1, -2, 0, 0), (0, 1, -1, 0)\}$,

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

- (c) $\mathcal{B}' = \{(1, -1, 1, -1), (2, -1, 0, 1), (-1, -1, -1, -1), (3, 2, 1, 0)\}$,

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

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

(c) $A = \begin{bmatrix} 1 & \frac{3}{2} & 0 & -\frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{3}{2} & 1 \end{bmatrix} + \begin{bmatrix} 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$.

5. D has matrix K and N has matrix L relative to \mathcal{E}_6 , where

$$K = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 & 0 & 0 \\ 1 & -1 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 & 0 & -2 \end{bmatrix} \text{ and } L = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -2 & 0 & 1 & 1 & 1 & 0 \\ 2 & 1 & -1 & -1 & -1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Index

A

Addition
of equations, 134
of linear transformations, 148, 177
of matrices, 117
of subsets, 16
of vectors, 2
Adjoint
of a linear operator, 318
of a matrix, 209
Algebraic multiplicity, 225
Annihilator, 245
Associated
direct sum decomposition, 327
eigenvector, 213, 215
Augmented matrix, 137

B

Basis, 29
cyclic, 348
dual, 241
orthogonal, 254
standard, 33, 54, 131
Bijective mapping, 128
Bilinear form, 270
complex, 283
hermitian complex, 286
matrix of, 271
rank of, 273
skew-symmetric, 282
symmetric, 277
Binary relation, 99

C

Canonical form
for quadratic form, 266

Jordan, 359
Cartesian product, 270
Cauchy-Schwarz inequality, 299
Change of variable
linear, 251
orthogonal, 255
Characteristic
equation, 214
matrix, 214
polynomial, 214
root, 213
value, 213
vector, 213
Coefficient matrix, 137
Cofactor, 196
Column
equivalent matrices, 99
matrix, 60
operation, 83
space, 106
vectors, 106
Companion matrix, 238
Complete set of projections, 327
Complex
bilinear form, 283
inner product space, 293
Components, 1
Conformable, 67
Congruence of matrices, 251
Conjugate
of a matrix, 259
transpose, 259
Conjunctive matrices, 285
Coordinate, 33, 62
matrix, 62
projections, 241

Corresponding
 direct sum decomposition, 327
 eigenvector, 213, 216
 Cramer's rule, 203
 Cross product, 28
 Cyclic
 basis, 348
 subspace, 348

D

Determinant, 191
 of order n , 192
 Diagonal
 block matrix, 348
 elements, 60
 matrix, 60
 Dimension, 33, 59, 115
 Direct sum, 39
 associated, 327
 Directed line segment, 19
 Distance, 301
 Division algorithm, 336
 Divisor, 336
 Dot product, 24
 Dual
 basis, 241
 space, 242

E

Eigenspace, 223
 Eigenvalue, 213–214
 Eigenvector, 213, 215
 Elementary
 column operation, 83
 matrix, 77
 operation on vectors, 42
 row operation, 91
 Equality
 of equations, 134
 of functions, 117
 of indexed sets, 29
 of mappings, 128
 of matrices, 60
 of polynomials, 116
 of vectors, 1

Equivalence, 106
 column, 99
 relation, 99
 row, 103
 Equivalent
 matrices, 106
 systems of equations, 135
 Euclidean space, 293

F

Field, 113
 Finite-dimensional vector space, 115
 Form
 bilinear, 270
 complex bilinear, 283
 hermitian, 287
 quadratic, 247, 280
 Function, 128

G

Gauss-Jordan elimination, 139
 Geometric multiplicity, 223
 Gram-Schmidt process, 255, 304
 Greatest common divisor, 336

H

Hamilton-Cayley Theorem, 338
 Hermitian
 complex bilinear form, 286
 congruent matrices, 285
 form, 287
 matrix, 286
 operator, 319

