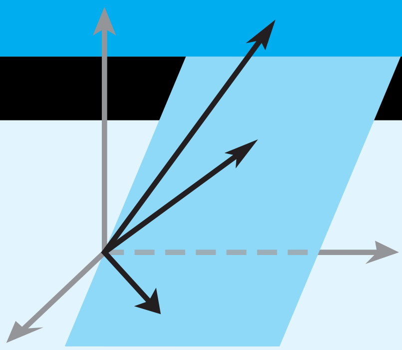


CHAPTER 5

Eigenvalues and Eigenvectors

CHAPTER CONTENTS

INTRODUCTION

In this chapter we will focus on classes of scalars and vectors known as "eigenvalues" and "eigenvectors," terms derived from the German word *eigen*, meaning "own," "peculiar to," "characteristic," or "individual." The underlying idea first appeared in the study of rotational motion but was later used to classify various kinds of surfaces and to describe solutions of certain differential equations. In the early 1900s it was applied to matrices and matrix transformations, and today it has applications in such diverse fields as computer graphics, mechanical vibrations, heat flow, population dynamics, quantum mechanics, and economics, to name just a few.

5.1 Eigenvalues and Eigenvectors

In this section we will define the notions of "eigenvalue" and "eigenvector" and discuss some of their basic properties.

Definition of Eigenvalue and Eigenvector

We begin with the main definition in this section.

DEFINITION 1 If A is an $n \times n$ matrix, then a nonzero vector $\mathbf{x}$ in R^n is called an ***eigenvector*** of A (or of the matrix operator T_A) if $A\mathbf{x}$ is a scalar multiple of $\mathbf{x}$; that is,

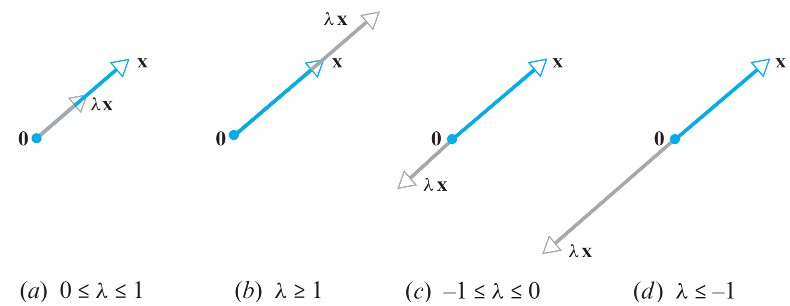
$$A\mathbf{x} = \lambda\mathbf{x}$$

for some scalar λ. The scalar λ is called an ***eigenvalue*** of A (or of T_A), and $\mathbf{x}$ is said to be an ***eigenvector corresponding to*** $\boldsymbol{\lambda}$.

> The requirement that an eigenvector be nonzero is imposed to avoid the unimportant case $A\mathbf{0} = \lambda\mathbf{0}$, which holds for every A and λ.

In general, the image of a vector $\mathbf{x}$ under multiplication by a square matrix A differs from $\mathbf{x}$ in both magnitude and direction. However, in the special case where $\mathbf{x}$ is an eigenvector of A, multiplication by A leaves the direction unchanged. For example, in R^2 or R^3 multiplication by A maps each eigenvector $\mathbf{x}$ of A (if any) along the same line through the origin as $\mathbf{x}$. Depending on the sign and magnitude of the eigenvalue λ

corresponding to $\mathbf{x}$, the operation $A\mathbf{x} = \lambda\mathbf{x}$ compresses or stretches $\mathbf{x}$ by a factor of λ, with a reversal of direction in the case where λ is negative (Figure 5.1.1).



▶ Figure 5.1.1

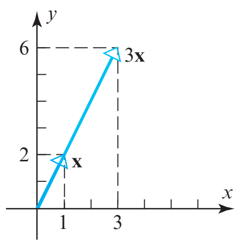
▶ EXAMPLE 1 Eigenvector of a 2 × 2 Matrix

The vector $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector of

$$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

corresponding to the eigenvalue $\lambda = 3$, since

$$A\mathbf{x} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3\mathbf{x}$$



▲ Figure 5.1.2

Geometrically, multiplication by A has stretched the vector $\mathbf{x}$ by a factor of 3 (Figure 5.1.2). ◀

Computing Eigenvalues and Eigenvectors

Our next objective is to obtain a general procedure for finding eigenvalues and eigenvectors of an $n \times n$ matrix A. We will begin with the problem of finding the eigenvalues of A. Note first that the equation $A\mathbf{x} = \lambda\mathbf{x}$ can be rewritten as $A\mathbf{x} = \lambda I\mathbf{x}$, or equivalently, as

$$(\lambda I - A)\mathbf{x} = \mathbf{0}$$

For λ to be an eigenvalue of A this equation must have a nonzero solution for $\mathbf{x}$. But it follows from parts (*b*) and (*g*) of Theorem 4.10.2 that this is so if and only if the coefficient matrix $\lambda I - A$ has a zero determinant. Thus, we have the following result.

Note that if $(A)_{ij} = a_{ij}$, then formula (1) can be written in expanded form as

$$\begin{vmatrix} \lambda - a_{11} & a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} = 0$$

THEOREM 5.1.1 *If A is an $n \times n$ matrix, then λ is an eigenvalue of A if and only if it satisfies the equation*

$$\det(\lambda I - A) = 0 \qquad (1)$$

This is called the ***characteristic equation*** *of A.*

▶ EXAMPLE 2 Finding Eigenvalues

In Example 1 we observed that $\lambda = 3$ is an eigenvalue of the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

but we did not explain how we found it. Use the characteristic equation to find all eigenvalues of this matrix.

Solution It follows from Formula (1) that the eigenvalues of A are the solutions of the equation $\det(\lambda I - A) = 0$, which we can write as

$$\begin{vmatrix} \lambda - 3 & 0 \\ -8 & \lambda + 1 \end{vmatrix} = 0$$

from which we obtain

$$(\lambda - 3)(\lambda + 1) = 0 \tag{2}$$

This shows that the eigenvalues of A are $\lambda = 3$ and $\lambda = -1$. Thus, in addition to the eigenvalue $\lambda = 3$ noted in Example 1, we have discovered a second eigenvalue $\lambda = -1$. ◀

When the determinant $\det(\lambda I - A)$ in (1) is expanded, the characteristic equation of A takes the form

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0 \tag{3}$$

where the left side of this equation is a polynomial of degree n in which the coefficient of λ^n is 1 (Exercise 37). The polynomial

$$p(\lambda) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n \tag{4}$$

is called the ***characteristic polynomial*** of A. For example, it follows from (2) that the characteristic polynomial of the 2×2 matrix in Example 2 is

$$p(\lambda) = (\lambda - 3)(\lambda + 1) = \lambda^2 - 2\lambda - 3$$

which is a polynomial of degree 2.

Since a polynomial of degree n has at most n distinct roots, it follows from (3) that the characteristic equation of an $n \times n$ matrix A has at most n distinct solutions and consequently the matrix has at most n distinct eigenvalues. Since some of these solutions may be complex numbers, it is possible for a matrix to have complex eigenvalues, even if that matrix itself has real entries. We will discuss this issue in more detail later, but for now we will focus on examples in which the eigenvalues are real numbers.

▶ EXAMPLE 3 Eigenvalues of a 3 × 3 Matrix

Find the eigenvalues of

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

Solution The characteristic polynomial of A is

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -4 & 17 & \lambda - 8 \end{bmatrix} = \lambda^3 - 8\lambda^2 + 17\lambda - 4$$

The eigenvalues of A must therefore satisfy the cubic equation

$$\lambda^3 - 8\lambda^2 + 17\lambda - 4 = 0 \tag{5}$$

To solve this equation, we will begin by searching for integer solutions. This task can be simplified by exploiting the fact that all integer solutions (if there are any) of a polynomial equation with *integer coefficients*

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

must be divisors of the constant term, c_n. Thus, the only possible integer solutions of (5) are the divisors of -4, that is, $\pm 1, \pm 2, \pm 4$. Successively substituting these values in (5) shows that $\lambda = 4$ is an integer solution and hence that $\lambda - 4$ is a factor of the left side of (5). Dividing $\lambda - 4$ into $\lambda^3 - 8\lambda^2 + 17\lambda - 4$ shows that (5) can be rewritten as

$$(\lambda - 4)(\lambda^2 - 4\lambda + 1) = 0$$

In applications involving large matrices it is often not feasible to compute the characteristic equation directly, so other methods must be used to find eigenvalues. We will consider such methods in Chapter 9.

Thus, the remaining solutions of (5) satisfy the quadratic equation

$$\lambda^2 - 4\lambda + 1 = 0$$

which can be solved by the quadratic formula. Thus, the eigenvalues of A are

$$\lambda = 4, \quad \lambda = 2 + \sqrt{3}, \quad \text{and} \quad \lambda = 2 - \sqrt{3}$$

▶ EXAMPLE 4 Eigenvalues of an Upper Triangular Matrix

Find the eigenvalues of the upper triangular matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

Solution Recalling that the determinant of a triangular matrix is the product of the entries on the main diagonal (Theorem 2.1.2), we obtain

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & -a_{14} \\ 0 & \lambda - a_{22} & -a_{23} & -a_{24} \\ 0 & 0 & \lambda - a_{33} & -a_{34} \\ 0 & 0 & 0 & \lambda - a_{44} \end{bmatrix}$$
$$= (\lambda - a_{11})(\lambda - a_{22})(\lambda - a_{33})(\lambda - a_{44})$$

Thus, the characteristic equation is

$$(\lambda - a_{11})(\lambda - a_{22})(\lambda - a_{33})(\lambda - a_{44}) = 0$$

and the eigenvalues are

$$\lambda = a_{11}, \quad \lambda = a_{22}, \quad \lambda = a_{33}, \quad \lambda = a_{44}$$

which are precisely the diagonal entries of A. ◀

The following general theorem should be evident from the computations in the preceding example.

THEOREM 5.1.2 *If A is an $n \times n$ triangular matrix (upper triangular, lower triangular, or diagonal), then the eigenvalues of A are the entries on the main diagonal of A.*

EXAMPLE 5 Eigenvalues of a Lower Triangular Matrix

By inspection, the eigenvalues of the lower triangular matrix

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

are $\lambda = \frac{1}{2}$, $\lambda = \frac{2}{3}$, and $\lambda = -\frac{1}{4}$. ◀

Had Theorem 5.1.2 been available earlier, we could have anticipated the result obtained in Example 2.

The following theorem gives some alternative ways of describing eigenvalues.

THEOREM 5.1.3 *If A is an $n \times n$ matrix, the following statements are equivalent.*

(*a*) *λ is an eigenvalue of A.*

(*b*) *λ is a solution of the characteristic equation* $\det(\lambda I - A) = 0$.

(*c*) *The system of equations $(\lambda I - A)\mathbf{x} = \mathbf{0}$ has nontrivial solutions.*

(*d*) *There is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$.*

Finding Eigenvectors and Bases for Eigenspaces

Now that we know how to find the eigenvalues of a matrix, we will consider the problem of finding the corresponding eigenvectors. By definition, the eigenvectors of A corresponding to an eigenvalue λ are the nonzero vectors that satisfy

$$(\lambda I - A)\mathbf{x} = \mathbf{0}$$

Thus, we can find the eigenvectors of A corresponding to λ by finding the nonzero vectors in the solution space of this linear system. This solution space, which is called the ***eigenspace*** of A corresponding to λ, can also be viewed as:

1. the null space of the matrix $\lambda I - A$
2. the kernel of the matrix operator $T_{\lambda I - A}\colon R^n \to R^n$
3. the set of vectors for which $A\mathbf{x} = \lambda\mathbf{x}$

Notice that $\mathbf{x} = \mathbf{0}$ is in every eigenspace but is not an eigenvector (see Definition 1). In the exercises we will ask you to show that this is the *only* vector that distinct eigenspaces have in common.

EXAMPLE 6 Bases for Eigenspaces

Find bases for the eigenspaces of the matrix

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

Historical Note Methods of linear algebra are used in the emerging field of computerized face recognition. Researchers are working with the idea that every human face in a racial group is a combination of a few dozen primary shapes. For example, by analyzing three-dimensional scans of many faces, researchers at Rockefeller University have produced both an average head shape in the Caucasian group—dubbed the ***meanhead*** (top row left in the figure to the left)—and a set of standardized variations from that shape, called ***eigenheads*** (15 of which are shown in the picture). These are so named because they are eigenvectors of a certain matrix that stores digitized facial information. Face shapes are represented mathematically as linear combinations of the eigenheads.

[*Image: © Dr. Joseph J. Atick, adapted from Scientific American*]

Solution The characteristic equation of A is

$$\begin{vmatrix} \lambda+1 & -3 \\ -2 & \lambda \end{vmatrix} = \lambda(\lambda+1) - 6 = (\lambda-2)(\lambda+3) = 0$$

so the eigenvalues of A are $\lambda = 2$ and $\lambda = -3$. Thus, there are two eigenspaces of A, one for each eigenvalue.

By definition,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is an eigenvector of A corresponding to an eigenvalue λ if and only if $(\lambda I - A)\mathbf{x} = \mathbf{0}$, that is,

$$\begin{bmatrix} \lambda+1 & -3 \\ -2 & \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

In the case where $\lambda = 2$ this equation becomes

$$\begin{bmatrix} 3 & -3 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

whose general solution is

$$x_1 = t, \quad x_2 = t$$

(verify). Since this can be written in matrix form as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

it follows that

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

is a basis for the eigenspace corresponding to $\lambda = 2$. We leave it for you to follow the pattern of these computations and show that

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

is a basis for the eigenspace corresponding to $\lambda = -3$. ◀

Figure 5.1.3 illustrates the geometric effect of multiplication by the matrix A in Example 6. The eigenspace corresponding to $\lambda = 2$ is the line L_1 through the origin and the point (1, 1), and the eigenspace corresponding to $\lambda = 3$ is the line L_2 through the origin and the point $(-\frac{3}{2}, 1)$. As indicated in the figure, multiplication by A maps each vector in L_1 back into L_1, scaling it by a factor of 2, and it maps each vector in L_2 back into L_2, scaling it by a factor of -3.

▶ EXAMPLE 7 Eigenvectors and Bases for Eigenspaces

Find bases for the eigenspaces of

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

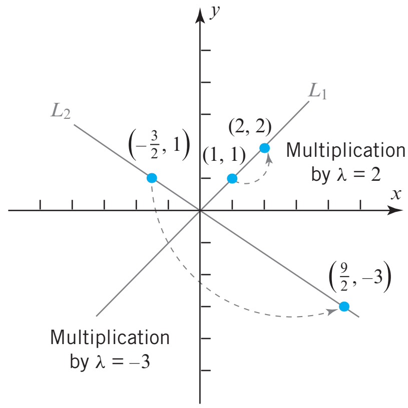


Figure 5.1.3

Solution The characteristic equation of A is $\lambda^3 - 5\lambda^2 + 8\lambda - 4 = 0$, or in factored form, $(\lambda - 1)(\lambda - 2)^2 = 0$ (verify). Thus, the distinct eigenvalues of A are $\lambda = 1$ and $\lambda = 2$, so there are two eigenspaces of A.

By definition,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

is an eigenvector of A corresponding to λ if and only if $\mathbf{x}$ is a nontrivial solution of $(\lambda I - A)\mathbf{x} = \mathbf{0}$, or in matrix form,

$$\begin{bmatrix} \lambda & 0 & 2 \\ -1 & \lambda - 2 & -1 \\ -1 & 0 & \lambda - 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{6}$$

In the case where $\lambda = 2$, Formula (6) becomes

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

Solving this system using Gaussian elimination yields (verify)

$$x_1 = -s, \quad x_2 = t, \quad x_3 = s$$

Thus, the eigenvectors of A corresponding to $\lambda = 2$ are the nonzero vectors of the form

$$\mathbf{x} = \begin{bmatrix} -s \\ t \\ s \end{bmatrix} = \begin{bmatrix} -s \\ 0 \\ s \end{bmatrix} + \begin{bmatrix} 0 \\ t \\ 0 \end{bmatrix} = s\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Since

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

are linearly independent (why?), these vectors form a basis for the eigenspace corresponding to $\lambda = 2$.

If $\lambda = 1$, then (6) becomes

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

Solving this system yields (verify)

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

Thus, the eigenvectors corresponding to $\lambda = 1$ are the nonzero vectors of the form

$$\begin{bmatrix} -2s \\ s \\ s \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad \text{so that} \quad \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = 1$. ◀

Eigenvalues and Invertibility

The next theorem establishes a relationship between the eigenvalues and the invertibility of a matrix.

THEOREM 5.1.4 *A square matrix A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A.*

Proof Assume that A is an $n \times n$ matrix and observe first that $\lambda = 0$ is a solution of the characteristic equation

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

if and only if the constant term c_n is zero. Thus, it suffices to prove that A is invertible if and only if $c_n \neq 0$. But

$$\det(\lambda I - A) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n$$

or, on setting $\lambda = 0$,

$$\det(-A) = c_n \quad \text{or} \quad (-1)^n \det(A) = c_n$$

It follows from the last equation that $\det(A) = 0$ if and only if $c_n = 0$, and this in turn implies that A is invertible if and only if $c_n \neq 0$. ◀

▶ EXAMPLE 8 Eigenvalues and Invertibility

The matrix A in Example 7 is invertible since it has eigenvalues $\lambda = 1$ and $\lambda = 2$, neither of which is zero. We leave it for you to check this conclusion by showing that $\det(A) \neq 0$. ◀

More on the Equivalence Theorem

As our final result in this section, we will use Theorem 5.1.4 to add one additional part to Theorem 4.10.2.

THEOREM 5.1.5 Equivalent Statements

If A is an $n \times n$ matrix, then the following statements are equivalent.

(a) *A is invertible.*

(b) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*

(c) *The reduced row echelon form of A is I_n.*

(d) *A is expressible as a product of elementary matrices.*

(e) *$A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.*

(f) *$A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.*

(g) $\det(A) \neq 0$.

(h) *The column vectors of A are linearly independent.*

(i) *The row vectors of A are linearly independent.*

(j) *The column vectors of A span R^n.*

(k) *The row vectors of A span R^n.*

(l) *The column vectors of A form a basis for R^n.*

(m) *The row vectors of A form a basis for R^n.*

(n) *A has rank n.*

(o) *A has nullity 0.*

(p) *The orthogonal complement of the null space of A is R^n.*

(q) *The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.*

(r) *The kernel of T_A is $\{\mathbf{0}\}$.*

(s) *The range of T_A is R^n.*

(t) *T_A is one-to-one.*

(u) *$\lambda = 0$ is not an eigenvalue of A.*

Eigenvalues of General Linear Transformations

Thus far, we have only defined eigenvalues and eigenvectors for matrices and linear operators on R^n. The following definition, which parallels Definition 1, extends this concept to general vector spaces.

DEFINITION 2 If $T: V \to V$ is a linear operator on a vector space V, then a nonzero vector $\mathbf{x}$ in V is called an ***eigenvector*** of T if $T(\mathbf{x})$ is a scalar multiple of $\mathbf{x}$; that is,

$$T(\mathbf{x}) = \lambda\mathbf{x}$$

for some scalar λ. The scalar λ is called an ***eigenvalue*** of T, and $\mathbf{x}$ is said to be an ***eigenvector corresponding to*** $\boldsymbol{\lambda}$.

As with matrix operators, we call the kernel of the operator $\lambda I - A$ the ***eigenspace*** of T corresponding to λ. Stated another way, this is the subspace of all vectors in V for which $T(\mathbf{x}) = \lambda\mathbf{x}$.

CALCULUS REQUIRED

In vector spaces of functions eigenvectors are commonly referred to as ***eigenfunctions***.

