CHAPTER 7

Linear Algebra: Matrices, Vectors, Determinants. Linear Systems

Linear algebra is a fairly extensive subject that covers vectors and matrices, determinants, systems of linear equations, vector spaces and linear transformations, eigenvalue problems, and other topics. As an area of study it has a broad appeal in that it has many applications in engineering, physics, geometry, computer science, economics, and other areas. It also contributes to a deeper understanding of mathematics itself.

Matrices, which are rectangular arrays of numbers or functions, and **vectors** are the main tools of linear algebra. Matrices are important because they let us express large amounts of data and functions in an organized and concise form. Furthermore, since matrices are single objects, we denote them by single letters and calculate with them directly. All these features have made matrices and vectors very popular for expressing scientific and mathematical ideas.

The chapter keeps a good mix between applications (electric networks, Markov processes, traffic flow, etc.) and theory. Chapter 7 is structured as follows: Sections 7.1 and 7.2 provide an intuitive introduction to matrices and vectors and their operations, including matrix multiplication. The next block of sections, that is, Secs. 7.3–7.5 provide the most important method for ***solving systems of linear equations*** by the Gauss elimination method. This method is a cornerstone of linear algebra, and the method itself and variants of it appear in different areas of mathematics and in many applications. It leads to a consideration of the behavior of solutions and concepts such as rank of a matrix, linear independence, and bases. We shift to determinants, a topic that has declined in importance, in Secs. 7.6 and 7.7. Section 7.8 covers inverses of matrices. The chapter ends with vector spaces, inner product spaces, linear transformations, and composition of linear transformations. Eigenvalue problems follow in Chap. 8.

COMMENT. ***Numeric linear algebra (Secs. 20.1–20.5) can be studied immediately after this chapter.***

Prerequisite: None.
Sections that may be omitted in a short course: 7.5, 7.9.
References and Answers to Problems: App. 1 Part B, and App. 2.

7.1 Matrices, Vectors: Addition and Scalar Multiplication

The basic concepts and rules of matrix and vector algebra are introduced in Secs. 7.1 and 7.2 and are followed by **linear systems** (systems of linear equations), a main application, in Sec. 7.3.

Let us first take a leisurely look at matrices before we formalize our discussion. A **matrix** is a rectangular array of numbers or functions which we will enclose in brackets. For example,

$$
(1) \qquad \begin{bmatrix} 0.3 & 1 & -5 \\ 0 & -0.2 & 16 \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},
$$

$$
\begin{bmatrix} e^{-x} & 2x^2 \\ e^{6x} & 4x \end{bmatrix}, \quad [a_1 \quad a_2 \quad a_3], \quad \begin{bmatrix} 4 \\ \frac{1}{2} \end{bmatrix}
$$

are matrices. The numbers (or functions) are called **entries** or, less commonly, *elements* of the matrix. The first matrix in (1) has two **rows**, which are the horizontal lines of entries. Furthermore, it has three **columns**, which are the vertical lines of entries. The second and third matrices are **square matrices**, which means that each has as many rows as columns—3 and 2, respectively. The entries of the second matrix have two indices, signifying their location within the matrix. The first index is the number of the row and the second is the number of the column, so that together the entry's position is uniquely identified. For example, a_{23} (read *a two three*) is in Row 2 and Column 3, etc. The notation is standard and applies to all matrices, including those that are not square.

Matrices having just a single row or column are called **vectors**. Thus, the fourth matrix in (1) has just one row and is called a **row vector**. The last matrix in (1) has just one column and is called a **column vector**. Because the goal of the indexing of entries was to uniquely identify the position of an element within a matrix, one index suffices for vectors, whether they are row or column vectors. Thus, the third entry of the row vector in (1) is denoted by a_3.

Matrices are handy for storing and processing data in applications. Consider the following two common examples.