I

Idempotent
 linear operator, 326
 principal, 345
 Identity
 matrix, 60, 74
 operation, 42
 Image, 128, 145
 inverse, 128, 145
 Index
 of a permutation, 188

- of an integer, 187
- of hermitian form, 288
- of nilpotency of a matrix, 355
- of nilpotency of an operator, 344
- of quadratic form, 266
- of real symmetric matrix, 267
- set, 13
- Inequality**
 - Cauchy-Schwarz, 299
- Infinite-dimensional vector space, 115
- Injective mapping, 128
- Inner product, 24, 293
 - generated by a matrix, 296
 - space, 293
 - standard, 295
- Invariant subspace, 341
- Inverse**
 - image, 128, 145
 - of elementary operation, 43
 - of linear transformation, 181
 - of matrix, 74
- Invertible linear transformation, 181
- Invertible matrix, 74
- Isometry, 309
- Isomorphic vector spaces, 128
- Isomorphism, 128
- J**
- Jordan
 - canonical form, 359
 - canonical matrix, 356
 - matrix, 356
- K**
- Kernel, 151
- Kronecker delta, 8, 41
- L**
- Leading one, 51, 54, 84, 93
- Leading variables, 139
- Length, 24, 298
- Linear
 - change of variable, 251
 - combination, 3
 - dependence, 7
- functional, 239
- operator, 146
- transformation, 145
- Linearly
 - dependent, 3, 5, 7
 - independent, 5
- Lower triangular matrix, 313
- M**
- Mapping, 128
 - bijective, 128
 - injective, 128
 - one-to-one, 128
 - onto, 128
 - surjective, 128
- Matrix, 59
 - augmented, 137
 - characteristic, 214
 - coefficient, 137
 - column, 60
 - companion, 238
 - congruence, 251
 - conjugate, 259
 - conjugate transpose, 259
 - coordinate, 62
 - diagonal, 60
 - diagonal block, 348
 - elementary, 77
 - hermitian, 286
 - identity, 60, 74
 - inverse, 74
 - invertible, 74
 - Jordan, 356
 - lower triangular, 313
 - nilpotent, 345, 355
 - nonsingular, 72
 - normal, 316
 - of bilinear form, 271
 - of complex bilinear form, 283
 - of constants, 137
 - of linear transformation, 157
 - of quadratic form, 249, 281
 - of unknowns, 137
 - orthogonal, 254
 - polynomial, 337

- positive definite, 269
- positive semidefinite, 269
- projection, 345
- reduced column-echelon, 84, 86
- row, 60
- scalar, 83
- similarity, 225
- singular, 72
- skew-hermitian, 292
- skew-symmetric, 92, 211, 253
- square, 60
- symmetric, 92
- transformation, 147
- transition, 61
- unitary, 304
- upper triangular, 313
- zero, 60, 117
- Metric, 301
 - space, 301
- Minimal polynomial, 338
- Minor, 196
- Monic polynomial, 336
- Multiplication
 - of a scalar and a matrix, 83, 117
 - of a scalar and a vector, 2, 114
 - of a scalar and an equation, 134
 - of linear transformations, 178
 - of matrices, 66
- Multiplicity
 - algebraic, 225
 - geometric, 223
- N**
- Natural ordering, 187
- Negative
 - definite hermitian form, 288
 - definite quadratic form, 267
 - semidefinite hermitian form, 288
 - semidefinite quadratic form, 267
- Nilpotent
 - linear operator, 344
 - matrix, 345, 355
- Nonnegative linear operator, 335
- Nonsingular
 - linear transformation, 181
- matrix, 72
- Norm, 24, 298, 301
- Normal
 - linear operator, 319
 - matrix, 316
- Normalized set of vectors, 254, 305
- Nullity, 151
- Number field, 114
- O**
- One-to-one mapping, 128
- Order
 - of a determinant, 192
 - of a matrix, 60
- Orthogonal
 - basis, 254
 - change of basis, 255
 - change of variable, 255
 - complement, 307
 - linear operator, 258, 309
 - matrix, 254
 - operator, 309
 - projection, 308, 325
 - set, 27, 303
 - set of projections, 327
 - similarity, 313
 - vectors, 26–27, 303
- Orthogonally similar matrices, 313
- Orthonormal set of vectors, 254, 303
- P**
- Parallelogram rule, 20
- Parameters, 139
- Permutation, 187
- Perp, 306
- Polynomial
 - characteristic, 214
 - matrix, 337
 - minimal, 338
 - monic, 336
- Positive
 - definite hermitian form, 288
 - definite quadratic form, 267
 - semidefinite hermitian form, 288
 - semidefinite quadratic form, 267