▶ EXAMPLE 9 Eigenvalue of a Differentiation Operator

If $D: C^\infty \to C^\infty$ is the differentiation operator on the vector space of functions with continuous derivatives of all orders on the interval $(-\infty, \infty)$, and if λ is a constant, then

$$D(e^{\lambda x}) = \lambda e^{\lambda x}$$

so that λ is an eigenvalue of D and $e^{\lambda x}$ is a corresponding eigenvector. ◀

Exercise Set 5.1

▶ In Exercises 1–4, confirm by multiplication that $\mathbf{x}$ is an eigenvector of A, and find the corresponding eigenvalue. ◀

1. $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

2. $A = \begin{bmatrix} 5 & -1 \\ 1 & 3 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

3. $A = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

4. $A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

▶ In each part of Exercises 5–6, find the characteristic equation, the eigenvalues, and bases for the eigenspaces of the matrix. ◀

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

(c) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$

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

(c) $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix}$

▶ In Exercises 7–12, find the characteristic equation, the eigenvalues, and bases for the eigenspaces of the matrix. ◀

7. $\begin{bmatrix} 4 & 0 & 1 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$ **8.** $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 4 \end{bmatrix}$

9. $\begin{bmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ 1 & 0 & -3 \end{bmatrix}$ **10.** $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

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

▶ In Exercises 13–14, find the characteristic equation of the matrix by inspection. ◀

13. $\begin{bmatrix} 3 & 0 & 0 \\ -2 & 7 & 0 \\ 4 & 8 & 1 \end{bmatrix}$ **14.** $\begin{bmatrix} 9 & -8 & 6 & 3 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}$

▶ In Exercises 15–16, find the eigenvalues and a basis for each eigenspace of the linear operator defined by the stated formula. [*Suggestion:* Work with the standard matrix for the operator.] ◀

15. $T(x, y) = (x + 4y, 2x + 3y)$

16. $T(x, y, z) = (2x - y - z,\ x - z,\ -x + y + 2z)$

17. (***Calculus required***) Let $D^2: C^\infty(-\infty, \infty) \to C^\infty(-\infty, \infty)$ be the operator that maps a function into its second derivative.

(a) Show that D^2 is linear.

(b) Show that if ω is a positive constant, then $\sin\sqrt{\omega}x$ and $\cos\sqrt{\omega}x$ are eigenvectors of D^2, and find their corresponding eigenvalues.

18. (***Calculus required***) Let $D^2: C^\infty \to C^\infty$ be the linear operator in Exercise 17. Show that if ω is a positive constant, then $\sinh\sqrt{\omega}x$ and $\cosh\sqrt{\omega}x$ are eigenvectors of D^2, and find their corresponding eigenvalues.

▶ In each part of Exercises 19–20, find the eigenvalues and the corresponding eigenspaces of the stated matrix operator on R^2. Refer to the tables in Section 4.9 and use geometric reasoning to find the answers. No computations are needed. ◀

19. (a) Reflection about the line $y = x$.

(b) Orthogonal projection onto the x-axis.

(c) Rotation about the origin through a positive angle of 90°.

(d) Contraction with factor k $(0 \le k < 1)$.

(e) Shear in the x-direction by a factor k $(k \ne 0)$.

20. (a) Reflection about the y-axis.

(b) Rotation about the origin through a positive angle of 180°.

(c) Dilation with factor k $(k > 1)$.

(d) Expansion in the y-direction with factor k $(k > 1)$.

(e) Shear in the y-direction by a factor k $(k \ne 0)$.

▶ In each part of Exercises 21–22, find the eigenvalues and the corresponding eigenspaces of the stated matrix operator on R^3. Refer to the tables in Section 4.9 and use geometric reasoning to find the answers. No computations are needed. ◀

21. (a) Reflection about the xy-plane.

(b) Orthogonal projection onto the xz-plane.

(c) Counterclockwise rotation about the positive x-axis through an angle of 90°.

(d) Contraction with factor k $(0 \le k < 1)$.

22. (a) Reflection about the xz-plane.

(b) Orthogonal projection onto the yz-plane.

(c) Counterclockwise rotation about the positive y-axis through an angle of 180°.

(d) Dilation with factor k $(k > 1)$.

23. Let A be a 2×2 matrix, and call a line through the origin of R^2 ***invariant*** under A if $A\mathbf{x}$ lies on the line when $\mathbf{x}$ does. Find equations for all lines in R^2, if any, that are invariant under the given matrix.

(a) $A = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

24. Find $\det(A)$ given that A has $p(\lambda)$ as its characteristic polynomial.

(a) $p(\lambda) = \lambda^3 - 2\lambda^2 + \lambda + 5$

(b) $p(\lambda) = \lambda^4 - \lambda^3 + 7$

[*Hint:* See the proof of Theorem 5.1.4.]

25. Suppose that the characteristic polynomial of some matrix A is found to be $p(\lambda) = (\lambda - 1)(\lambda - 3)^2(\lambda - 4)^3$. In each part, answer the question and explain your reasoning.

(a) What is the size of A?

(b) Is A invertible?

(c) How many eigenspaces does A have?

26. The eigenvectors that we have been studying are sometimes called ***right eigenvectors*** to distinguish them from ***left eigenvectors***, which are $n \times 1$ column matrices $\mathbf{x}$ that satisfy the equation $\mathbf{x}^T A = \mu \mathbf{x}^T$ for some scalar μ. For a given matrix A, how are the right eigenvectors and their corresponding eigenvalues related to the left eigenvectors and their corresponding eigenvalues?

27. Find a 3×3 matrix A that has eigenvalues 1, -1, and 0, and for which

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

are their corresponding eigenvectors.

Working with Proofs

28. Prove that the characteristic equation of a 2×2 matrix A can be expressed as $\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0$, where $\operatorname{tr}(A)$ is the trace of A.

29. Use the result in Exercise 28 to show that if

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

then the solutions of the characteristic equation of A are

$$\lambda = \tfrac{1}{2}\left[(a + d) \pm \sqrt{(a - d)^2 + 4bc}\,\right]$$

Use this result to show that A has

(a) two distinct real eigenvalues if $(a - d)^2 + 4bc > 0$.

(b) two repeated real eigenvalues if $(a - d)^2 + 4bc = 0$.

(c) complex conjugate eigenvalues if $(a - d)^2 + 4bc < 0$.

30. Let A be the matrix in Exercise 29. Show that if $b \neq 0$, then

$$\mathbf{x}_1 = \begin{bmatrix} -b \\ a - \lambda_1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -b \\ a - \lambda_2 \end{bmatrix}$$

are eigenvectors of A that correspond, respectively, to the eigenvalues

$$\lambda_1 = \tfrac{1}{2}\left[(a + d) + \sqrt{(a - d)^2 + 4bc}\,\right]$$

and

$$\lambda_2 = \tfrac{1}{2}\left[(a + d) - \sqrt{(a - d)^2 + 4bc}\,\right]$$

31. Use the result of Exercise 28 to prove that if

$$p(\lambda) = \lambda^2 + c_1\lambda + c_2$$

is the characteristic polynomial of a 2×2 matrix, then

$$p(A) = A^2 + c_1 A + c_2 I = 0$$

(Stated informally, A satisfies its characteristic equation. This result is true as well for $n \times n$ matrices.)

32. Prove: If a, b, c, and d are integers such that $a + b = c + d$, then

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

has integer eigenvalues.

33. Prove: If λ is an eigenvalue of an invertible matrix A and $\mathbf{x}$ is a corresponding eigenvector, then $1/\lambda$ is an eigenvalue of A^{-1} and $\mathbf{x}$ is a corresponding eigenvector.

34. Prove: If λ is an eigenvalue of A, $\mathbf{x}$ is a corresponding eigenvector, and s is a scalar, then $\lambda - s$ is an eigenvalue of $A - sI$ and $\mathbf{x}$ is a corresponding eigenvector.

35. Prove: If λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, then $s\lambda$ is an eigenvalue of sA for every scalar s and $\mathbf{x}$ is a corresponding eigenvector.

36. Find the eigenvalues and bases for the eigenspaces of

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

and then use Exercises 33 and 34 to find the eigenvalues and bases for the eigenspaces of

(a) A^{-1} (b) $A - 3I$ (c) $A + 2I$

37. Prove that the characteristic polynomial of an $n \times n$ matrix A has degree n and that the coefficient of λ^n in that polynomial is 1.

38. (a) Prove that if A is a square matrix, then A and A^T have the same eigenvalues. [*Hint:* Look at the characteristic equation $\det(\lambda I - A) = 0$.]

(b) Show that A and A^T need not have the same eigenspaces. [*Hint:* Use the result in Exercise 30 to find a 2×2 matrix for which A and A^T have different eigenspaces.]

39. Prove that the intersection of any two distinct eigenspaces of a matrix A is $\{\mathbf{0}\}$.

True-False Exercises

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

(a) If A is a square matrix and $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero scalar λ, then $\mathbf{x}$ is an eigenvector of A.

(b) If λ is an eigenvalue of a matrix A, then the linear system $(\lambda I - A)\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c) If the characteristic polynomial of a matrix A is $p(\lambda) = \lambda^2 + 1$, then A is invertible.

(d) If λ is an eigenvalue of a matrix A, then the eigenspace of A corresponding to λ is the set of eigenvectors of A corresponding to λ.

(e) The eigenvalues of a matrix A are the same as the eigenvalues of the reduced row echelon form of A.

(f) If 0 is an eigenvalue of a matrix A, then the set of columns of A is linearly independent.

Working with Technology

T1. For the given matrix A, find the characteristic polynomial and the eigenvalues, and then use the method of Example 7 to find bases for the eigenspaces.

$$A = \begin{bmatrix} -8 & 33 & 38 & 173 & -30 \\ 0 & 0 & -1 & -4 & 0 \\ 0 & 0 & -5 & -25 & 1 \\ 0 & 0 & 1 & 5 & 0 \\ 4 & -16 & -19 & -86 & 15 \end{bmatrix}$$

T2. The Cayley–Hamilton Theorem states that every square matrix satisfies its characteristic equation; that is, if A is an $n \times n$ matrix whose characteristic equation is

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

then $A^n + c_1A^{n-1} + \cdots + c_n = 0$.

(a) Verify the Cayley–Hamilton Theorem for the matrix

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

(b) Use the result in Exercise 28 to prove the Cayley–Hamilton Theorem for 2×2 matrices.

5.2 Diagonalization

In this section we will be concerned with the problem of finding a basis for R^n that consists of eigenvectors of an $n \times n$ matrix A. Such bases can be used to study geometric properties of A and to simplify various numerical computations. These bases are also of physical significance in a wide variety of applications, some of which will be considered later in this text.

The Matrix Diagonalization Problem

Products of the form $P^{-1}AP$ in which A and P are $n \times n$ matrices and P is invertible will be our main topic of study in this section. There are various ways to think about such products, one of which is to view them as transformations

$$A \rightarrow P^{-1}AP$$

in which the matrix A is mapped into the matrix $P^{-1}AP$. These are called ***similarity transformations***. Such transformations are important because they preserve many properties of the matrix A. For example, if we let $B = P^{-1}AP$, then A and B have the same determinant since

$$\begin{aligned} \det(B) &= \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) \\ &= \frac{1}{\det(P)}\det(A)\det(P) = \det(A) \end{aligned}$$

In general, any property that is preserved by a similarity transformation is called a ***similarity invariant*** and is said to be ***invariant under similarity***. Table 1 lists the most important similarity invariants. The proofs of some of these are given as exercises.

Table 1 Similarity Invariants

Property	**Description**
Determinant	A and $P^{-1}AP$ have the same determinant.
Invertibility	A is invertible if and only if $P^{-1}AP$ is invertible.
Rank	A and $P^{-1}AP$ have the same rank.
Nullity	A and $P^{-1}AP$ have the same nullity.
Trace	A and $P^{-1}AP$ have the same trace.
Characteristic polynomial	A and $P^{-1}AP$ have the same characteristic polynomial.
Eigenvalues	A and $P^{-1}AP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of A (and hence of $P^{-1}AP$) then the eigenspace of A corresponding to λ and the eigenspace of $P^{-1}AP$ corresponding to λ have the same dimension.

We will find the following terminology useful in our study of similarity transformations.

DEFINITION 1 If A and B are square matrices, then we say that ***B is similar to A*** if there is an invertible matrix P such that $B = P^{-1}AP$.

Note that if B is similar to A, then it is also true that A is similar to B since we can express A as $A = Q^{-1}BQ$ by taking $Q = P^{-1}$. This being the case, we will usually say that A and B are ***similar matrices*** if either is similar to the other.

Because diagonal matrices have such a simple form, it is natural to inquire whether a given $n \times n$ matrix A is similar to a matrix of this type. Should this turn out to be the case, and should we be able to actually find a diagonal matrix D that is similar to A, then we would be able to ascertain many of the similarity invariant properties of A directly from the diagonal entries of D. For example, the diagonal entries of D will be the eigenvalues of A (Theorem 5.1.2), and the product of the diagonal entries of D will be the determinant of A (Theorem 2.1.2). This leads us to introduce the following terminology.

DEFINITION 2 A square matrix A is said to be ***diagonalizable*** if it is similar to some diagonal matrix; that is, if there exists an invertible matrix P such that $P^{-1}AP$ is diagonal. In this case the matrix P is said to ***diagonalize*** A.

The following theorem and the ideas used in its proof will provide us with a roadmap for devising a technique for determining whether a matrix is diagonalizable and, if so, for finding a matrix P that will perform the diagonalization.

Part (b) of Theorem 5.2.1 is equivalent to saying that there is a *basis* for R^n consisting of eigenvectors of A. Why?

THEOREM 5.2.1 *If A is an $n \times n$ matrix, the following statements are equivalent.*

(a) *A is diagonalizable.*

(b) *A has n linearly independent eigenvectors.*

Proof (a) ⇒ (b) Since A is assumed to be diagonalizable, it follows that there exist an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$ or, equivalently,

$$AP = PD \tag{1}$$

If we denote the column vectors of P by $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, and if we assume that the diagonal entries of D are $\lambda_1, \lambda_2, \ldots, \lambda_n$, then by Formula (6) of Section 1.3 the left side of (1) can be expressed as

$$AP = A[\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n] = [A\mathbf{p}_1 \quad A\mathbf{p}_2 \quad \cdots \quad A\mathbf{p}_n]$$

and, as noted in the comment following Example 1 of Section 1.7, the right side of (1) can be expressed as

$$PD = [\lambda_1\mathbf{p}_1 \quad \lambda_2\mathbf{p}_2 \quad \cdots \quad \lambda_n\mathbf{p}_n]$$

Thus, it follows from (1) that

$$A\mathbf{p}_1 = \lambda_1\mathbf{p}_1, \quad A\mathbf{p}_2 = \lambda_2\mathbf{p}_2, \ldots, \quad A\mathbf{p}_n = \lambda_n\mathbf{p}_n \tag{2}$$

Since P is invertible, we know from Theorem 5.1.5 that its column vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ are linearly independent (and hence nonzero). Thus, it follows from (2) that these n column vectors are eigenvectors of A.

Proof (b) ⇒ (a) Assume that A has n linearly independent eigenvectors, $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, and that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues. If we let

$$P = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n]$$

and if we let D be the diagonal matrix that has $\lambda_1, \lambda_2, \ldots, \lambda_n$ as its successive diagonal entries, then

$$\begin{aligned} AP &= A[\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n] = [A\mathbf{p}_1 \quad A\mathbf{p}_2 \quad \cdots \quad A\mathbf{p}_n] \\ &= [\lambda_1\mathbf{p}_1 \quad \lambda_2\mathbf{p}_2 \quad \cdots \quad \lambda_n\mathbf{p}_n] = PD \end{aligned}$$

Since the column vectors of P are linearly independent, it follows from Theorem 5.1.5 that P is invertible, so that this last equation can be rewritten as $P^{-1}AP = D$, which shows that A is diagonalizable. ◀

Whereas Theorem 5.2.1 tells us that we need to find n linearly independent eigenvectors to diagonalize a matrix, the following theorem tells us where such vectors might be found. Part (a) is proved at the end of this section, and part (b) is an immediate consequence of part (a) and Theorem 5.2.1 (why?).

THEOREM 5.2.2

(a) *If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of a matrix A, and if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are corresponding eigenvectors, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent set.*

(b) *An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.*

Remark Part (a) of Theorem 5.2.2 is a special case of a more general result: Specifically, if $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues, and if $S_1, S_2, \ldots, S_k$ are corresponding sets of linearly independent eigenvectors, then the *union* of these sets is linearly independent.

Procedure for Diagonalizing a Matrix

Theorem 5.2.1 guarantees that an $n \times n$ matrix A with n linearly independent eigenvectors is diagonalizable, and the proof of that theorem together with Theorem 5.2.2 suggests the following procedure for diagonalizing A.

A Procedure for Diagonalizing an $n \times n$ Matrix

Step 1. Determine first whether the matrix is actually diagonalizable by searching for n linearly independent eigenvectors. One way to do this is to find a basis for each eigenspace and count the total number of vectors obtained. If there is a total of n vectors, then the matrix is diagonalizable, and if the total is less than n, then it is not.

Step 2. If you ascertained that the matrix is diagonalizable, then form the matrix $P = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_n]$ whose column vectors are the n basis vectors you obtained in Step 1.

Step 3. $P^{-1}AP$ will be a diagonal matrix whose successive diagonal entries are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ that correspond to the successive columns of P.

EXAMPLE 1 Finding a Matrix P That Diagonalizes a Matrix A

Find a matrix P that diagonalizes

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

Solution In Example 7 of the preceding section we found the characteristic equation of A to be

$$(\lambda - 1)(\lambda - 2)^2 = 0$$

and we found the following bases for the eigenspaces:

$$\lambda = 2: \quad \mathbf{p}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \qquad \lambda = 1: \quad \mathbf{p}_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

There are three basis vectors in total, so the matrix

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

diagonalizes A. As a check, you should verify that

$$P^{-1}AP = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In general, there is no preferred order for the columns of P. Since the ith diagonal entry of $P^{-1}AP$ is an eigenvalue for the ith column vector of P, changing the order of the columns of P just changes the order of the eigenvalues on the diagonal of $P^{-1}AP$. Thus, had we written

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

in the preceding example, we would have obtained

$$P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

▶ EXAMPLE 2 A Matrix That Is Not Diagonalizable

Show that the following matrix is not diagonalizable:

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

Solution The characteristic polynomial of A is

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & 0 & 0 \\ -1 & \lambda - 2 & 0 \\ 3 & -5 & \lambda - 2 \end{vmatrix} = (\lambda - 1)(\lambda - 2)^2$$

so the characteristic equation is

$$(\lambda - 1)(\lambda - 2)^2 = 0$$

and the distinct eigenvalues of A are $\lambda = 1$ and $\lambda = 2$. We leave it for you to show that bases for the eigenspaces are

$$\lambda = 1: \quad \mathbf{p}_1 = \begin{bmatrix} \frac{1}{8} \\ -\frac{1}{8} \\ 1 \end{bmatrix}; \qquad \lambda = 2: \quad \mathbf{p}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Since A is a 3×3 matrix and there are only two basis vectors in total, A is not diagonalizable.

Alternative Solution If you are concerned only in determining whether a matrix is diagonalizable and not with actually finding a diagonalizing matrix P, then it is not necessary to compute bases for the eigenspaces—it suffices to find the dimensions of the eigenspaces. For this example, the eigenspace corresponding to $\lambda = 1$ is the solution space of the system

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

Since the coefficient matrix has rank 2 (verify), the nullity of this matrix is 1 by Theorem 4.8.2, and hence the eigenspace corresponding to $\lambda = 1$ is one-dimensional.

The eigenspace corresponding to $\lambda = 2$ is the solution space of the system

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

This coefficient matrix also has rank 2 and nullity 1 (verify), so the eigenspace corresponding to $\lambda = 2$ is also one-dimensional. Since the eigenspaces produce a total of two basis vectors, and since three are needed, the matrix A is not diagonalizable.