EXAMPLE 1 Linear Systems, a Major Application of Matrices

We are given a system of linear equations, briefly a **linear system**, such as

$$
\begin{aligned}
4x_1 + 6x_2 + 9x_3 &= 6 \\
6x_1 \qquad\;\; - 2x_3 &= 20 \\
5x_1 - 8x_2 + \;\; x_3 &= 10
\end{aligned}
$$

where x_1, x_2, x_3 are the **unknowns**. We form the **coefficient matrix**, call it $\mathbf{A}$, by listing the coefficients of the unknowns in the position in which they appear in the linear equations. In the second equation, there is no unknown x_2, which means that the coefficient of x_2 is 0 and hence in matrix $\mathbf{A}$, $a_{22} = 0$, Thus,

$$\mathbf{A} = \begin{bmatrix} 4 & 6 & 9 \\ 6 & 0 & -2 \\ 5 & -8 & 1 \end{bmatrix}. \quad \text{We form another matrix} \quad \tilde{\mathbf{A}} = \begin{bmatrix} 4 & 6 & 9 & 6 \\ 6 & 0 & -2 & 20 \\ 5 & -8 & 1 & 10 \end{bmatrix}$$

by augmenting **A** with the right sides of the linear system and call it the augmented matrix of the system.

Since we can go back and recapture the system of linear equations directly from the augmented matrix $\tilde{\mathbf{A}}$, $\tilde{\mathbf{A}}$ contains all the information of the system and can thus be used to solve the linear system. This means that we can just use the augmented matrix to do the calculations needed to solve the system. We shall explain this in detail in Sec. 7.3. Meanwhile you may verify by substitution that the solution is $x_1 = 3, x_2 = \frac{1}{2}, x_3 = -1$.

The notation x_1, x_2, x_3 for the unknowns is practical but not essential; we could choose x, y, z or some other letters. ■

EXAMPLE 2 Sales Figures in Matrix Form

Sales figures for three products I, II, III in a store on Monday (Mon), Tuesday (Tues), $\cdots$ may for each week be arranged in a matrix

$$\begin{array}{cc} & \begin{array}{ccccccc} \text{Mon} & \text{Tues} & \text{Wed} & \text{Thur} & \text{Fri} & \text{Sat} & \text{Sun} \end{array} \\ \mathbf{A} = & \begin{bmatrix} 40 & 33 & 81 & 0 & 21 & 47 & 33 \\ 0 & 12 & 78 & 50 & 50 & 96 & 90 \\ 10 & 0 & 0 & 27 & 43 & 78 & 56 \end{bmatrix} \cdot \begin{array}{c} \text{I} \\ \text{II} \\ \text{III} \end{array} \end{array}$$

If the company has 10 stores, we can set up 10 such matrices, one for each store. Then, by adding corresponding entries of these matrices, we can get a matrix showing the total sales of each product on each day. Can you think of other data which can be stored in matrix form? For instance, in transportation or storage problems? Or in listing distances in a network of roads? ■

General Concepts and Notations

Let us formalize what we just have discussed. We shall denote matrices by capital boldface letters **A**, **B**, **C**, $\cdots$, or by writing the general entry in brackets; thus $\mathbf{A} = [a_{jk}]$, and so on. By an ***m* × *n* matrix** (read *m by n matrix*) we mean a matrix with m rows and n columns—rows always come first! $m \times n$ is called the **size** of the matrix. Thus an $m \times n$ matrix is of the form

$$\mathbf{A} = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \tag{2}$$

The matrices in (1) are of sizes 2×3, 3×3, 2×2, 1×3, and 2×1, respectively.

Each entry in (2) has two subscripts. The first is the *row number* and the second is the *column number*. Thus a_{21} is the entry in Row 2 and Column 1.

If $m = n$, we call **A** an $n \times n$ **square matrix**. Then its diagonal containing the entries $a_{11}, a_{22}, \cdots, a_{nn}$ is called the **main diagonal** of **A**. Thus the main diagonals of the two square matrices in (1) are a_{11}, a_{22}, a_{33} and e^{-x}, $4x$, respectively.