Principal idempotents, 345

Product

cartesian, 270

cross, 28

dot, 24

inner, 24, 293

of a scalar and a linear transformation, 148, 177

of a scalar and a matrix, 83, 117

of a scalar and a vector, 2, 114

of linear transformations, 178

of matrices, 66

scalar, 24

Projection, 26, 241, 325–326

matrices, 345

orthogonal, 308, 325

scalar, 26

Proper

number, 213

value, 213

vector, 213

Q

Quadratic form, 247, 280

index of, 266

matrix of, 281

negative definite, 267

negative semidefinite, 267

positive definite, 267

positive semidefinite, 267

rank of, 264

real, 247

signature of, 266

R

Range, 150

Rank

of bilinear form, 273

of linear transformation, 150

of matrix, 106

of real quadratic form, 264

Real

coordinate space, 1

coordinate vectors, 1

inner product space, 293

quadratic form, 247

Reduced column-echelon form, 84, 86

Reduced row-echelon form, 93, 95

Reflexive property, 99

Relation

binary, 99

equivalence, 99

Restriction of a linear operator, 341

Row

equivalent matrices, 103

matrix, 60

operation, 91

S

Scalar, 1, 114

component, 26

matrix, 83

multiple, 2, 83

multiplication, 2, 114, 134

product, 24

projection, 26

Self-adjoint linear operator, 319

Signature

of hermitian form, 288

of quadratic form, 266

Similar matrices, 225

Singular matrix, 72

Skew-adjoint linear operator, 323

Skew-hermitian

linear operator, 323

matrix, 292

Skew-symmetric

bilinear form, 282

linear operator, 323

matrix, 92, 211, 253

Solution

of a system of equations, 134

set, 134

Space

column, 106

complex inner product, 293

dual, 242

Euclidean, 293

metric, 301

real coordinate, 1
 real inner product, 293
 unitary, 293
 vector, 114

Span, 13

Spectral decomposition
 of a linear operator, 332
 of a matrix, 345

Spectrum, 213–214

Square matrix, 60

Square root of a linear operator, 335

Standard

basis, 33
 basis of subspace, 54, 131
 inner products, 295

Submatrix, 211

Subspace, 10, 122

cyclic, 348
 invariant, 341
 spanned by a set, 15, 123

Sum

direct, 39
 of linear transformations, 148
 of matrices, 117
 of subsets, 16
 of vectors, 2

Superdiagonal elements, 350

Surjective mapping, 128

Symmetric

bilinear form, 277

Symmetric

matrix, 92
 operator, 319
 property, 99

T

Trace, 155, 231, 240, 297

Transformation, 128

linear, 145

matrix, 147

Transition matrix, 61

Transitive property, 99

Transpose, 92

U

Unit vector, 27

Unitarily similar matrices, 313

Unitary

matrix, 304
 operator, 309
 similarity, 313
 space, 293

Upper triangular matrix, 313

V

Vector, 1, 114

column, 106
 component, 1, 26
 geometric interpretation, 19
 norm, 24, 298
 projection, 26
 space, 114
 unit, 27

Z

Zero

linear transformation, 146

matrix, 60, 117

subspace, 11