EXAMPLE 3 Recognizing Diagonalizability

We saw in Example 3 of the preceding section that

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

has three distinct eigenvalues: $\lambda = 4$, $\lambda = 2 + \sqrt{3}$, and $\lambda = 2 - \sqrt{3}$. Therefore, A is diagonalizable and

$$P^{-1}AP = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2+\sqrt{3} & 0 \\ 0 & 0 & 2-\sqrt{3} \end{bmatrix}$$

for some invertible matrix P. If needed, the matrix P can be found using the method shown in Example 1 of this section.

EXAMPLE 4 Diagonalizability of Triangular Matrices

From Theorem 5.1.2, the eigenvalues of a triangular matrix are the entries on its main diagonal. Thus, a triangular matrix with distinct entries on the main diagonal is diagonalizable. For example,

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

is a diagonalizable matrix with eigenvalues $\lambda_1 = -1$, $\lambda_2 = 3$, $\lambda_3 = 5$, $\lambda_4 = -2$. ◀

Eigenvalues of Powers of a Matrix

Since there are many applications in which it is necessary to compute high powers of a square matrix A, we will now turn our attention to that important problem. As we will see, the most efficient way to compute A^k, particularly for large values of k, is to first diagonalize A. But because diagonalizing a matrix A involves finding its eigenvalues and eigenvectors, we will need to know how these quantities are related to those of A^k. As an illustration, suppose that λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector. Then

$$A^2\mathbf{x} = A(A\mathbf{x}) = A(\lambda\mathbf{x}) = \lambda(A\mathbf{x}) = \lambda(\lambda\mathbf{x}) = \lambda^2\mathbf{x}$$

which shows not only that λ^2 is a eigenvalue of A^2 but that $\mathbf{x}$ is a corresponding eigenvector. In general, we have the following result.

Note that diagonalizability is not a requirement in Theorem 5.2.3.

THEOREM 5.2.3 *If k is a positive integer, λ is an eigenvalue of a matrix A, and $\mathbf{x}$ is a corresponding eigenvector, then λ^k is an eigenvalue of A^k and $\mathbf{x}$ is a corresponding eigenvector.*

EXAMPLE 5 Eigenvalues and Eigenvectors of Matrix Powers

In Example 2 we found the eigenvalues and corresponding eigenvectors of the matrix

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

Do the same for A^7.

Solution We know from Example 2 that the eigenvalues of A are $\lambda = 1$ and $\lambda = 2$, so the eigenvalues of A^7 are $\lambda = 1^7 = 1$ and $\lambda = 2^7 = 128$. The eigenvectors $\mathbf{p}_1$ and $\mathbf{p}_2$ obtained in Example 1 corresponding to the eigenvalues $\lambda = 1$ and $\lambda = 2$ of A are also the eigenvectors corresponding to the eigenvalues $\lambda = 1$ and $\lambda = 128$ of A^7. ◀

Computing Powers of a Matrix

The problem of computing powers of a matrix is greatly simplified when the matrix is diagonalizable. To see why this is so, suppose that A is a diagonalizable $n \times n$ matrix, that P diagonalizes A, and that

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = D$$

Squaring both sides of this equation yields

$$(P^{-1}AP)^2 = \begin{bmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^2 \end{bmatrix} = D^2$$

We can rewrite the left side of this equation as

$$(P^{-1}AP)^2 = P^{-1}APP^{-1}AP = P^{-1}AIAP = P^{-1}A^2P$$

from which we obtain the relationship $P^{-1}A^2P = D^2$. More generally, if k is a positive integer, then a similar computation will show that

$$P^{-1}A^kP = D^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix}$$

which we can rewrite as

$$A^k = PD^kP^{-1} = P\begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix}P^{-1} \tag{3}$$

> Formula (3) reveals that raising a diagonalizable matrix A to a positive integer power has the effect of raising its eigenvalues to that power.

▶ **EXAMPLE 6 Powers of a Matrix**

Use (3) to find A^{13}, where

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

Solution We showed in Example 1 that the matrix A is diagonalized by

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

and that

$$D = P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus, it follows from (3) that

$$A^{13} = PD^{13}P^{-1} = \begin{bmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2^{13} & 0 & 0 \\ 0 & 2^{13} & 0 \\ 0 & 0 & 1^{13} \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} -8190 & 0 & -16382 \\ 8191 & 8192 & 8191 \\ 8191 & 0 & 16383 \end{bmatrix} \blacktriangleleft$$

Remark With the method in the preceding example, most of the work is in diagonalizing A. Once that work is done, it can be used to compute any power of A. Thus, to compute A^{1000} we need only change the exponents from 13 to 1000 in (4).

Geometric and Algebraic Multiplicity

Theorem 5.2.2(*b*) does not completely settle the diagonalizability question since it only guarantees that a square matrix with n distinct eigenvalues is diagonalizable; it does not preclude the possibility that there may exist diagonalizable matrices with fewer than n distinct eigenvalues. The following example shows that this is indeed the case.

▶ EXAMPLE 7 The Converse of Theorem 5.2.2(*b*) Is False

Consider the matrices

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

It follows from Theorem 5.1.2 that both of these matrices have only one distinct eigenvalue, namely $\lambda = 1$, and hence only one eigenspace. We leave it as an exercise for you to solve the characteristic equations

$$(\lambda I - I)\mathbf{x} = \mathbf{0} \quad \text{and} \quad (\lambda I - J)\mathbf{x} = \mathbf{0}$$

with $\lambda = 1$ and show that for I the eigenspace is three-dimensional (all of R^3) and for J it is one-dimensional, consisting of all scalar multiples of

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

This shows that the converse of Theorem 5.2.2(*b*) is false, since we have produced two 3×3 matrices with fewer than three distinct eigenvalues, one of which is diagonalizable and the other of which is not. ◀

A full excursion into the study of diagonalizability is left for more advanced courses, but we will touch on one theorem that is important for a fuller understanding of diagonalizability. It can be proved that if λ_0 is an eigenvalue of A, then the dimension of the eigenspace corresponding to λ_0 cannot exceed the number of times that $\lambda - \lambda_0$ appears as a factor of the characteristic polynomial of A. For example, in Examples 1 and 2 the characteristic polynomial is

$$(\lambda - 1)(\lambda - 2)^2$$

Thus, the eigenspace corresponding to $\lambda = 1$ is at most (hence exactly) one-dimensional, and the eigenspace corresponding to $\lambda = 2$ is at most two-dimensional. In Example 1 the eigenspace corresponding to $\lambda = 2$ actually had dimension 2, resulting in diagonalizability, but in Example 2 the eigenspace corresponding to $\lambda = 2$ had only dimension 1, resulting in nondiagonalizability.

There is some terminology that is related to these ideas. If λ_0 is an eigenvalue of an $n \times n$ matrix A, then the dimension of the eigenspace corresponding to λ_0 is called the ***geometric multiplicity*** of λ_0, and the number of times that $\lambda - \lambda_0$ appears as a factor in the characteristic polynomial of A is called the ***algebraic multiplicity*** of λ_0. The following theorem, which we state without proof, summarizes the preceding discussion.

THEOREM 5.2.4 Geometric and Algebraic Multiplicity

If A is a square matrix, then:

(*a*) *For every eigenvalue of A, the geometric multiplicity is less than or equal to the algebraic multiplicity.*

(*b*) *A is diagonalizable if and only if the geometric multiplicity of every eigenvalue is equal to the algebraic multiplicity.*

We will complete this section with an optional proof of Theorem 5.2.2(*a*).

OPTIONAL

Proof of Theorem 5.2.2(a) Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. We will assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly dependent and obtain a contradiction. We can then conclude that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent.

Since an eigenvector is nonzero by definition, $\{\mathbf{v}_1\}$ is linearly independent. Let r be the largest integer such that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is linearly independent. Since we are assuming that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly dependent, r satisfies $1 \le r < k$. Moreover, by the definition of r, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r+1}\}$ is linearly dependent. Thus, there are scalars $c_1, c_2, \ldots, c_{r+1}$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{r+1}\mathbf{v}_{r+1} = \mathbf{0} \tag{5}$$

Multiplying both sides of (5) by A and using the fact that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \ldots, \quad A\mathbf{v}_{r+1} = \lambda_{r+1}\mathbf{v}_{r+1}$$

we obtain

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_{r+1}\lambda_{r+1}\mathbf{v}_{r+1} = \mathbf{0} \tag{6}$$

If we now multiply both sides of (5) by λ_{r+1} and subtract the resulting equation from (6) we obtain

$$c_1(\lambda_1 - \lambda_{r+1})\mathbf{v}_1 + c_2(\lambda_2 - \lambda_{r+1})\mathbf{v}_2 + \cdots + c_r(\lambda_r - \lambda_{r+1})\mathbf{v}_r = \mathbf{0}$$

Since $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly independent set, this equation implies that

$$c_1(\lambda_1 - \lambda_{r+1}) = c_2(\lambda_2 - \lambda_{r+1}) = \cdots = c_r(\lambda_r - \lambda_{r+1}) = 0$$

and since $\lambda_1, \lambda_2, \ldots, \lambda_{r+1}$ are assumed to be distinct, it follows that

$$c_1 = c_2 = \cdots = c_r = 0 \tag{7}$$

Substituting these values in (5) yields

$$c_{r+1}\mathbf{v}_{r+1} = \mathbf{0}$$

Since the eigenvector $\mathbf{v}_{r+1}$ is nonzero, it follows that

$$c_{r+1} = 0 \tag{8}$$

But equations (7) and (8) contradict the fact that $c_1, c_2, \ldots, c_{r+1}$ are not all zero so the proof is complete. ◀

Exercise Set 5.2

▶ In Exercises 1–4, show that A and B are not similar matrices. ◀

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

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

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

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

▶ In Exercises 5–8, find a matrix P that diagonalizes A, and check your work by computing $P^{-1}AP$. ◀

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

6. $A = \begin{bmatrix} -14 & 12 \\ -20 & 17 \end{bmatrix}$

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

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

9. Let

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

(a) Find the eigenvalues of A.

(b) For each eigenvalue λ, find the rank of the matrix $\lambda I - A$.

(c) Is A diagonalizable? Justify your conclusion.

10. Follow the directions in Exercise 9 for the matrix

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

▶ In Exercises 11–14, find the geometric and algebraic multiplicity of each eigenvalue of the matrix A, and determine whether A is diagonalizable. If A is diagonalizable, then find a matrix P that diagonalizes A, and find $P^{-1}AP$. ◀

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

12. $A = \begin{bmatrix} 19 & -9 & -6 \\ 25 & -11 & -9 \\ 17 & -9 & -4 \end{bmatrix}$

13. $A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 0 & 1 \end{bmatrix}$

14. $A = \begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix}$

▶ In each part of Exercises 15–16, the characteristic equation of a matrix A is given. Find the size of the matrix and the possible dimensions of its eigenspaces. ◀

15. (a) $(\lambda - 1)(\lambda + 3)(\lambda - 5) = 0$

(b) $\lambda^2(\lambda - 1)(\lambda - 2)^3 = 0$

16. (a) $\lambda^3(\lambda^2 - 5\lambda - 6) = 0$

(b) $\lambda^3 - 3\lambda^2 + 3\lambda - 1 = 0$

▶ In Exercises 17–18, use the method of Example 6 to compute the matrix A^{10}. ◀

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

18. $A = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$

19. Let

$$A = \begin{bmatrix} -1 & 7 & -1 \\ 0 & 1 & 0 \\ 0 & 15 & -2 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 5 \end{bmatrix}$$

Confirm that P diagonalizes A, and then compute A^{11}.

20. Let

$$A = \begin{bmatrix} 1 & -2 & 8 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & -4 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Confirm that P diagonalizes A, and then compute each of the following powers of A.

(a) A^{1000} (b) A^{-1000} (c) A^{2301} (d) A^{-2301}

21. Find A^n if n is a positive integer and

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

22. Show that the matrices

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

are similar.

23. We know from Table 1 that similar matrices have the same rank. Show that the converse is false by showing that the matrices

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

have the same rank but are not similar. [*Suggestion:* If they were similar, then there would be an invertible 2×2 matrix P for which $AP = PB$. Show that there is no such matrix.]

24. We know from Table 1 that similar matrices have the same eigenvalues. Use the method of Exercise 23 to show that the converse is false by showing that the matrices

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

have the same eigenvalues but are not similar.

25. If A, B, and C are $n \times n$ matrices such that A is similar to B and B is similar to C, do you think that A must be similar to C? Justify your answer.

26. (a) Is it possible for an $n \times n$ matrix to be similar to itself? Justify your answer.

(b) What can you say about an $n \times n$ matrix that is similar to $0_{n \times n}$? Justify your answer.

(c) Is is possible for a nonsingular matrix to be similar to a singular matrix? Justify your answer.

27. Suppose that the characteristic polynomial of some matrix A is found to be $p(\lambda) = (\lambda - 1)(\lambda - 3)^2(\lambda - 4)^3$. In each part, answer the question and explain your reasoning.

(a) What can you say about the dimensions of the eigenspaces of A?

(b) What can you say about the dimensions of the eigenspaces if you know that A is diagonalizable?

(c) If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of eigenvectors of A, all of which correspond to the same eigenvalue of A, what can you say about that eigenvalue?

28. Let

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

Show that

(a) A is diagonalizable if $(a - d)^2 + 4bc > 0$.

(b) A is not diagonalizable if $(a - d)^2 + 4bc < 0$.

[*Hint:* See Exercise 29 of Section 5.1.]

29. In the case where the matrix A in Exercise 28 is diagonalizable, find a matrix P that diagonalizes A. [*Hint:* See Exercise 30 of Section 5.1.]

▶ In Exercises 30–33, find the standard matrix A for the given linear operator, and determine whether that matrix is diagonalizable. If diagonalizable, find a matrix P that diagonalizes A. ◀

30. $T(x_1, x_2) = (2x_1 - x_2, x_1 + x_2)$

31. $T(x_1, x_2) = (-x_2, -x_1)$

32. $T(x_1, x_2, x_3) = (8x_1 + 3x_2 - 4x_3, -3x_1 + x_2 + 3x_3, 4x_1 + 3x_2)$

33. $T(x_1, x_2, x_3) = (3x_1, x_2, x_1 - x_2)$

34. If P is a fixed $n \times n$ matrix, then the similarity transformation

$$A \to P^{-1}AP$$

can be viewed as an operator $S_P(A) = P^{-1}AP$ on the vector space M_{nn} of $n \times n$ matrices.

(a) Show that S_P is a linear operator.

(b) Find the kernel of S_P.

(c) Find the rank of S_P.

Working with Proofs

35. Prove that similar matrices have the same rank and nullity.

36. Prove that similar matrices have the same trace.

37. Prove that if A is diagonalizable, then so is A^k for every positive integer k.

38. We know from Table 1 that similar matrices, A and B, have the same eigenvalues. However, it is not true that those eigenvalues have the same corresponding eigenvectors for the two matrices. Prove that if $B = P^{-1}AP$, and $\mathbf{v}$ is an eigenvector of B corresponding to the eigenvalue λ, then $P\mathbf{v}$ is the eigenvector of A corresponding to λ.

39. Let A be an $n \times n$ matrix, and let $q(A)$ be the matrix

$$q(A) = a_n A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 I_n$$

(a) Prove that if $B = P^{-1}AP$, then $q(B) = P^{-1}q(A)P$.

(b) Prove that if A is diagonalizable, then so is $q(A)$.

40. Prove that if A is a diagonalizable matrix, then the rank of A is the number of nonzero eigenvalues of A.

41. This problem will lead you through a proof of the fact that the algebraic multiplicity of an eigenvalue of an $n \times n$ matrix A is greater than or equal to the geometric multiplicity. For this purpose, assume that λ_0 is an eigenvalue with geometric multiplicity k.

(a) Prove that there is a basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for R^n in which the first k vectors of B form a basis for the eigenspace corresponding to λ_0.

(b) Let P be the matrix having the vectors in B as columns. Prove that the product AP can be expressed as

$$AP = P\begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix}$$

[*Hint:* Compare the first k column vectors on both sides.]

(c) Use the result in part (b) to prove that A is similar to

$$C = \begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix}$$

and hence that A and C have the same characteristic polynomial.

(d) By considering $\det(\lambda I - C)$, prove that the characteristic polynomial of C (and hence A) contains the factor $(\lambda - \lambda_0)$ at least k times, thereby proving that the algebraic multiplicity of λ_0 is greater than or equal to the geometric multiplicity k.

True-False Exercises

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

(a) An $n \times n$ matrix with fewer than n distinct eigenvalues is not diagonalizable.

(b) An $n \times n$ matrix with fewer than n linearly independent eigenvectors is not diagonalizable.

(c) If A and B are similar $n \times n$ matrices, then there exists an invertible $n \times n$ matrix P such that $PA = BP$.

(d) If A is diagonalizable, then there is a unique matrix P such that $P^{-1}AP$ is diagonal.

(e) If A is diagonalizable and invertible, then A^{-1} is diagonalizable.

(f) If A is diagonalizable, then A^T is diagonalizable.

(g) If there is a basis for R^n consisting of eigenvectors of an $n \times n$ matrix A, then A is diagonalizable.

(h) If every eigenvalue of a matrix A has algebraic multiplicity 1, then A is diagonalizable.

(i) If 0 is an eigenvalue of a matrix A, then A^2 is singular.

Working with Technology

T1. Generate a random 4×4 matrix A and an invertible 4×4 matrix P and then confirm, as stated in Table 1, that $P^{-1}AP$ and A have the same

(a) determinant.

(b) rank.

(c) nullity.

(d) trace.

(e) characteristic polynomial.

(f) eigenvalues.

T2. (a) Use Theorem 5.2.1 to show that the following matrix is diagonalizable.

$$A = \begin{bmatrix} -13 & -60 & -60 \\ 10 & 42 & 40 \\ -5 & -20 & -18 \end{bmatrix}$$

(b) Find a matrix P that diagonalizes A.

(c) Use the method of Example 6 to compute A^{10}, and check your result by computing A^{10} directly.

T3. Use Theorem 5.2.1 to show that the following matrix is not diagonalizable.

$$A = \begin{bmatrix} -10 & 11 & -6 \\ -15 & 16 & -10 \\ -3 & 3 & -2 \end{bmatrix}$$

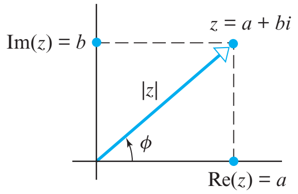
5.3 Complex Vector Spaces

Because the characteristic equation of any square matrix can have complex solutions, the notions of complex eigenvalues and eigenvectors arise naturally, even within the context of matrices with real entries. In this section we will discuss this idea and apply our results to study symmetric matrices in more detail. A review of the essentials of complex numbers appears in the back of this text.

Review of Complex Numbers

Recall that if $z = a + bi$ is a complex number, then:

- $\text{Re}(z) = a$ and $\text{Im}(z) = b$ are called the ***real part*** of z and the ***imaginary part*** of z, respectively,
- $|z| = \sqrt{a^2 + b^2}$ is called the ***modulus*** (or ***absolute value***) of z,
- $\overline{z} = a - bi$ is called the ***complex conjugate*** of z,



▲ Figure 5.3.1

- $z\bar{z} = a^2 + b^2 = |z|^2$,
- the angle ϕ in Figure 5.3.1 is called an ***argument*** of z,
- $\text{Re}(z) = |z| \cos \phi$
- $\text{Im}(z) = |z| \sin \phi$
- $z = |z|(\cos \phi + i \sin \phi)$ is called the ***polar form*** of z.