Square matrices are particularly important, as we shall see. A matrix of any size $m \times n$ is called a **rectangular matrix**; this includes square matrices as a special case.

Vectors

A **vector** is a matrix with only one row or column. Its entries are called the **components** of the vector. We shall denote vectors by ***lowercase*** boldface letters **a**, **b**, $\cdots$ or by its general component in brackets, $\mathbf{a} = [a_j]$, and so on. Our special vectors in (1) suggest that a (general) **row vector** is of the form

$$\mathbf{a} = [a_1 \quad a_2 \quad \cdots \quad a_n]. \qquad \text{For instance,} \qquad \mathbf{a} = [-2 \quad 5 \quad 0.8 \quad 0 \quad 1].$$

A **column vector** is of the form

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \qquad \text{For instance,} \qquad \mathbf{b} = \begin{bmatrix} 4 \\ 0 \\ -7 \end{bmatrix}.$$

Addition and Scalar Multiplication of Matrices and Vectors

What makes matrices and vectors really useful and particularly suitable for computers is the fact that we can calculate with them almost as easily as with numbers. Indeed, we now introduce rules for addition and for scalar multiplication (multiplication by numbers) that were suggested by practical applications. (Multiplication of matrices by matrices follows in the next section.) We first need the concept of equality.

DEFINITION

Equality of Matrices

Two matrices $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ are **equal**, written $\mathbf{A} = \mathbf{B}$, if and only if they have the same size and the corresponding entries are equal, that is, $a_{11} = b_{11}$, $a_{12} = b_{12}$, and so on. Matrices that are not equal are called **different**. Thus, matrices of different sizes are always different.

EXAMPLE 3 **Equality of Matrices**

Let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} 4 & 0 \\ 3 & -1 \end{bmatrix}.$$

Then

$$\mathbf{A} = \mathbf{B} \qquad \text{if and only if} \qquad \begin{matrix} a_{11} = 4, & a_{12} = 0, \\ a_{21} = 3, & a_{22} = -1. \end{matrix}$$

The following matrices are all different. Explain!

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

DEFINITION

Addition of Matrices

The **sum** of two matrices $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ ***of the same size*** is written $\mathbf{A} + \mathbf{B}$ and has the entries $a_{jk} + b_{jk}$ obtained by adding the corresponding entries of $\mathbf{A}$ and $\mathbf{B}$. Matrices of different sizes cannot be added.

As a special case, the **sum** $\mathbf{a} + \mathbf{b}$ of two row vectors or two column vectors, which must have the same number of components, is obtained by adding the corresponding components.

EXAMPLE 4 **Addition of Matrices and Vectors**

If $\quad \mathbf{A} = \begin{bmatrix} -4 & 6 & 3 \\ 0 & 1 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 5 & -1 & 0 \\ 3 & 1 & 0 \end{bmatrix}$, then $\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 5 & 3 \\ 3 & 2 & 2 \end{bmatrix}$.

$\mathbf{A}$ in Example 3 and our present $\mathbf{A}$ cannot be added. If $\mathbf{a} = [5 \quad 7 \quad 2]$ and $\mathbf{b} = [-6 \quad 2 \quad 0]$, then $\mathbf{a} + \mathbf{b} = [-1 \quad 9 \quad 2]$.

An application of matrix addition was suggested in Example 2. Many others will follow. ■

DEFINITION

Scalar Multiplication (Multiplication by a Number)

The **product** of any $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ and any **scalar** c (number c) is written $c\mathbf{A}$ and is the $m \times n$ matrix $c\mathbf{A} = [ca_{jk}]$ obtained by multiplying each entry of $\mathbf{A}$ by c.

Here $(-1)\mathbf{A}$ is simply written $-\mathbf{A}$ and is called the **negative** of $\mathbf{A}$. Similarly, $(-k)\mathbf{A}$ is written $-k\mathbf{A}$. Also, $\mathbf{A} + (-\mathbf{B})$ is written $\mathbf{A} - \mathbf{B}$ and is called the **difference** of $\mathbf{A}$ and $\mathbf{B}$ (which must have the same size!).