Complex Eigenvalues

In Formula (3) of Section 5.1 we observed that the characteristic equation of a general $n \times n$ matrix A has the form

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0 \tag{1}$$

in which the highest power of λ has a coefficient of 1. Up to now we have limited our discussion to matrices in which the solutions of (1) are real numbers. However, it is possible for the characteristic equation of a matrix A with real entries to have imaginary solutions; for example, the characteristic equation of the matrix

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

is

$$\begin{vmatrix} \lambda + 2 & 1 \\ -5 & \lambda - 2 \end{vmatrix} = \lambda^2 + 1 = 0$$

which has the imaginary solutions $\lambda = i$ and $\lambda = -i$. To deal with this case we will need to explore the notion of a complex vector space and some related ideas.

Vectors in C^n

A vector space in which scalars are allowed to be complex numbers is called a ***complex vector space***. In this section we will be concerned only with the following complex generalization of the real vector space R^n.

DEFINITION 1 If n is a positive integer, then a ***complex n-tuple*** is a sequence of n complex numbers $(v_1, v_2, \dots, v_n)$. The set of all complex n-tuples is called ***complex n-space*** and is denoted by C^n. Scalars are complex numbers, and the operations of addition, subtraction, and scalar multiplication are performed componentwise.

The terminology used for n-tuples of real numbers applies to complex n-tuples without change. Thus, if $v_1, v_2, \dots, v_n$ are complex numbers, then we call $\mathbf{v} = (v_1, v_2, \dots, v_n)$ a ***vector*** in C^n and $v_1, v_2, \dots, v_n$ its ***components***. Some examples of vectors in C^3 are

$$\mathbf{u} = (1 + i, -4i, 3 + 2i), \quad \mathbf{v} = (0, i, 5), \quad \mathbf{w} = \left(6 - \sqrt{2}\,i, 9 + \tfrac{1}{2}i, \pi i\right)$$

Every vector

$$\mathbf{v} = (v_1, v_2, \dots, v_n) = (a_1 + b_1 i, a_2 + b_2 i, \dots, a_n + b_n i)$$

in C^n can be split into ***real*** and ***imaginary parts*** as

$$\mathbf{v} = (a_1, a_2, \dots, a_n) + i(b_1, b_2, \dots, b_n)$$

which we also denote as

$$\mathbf{v} = \text{Re}(\mathbf{v}) + i\,\text{Im}(\mathbf{v})$$

where

$$\text{Re}(\mathbf{v}) = (a_1, a_2, \ldots, a_n) \quad \text{and} \quad \text{Im}(\mathbf{v}) = (b_1, b_2, \ldots, b_n)$$

The vector

$$\overline{\mathbf{v}} = (\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n) = (a_1 - b_1 i, a_2 - b_2 i, \ldots, a_n - b_n i)$$

is called the ***complex conjugate*** of $\mathbf{v}$ and can be expressed in terms of $\text{Re}(\mathbf{v})$ and $\text{Im}(\mathbf{v})$ as

$$\overline{\mathbf{v}} = (a_1, a_2, \ldots, a_n) - i(b_1, b_2, \ldots, b_n) = \text{Re}(\mathbf{v}) - i\,\text{Im}(\mathbf{v}) \tag{2}$$

It follows that the vectors in R^n can be viewed as those vectors in C^n whose imaginary part is zero; or stated another way, a vector $\mathbf{v}$ in C^n is in R^n if and only if $\overline{\mathbf{v}} = \mathbf{v}$.

In this section we will need to distinguish between matrices whose entries *must* be real numbers, called ***real matrices***, and matrices whose entries may be *either* real numbers or complex numbers, called ***complex matrices***. When convenient, you can think of a real matrix as a complex matrix each of whose entries has a zero imaginary part. The standard operations on real matrices carry over without change to complex matrices, and all of the familiar properties of matrices continue to hold.

If A is a complex matrix, then $\text{Re}(A)$ and $\text{Im}(A)$ are the matrices formed from the real and imaginary parts of the entries of A, and $\overline{A}$ is the matrix formed by taking the complex conjugate of each entry in A.

EXAMPLE 1 Real and Imaginary Parts of Vectors and Matrices

As you might expect, if A is a complex matrix, then A and $\overline{A}$ can be expressed in terms of $\text{Re}(A)$ and $\text{Im}(A)$ as

$$A = \text{Re}(A) + i\,\text{Im}(A)$$
$$\overline{A} = \text{Re}(A) - i\,\text{Im}(A)$$

Let

$$\mathbf{v} = (3+i, -2i, 5) \quad \text{and} \quad A = \begin{bmatrix} 1+i & -i \\ 4 & 6-2i \end{bmatrix}$$

Then

$$\overline{\mathbf{v}} = (3-i, 2i, 5), \quad \text{Re}(\mathbf{v}) = (3, 0, 5), \quad \text{Im}(\mathbf{v}) = (1, -2, 0)$$

$$\overline{A} = \begin{bmatrix} 1-i & i \\ 4 & 6+2i \end{bmatrix}, \quad \text{Re}(A) = \begin{bmatrix} 1 & 0 \\ 4 & 6 \end{bmatrix}, \quad \text{Im}(A) = \begin{bmatrix} 1 & -1 \\ 0 & -2 \end{bmatrix}$$

$$\det(A) = \begin{vmatrix} 1+i & -i \\ 4 & 6-2i \end{vmatrix} = (1+i)(6-2i) - (-i)(4) = 8 + 8i$$

Algebraic Properties of the Complex Conjugate

The next two theorems list some properties of complex vectors and matrices that we will need in this section. Some of the proofs are given as exercises.

THEOREM 5.3.1 *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* C^n, *and if* k *is a scalar, then*:

(*a*) $\overline{\overline{\mathbf{u}}} = \mathbf{u}$

(*b*) $\overline{k\mathbf{u}} = \overline{k}\,\overline{\mathbf{u}}$

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

(*d*) $\overline{\mathbf{u} - \mathbf{v}} = \overline{\mathbf{u}} - \overline{\mathbf{v}}$

THEOREM 5.3.2 *If* A *is an* $m \times k$ *complex matrix and* B *is a* $k \times n$ *complex matrix, then*:

(*a*) $\overline{\overline{A}} = A$

(*b*) $(\overline{A^T}) = (\overline{A})^T$

(*c*) $\overline{AB} = \overline{A}\,\overline{B}$

The Complex Euclidean Inner Product

The following definition extends the notions of dot product and norm to C^n.

DEFINITION 2 If $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ are vectors in C^n, then the ***complex Euclidean inner product*** of $\mathbf{u}$ and $\mathbf{v}$ (also called the ***complex dot product***) is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1\overline{v}_1 + u_2\overline{v}_2 + \cdots + u_n\overline{v}_n \tag{3}$$

We also define the ***Euclidean norm*** on C^n to be

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2} \tag{4}$$

The complex conjugates in (3) ensure that $\|\mathbf{v}\|$ is a real number, for without them the quantity $\mathbf{v} \cdot \mathbf{v}$ in (4) might be imaginary.

As in the real case, we call $\mathbf{v}$ a ***unit vector*** in C^n if $\|\mathbf{v}\| = 1$, and we say two vectors $\mathbf{u}$ and $\mathbf{v}$ are ***orthogonal*** if $\mathbf{u} \cdot \mathbf{v} = 0$.

EXAMPLE 2 Complex Euclidean Inner Product and Norm

Find $\mathbf{u} \cdot \mathbf{v}$, $\mathbf{v} \cdot \mathbf{u}$, $\|\mathbf{u}\|$, and $\|\mathbf{v}\|$ for the vectors

$$\mathbf{u} = (1+i, i, 3-i) \quad \text{and} \quad \mathbf{v} = (1+i, 2, 4i)$$

Solution

$$\mathbf{u} \cdot \mathbf{v} = (1+i)(\overline{1+i}) + i(\overline{2}) + (3-i)(\overline{4i}) = (1+i)(1-i) + 2i + (3-i)(-4i) = -2 - 10i$$

$$\mathbf{v} \cdot \mathbf{u} = (1+i)(\overline{1+i}) + 2(\overline{i}) + (4i)(\overline{3-i}) = (1+i)(1-i) - 2i + 4i(3+i) = -2 + 10i$$

$$\|\mathbf{u}\| = \sqrt{|1+i|^2 + |i|^2 + |3-i|^2} = \sqrt{2+1+10} = \sqrt{13}$$

$$\|\mathbf{v}\| = \sqrt{|1+i|^2 + |2|^2 + |4i|^2} = \sqrt{2+4+16} = \sqrt{22}$$

Recall from Table 1 of Section 3.2 that if $\mathbf{u}$ and $\mathbf{v}$ are *column vectors* in R^n, then their dot product can be expressed as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$$

The analogous formulas in C^n are (verify)

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\overline{\mathbf{v}} = \overline{\mathbf{v}}^T\mathbf{u} \tag{5}$$

Example 2 reveals a major difference between the dot product on R^n and the complex dot product on C^n. For the dot product on R^n we always have $\mathbf{v} \cdot \mathbf{u} = \mathbf{u} \cdot \mathbf{v}$ (the *symmetry property*), but for the complex dot product the corresponding relationship is given by $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$, which is called its ***antisymmetry*** property. The following theorem is an analog of Theorem 3.2.2.

THEOREM 5.3.3 *If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in C^n, and if k is a scalar, then the complex Euclidean inner product has the following properties:*

(*a*) $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$ **[Antisymmetry property]**

(*b*) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ **[Distributive property]**

(*c*) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$ **[Homogeneity property]**

(*d*) $\mathbf{u} \cdot k\mathbf{v} = \overline{k}(\mathbf{u} \cdot \mathbf{v})$ **[Antihomogeneity property]**

(*e*) $\mathbf{v} \cdot \mathbf{v} \ge 0$ *and* $\mathbf{v} \cdot \mathbf{v} = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$. **[Positivity property]**

Parts (c) and (d) of this theorem state that a scalar multiplying a complex Euclidean inner product can be regrouped with the first vector, but to regroup it with the second vector you must first take its complex conjugate. We will prove part (d), and leave the others as exercises.

Proof (d)

$$k(\mathbf{u}\cdot\mathbf{v}) = k(\overline{\mathbf{v}\cdot\mathbf{u}}) = \overline{\overline{k}}\,(\overline{\mathbf{v}\cdot\mathbf{u}}) = \overline{\overline{k}\,(\mathbf{v}\cdot\mathbf{u})} = \overline{(\overline{k}\mathbf{v})\cdot\mathbf{u}} = \mathbf{u}\cdot(\overline{k}\mathbf{v})$$

To complete the proof, substitute $\overline{k}$ for k and use the fact that $\overline{\overline{k}} = k$. ◀

Vector Concepts in C^n

Is R^n a subspace of C^n? Explain.

Except for the use of complex scalars, the notions of linear combination, linear independence, subspace, spanning, basis, and dimension carry over without change to C^n.

Eigenvalues and eigenvectors are defined for complex matrices exactly as for real matrices. If A is an $n \times n$ matrix with complex entries, then the complex roots of the characteristic equation $\det(\lambda I - A) = 0$ are called ***complex eigenvalues*** of A. As in the real case, λ is a complex eigenvalue of A if and only if there exists a nonzero vector $\mathbf{x}$ in C^n such that $A\mathbf{x} = \lambda\mathbf{x}$. Each such $\mathbf{x}$ is called a ***complex eigenvector*** of A corresponding to λ. The complex eigenvectors of A corresponding to λ are the nonzero solutions of the linear system $(\lambda I - A)\mathbf{x} = \mathbf{0}$, and the set of all such solutions is a subspace of C^n, called the ***complex eigenspace*** of A corresponding to λ.

The following theorem states that if a *real matrix* has complex eigenvalues, then those eigenvalues and their corresponding eigenvectors occur in conjugate pairs.

THEOREM 5.3.4 *If λ is an eigenvalue of a real $n \times n$ matrix A, and if $\mathbf{x}$ is a corresponding eigenvector, then $\overline{\lambda}$ is also an eigenvalue of A, and $\overline{\mathbf{x}}$ is a corresponding eigenvector.*

Proof Since λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, we have

$$\overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\,\overline{\mathbf{x}} \tag{6}$$

However, $\overline{A} = A$, since A has real entries, so it follows from part (c) of Theorem 5.3.2 that

$$\overline{A\mathbf{x}} = \overline{A}\,\overline{\mathbf{x}} = A\overline{\mathbf{x}} \tag{7}$$

Equations (6) and (7) together imply that

$$A\overline{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda}\,\overline{\mathbf{x}}$$

in which $\overline{\mathbf{x}} \neq \mathbf{0}$ (why?); this tells us that $\overline{\lambda}$ is an eigenvalue of A and $\overline{\mathbf{x}}$ is a corresponding eigenvector. ◀

▶ EXAMPLE 3 Complex Eigenvalues and Eigenvectors

Find the eigenvalues and bases for the eigenspaces of

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

Solution The characteristic polynomial of A is

$$\begin{vmatrix} \lambda + 2 & 1 \\ -5 & \lambda - 2 \end{vmatrix} = \lambda^2 + 1 = (\lambda - i)(\lambda + i)$$

so the eigenvalues of A are $\lambda = i$ and $\lambda = -i$. Note that these eigenvalues are complex conjugates, as guaranteed by Theorem 5.3.4. To find the eigenvectors we must solve the system

$$\begin{bmatrix} \lambda+2 & 1 \\ -5 & \lambda-2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

with $\lambda = i$ and then with $\lambda = -i$. With $\lambda = i$, this system becomes

$$\begin{bmatrix} i+2 & 1 \\ -5 & i-2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8}$$

We could solve this system by reducing the augmented matrix

$$\begin{bmatrix} i+2 & 1 & 0 \\ -5 & i-2 & 0 \end{bmatrix} \tag{9}$$

to reduced row echelon form by Gauss–Jordan elimination, though the complex arithmetic is somewhat tedious. A simpler procedure here is first to observe that the reduced row echelon form of (9) must have a row of zeros because (8) has nontrivial solutions. This being the case, each row of (9) must be a scalar multiple of the other, and hence the first row can be made into a row of zeros by adding a suitable multiple of the second row to it. Accordingly, we can simply set the entries in the first row to zero, then interchange the rows, and then multiply the new first row by $-\frac{1}{5}$ to obtain the reduced row echelon form

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

Thus, a general solution of the system is

$$x_1 = \left(-\tfrac{2}{5} + \tfrac{1}{5}i\right)t, \quad x_2 = t$$

This tells us that the eigenspace corresponding to $\lambda = i$ is one-dimensional and consists of all complex scalar multiples of the basis vector

$$\mathbf{x} = \begin{bmatrix} -\frac{2}{5} + \frac{1}{5}i \\ 1 \end{bmatrix} \tag{10}$$

As a check, let us confirm that $A\mathbf{x} = i\mathbf{x}$. We obtain

$$A\mathbf{x} = \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} -\frac{2}{5} + \frac{1}{5}i \\ 1 \end{bmatrix} = \begin{bmatrix} -2\left(-\frac{2}{5} + \frac{1}{5}i\right) - 1 \\ 5\left(-\frac{2}{5} + \frac{1}{5}i\right) + 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} - \frac{2}{5}i \\ i \end{bmatrix} = i\mathbf{x}$$

We could find a basis for the eigenspace corresponding to $\lambda = -i$ in a similar way, but the work is unnecessary since Theorem 5.3.4 implies that

$$\overline{\mathbf{x}} = \begin{bmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{bmatrix} \tag{11}$$

must be a basis for this eigenspace. The following computations confirm that $\overline{\mathbf{x}}$ is an eigenvector of A corresponding to $\lambda = -i$:

$$\begin{aligned} A\overline{\mathbf{x}} &= \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -2\left(-\frac{2}{5} - \frac{1}{5}i\right) - 1 \\ 5\left(-\frac{2}{5} - \frac{1}{5}i\right) + 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} + \frac{2}{5}i \\ -i \end{bmatrix} = -i\overline{\mathbf{x}} \end{aligned}$$ ◀

Since a number of our subsequent examples will involve 2×2 matrices with real entries, it will be useful to discuss some general results about the eigenvalues of such matrices. Observe first that the characteristic polynomial of the matrix

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

is

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - a & -b \\ -c & \lambda - d \end{vmatrix} = (\lambda - a)(\lambda - d) - bc = \lambda^2 - (a + d)\lambda + (ad - bc)$$

We can express this in terms of the trace and determinant of A as

$$\det(\lambda I - A) = \lambda^2 - \operatorname{tr}(A)\lambda + \det(A) \tag{12}$$

from which it follows that the characteristic equation of A is

$$\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0 \tag{13}$$

Now recall from algebra that if $ax^2 + bx + c = 0$ is a quadratic equation with real coefficients, then the ***discriminant*** $b^2 - 4ac$ determines the nature of the roots:

$$b^2 - 4ac > 0 \quad \text{[Two distinct real roots]}$$
$$b^2 - 4ac = 0 \quad \text{[One repeated real root]}$$
$$b^2 - 4ac < 0 \quad \text{[Two conjugate imaginary roots]}$$

Applying this to (13) with $a = 1$, $b = -\operatorname{tr}(A)$, and $c = \det(A)$ yields the following theorem.

THEOREM 5.3.5 *If A is a 2×2 matrix with real entries, then the characteristic equation of A is $\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0$ and*

(*a*) *A has two distinct real eigenvalues if* $\operatorname{tr}(A)^2 - 4\det(A) > 0$;

(*b*) *A has one repeated real eigenvalue if* $\operatorname{tr}(A)^2 - 4\det(A) = 0$;

(*c*) *A has two complex conjugate eigenvalues if* $\operatorname{tr}(A)^2 - 4\det(A) < 0$.

▶ EXAMPLE 4 Eigenvalues of a 2 × 2 Matrix

In each part, use Formula (13) for the characteristic equation to find the eigenvalues of

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

Olga Taussky-Todd (1906–1995)

Historical Note Olga Taussky-Todd was one of the pioneering women in matrix analysis and the first woman appointed to the faculty at the California Institute of Technology. She worked at the National Physical Laboratory in London during World War II, where she was assigned to study flutter in supersonic aircraft. While there, she realized that some results about the eigenvalues of a certain 6×6 complex matrix could be used to answer key questions about the flutter problem that would otherwise have required laborious calculation. After World War II Olga Taussky-Todd continued her work on matrix-related subjects and helped to draw many known but disparate results about matrices into the coherent subject that we now call matrix theory.

[*Image: Courtesy of the Archives, California Institute of Technology*]

Solution (a) We have $\text{tr}(A) = 7$ and $\det(A) = 12$, so the characteristic equation of A is

$$\lambda^2 - 7\lambda + 12 = 0$$

Factoring yields $(\lambda - 4)(\lambda - 3) = 0$, so the eigenvalues of A are $\lambda = 4$ and $\lambda = 3$.

Solution (b) We have $\text{tr}(A) = 2$ and $\det(A) = 1$, so the characteristic equation of A is

$$\lambda^2 - 2\lambda + 1 = 0$$

Factoring this equation yields $(\lambda - 1)^2 = 0$, so $\lambda = 1$ is the only eigenvalue of A; it has algebraic multiplicity 2.

Solution (c) We have $\text{tr}(A) = 4$ and $\det(A) = 13$, so the characteristic equation of A is

$$\lambda^2 - 4\lambda + 13 = 0$$

Solving this equation by the quadratic formula yields

$$\lambda = \frac{4 \pm \sqrt{(-4)^2 - 4(13)}}{2} = \frac{4 \pm \sqrt{-36}}{2} = 2 \pm 3i$$

Thus, the eigenvalues of A are $\lambda = 2 + 3i$ and $\lambda = 2 - 3i$. ◀

Symmetric Matrices Have Real Eigenvalues

Our next result, which is concerned with the eigenvalues of real symmetric matrices, is important in a wide variety of applications. The key to its proof is to think of a real symmetric matrix as a complex matrix whose entries have an imaginary part of zero.

THEOREM 5.3.6 *If A is a real symmetric matrix, then A has real eigenvalues.*

Proof Suppose that λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, where we allow for the possibility that λ is complex and $\mathbf{x}$ is in C^n. Thus,

$$A\mathbf{x} = \lambda\mathbf{x}$$

where $\mathbf{x} \neq \mathbf{0}$. If we multiply both sides of this equation by $\overline{\mathbf{x}}^T$ and use the fact that

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T(\lambda\mathbf{x}) = \lambda(\overline{\mathbf{x}}^T\mathbf{x}) = \lambda(\mathbf{x} \cdot \mathbf{x}) = \lambda\|\mathbf{x}\|^2$$

then we obtain

$$\lambda = \frac{\overline{\mathbf{x}}^T A\mathbf{x}}{\|\mathbf{x}\|^2}$$

Since the denominator in this expression is real, we can prove that λ is real by showing that

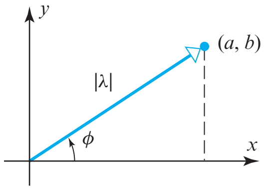
$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\mathbf{x}}^T A\mathbf{x} \tag{14}$$

But A is symmetric and has real entries, so it follows from the second equality in (5) and properties of the conjugate that

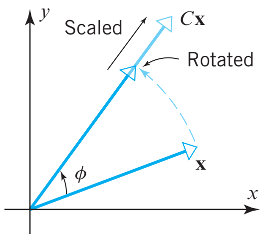
$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\overline{\mathbf{x}}}^T\,\overline{A\mathbf{x}} = \mathbf{x}^T\,\overline{A\mathbf{x}} = (\overline{A\mathbf{x}})^T\mathbf{x} = (\overline{A}\overline{\mathbf{x}})^T\mathbf{x} = (A\overline{\mathbf{x}})^T\mathbf{x} = \overline{\mathbf{x}}^T A^T\mathbf{x} = \overline{\mathbf{x}}^T A\mathbf{x}$$ ◀

A Geometric Interpretation of Complex Eigenvalues

The following theorem is the key to understanding the geometric significance of complex eigenvalues of real 2×2 matrices.



▲ Figure 5.3.2



▲ Figure 5.3.3

THEOREM 5.3.7 *The eigenvalues of the real matrix*

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{15}$$

are $\lambda = a \pm bi$. If a and b are not both zero, then this matrix can be factored as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \tag{16}$$

where ϕ is the angle from the positive x-axis to the ray that joins the origin to the point (a, b) (Figure 5.3.2).

Geometrically, this theorem states that multiplication by a matrix of form (15) can be viewed as a rotation through the angle ϕ followed by a scaling with factor $|\lambda|$ (Figure 5.3.3).

Proof The characteristic equation of C is $(\lambda - a)^2 + b^2 = 0$ (verify), from which it follows that the eigenvalues of C are $\lambda = a \pm bi$. Assuming that a and b are not both zero, let ϕ be the angle from the positive x-axis to the ray that joins the origin to the point (a, b). The angle ϕ is an argument of the eigenvalue $\lambda = a + bi$, so we see from Figure 5.3.2 that

$$a = |\lambda|\cos\phi \quad \text{and} \quad b = |\lambda|\sin\phi$$

It follows from this that the matrix in (15) can be written as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \dfrac{a}{|\lambda|} & -\dfrac{b}{|\lambda|} \\ \dfrac{b}{|\lambda|} & \dfrac{a}{|\lambda|} \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$ ◀

The following theorem, whose proof is considered in the exercises, shows that every real 2×2 matrix with complex eigenvalues is similar to a matrix of form (15).

THEOREM 5.3.8 *Let A be a real 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$ (where $b \neq 0$). If $\mathbf{x}$ is an eigenvector of A corresponding to $\lambda = a - bi$, then the matrix $P = [\mathrm{Re}(\mathbf{x}) \quad \mathrm{Im}(\mathbf{x})]$ is invertible and*

$$A = P \begin{bmatrix} a & -b \\ b & a \end{bmatrix} P^{-1} \tag{17}$$

▶ EXAMPLE 5 A Matrix Factorization Using Complex Eigenvalues

Factor the matrix in Example 3 into form (17) using the eigenvalue $\lambda = -i$ and the corresponding eigenvector that was given in (11).

Solution For consistency with the notation in Theorem 5.3.8, let us denote the eigenvector in (11) that corresponds to $\lambda = -i$ by $\mathbf{x}$ (rather than $\overline{\mathbf{x}}$ as before). For this λ and $\mathbf{x}$ we have

$$a = 0, \quad b = 1, \quad \mathrm{Re}(\mathbf{x}) = \begin{bmatrix} -\frac{2}{5} \\ 1 \end{bmatrix}, \quad \mathrm{Im}(\mathbf{x}) = \begin{bmatrix} -\frac{1}{5} \\ 0 \end{bmatrix}$$

Thus,

$$P = [\operatorname{Re}(\mathbf{x}) \quad \operatorname{Im}(\mathbf{x})] = \begin{bmatrix} -\frac{2}{5} & -\frac{1}{5} \\ 1 & 0 \end{bmatrix}$$

so A can be factored in form (17) as

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

You may want to confirm this by multiplying out the right side. ◀

A Geometric Interpretation of Theorem 5.3.8

To clarify what Theorem 5.3.8 says geometrically, let us denote the matrices on the right side of (16) by S and R_ϕ, respectively, and then use (16) to rewrite (17) as

$$A = PSR_\phi P^{-1} = P \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} P^{-1} \tag{18}$$

If we now view P as the transition matrix from the basis $B = \{\operatorname{Re}(\mathbf{x}), \operatorname{Im}(\mathbf{x})\}$ to the standard basis, then (18) tells us that computing a product $A\mathbf{x}_0$ can be broken down into a three-step process:

Interpreting Formula (18)

Step 1. Map $\mathbf{x}_0$ from standard coordinates into B-coordinates by forming the product $P^{-1}\mathbf{x}_0$.

Step 2. Rotate and scale the vector $P^{-1}\mathbf{x}_0$ by forming the product $SR_\phi P^{-1}\mathbf{x}_0$.

Step 3. Map the rotated and scaled vector back to standard coordinates to obtain $A\mathbf{x}_0 = PSR_\phi P^{-1}\mathbf{x}_0$.

Power Sequences

There are many problems in which one is interested in how successive applications of a matrix transformation affect a specific vector. For example, if A is the standard matrix for an operator on R^n and $\mathbf{x}_0$ is some fixed vector in R^n, then one might be interested in the behavior of the power sequence

$$\mathbf{x}_0, \quad A\mathbf{x}_0, \quad A^2\mathbf{x}_0, \ldots, \quad A^k\mathbf{x}_0, \ldots$$

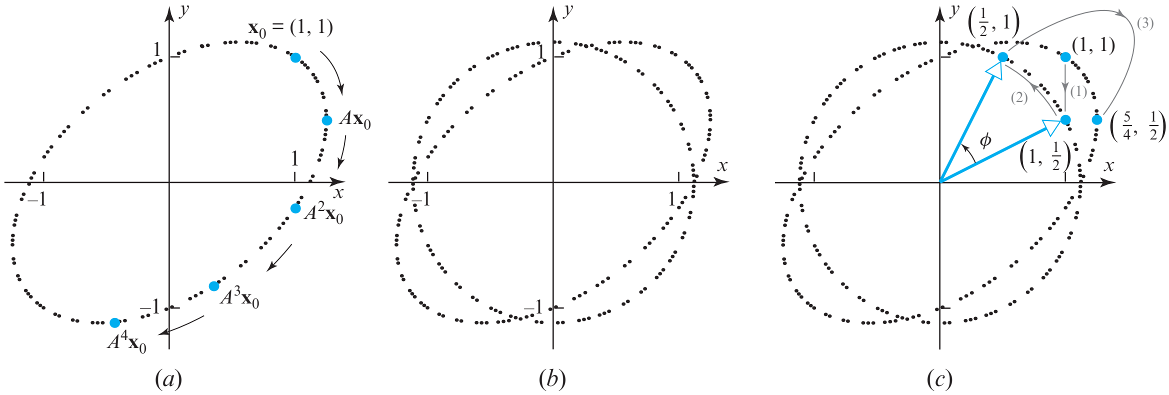
For example, if

$$A = \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix} \quad \text{and} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

then with the help of a computer or calculator one can show that the first four terms in the power sequence are

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A\mathbf{x}_0 = \begin{bmatrix} 1.25 \\ 0.5 \end{bmatrix}, \quad A^2\mathbf{x}_0 = \begin{bmatrix} 1.0 \\ -0.2 \end{bmatrix}, \quad A^3\mathbf{x}_0 = \begin{bmatrix} 0.35 \\ -0.82 \end{bmatrix}$$

With the help of MATLAB or a computer algebra system one can show that if the first 100 terms are plotted as ordered pairs (x, y), then the points move along the elliptical path shown in Figure 5.3.4*a*.



▲ Figure 5.3.4

To understand why the points move along an elliptical path, we will need to examine the eigenvalues and eigenvectors of A. We leave it for you to show that the eigenvalues of A are $\lambda = \frac{4}{5} \pm \frac{3}{5}i$ and that the corresponding eigenvectors are

$$\lambda_1 = \frac{4}{5} - \frac{3}{5}i: \quad \mathbf{v}_1 = \left(\frac{1}{2} + i, 1\right) \quad \text{and} \quad \lambda_2 = \frac{4}{5} + \frac{3}{5}i: \quad \mathbf{v}_2 = \left(\frac{1}{2} - i, 1\right)$$

If we take $\lambda = \lambda_1 = \frac{4}{5} - \frac{3}{5}i$ and $\mathbf{x} = \mathbf{v}_1 = \left(\frac{1}{2} + i, 1\right)$ in (17) and use the fact that $|\lambda| = 1$, then we obtain the factorization

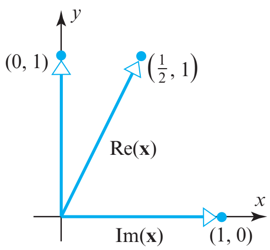
$$\underset{A}{\begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix}} = \underset{P}{\begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}} \underset{R_\phi}{\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}} \underset{P^{-1}}{\begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}} \tag{19}$$

where R_ϕ is a rotation about the origin through the angle ϕ whose tangent is

$$\tan\phi = \frac{\sin\phi}{\cos\phi} = \frac{3/5}{4/5} = \frac{3}{4} \qquad \left(\phi = \tan^{-1}\frac{3}{4} \approx 36.9°\right)$$

The matrix P in (19) is the transition matrix from the basis

$$B = \{\text{Re}(\mathbf{x}), \text{Im}(\mathbf{x})\} = \left\{\left(\frac{1}{2}, 1\right), (1, 0)\right\}$$



▲ Figure 5.3.5

to the standard basis, and P^{-1} is the transition matrix from the standard basis to the basis B (Figure 5.3.5). Next, observe that if n is a positive integer, then (19) implies that

$$A^n\mathbf{x}_0 = (PR_\phi P^{-1})^n\mathbf{x}_0 = PR_\phi^n P^{-1}\mathbf{x}_0$$

so the product $A^n\mathbf{x}_0$ can be computed by first mapping $\mathbf{x}_0$ into the point $P^{-1}\mathbf{x}_0$ in B-coordinates, then multiplying by R_ϕ^n to rotate this point about the origin through the angle $n\phi$, and then multiplying $R_\phi^n P^{-1}\mathbf{x}_0$ by P to map the resulting point back to standard coordinates. We can now see what is happening geometrically: In B-coordinates each successive multiplication by A causes the point $P^{-1}\mathbf{x}_0$ to advance through an angle ϕ, thereby tracing a circular orbit about the origin. However, the basis B is *skewed* (not orthogonal), so when the points on the circular orbit are transformed back to standard coordinates, the effect is to distort the circular orbit into the elliptical orbit traced by $A^n\mathbf{x}_0$ (Figure 5.3.4*b*). Here are the computations for the first step (successive steps are illustrated in Figure 5.3.4*c*):

$$\begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix} \quad [\mathbf{x}_0 \text{ is mapped to } B\text{-coordinates.}]$$

$$= \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \quad [\text{The point } (1, \tfrac{1}{2}) \text{ is rotated through the angle } \phi.]$$

$$= \begin{bmatrix} \frac{5}{4} \\ \frac{1}{2} \end{bmatrix} \quad [\text{The point } (\tfrac{1}{2}, 1) \text{ is mapped to standard coordinates.}]$$

Exercise Set 5.3

▶ In Exercises 1–2, find $\bar{\mathbf{u}}$, $\operatorname{Re}(\mathbf{u})$, $\operatorname{Im}(\mathbf{u})$, and $\|\mathbf{u}\|$. ◀

1. $\mathbf{u} = (2 - i, 4i, 1 + i)$ **2.** $\mathbf{u} = (6, 1 + 4i, 6 - 2i)$

▶ In Exercises 3–4, show that $\mathbf{u}$, $\mathbf{v}$, and k satisfy Theorem 5.3.1. ◀

3. $\mathbf{u} = (3 - 4i, 2 + i, -6i)$, $\mathbf{v} = (1 + i, 2 - i, 4)$, $k = i$

4. $\mathbf{u} = (6, 1 + 4i, 6 - 2i)$, $\mathbf{v} = (4, 3 + 2i, i - 3)$, $k = -i$

5. Solve the equation $i\mathbf{x} - 3\mathbf{v} = \bar{\mathbf{u}}$ for $\mathbf{x}$, where $\mathbf{u}$ and $\mathbf{v}$ are the vectors in Exercise 3.

6. Solve the equation $(1 + i)\mathbf{x} + 2\mathbf{u} = \bar{\mathbf{v}}$ for $\mathbf{x}$, where $\mathbf{u}$ and $\mathbf{v}$ are the vectors in Exercise 4.

▶ In Exercises 7–8, find $\overline{A}$, $\operatorname{Re}(A)$, $\operatorname{Im}(A)$, $\det(A)$, and $\operatorname{tr}(A)$. ◀

7. $A = \begin{bmatrix} -5i & 4 \\ 2 - i & 1 + 5i \end{bmatrix}$ **8.** $A = \begin{bmatrix} 4i & 2 - 3i \\ 2 + 3i & 1 \end{bmatrix}$

9. Let A be the matrix given in Exercise 7, and let B be the matrix

$$B = \begin{bmatrix} 1 - i \\ 2i \end{bmatrix}$$

Confirm that these matrices have the properties stated in Theorem 5.3.2.

10. Let A be the matrix given in Exercise 8, and let B be the matrix

$$B = \begin{bmatrix} 5i \\ 1 - 4i \end{bmatrix}$$

Confirm that these matrices have the properties stated in Theorem 5.3.2.

▶ In Exercises 11–12, compute $\mathbf{u} \cdot \mathbf{v}$, $\mathbf{u} \cdot \mathbf{w}$, and $\mathbf{v} \cdot \mathbf{w}$, and show that the vectors satisfy Formula (5) and parts (*a*), (*b*), and (*c*) of Theorem 5.3.3. ◀

11. $\mathbf{u} = (i, 2i, 3)$, $\mathbf{v} = (4, -2i, 1 + i)$, $\mathbf{w} = (2 - i, 2i, 5 + 3i)$, $k = 2i$

12. $\mathbf{u} = (1 + i, 4, 3i)$, $\mathbf{v} = (3, -4i, 2 + 3i)$, $\mathbf{w} = (1 - i, 4i, 4 - 5i)$, $k = 1 + i$

13. Compute $\overline{(\mathbf{u} \cdot \bar{\mathbf{v}})} - \overline{\mathbf{w} \cdot \mathbf{u}}$ for the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in Exercise 11.

14. Compute $\overline{(i\bar{\mathbf{u}} \cdot \mathbf{w}) + (\|\mathbf{u}\|\mathbf{v}) \cdot \mathbf{u}}$ for the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in Exercise 12.

▶ In Exercises 15–18, find the eigenvalues and bases for the eigenspaces of A. ◀

15. $A = \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}$ **16.** $A = \begin{bmatrix} -1 & -5 \\ 4 & 7 \end{bmatrix}$

17. $A = \begin{bmatrix} 5 & -2 \\ 1 & 3 \end{bmatrix}$ **18.** $A = \begin{bmatrix} 8 & 6 \\ -3 & 2 \end{bmatrix}$

▶ In Exercises 19–22, each matrix C has form (15). Theorem 5.3.7 implies that C is the product of a scaling matrix with factor $|\lambda|$ and a rotation matrix with angle ϕ. Find $|\lambda|$ and ϕ for which $-\pi < \phi \le \pi$. ◀

19. $C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ **20.** $C = \begin{bmatrix} 0 & 5 \\ -5 & 0 \end{bmatrix}$

21. $C = \begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix}$ **22.** $C = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{bmatrix}$

▶ In Exercises 23–26, find an invertible matrix P and a matrix C of form (15) such that $A = PCP^{-1}$. ◀

23. $A = \begin{bmatrix} -1 & -5 \\ 4 & 7 \end{bmatrix}$ **24.** $A = \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}$

25. $A = \begin{bmatrix} 8 & 6 \\ -3 & 2 \end{bmatrix}$ **26.** $A = \begin{bmatrix} 5 & -2 \\ 1 & 3 \end{bmatrix}$

27. Find all complex scalars k, if any, for which $\mathbf{u}$ and $\mathbf{v}$ are orthogonal in C^3.

(a) $\mathbf{u} = (2i, i, 3i),\ \ \mathbf{v} = (i, 6i, k)$

(b) $\mathbf{u} = (k, k, 1+i),\ \ \mathbf{v} = (1, -1, 1-i)$

28. Show that if A is a real $n \times n$ matrix and $\mathbf{x}$ is a column vector in C^n, then $\text{Re}(A\mathbf{x}) = A(\text{Re}(\mathbf{x}))$ and $\text{Im}(A\mathbf{x}) = A(\text{Im}(\mathbf{x}))$.

29. The matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},\quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

called ***Pauli spin matrices***, are used in quantum mechanics to study particle spin. The ***Dirac matrices***, which are also used in quantum mechanics, are expressed in terms of the Pauli spin matrices and the 2×2 identity matrix I_2 as

$$\beta = \begin{bmatrix} I_2 & 0 \\ 0 & -I_2 \end{bmatrix},\quad \alpha_x = \begin{bmatrix} 0 & \sigma_1 \\ \sigma_1 & 0 \end{bmatrix},$$

$$\alpha_y = \begin{bmatrix} 0 & \sigma_2 \\ \sigma_2 & 0 \end{bmatrix},\quad \alpha_z = \begin{bmatrix} 0 & \sigma_3 \\ \sigma_3 & 0 \end{bmatrix}$$

(a) Show that $\beta^2 = \alpha_x^2 = \alpha_y^2 = \alpha_z^2$.

(b) Matrices A and B for which $AB = -BA$ are said to be ***anticommutative***. Show that the Dirac matrices are anticommutative.

30. If k is a real scalar and $\mathbf{v}$ is a vector in R^n, then Theorem 3.2.1 states that $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$. Is this relationship also true if k is a complex scalar and $\mathbf{v}$ is a vector in C^n? Justify your answer.

Working with Proofs

31. Prove part (c) of Theorem 5.3.1.

32. Prove Theorem 5.3.2.

33. Prove that if $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, then

$$\mathbf{u}\cdot\mathbf{v} = \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2 + \frac{i}{4}\|\mathbf{u}+i\mathbf{v}\|^2 - \frac{i}{4}\|\mathbf{u}-i\mathbf{v}\|^2$$

34. It follows from Theorem 5.3.7 that the eigenvalues of the rotation matrix

$$R_\phi = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

are $\lambda = \cos\phi \pm i\sin\phi$. Prove that if $\mathbf{x}$ is an eigenvector corresponding to either eigenvalue, then $\text{Re}(\mathbf{x})$ and $\text{Im}(\mathbf{x})$ are orthogonal and have the same length. [*Note:* This implies that $P = [\text{Re}(\mathbf{x}) \mid \text{Im}(\mathbf{x})]$ is a real scalar multiple of an orthogonal matrix.]