EXAMPLE 5 **Scalar Multiplication**

$$\text{If} \quad \mathbf{A} = \begin{bmatrix} 2.7 & -1.8 \\ 0 & 0.9 \\ 9.0 & -4.5 \end{bmatrix}, \quad \text{then} \quad -\mathbf{A} = \begin{bmatrix} -2.7 & 1.8 \\ 0 & -0.9 \\ -9.0 & 4.5 \end{bmatrix}, \quad \frac{10}{9}\mathbf{A} = \begin{bmatrix} 3 & -2 \\ 0 & 1 \\ 10 & -5 \end{bmatrix}, \quad 0\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

If a matrix $\mathbf{B}$ shows the distances between some cities in miles, $1.609\mathbf{B}$ gives these distances in kilometers. ■

Rules for Matrix Addition and Scalar Multiplication. From the familiar laws for the addition of numbers we obtain similar laws for the addition of matrices of the same size $m \times n$, namely,

$$\textbf{(3)} \quad \begin{aligned} &\text{(a)} && \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \\ &\text{(b)} && (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \qquad (\text{written } \mathbf{A} + \mathbf{B} + \mathbf{C}) \\ &\text{(c)} && \mathbf{A} + \mathbf{0} = \mathbf{A} \\ &\text{(d)} && \mathbf{A} + (-\mathbf{A}) = \mathbf{0}. \end{aligned}$$

Here $\mathbf{0}$ denotes the **zero matrix** (of size $m \times n$), that is, the $m \times n$ matrix with all entries zero. If $m = 1$ or $n = 1$, this is a vector, called a **zero vector**.

Hence matrix addition is *commutative* and *associative* [by (3a) and (3b)].
Similarly, for scalar multiplication we obtain the rules

$$\textbf{(4)}\qquad \begin{aligned} &\text{(a)} & c(\mathbf{A}+\mathbf{B}) &= c\mathbf{A}+c\mathbf{B} \\ &\text{(b)} & (c+k)\mathbf{A} &= c\mathbf{A}+k\mathbf{A} \\ &\text{(c)} & c(k\mathbf{A}) &= (ck)\mathbf{A} \qquad \text{(written } ck\mathbf{A}\text{)} \\ &\text{(d)} & 1\mathbf{A} &= \mathbf{A}. \end{aligned}$$

PROBLEM SET 7.1

1–7 GENERAL QUESTIONS

1. **Equality.** Give reasons why the five matrices in Example 3 are all different.
2. **Double subscript notation.** If you write the matrix in Example 2 in the form $\mathbf{A} = [a_{jk}]$, what is a_{31}? a_{13}? a_{26}? a_{33}?
3. **Sizes.** What sizes do the matrices in Examples 1, 2, 3, and 5 have?
4. **Main diagonal.** What is the main diagonal of $\mathbf{A}$ in Example 1? Of $\mathbf{A}$ and $\mathbf{B}$ in Example 3?
5. **Scalar multiplication.** If $\mathbf{A}$ in Example 2 shows the number of items sold, what is the matrix $\mathbf{B}$ of units sold if a unit consists of **(a)** 5 items and **(b)** 10 items?
6. If a 12×12 matrix $\mathbf{A}$ shows the distances between 12 cities in kilometers, how can you obtain from $\mathbf{A}$ the matrix $\mathbf{B}$ showing these distances in miles?
7. **Addition of vectors.** Can you add: A row and a column vector with different numbers of components? With the same number of components? Two row vectors with the same number of components but different numbers of zeros? A vector and a scalar? A vector with four components and a 2×2 matrix?