35. The two parts of this exercise lead you through a proof of Theorem 5.3.8.

(a) For notational simplicity, let

$$M = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

and let $\mathbf{u} = \text{Re}(\mathbf{x})$ and $\mathbf{v} = \text{Im}(\mathbf{x})$, so $P = [\mathbf{u} \mid \mathbf{v}]$. Show that the relationship $A\mathbf{x} = \lambda\mathbf{x}$ implies that

$$A\mathbf{x} = (a\mathbf{u} + b\mathbf{v}) + i(-b\mathbf{u} + a\mathbf{v})$$

and then equate real and imaginary parts in this equation to show that

$$AP = [A\mathbf{u} \mid A\mathbf{v}] = [a\mathbf{u} + b\mathbf{v} \mid -b\mathbf{u} + a\mathbf{v}] = PM$$

(b) Show that P is invertible, thereby completing the proof, since the result in part (a) implies that $A = PMP^{-1}$. [*Hint:* If P is not invertible, then one of its column vectors is a real scalar multiple of the other, say $\mathbf{v} = c\mathbf{u}$. Substitute this into the equations $A\mathbf{u} = a\mathbf{u} + b\mathbf{v}$ and $A\mathbf{v} = -b\mathbf{u} + a\mathbf{v}$ obtained in part (a), and show that $(1+c^2)b\mathbf{u} = \mathbf{0}$. Finally, show that this leads to a contradiction, thereby proving that P is invertible.]

36. In this problem you will prove the complex analog of the Cauchy–Schwarz inequality.

(a) Prove: If k is a complex number, and $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, then

$$(\mathbf{u} - k\mathbf{v})\cdot(\mathbf{u} - k\mathbf{v}) = \mathbf{u}\cdot\mathbf{u} - \bar{k}(\mathbf{u}\cdot\mathbf{v}) - k\overline{(\mathbf{u}\cdot\mathbf{v})} + k\bar{k}(\mathbf{v}\cdot\mathbf{v})$$

(b) Use the result in part (a) to prove that

$$0 \le \mathbf{u}\cdot\mathbf{u} - \bar{k}(\mathbf{u}\cdot\mathbf{v}) - k\overline{(\mathbf{u}\cdot\mathbf{v})} + k\bar{k}(\mathbf{v}\cdot\mathbf{v})$$

(c) Take $k = (\mathbf{u}\cdot\mathbf{v})/(\mathbf{v}\cdot\mathbf{v})$ in part (b) to prove that

$$|\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$$

True-False Exercises

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

(a) There is a real 5×5 matrix with no real eigenvalues.

(b) The eigenvalues of a 2×2 complex matrix are the solutions of the equation $\lambda^2 - \text{tr}(A)\lambda + \det(A) = 0$.

(c) A 2×2 matrix A with real entries has two distinct eigenvalues if and only if $\text{tr}(A)^2 \ne 4\det(A)$.

(d) If λ is a complex eigenvalue of a real matrix A with a corresponding complex eigenvector $\mathbf{v}$, then $\bar{\lambda}$ is a complex eigenvalue of A and $\bar{\mathbf{v}}$ is a complex eigenvector of A corresponding to $\bar{\lambda}$.

(e) Every eigenvalue of a complex symmetric matrix is real.

(f) If a 2×2 real matrix A has complex eigenvalues and $\mathbf{x}_0$ is a vector in R^2, then the vectors $\mathbf{x}_0, A\mathbf{x}_0, A^2\mathbf{x}_0, \ldots, A^n\mathbf{x}_0, \ldots$ lie on an ellipse.

5.4 Differential Equations

Many laws of physics, chemistry, biology, engineering, and economics are described in terms of "differential equations"—that is, equations involving functions and their derivatives. In this section we will illustrate one way in which matrix diagonalization can be used to solve systems of differential equations. Calculus is a prerequisite for this section.

Terminology

Recall from calculus that a ***differential equation*** is an equation involving unknown functions and their derivatives. The ***order*** of a differential equation is the order of the highest derivative it contains. The simplest differential equations are the first-order equations of the form

$$y' = ay \tag{1}$$

where $y = f(x)$ is an unknown differentiable function to be determined, $y' = dy/dx$ is its derivative, and a is a constant. As with most differential equations, this equation has infinitely many solutions; they are the functions of the form

$$y = ce^{ax} \tag{2}$$

where c is an arbitrary constant. That every function of this form is a solution of (1) follows from the computation

$$y' = cae^{ax} = ay$$

and that these are the only solution is shown in the exercises. Accordingly, we call (2) the ***general solution*** of (1). As an example, the general solution of the differential equation $y' = 5y$ is

$$y = ce^{5x} \tag{3}$$

Often, a physical problem that leads to a differential equation imposes some conditions that enable us to isolate one particular solution from the general solution. For example, if we require that solution (3) of the equation $y' = 5y$ satisfy the added condition

$$y(0) = 6 \tag{4}$$

(that is, $y = 6$ when $x = 0$), then on substituting these values in (3), we obtain $6 = ce^0 = c$, from which we conclude that

$$y = 6e^{5x}$$

is the only solution $y' = 5y$ that satisfies (4).

A condition such as (4), which specifies the value of the general solution at a point, is called an ***initial condition***, and the problem of solving a differential equation subject to an initial condition is called an ***initial-value problem***.

First-Order Linear Systems

In this section we will be concerned with solving systems of differential equations of the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \\ &\vdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n \end{aligned} \tag{5}$$

where $y_1 = f_1(x), y_2 = f_2(x), \ldots, y_n = f_n(x)$ are functions to be determined, and the a_{ij}'s are constants. In matrix notation, (5) can be written as

$$\begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

or more briefly as

$$\mathbf{y}' = A\mathbf{y} \tag{6}$$

where the notation $\mathbf{y}'$ denotes the vector obtained by differentiating each component of $\mathbf{y}$.

We call (5) or its matrix form (6) a ***constant coefficient first-order homogeneous linear system***. It is of first order because all derivatives are of that order, it is linear because differentiation and matrix multiplication are linear transformations, and it is homogeneous because

$$y_1 = y_2 = \cdots = y_n = 0$$

is a solution regardless of the values of the coefficients. As expected, this is called the ***trivial solution***. In this section we will work primarily with the matrix form. Here is an example.

▶ EXAMPLE 1 Solution of a Linear System with Initial Conditions

(a) Write the following system in matrix form:

$$\begin{aligned} y_1' &= 3y_1 \\ y_2' &= -2y_2 \\ y_3' &= 5y_3 \end{aligned} \tag{7}$$

(b) Solve the system.

(c) Find a solution of the system that satisfies the initial conditions $y_1(0) = 1$, $y_2(0) = 4$, and $y_3(0) = -2$.

Solution (a)

$$\begin{bmatrix} y_1' \\ y_2' \\ y_3' \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{8}$$

or

$$\mathbf{y}' = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \mathbf{y} \tag{9}$$

Solution (b) Because each equation in (7) involves only one unknown function, we can solve the equations individually. It follows from (2) that these solutions are

$$\begin{aligned} y_1 &= c_1 e^{3x} \\ y_2 &= c_2 e^{-2x} \\ y_3 &= c_3 e^{5x} \end{aligned}$$

or, in matrix notation,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} c_1 e^{3x} \\ c_2 e^{-2x} \\ c_3 e^{5x} \end{bmatrix} \tag{10}$$

Solution (c) From the given initial conditions, we obtain

$$\begin{aligned} 1 &= y_1(0) = c_1 e^0 = c_1 \\ 4 &= y_2(0) = c_2 e^0 = c_2 \\ -2 &= y_3(0) = c_3 e^0 = c_3 \end{aligned}$$

so the solution satisfying these conditions is

$$y_1 = e^{3x}, \quad y_2 = 4e^{-2x}, \quad y_3 = -2e^{5x}$$

or, in matrix notation,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} e^{3x} \\ 4e^{-2x} \\ -2e^{5x} \end{bmatrix} \blacktriangleleft$$

Solution by Diagonalization

What made the system in Example 1 easy to solve was the fact that each equation involved only one of the unknown functions, so its matrix formulation, $\mathbf{y}' = A\mathbf{y}$, had a *diagonal* coefficient matrix A [Formula (9)]. A more complicated situation occurs when some or all of the equations in the system involve more than one of the unknown functions, for in this case the coefficient matrix is not diagonal. Let us now consider how we might solve such a system.

The basic idea for solving a system $\mathbf{y}' = A\mathbf{y}$ whose coefficient matrix A is not diagonal is to introduce a new unknown vector $\mathbf{u}$ that is related to the unknown vector $\mathbf{y}$ by an equation of the form $\mathbf{y} = P\mathbf{u}$ in which P is an invertible matrix that diagonalizes A. Of course, such a matrix may or may not exist, but if it does, then we can rewrite the equation $\mathbf{y}' = A\mathbf{y}$ as

$$P\mathbf{u}' = A(P\mathbf{u})$$

or alternatively as

$$\mathbf{u}' = (P^{-1}AP)\mathbf{u}$$

Since P is assumed to diagonalize A, this equation has the form

$$\mathbf{u}' = D\mathbf{u}$$

where D is diagonal. We can now solve this equation for $\mathbf{u}$ using the method of Example 1, and then obtain $\mathbf{y}$ by matrix multiplication using the relationship $\mathbf{y} = P\mathbf{u}$.

In summary, we have the following procedure for solving a system $\mathbf{y}' = A\mathbf{y}$ in the case were A is diagonalizable.

A Procedure for Solving $\mathbf{y}' = A\mathbf{y}$ If A Is Diagonalizable

Step 1. Find a matrix P that diagonalizes A.

Step 2. Make the substitutions $\mathbf{y} = P\mathbf{u}$ and $\mathbf{y}' = P\mathbf{u}'$ to obtain a new "diagonal system" $\mathbf{u}' = D\mathbf{u}$, where $D = P^{-1}AP$.

Step 3. Solve $\mathbf{u}' = D\mathbf{u}$.

Step 4. Determine $\mathbf{y}$ from the equation $\mathbf{y} = P\mathbf{u}$.

▶ EXAMPLE 2 Solution Using Diagonalization

(a) Solve the system

$$\begin{aligned} y_1' &= y_1 + y_2 \\ y_2' &= 4y_1 - 2y_2 \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = 1$, $y_2(0) = 6$.

Solution (a) The coefficient matrix for the system is

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

As discussed in Section 5.2, A will be diagonalized by any matrix P whose columns are linearly independent eigenvectors of A. Since

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -1 \\ -4 & \lambda + 2 \end{vmatrix} = \lambda^2 + \lambda - 6 = (\lambda + 3)(\lambda - 2)$$

the eigenvalues of A are $\lambda = 2$ and $\lambda = -3$. By definition,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is an eigenvector of A corresponding to λ if and only if $\mathbf{x}$ is a nontrivial solution of

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

If $\lambda = 2$, this system becomes

$$\begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving this system yields $x_1 = t$, $x_2 = t$, so

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus,

$$\mathbf{p}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = 2$. Similarly, you can show that

$$\mathbf{p}_2 = \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = -3$. Thus,

$$P = \begin{bmatrix} 1 & -\frac{1}{4} \\ 1 & 1 \end{bmatrix}$$

diagonalizes A, and

$$D = P^{-1}AP = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$$

Thus, as noted in Step 2 of the procedure stated above, the substitution

$$\mathbf{y} = P\mathbf{u} \quad \text{and} \quad \mathbf{y}' = P\mathbf{u}'$$

yields the "diagonal system"

$$\mathbf{u}' = D\mathbf{u} = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix} \mathbf{u} \quad \text{or} \quad \begin{aligned} u_1' &= 2u_1 \\ u_2' &= -3u_2 \end{aligned}$$

From (2) the solution of this system is

$$\begin{aligned} u_1 &= c_1 e^{2x} \\ u_2 &= c_2 e^{-3x} \end{aligned} \quad \text{or} \quad \mathbf{u} = \begin{bmatrix} c_1 e^{2x} \\ c_2 e^{-3x} \end{bmatrix}$$

so the equation $\mathbf{y} = P\mathbf{u}$ yields, as the solution for $\mathbf{y}$,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{4} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1e^{2x} \\ c_2e^{-3x} \end{bmatrix} = \begin{bmatrix} c_1e^{2x} - \frac{1}{4}c_2e^{-3x} \\ c_1e^{2x} + c_2e^{-3x} \end{bmatrix}$$

or

$$\begin{aligned} y_1 &= c_1e^{2x} - \tfrac{1}{4}c_2e^{-3x} \\ y_2 &= c_1e^{2x} + c_2e^{-3x} \end{aligned} \tag{11}$$

Solution (b) If we substitute the given initial conditions in (11), we obtain

$$\begin{aligned} c_1 - \tfrac{1}{4}c_2 &= 1 \\ c_1 + c_2 &= 6 \end{aligned}$$

Solving this system, we obtain $c_1 = 2$, $c_2 = 4$, so it follows from (11) that the solution satisfying the initial conditions is

$$\begin{aligned} y_1 &= 2e^{2x} - e^{-3x} \\ y_2 &= 2e^{2x} + 4e^{-3x} \end{aligned}$$

Remark Keep in mind that the method of Example 2 works because the coefficient matrix of the system is diagonalizable. In cases where this is not so, other methods are required. These are typically discussed in books devoted to differential equations.

Exercise Set 5.4

1. (a) Solve the system

$$\begin{aligned} y_1' &= y_1 + 4y_2 \\ y_2' &= 2y_1 + 3y_2 \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = 0$, $y_2(0) = 0$.

2. (a) Solve the system

$$\begin{aligned} y_1' &= y_1 + 3y_2 \\ y_2' &= 4y_1 + 5y_2 \end{aligned}$$

(b) Find the solution that satisfies the conditions $y_1(0) = 2$, $y_2(0) = 1$.

3. (a) Solve the system

$$\begin{aligned} y_1' &= 4y_1 + y_3 \\ y_2' &= -2y_1 + y_2 \\ y_3' &= -2y_1 + y_3 \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = -1$, $y_2(0) = 1$, $y_3(0) = 0$.

4. Solve the system

$$\begin{aligned} y_1' &= 4y_1 + 2y_2 + 2y_3 \\ y_2' &= 2y_1 + 4y_2 + 2y_3 \\ y_3' &= 2y_1 + 2y_2 + 4y_3 \end{aligned}$$

5. Show that every solution of $y' = ay$ has the form $y = ce^{ax}$. [*Hint:* Let $y = f(x)$ be a solution of the equation, and show that $f(x)e^{-ax}$ is constant.]

6. Show that if A is diagonalizable and

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

is a solution of the system $\mathbf{y}' = A\mathbf{y}$, then each y_i is a linear combination of $e^{\lambda_1 x}, e^{\lambda_2 x}, \ldots, e^{\lambda_n x}$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

7. Sometimes it is possible to solve a single higher-order linear differential equation with constant coefficients by expressing it as a system and applying the methods of this section. For the differential equation $y'' - y' - 6y = 0$, show that the substitutions $y_1 = y$ and $y_2 = y'$ lead to the system

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= 6y_1 + y_2 \end{aligned}$$

Solve this system, and use the result to solve the original differential equation.

8. Use the procedure in Exercise 7 to solve $y'' + y' - 12y = 0$.

9. Explain how you might use the procedure in Exercise 7 to solve $y''' - 6y'' + 11y' - 6y = 0$. Use that procedure to solve the equation.

10. Solve the nondiagonalizable system

$$y_1' = y_1 + y_2$$
$$y_2' = y_2$$

[*Hint:* Solve the second equation for y_2, substitute in the first equation, and then multiply both sides of the resulting equation by e^{-x}.]

11. Consider a system of differential equations $\mathbf{y}' = A\mathbf{y}$, where A is a 2×2 matrix. For what values of $a_{11}, a_{12}, a_{21}, a_{22}$ do the component solutions $y_1(t), y_2(t)$ tend to zero as $t \to \infty$? In particular, what must be true about the determinant and the trace of A for this to happen?

12. (a) By rewriting (11) in matrix form, show that the solution of the system in Example 2 can be expressed as

$$\mathbf{y} = c_1 e^{2x} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 e^{-3x} \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$$

This is called the ***general solution*** of the system.

(b) Note that in part (a), the vector in the first term is an eigenvector corresponding to the eigenvalue $\lambda_1 = 2$, and the vector in the second term is an eigenvector corresponding to the eigenvalue $\lambda_2 = -3$. This is a special case of the following general result:

> ***Theorem.*** *If the coefficient matrix A of the system $\mathbf{y}' = A\mathbf{y}$ is diagonalizable, then the general solution of the system can be expressed as*
>
> $$\mathbf{y} = c_1 e^{\lambda_1 x}\mathbf{x}_1 + c_2 e^{\lambda_2 x}\mathbf{x}_2 + \cdots + c_n e^{\lambda_n x}\mathbf{x}_n$$
>
> *where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A, and $\mathbf{x}_i$ is an eigenvector of A corresponding to λ_i.*

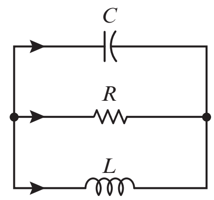
13. The electrical circuit in the accompanying figure is called a ***parallel LRC circuit***; it contains a resistor with resistance R ohms (Ω), an inductor with inductance L henries (H), and a capacitor with capacitance C farads (F). It is shown in electrical circuit analysis that at time t the current i_L through the inductor and the voltage v_C across the capacitor are solutions of the system

$$\begin{bmatrix} i_L'(t) \\ v_C'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1/L \\ -1/C & -1/(RC) \end{bmatrix} \begin{bmatrix} i_L(t) \\ v_C(t) \end{bmatrix}$$

(a) Find the general solution of this system in the case where $R = 1$ ohm, $L = 1$ henry, and $C = 0.5$ farad.

(b) Find $i_L(t)$ and $v_C(t)$ subject to the initial conditions $i_L(0) = 2$ amperes and $v_C(0) = 1$ volt.

(c) What can you say about the current and voltage in part (b) over the "long term" (that is, as $t \to \infty$)?



◀ Figure Ex-13

▶ In Exercises 14–15, a mapping

$$L: C^\infty(-\infty, \infty) \to C^\infty(-\infty, \infty)$$

is given.

(a) Show that L is a linear operator.

(b) Use the ideas in Exercises 7 and 9 to solve the differential equation $L(y) = 0$. ◀

14. $L(y) = y'' + 2y' - 3y$

15. $L(y) = y''' - 2y'' - y' + 2y$

Working with Proofs

16. Prove the theorem in Exercise 12 by tracing through the four-step procedure preceding Example 2 with

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad \text{and} \quad P = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_n]$$

True-False Exercises

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

(a) Every system of differential equations $\mathbf{y}' = A\mathbf{y}$ has a solution.

(b) If $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{y}' = A\mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.

(c) If $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{y}' = A\mathbf{y}$, then $(c\mathbf{x} + d\mathbf{y})' = A(c\mathbf{x} + d\mathbf{y})$ for all scalars c and d.

(d) If A is a square matrix with distinct real eigenvalues, then it is possible to solve $\mathbf{x}' = A\mathbf{x}$ by diagonalization.

(e) If A and P are similar matrices, then $\mathbf{y}' = A\mathbf{y}$ and $\mathbf{u}' = P\mathbf{u}$ have the same solutions.

Working with Technology

T1. (a) Find the general solution of the following system by computing appropriate eigenvalues and eigenvectors.

$$\begin{aligned} y_1' &= 3y_1 + 2y_2 + 2y_3 \\ y_2' &= y_1 + 4y_2 + y_3 \\ y_3' &= -2y_1 - 4y_2 - y_3 \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = 0$, $y_2(0) = 1$, $y_3(0) = -3$. [*Technology not required.*]

T2. It is shown in electrical circuit theory that for the *LRC* circuit in Figure Ex-13 the current I in amperes (A) through the inductor and the voltage drop V in volts (V) across the capacitor satisfy the system of differential equations

$$\frac{dI}{dt} = \frac{V}{L}$$

$$\frac{dV}{dt} = -\frac{I}{C} - \frac{V}{RC}$$

where the derivatives are with respect to the time t. Find I and V as functions of t if $L = 0.5$ H, $C = 0.2$ F, $R = 2\ \Omega$, and the initial values of V and I are $V(0) = 1$ V and $I(0) = 2$ A.

5.5 Dynamical Systems and Markov Chains

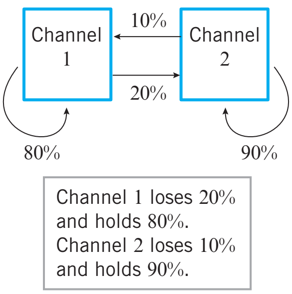
In this optional section we will show how matrix methods can be used to analyze the behavior of physical systems that evolve over time. The methods that we will study here have been applied to problems in business, ecology, demographics, sociology, and most of the physical sciences.

Dynamical Systems

A ***dynamical system*** is a finite set of variables whose values change with time. The value of a variable at a point in time is called the ***state of the variable*** at that time, and the vector formed from these states is called the ***state vector*** of the dynamical system at that time. Our primary objective in this section is to analyze how the state vector of a dynamical system changes with time. Let us begin with an example.

▶ EXAMPLE 1 Market Share as a Dynamical System

Suppose that two competing television channels, channel 1 and channel 2, each have 50% of the viewer market at some initial point in time. Assume that over each one-year period channel 1 captures 10% of channel 2's share, and channel 2 captures 20% of channel 1's share (see Figure 5.5.1). What is each channel's market share after one year?



▲ Figure 5.5.1

Solution Let us begin by introducing the time-dependent variables

$$x_1(t) = \text{fraction of the market held by channel 1 at time } t$$
$$x_2(t) = \text{fraction of the market held by channel 2 at time } t$$

and the column vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \begin{matrix} \leftarrow \textbf{Channel 1's fraction of the market at time } t \textbf{ in years} \\ \leftarrow \textbf{Channel 2's fraction of the market at time } t \textbf{ in years} \end{matrix}$$

The variables $x_1(t)$ and $x_2(t)$ form a dynamical system whose state at time t is the vector $\mathbf{x}(t)$. If we take $t = 0$ to be the starting point at which the two channels had 50% of the market, then the state of the system at that time is

$$\mathbf{x}(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \begin{matrix} \leftarrow \textbf{Channel 1's fraction of the market at time } t = 0 \\ \leftarrow \textbf{Channel 2's fraction of the market at time } t = 0 \end{matrix} \tag{1}$$

Now let us try to find the state of the system at time $t = 1$ (one year later). Over the one-year period, channel 1 retains 80% of its initial 50%, and it gains 10% of channel 2's initial 50%. Thus,

$$\mathbf{x}_1(1) = 0.8(0.5) + 0.1(0.5) = 0.45 \tag{2}$$

Similarly, channel 2 gains 20% of channel 1's initial 50%, and retains 90% of its initial 50%. Thus,

$$x_2(1) = 0.2(0.5) + 0.9(0.5) = 0.55 \tag{3}$$

Therefore, the state of the system at time $t = 1$ is

$$\mathbf{x}(1) = \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix} \begin{matrix} \leftarrow \textbf{Channel 1's fraction of the market at time } t = 1 \\ \leftarrow \textbf{Channel 2's fraction of the market at time } t = 1 \end{matrix} \tag{4}$$

▶ EXAMPLE 2 Evolution of Market Share over Five Years

Track the market shares of channels 1 and 2 in Example 1 over a five-year period.

Solution To solve this problem suppose that we have already computed the market share of each channel at time $t = k$ and we are interested in using the known values of $x_1(k)$ and $x_2(k)$ to compute the market shares $x_1(k+1)$ and $x_2(k+1)$ one year later. The analysis is exactly the same as that used to obtain Equations (2) and (3). Over the one-year period, channel 1 retains 80% of its starting fraction $x_1(k)$ and gains 10% of channel 2's starting fraction $x_2(k)$. Thus,

$$x_1(k+1) = (0.8)x_1(k) + (0.1)x_2(k) \tag{5}$$

Similarly, channel 2 gains 20% of channel 1's starting fraction $x_1(k)$ and retains 90% of its own starting fraction $x_2(k)$. Thus,

$$x_2(k+1) = (0.2)x_1(k) + (0.9)x_2(k) \tag{6}$$

Equations (5) and (6) can be expressed in matrix form as

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \tag{7}$$

which provides a way of using matrix multiplication to compute the state of the system at time $t = k+1$ from the state at time $t = k$. For example, using (1) and (7) we obtain

$$\mathbf{x}(1) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \mathbf{x}(0) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix}$$

which agrees with (4). Similarly,

$$\mathbf{x}(2) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \mathbf{x}(1) = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix} = \begin{bmatrix} 0.415 \\ 0.585 \end{bmatrix}$$

We can now continue this process, using Formula (7) to compute $\mathbf{x}(3)$ from $\mathbf{x}(2)$, then $\mathbf{x}(4)$ from $\mathbf{x}(3)$, and so on. This yields (verify)

$$\mathbf{x}(3) = \begin{bmatrix} 0.3905 \\ 0.6095 \end{bmatrix}, \quad \mathbf{x}(4) = \begin{bmatrix} 0.37335 \\ 0.62665 \end{bmatrix}, \quad \mathbf{x}(5) = \begin{bmatrix} 0.361345 \\ 0.638655 \end{bmatrix} \tag{8}$$

Thus, after five years, channel 1 will hold about 36% of the market and channel 2 will hold about 64% of the market. ◀

If desired, we can continue the market analysis in the last example beyond the five-year period and explore what happens to the market share over the long term. We did so, using a computer, and obtained the following state vectors (rounded to six decimal places):

$$\mathbf{x}(10) \approx \begin{bmatrix} 0.338041 \\ 0.661959 \end{bmatrix}, \quad \mathbf{x}(20) \approx \begin{bmatrix} 0.333466 \\ 0.666534 \end{bmatrix}, \quad \mathbf{x}(40) \approx \begin{bmatrix} 0.333333 \\ 0.666667 \end{bmatrix} \tag{9}$$

All subsequent state vectors, when rounded to six decimal places, are the same as $\mathbf{x}(40)$, so we see that the market shares eventually stabilize with channel 1 holding about one-third of the market and channel 2 holding about two-thirds. Later in this section, we will explain why this stabilization occurs.

Markov Chains

In many dynamical systems the states of the variables are not known with certainty but can be expressed as probabilities; such dynamical systems are called ***stochastic processes*** (from the Greek word *stochastikos*, meaning "proceeding by guesswork"). A detailed study of stochastic processes requires a precise definition of the term *probability*, which is outside the scope of this course. However, the following interpretation will suffice for our present purposes:

> *Stated informally, the **probability** that an experiment or observation will have a certain outcome is the fraction of the time that the outcome would occur if the experiment could be repeated indefinitely under constant conditions—the greater the number of actual repetitions, the more accurately the probability describes the fraction of time that the outcome occurs.*

For example, when we say that the probability of tossing heads with a fair coin is $\frac{1}{2}$, we mean that if the coin were tossed many times under constant conditions, then we would expect about half of the outcomes to be heads. Probabilities are often expressed as decimals or percentages. Thus, the probability of tossing heads with a fair coin can also be expressed as 0.5 or 50%.

If an experiment or observation has n possible outcomes, then the probabilities of those outcomes must be nonnegative fractions whose sum is 1. The probabilities are nonnegative because each describes the fraction of occurrences of an outcome over the long term, and the sum is 1 because they account for all possible outcomes. For example, if a box containing 10 balls has one red ball, three green balls, and six yellow balls, and if a ball is drawn at random from the box, then the probabilities of the various outcomes are

$$\begin{aligned} p_1 &= \text{prob(red)} = 1/10 = 0.1 \\ p_2 &= \text{prob(green)} = 3/10 = 0.3 \\ p_3 &= \text{prob(yellow)} = 6/10 = 0.6 \end{aligned}$$

Each probability is a nonnegative fraction and

$$p_1 + p_2 + p_3 = 0.1 + 0.3 + 0.6 = 1$$

In a stochastic process with n possible states, the state vector at each time t has the form

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \begin{matrix} \textbf{Probability that the system is in state 1} \\ \textbf{Probability that the system is in state 2} \\ \vdots \\ \textbf{Probability that the system is in state } \boldsymbol{n} \end{matrix}$$

The entries in this vector must add up to 1 since they account for all n possibilities. In general, a vector with nonnegative entries that add up to 1 is called a ***probability vector***.

▶ EXAMPLE 3 Example 1 Revisited from the Probability Viewpoint

Observe that the state vectors in Examples 1 and 2 are all probability vectors. This is to be expected since the entries in each state vector are the fractional market shares of the channels, and together they account for the entire market. In practice, it is preferable to interpret the entries in the state vectors as probabilities rather than exact market fractions, since market information is usually obtained by statistical sampling procedures with intrinsic uncertainties. Thus, for example, the state vector

$$\mathbf{x}(1) = \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix}$$

which we interpreted in Example 1 to mean that channel 1 has 45% of the market and channel 2 has 55%, can also be interpreted to mean that an individual picked at random from the market will be a channel 1 viewer with probability 0.45 and a channel 2 viewer with probability 0.55. ◀

A square matrix, each of whose columns is a probability vector, is called a ***stochastic matrix***. Such matrices commonly occur in formulas that relate successive states of a stochastic process. For example, the state vectors $\mathbf{x}(k+1)$ and $\mathbf{x}(k)$ in (7) are related by an equation of the form $\mathbf{x}(k+1) = P\mathbf{x}(k)$ in which

$$P = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \tag{10}$$

is a stochastic matrix. It should not be surprising that the column vectors of P are probability vectors, since the entries in each column provide a breakdown of what happens to each channel's market share over the year—the entries in column 1 convey that each year channel 1 retains 80% of its market share and loses 20%; and the entries in column 2 convey that each year channel 2 retains 90% of its market share and loses 10%. The entries in (10) can also be viewed as probabilities:

$p_{11} = 0.8 =$ probability that a channel 1 viewer remains a channel 1 viewer
$p_{21} = 0.2 =$ probability that a channel 1 viewer becomes a channel 2 viewer
$p_{12} = 0.1 =$ probability that a channel 2 viewer becomes a channel 1 viewer
$p_{22} = 0.9 =$ probability that a channel 2 viewer remains a channel 2 viewer

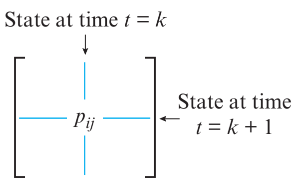
Example 1 is a special case of a large class of stochastic processes called *Markov chains*.

DEFINITION 1 A ***Markov chain*** is a dynamical system whose state vectors at a succession of equally spaced times are probability vectors and for which the state vectors at successive times are related by an equation of the form

$$\mathbf{x}(k+1) = P\mathbf{x}(k)$$

in which $P = [p_{ij}]$ is a stochastic matrix and p_{ij} is the probability that the system will be in state i at time $t = k+1$ if it is in state j at time $t = k$. The matrix P is called the ***transition matrix*** for the system.

WARNING Note that in this definition the row index i corresponds to the later state and the column index j to the earlier state (Figure 5.5.2).



The entry p_{ij} is the probability that the system is in state i at time $t = k + 1$ if it is in state j at time $t = k$.

▲ Figure 5.5.2

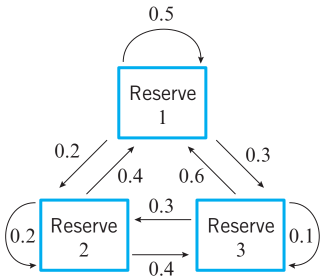


Figure 5.5.3

▶ EXAMPLE 4 Wildlife Migration as a Markov Chain

Suppose that a tagged lion can migrate over three adjacent game reserves in search of food, reserve 1, reserve 2, and reserve 3. Based on data about the food resources, researchers conclude that the monthly migration pattern of the lion can be modeled by a Markov chain with transition matrix

$$\begin{array}{c} \text{Reserve at time } t = k \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \end{array}$$

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.6 \\ 0.2 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \quad \text{Reserve at time } t = k+1$$

(see Figure 5.5.3). That is,

$p_{11} = 0.5 =$ probability that the lion will stay in reserve 1 when it is in reserve 1
$p_{12} = 0.4 =$ probability that the lion will move from reserve 2 to reserve 1
$p_{13} = 0.6 =$ probability that the lion will move from reserve 3 to reserve 1
$p_{21} = 0.2 =$ probability that the lion will move from reserve 1 to reserve 2
$p_{22} = 0.2 =$ probability that the lion will stay in reserve 2 when it is in reserve 2
$p_{23} = 0.3 =$ probability that the lion will move from reserve 3 to reserve 2
$p_{31} = 0.3 =$ probability that the lion will move from reserve 1 to reserve 3
$p_{32} = 0.4 =$ probability that the lion will move from reserve 2 to reserve 3
$p_{33} = 0.1 =$ probability that the lion will stay in reserve 3 when it is in reserve 3

Assuming that t is in months and the lion is released in reserve 2 at time $t = 0$, track its probable locations over a six-month period.

Solution Let $x_1(k)$, $x_2(k)$, and $x_3(k)$ be the probabilities that the lion is in reserve 1, 2, or 3, respectively, at time $t = k$, and let

$$\mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

be the state vector at that time. Since we know with certainty that the lion is in reserve 2 at time $t = 0$, the initial state vector is

$$\mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Andrei Andreyevich Markov (1856–1922)

Historical Note Markov chains are named in honor of the Russian mathematician A. A. Markov, a lover of poetry, who used them to analyze the alternation of vowels and consonants in the poem *Eugene Onegin* by Pushkin. Markov believed that the only applications of his chains were to the analysis of literary works, so he would be astonished to learn that his discovery is used today in the social sciences, quantum theory, and genetics!

[*Image: SPL/Science Source*]

We leave it for you to show that the state vectors over a six-month period are

$$\mathbf{x}(1) = P\mathbf{x}(0) = \begin{bmatrix} 0.400 \\ 0.200 \\ 0.400 \end{bmatrix}, \quad \mathbf{x}(2) = P\mathbf{x}(1) = \begin{bmatrix} 0.520 \\ 0.240 \\ 0.240 \end{bmatrix}, \quad \mathbf{x}(3) = P\mathbf{x}(2) = \begin{bmatrix} 0.500 \\ 0.224 \\ 0.276 \end{bmatrix}$$

$$\mathbf{x}(4) = P\mathbf{x}(3) \approx \begin{bmatrix} 0.505 \\ 0.228 \\ 0.267 \end{bmatrix}, \quad \mathbf{x}(5) = P\mathbf{x}(4) \approx \begin{bmatrix} 0.504 \\ 0.227 \\ 0.269 \end{bmatrix}, \quad \mathbf{x}(6) = P\mathbf{x}(5) \approx \begin{bmatrix} 0.504 \\ 0.227 \\ 0.269 \end{bmatrix}$$

As in Example 2, the state vectors here seem to stabilize over time with a probability of approximately 0.504 that the lion is in reserve 1, a probability of approximately 0.227 that it is in reserve 2, and a probability of approximately 0.269 that it is in reserve 3. ◀

Markov Chains in Terms of Powers of the Transition Matrix

In a Markov chain with an initial state of $\mathbf{x}(0)$, the successive state vectors are

$$\mathbf{x}(1) = P\mathbf{x}(0), \quad \mathbf{x}(2) = P\mathbf{x}(1), \quad \mathbf{x}(3) = P\mathbf{x}(2), \quad \mathbf{x}(4) = P\mathbf{x}(3), \ldots$$

For brevity, it is common to denote $\mathbf{x}(k)$ by $\mathbf{x}_k$, which allows us to write the successive state vectors more briefly as

$$\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P\mathbf{x}_1, \quad \mathbf{x}_3 = P\mathbf{x}_2, \quad \mathbf{x}_4 = P\mathbf{x}_3, \ldots \tag{11}$$

Alternatively, these state vectors can be expressed in terms of the initial state vector $\mathbf{x}_0$ as

$$\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P(P\mathbf{x}_0) = P^2\mathbf{x}_0, \quad \mathbf{x}_3 = P(P^2\mathbf{x}_0) = P^3\mathbf{x}_0, \quad \mathbf{x}_4 = P(P^3\mathbf{x}_0) = P^4\mathbf{x}_0, \ldots$$

from which it follows that

$$\mathbf{x}_k = P^k\mathbf{x}_0 \tag{12}$$

Note that Formula (12) makes it possible to compute the state vector $\mathbf{x}_k$ without first computing the earlier state vectors as required in Formula (11).

▶ EXAMPLE 5 Finding a State Vector Directly from $\mathbf{x}_0$

Use Formula (12) to find the state vector $\mathbf{x}(3)$ in Example 2.

Solution From (1) and (7), the initial state vector and transition matrix are

$$\mathbf{x}_0 = \mathbf{x}(0) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}$$

We leave it for you to calculate P^3 and show that

$$\mathbf{x}(3) = \mathbf{x}_3 = P^3\mathbf{x}_0 = \begin{bmatrix} 0.562 & 0.219 \\ 0.438 & 0.781 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.3905 \\ 0.6095 \end{bmatrix}$$

which agrees with the result in (8). ◀

Long-Term Behavior of a Markov Chain

We have seen two examples of Markov chains in which the state vectors seem to stabilize after a period of time. Thus, it is reasonable to ask whether all Markov chains have this property. The following example shows that this is not the case.

▶ EXAMPLE 6 A Markov Chain That Does Not Stabilize

The matrix

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

is stochastic and hence can be regarded as the transition matrix for a Markov chain. A simple calculation shows that $P^2 = I$, from which it follows that

$$I = P^2 = P^4 = P^6 = \cdots \quad \text{and} \quad P = P^3 = P^5 = P^7 = \cdots$$

Thus, the successive states in the Markov chain with initial vector $\mathbf{x}_0$ are

$$\mathbf{x}_0, \quad P\mathbf{x}_0, \quad \mathbf{x}_0, \quad P\mathbf{x}_0, \quad \mathbf{x}_0, \ldots$$

which oscillate between $\mathbf{x}_0$ and $P\mathbf{x}_0$. Thus, the Markov chain does not stabilize unless both components of $\mathbf{x}_0$ are $\frac{1}{2}$ (verify). ◀

A precise definition of what it means for a sequence of numbers or vectors to stabilize is given in calculus; however, that level of precision will not be needed here. Stated informally, we will say that a sequence of vectors

$$\mathbf{x}_1, \quad \mathbf{x}_2, \ldots, \quad \mathbf{x}_k, \ldots$$

approaches a ***limit*** $\mathbf{q}$ or that it ***converges*** to $\mathbf{q}$ if all entries in $\mathbf{x}_k$ can be made as close as we like to the corresponding entries in the vector $\mathbf{q}$ by taking k sufficiently large. We denote this by writing $\mathbf{x}_k \to \mathbf{q}$ as $k \to \infty$. Similarly, we say that a sequence of matrices

$$P_1, P_2, P_3, \ldots, P_k, \ldots$$

converges to a matrix Q, written $P_k \to Q$ as $k \to \infty$, if each entry of P_k can be made as close as we like to the corresponding entry of Q by taking k sufficiently large.

We saw in Example 6 that the state vectors of a Markov chain need not approach a limit in all cases. However, by imposing a mild condition on the transition matrix of a Markov chain, we can guarantee that the state vectors will approach a limit.

DEFINITION 2 A stochastic matrix P is said to be ***regular*** if P or some positive power of P has all positive entries, and a Markov chain whose transition matrix is regular is said to be a ***regular Markov chain***.

▶ EXAMPLE 7 Regular Stochastic Matrices

The transition matrices in Examples 2 and 4 are regular because their entries are positive. The matrix

$$P = \begin{bmatrix} 0.5 & 1 \\ 0.5 & 0 \end{bmatrix}$$

is regular because

$$P^2 = \begin{bmatrix} 0.75 & 0.5 \\ 0.25 & 0.5 \end{bmatrix}$$

has positive entries. The matrix P in Example 6 is not regular because P and every positive power of P have some zero entries (verify). ◀

The following theorem, which we state without proof, is the fundamental result about the long-term behavior of Markov chains.

THEOREM 5.5.1 *If P is the transition matrix for a regular Markov chain, then:*

(*a*) *There is a unique probability vector* $\mathbf{q}$ *with positive entries such that* $P\mathbf{q} = \mathbf{q}$.

(*b*) *For any initial probability vector* $\mathbf{x}_0$, *the sequence of state vectors*

$$\mathbf{x}_0, \quad P\mathbf{x}_0, \ldots, \quad P^k\mathbf{x}_0, \ldots$$

converges to $\mathbf{q}$.

(*c*) *The sequence* $P, P^2, P^3, \ldots, P^k, \ldots$ *converges to the matrix Q each of whose column vectors is* $\mathbf{q}$.

The vector $\mathbf{q}$ in Theorem 5.5.1 is called the ***steady-state*** vector of the Markov chain. Because it is a nonzero vector that satisfies the equation $P\mathbf{q} = \mathbf{q}$, it is an eigenvector corresponding to the eigenvalue $\lambda = 1$ of P. Thus, $\mathbf{q}$ can be found by solving the linear system

$$(I - P)\mathbf{q} = \mathbf{0} \tag{13}$$

subject to the requirement that $\mathbf{q}$ be a probability vector. Here are some examples.

▶ EXAMPLE 8 Examples 1 and 2 Revisited

The transition matrix for the Markov chain in Example 2 is

$$P = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}$$

Since the entries of P are positive, the Markov chain is regular and hence has a unique steady-state vector $\mathbf{q}$. To find $\mathbf{q}$ we will solve the system $(I - P)\mathbf{q} = \mathbf{0}$, which we can write as

$$\begin{bmatrix} 0.2 & -0.1 \\ -0.2 & 0.1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The general solution of this system is

$$q_1 = 0.5s, \quad q_2 = s$$

(verify), which we can write in vector form as

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} 0.5s \\ s \end{bmatrix} = \begin{bmatrix} \frac{1}{2}s \\ s \end{bmatrix} \tag{14}$$

For $\mathbf{q}$ to be a probability vector, we must have

$$1 = q_1 + q_2 = \tfrac{3}{2}s$$

which implies that $s = \frac{2}{3}$. Substituting this value in (14) yields the steady-state vector

$$\mathbf{q} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$$

which is consistent with the numerical results obtained in (9).

▶ EXAMPLE 9 Example 4 Revisited

The transition matrix for the Markov chain in Example 4 is

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.6 \\ 0.2 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 \end{bmatrix}$$

Since the entries of P are positive, the Markov chain is regular and hence has a unique steady-state vector $\mathbf{q}$. To find $\mathbf{q}$ we will solve the system $(I - P)\mathbf{q} = \mathbf{0}$, which we can write (using fractions) as

$$\begin{bmatrix} \frac{1}{2} & -\frac{2}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{4}{5} & -\frac{3}{10} \\ -\frac{3}{10} & -\frac{2}{5} & \frac{9}{10} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{15}$$

(We have converted to fractions to avoid roundoff error in this illustrative example.) We leave it for you to confirm that the reduced row echelon form of the coefficient matrix is

$$\begin{bmatrix} 1 & 0 & -\frac{15}{8} \\ 0 & 1 & -\frac{27}{32} \\ 0 & 0 & 0 \end{bmatrix}$$

and that the general solution of (15) is

$$q_1 = \tfrac{15}{8}s, \quad q_2 = \tfrac{27}{32}s, \quad q_3 = s \tag{16}$$

For $\mathbf{q}$ to be a probability vector we must have $q_1 + q_2 + q_3 = 1$, from which it follows that $s = \frac{32}{119}$ (verify). Substituting this value in (16) yields the steady-state vector

$$\mathbf{q} = \begin{bmatrix} \frac{60}{119} \\ \frac{27}{119} \\ \frac{32}{119} \end{bmatrix} \approx \begin{bmatrix} 0.5042 \\ 0.2269 \\ 0.2689 \end{bmatrix}$$

(verify), which is consistent with the results obtained in Example 4. ◀

Exercise Set 5.5

▶ In Exercises 1–2, determine whether A is a stochastic matrix. If A is not stochastic, then explain why not. ◀

1. (a) $A = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}$ (b) $A = \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$ (d) $A = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix}$

2. (a) $A = \begin{bmatrix} 0.2 & 0.9 \\ 0.8 & 0.1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 0.2 & 0.8 \\ 0.9 & 0.1 \end{bmatrix}$

(c) $A = \begin{bmatrix} \frac{1}{12} & \frac{1}{9} & \frac{1}{6} \\ \frac{1}{2} & 0 & \frac{5}{6} \\ \frac{5}{12} & \frac{8}{9} & 0 \end{bmatrix}$ (d) $A = \begin{bmatrix} -1 & \frac{1}{3} & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{1}{2} \\ 2 & \frac{1}{3} & 0 \end{bmatrix}$

▶ In Exercises 3–4, use Formulas (11) and (12) to compute the state vector $\mathbf{x}_4$ in two different ways. ◀

3. $P = \begin{bmatrix} 0.5 & 0.6 \\ 0.5 & 0.4 \end{bmatrix}$; $\mathbf{x}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

4. $P = \begin{bmatrix} 0.8 & 0.5 \\ 0.2 & 0.5 \end{bmatrix}$; $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

▶ In Exercises 5–6, determine whether P is a regular stochastic matrix. ◀

5. (a) $P = \begin{bmatrix} \frac{1}{5} & \frac{1}{7} \\ \frac{4}{5} & \frac{6}{7} \end{bmatrix}$ (b) $P = \begin{bmatrix} \frac{1}{5} & 0 \\ \frac{4}{5} & 1 \end{bmatrix}$ (c) $P = \begin{bmatrix} \frac{1}{5} & 1 \\ \frac{4}{5} & 0 \end{bmatrix}$

6. (a) $P = \begin{bmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{bmatrix}$ (b) $P = \begin{bmatrix} 1 & \frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}$ (c) $P = \begin{bmatrix} \frac{3}{4} & \frac{1}{3} \\ \frac{1}{4} & \frac{2}{3} \end{bmatrix}$

▶ In Exercises 7–10, verify that P is a regular stochastic matrix, and find the steady-state vector for the associated Markov chain. ◀

7. $P = \begin{bmatrix} \frac{1}{4} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{3} \end{bmatrix}$

8. $P = \begin{bmatrix} 0.2 & 0.6 \\ 0.8 & 0.4 \end{bmatrix}$

9. $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{2}{3} \end{bmatrix}$

10. $P = \begin{bmatrix} \frac{1}{3} & \frac{1}{4} & \frac{2}{5} \\ 0 & \frac{3}{4} & \frac{2}{5} \\ \frac{2}{3} & 0 & \frac{1}{5} \end{bmatrix}$

11. Consider a Markov process with transition matrix

$$\begin{array}{c} \\ \textbf{State 1} \\ \textbf{State 2} \end{array} \begin{array}{c} \textbf{State 1} \quad \textbf{State 2} \\ \begin{bmatrix} 0.2 & 0.1 \\ 0.8 & 0.9 \end{bmatrix} \end{array}$$

(a) What does the entry 0.2 represent?

(b) What does the entry 0.1 represent?

(c) If the system is in state 1 initially, what is the probability that it will be in state 2 at the next observation?

(d) If the system has a 50% chance of being in state 1 initially, what is the probability that it will be in state 2 at the next observation?

12. Consider a Markov process with transition matrix

$$\begin{array}{c} \\ \textbf{State 1} \\ \textbf{State 2} \end{array} \begin{array}{c} \textbf{State 1} \quad \textbf{State 2} \\ \begin{bmatrix} 0 & \frac{1}{7} \\ 1 & \frac{6}{7} \end{bmatrix} \end{array}$$

(a) What does the entry $\frac{6}{7}$ represent?

(b) What does the entry 0 represent?

(c) If the system is in state 1 initially, what is the probability that it will be in state 1 at the next observation?

(d) If the system has a 50% chance of being in state 1 initially, what is the probability that it will be in state 2 at the next observation?

13. On a given day the air quality in a certain city is either good or bad. Records show that when the air quality is good on one day, then there is a 95% chance that it will be good the next day, and when the air quality is bad on one day, then there is a 45% chance that it will be bad the next day.

(a) Find a transition matrix for this phenomenon.

(b) If the air quality is good today, what is the probability that it will be good two days from now?

(c) If the air quality is bad today, what is the probability that it will be bad three days from now?

(d) If there is a 20% chance that the air quality will be good today, what is the probability that it will be good tomorrow?

14. In a laboratory experiment, a mouse can choose one of two food types each day, type I or type II. Records show that if the mouse chooses type I on a given day, then there is a 75% chance that it will choose type I the next day, and if it chooses type II on one day, then there is a 50% chance that it will choose type II the next day.

(a) Find a transition matrix for this phenomenon.

(b) If the mouse chooses type I today, what is the probability that it will choose type I two days from now?

(c) If the mouse chooses type II today, what is the probability that it will choose type II three days from now?

(d) If there is a 10% chance that the mouse will choose type I today, what is the probability that it will choose type I tomorrow?

15. Suppose that at some initial point in time 100,000 people live in a certain city and 25,000 people live in its suburbs. The Regional Planning Commission determines that each year 5% of the city population moves to the suburbs and 3% of the suburban population moves to the city.

(a) Assuming that the total population remains constant, make a table that shows the populations of the city and its suburbs over a five-year period (round to the nearest integer).

(b) Over the long term, how will the population be distributed between the city and its suburbs?

16. Suppose that two competing television stations, station 1 and station 2, each have 50% of the viewer market at some initial point in time. Assume that over each one-year period station 1 captures 5% of station 2's market share and station 2 captures 10% of station 1's market share.

(a) Make a table that shows the market share of each station over a five-year period.

(b) Over the long term, how will the market share be distributed between the two stations?

17. Fill in the missing entries of the stochastic matrix

$$P = \begin{bmatrix} \frac{7}{10} & * & \frac{1}{5} \\ * & \frac{3}{10} & * \\ \frac{1}{10} & \frac{3}{5} & \frac{3}{10} \end{bmatrix}$$

and find its steady-state vector.

18. If P is an $n \times n$ stochastic matrix, and if M is a $1 \times n$ matrix whose entries are all 1's, then $MP =$ ________.

19. If P is a regular stochastic matrix with steady-state vector $\mathbf{q}$, what can you say about the sequence of products

$$P\mathbf{q}, \quad P^2\mathbf{q}, \quad P^3\mathbf{q}, \ldots, \quad P^k\mathbf{q}, \ldots$$

as $k \to \infty$?

20. (a) If P is a regular $n \times n$ stochastic matrix with steady-state vector $\mathbf{q}$, and if $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the standard unit vectors in column form, what can you say about the behavior of the sequence

$$P\mathbf{e}_i, \quad P^2\mathbf{e}_i, \quad P^3\mathbf{e}_i, \ldots, \quad P^k\mathbf{e}_i, \ldots$$

as $k \to \infty$ for each $i = 1, 2, \ldots, n$?

(b) What does this tell you about the behavior of the column vectors of P^k as $k \to \infty$?

Working with Proofs

21. Prove that the product of two stochastic matrices with the same size is a stochastic matrix. [*Hint:* Write each column of the product as a linear combination of the columns of the first factor.]

22. Prove that if P is a stochastic matrix whose entries are all greater than or equal to ρ, then the entries of P^2 are greater than or equal to ρ.

True-False Exercises

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

(a) The vector $\begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}$ is a probability vector.

(b) The matrix $\begin{bmatrix} 0.2 & 1 \\ 0.8 & 0 \end{bmatrix}$ is a regular stochastic matrix.

(c) The column vectors of a transition matrix are probability vectors.

(d) A steady-state vector for a Markov chain with transition matrix P is any solution of the linear system $(I - P)\mathbf{q} = \mathbf{0}$.

(e) The square of every regular stochastic matrix is stochastic.

(f) A vector with real entries that sum to 1 is a probability vector.

(g) Every regular stochastic matrix has $\lambda = 1$ as an eigenvalue.

Working with Technology

T1. In Examples 4 and 9 we considered the Markov chain with transition matrix P and initial state vector $\mathbf{x}(0)$ where

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.6 \\ 0.2 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

(a) Confirm the numerical values of $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(6)$ obtained in Example 4 using the method given in that example.

(b) As guaranteed by part (*c*) of Theorem 5.5.1, confirm that the sequence $P, P^2, P^3, \ldots, P^k, \ldots$ converges to the matrix Q each of whose column vectors is the steady-state vector $\mathbf{q}$ obtained in Example 9.

T2. Suppose that a car rental agency has three locations, numbered 1, 2, and 3. A customer may rent a car from any of the three locations and return it to any of the three locations. Records show that cars are rented and returned in accordance with the following probabilities:

		Rented from Location		
		1	**2**	**3**
Returned to Location	**1**	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{5}$
	2	$\frac{4}{5}$	$\frac{3}{10}$	$\frac{1}{5}$
	3	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{1}{5}$

(a) Assuming that a car is rented from location 1, what is the probability that it will be at location 1 after two rentals?

(b) Assuming that this dynamical system can be modeled as a Markov chain, find the steady-state vector.

(c) If the rental agency owns 120 cars, how many parking spaces should it allocate at each location to be reasonably certain that it will have enough spaces for the cars over the long term? Explain your reasoning.

T3. Physical traits are determined by the genes that an offspring receives from its parents. In the simplest case a trait in the offspring is determined by one pair of genes, one member of the pair inherited from the male parent and the other from the female parent. Typically, each gene in a pair can assume one of two forms, called ***alleles***, denoted by A and a. This leads to three possible pairings:

$$AA, \quad Aa, \quad aa$$

called ***genotypes*** (the pairs Aa and aA determine the same trait and hence are not distinguished from one another). It is shown in the study of heredity that if a parent of known genotype is crossed with a random parent of unknown genotype, then the offspring will have the genotype probabilities given in the following table, which can be viewed as a transition matrix for a Markov process:

		Genotype of Parent		
		AA	***Aa***	***aa***
Genotype of Offspring	***AA***	$\frac{1}{2}$	$\frac{1}{4}$	0
	Aa	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
	aa	0	$\frac{1}{4}$	$\frac{1}{2}$

Thus, for example, the offspring of a parent of genotype AA that is crossed at random with a parent of unknown genotype will have a 50% chance of being AA, a 50% chance of being Aa, and no chance of being aa.

(a) Show that the transition matrix is regular.

(b) Find the steady-state vector, and discuss its physical interpretation.

Chapter 5 Supplementary Exercises

1. (a) Show that if $0 < \theta < \pi$, then

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

has no real eigenvalues and consequently no real eigenvectors.

(b) Give a geometric explanation of the result in part (a).

2. Find the eigenvalues of

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

3. (a) Show that if D is a diagonal matrix with nonnegative entries on the main diagonal, then there is a matrix S such that $S^2 = D$.

(b) Show that if A is a diagonalizable matrix with nonnegative eigenvalues, then there is a matrix S such that $S^2 = A$.

(c) Find a matrix S such that $S^2 = A$, given that

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

4. Given that A and B are similar matrices, in each part determine whether the given matrices are also similar.

(a) A^T and B^T

(b) A^k and B^k (k a positive integer)

(c) A^{-1} and B^{-1} (if A is invertible)

5. Prove: If A is a square matrix and $p(\lambda) = \det(\lambda I - A)$ is the characteristic polynomial of A, then the coefficient of λ^{n-1} in $p(\lambda)$ is the negative of the trace of A.

6. Prove: If $b \neq 0$, then

$$A = \begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$$

is not diagonalizable.

7. In advanced linear algebra, one proves the ***Cayley–Hamilton Theorem***, which states that a square matrix A satisfies its characteristic equation; that is, if

$$c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n = 0$$

is the characteristic equation of A, then

$$c_0 I + c_1 A + c_2 A^2 + \cdots + c_{n-1}A^{n-1} + A^n = 0$$

Verify this result for

(a) $A = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix}$ (b) $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}$

In Exercises 8–10, use the Cayley–Hamilton Theorem, stated in Exercise 7.

8. (a) Use Exercise 28 of Section 5.1 to establish the Cayley–Hamilton Theorem for 2×2 matrices.

(b) Prove the Cayley–Hamilton Theorem for $n \times n$ diagonalizable matrices.

9. The Cayley–Hamilton Theorem provides a method for calculating powers of a matrix. For example, if A is a 2×2 matrix with characteristic equation

$$c_0 + c_1\lambda + \lambda^2 = 0$$

then $c_0 I + c_1 A + A^2 = 0$, so

$$A^2 = -c_1 A - c_0 I$$

Multiplying through by A yields $A^3 = -c_1 A^2 - c_0 A$, which expresses A^3 in terms of A^2 and A, and multiplying through by A^2 yields $A^4 = -c_1 A^3 - c_0 A^2$, which expresses A^4 in terms of A^3 and A^2. Continuing in this way, we can calculate successive powers of A by expressing them in terms of lower powers. Use this procedure to calculate A^2, A^3, A^4, and A^5 for

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

10. Use the method of the preceding exercise to calculate A^3 and A^4 for

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

11. Find the eigenvalues of the matrix

$$A = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_n \\ \vdots & \vdots & & \vdots \\ c_1 & c_2 & \cdots & c_n \end{bmatrix}$$

12. (a) It was shown in Exercise 37 of Section 5.1 that if A is an $n \times n$ matrix, then the coefficient of λ^n in the characteristic polynomial of A is 1. (A polynomial with this property is called ***monic***.) Show that the matrix

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix}$$

has characteristic polynomial

$$p(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n$$

This shows that every monic polynomial is the characteristic polynomial of some matrix. The matrix in this example is called the ***companion matrix*** of $p(\lambda)$. [*Hint:* Evaluate all determinants in the problem by adding a multiple of the second row to the first to introduce a zero at the top of the first column, and then expanding by cofactors along the first column.]

(b) Find a matrix with characteristic polynomial

$$p(\lambda) = 1 - 2\lambda + \lambda^2 + 3\lambda^3 + \lambda^4$$

13. A square matrix A is called ***nilpotent*** if $A^n = 0$ for some positive integer n. What can you say about the eigenvalues of a nilpotent matrix?

14. Prove: If A is an $n \times n$ matrix and n is odd, then A has at least one real eigenvalue.

15. Find a 3×3 matrix A that has eigenvalues $\lambda = 0, 1$, and -1 with corresponding eigenvectors

$$\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

respectively.

16. Suppose that a 4×4 matrix A has eigenvalues $\lambda_1 = 1$, $\lambda_2 = -2$, $\lambda_3 = 3$, and $\lambda_4 = -3$.

(a) Use the method of Exercise 24 of Section 5.1 to find $\det(A)$.

(b) Use Exercise 5 above to find $\operatorname{tr}(A)$.

17. Let A be a square matrix such that $A^3 = A$. What can you say about the eigenvalues of A?

18. (a) Solve the system

$$\begin{aligned} y_1' &= y_1 + 3y_2 \\ y_2' &= 2y_1 + 4y_2 \end{aligned}$$

(b) Find the solution satisfying the initial conditions $y_1(0) = 5$ and $y_2(0) = 6$.

19. Let A be a 3×3 matrix, one of whose eigenvalues is 1. Given that both the sum and the product of all three eigenvalues is 6, what are the possible values for the remaining two eigenvalues?

20. Show that the matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$

are similar if

$$d_k = \cos\frac{2\pi k}{3} + i\sin\frac{2\pi k}{3} \quad (k = 1, 2, 3)$$