8–16 ADDITION AND SCALAR MULTIPLICATION OF MATRICES AND VECTORS

Let

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

$$\mathbf{C} = \begin{bmatrix} 5 & 2 \\ -2 & 4 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} -4 & 1 \\ 5 & 0 \\ 2 & -1 \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} 0 & 2 \\ 3 & 4 \\ 3 & -1 \end{bmatrix}$$

$$\mathbf{u} = \begin{bmatrix} 1.5 \\ 0 \\ -3.0 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} -5 \\ -30 \\ 10 \end{bmatrix}.$$

Find the following expressions, indicating which of the rules in (3) or (4) they illustrate, or give reasons why they are not defined.

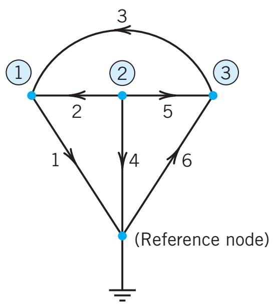
8. $2\mathbf{A} + 4\mathbf{B}$, $4\mathbf{B} + 2\mathbf{A}$, $0\mathbf{A} + \mathbf{B}$, $0.4\mathbf{B} - 4.2\mathbf{A}$
9. $3\mathbf{A}$, $0.5\mathbf{B}$, $3\mathbf{A} + 0.5\mathbf{B}$, $3\mathbf{A} + 0.5\mathbf{B} + \mathbf{C}$
10. $(4 \cdot 3)\mathbf{A}$, $4(3\mathbf{A})$, $14\mathbf{B} - 3\mathbf{B}$, $11\mathbf{B}$
11. $8\mathbf{C} + 10\mathbf{D}$, $2(5\mathbf{D} + 4\mathbf{C})$, $0.6\mathbf{C} - 0.6\mathbf{D}$, $0.6(\mathbf{C} - \mathbf{D})$
12. $(\mathbf{C} + \mathbf{D}) + \mathbf{E}$, $(\mathbf{D} + \mathbf{E}) + \mathbf{C}$, $0(\mathbf{C} - \mathbf{E}) + 4\mathbf{D}$, $\mathbf{A} - 0\mathbf{C}$
13. $(2 \cdot 7)\mathbf{C}$, $2(7\mathbf{C})$, $-\mathbf{D} + 0\mathbf{E}$, $\mathbf{E} - \mathbf{D} + \mathbf{C} + \mathbf{u}$
14. $(5\mathbf{u} + 5\mathbf{v}) - \frac{1}{2}\mathbf{w}$, $-20(\mathbf{u} + \mathbf{v}) + 2\mathbf{w}$, $\mathbf{E} - (\mathbf{u} + \mathbf{v})$, $10(\mathbf{u} + \mathbf{v}) + \mathbf{w}$
15. $(\mathbf{u} + \mathbf{v}) - \mathbf{w}$, $\mathbf{u} + (\mathbf{v} - \mathbf{w})$, $\mathbf{C} + 0\mathbf{w}$, $0\mathbf{E} + \mathbf{u} - \mathbf{v}$
16. $15\mathbf{v} - 3\mathbf{w} - 0\mathbf{u}$, $-3\mathbf{w} + 15\mathbf{v}$, $\mathbf{D} - \mathbf{u} + 3\mathbf{C}$, $8.5\mathbf{w} - 11.1\mathbf{u} + 0.4\mathbf{v}$
17. **Resultant of forces.** If the above vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ represent forces in space, their sum is called their *resultant*. Calculate it.
18. **Equilibrium.** By definition, forces are *in equilibrium* if their resultant is the zero vector. Find a force $\mathbf{p}$ such that the above $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{p}$ are in equilibrium.
19. **General rules.** Prove (3) and (4) for general 2×3 matrices and scalars c and k.

20. TEAM PROJECT. Matrices for Networks. Matrices have various engineering applications, as we shall see. For instance, they can be used to characterize connections in electrical networks, in nets of roads, in production processes, etc., as follows.

(a) Nodal Incidence Matrix. The network in Fig. 155 consists of six *branches* (connections) and four *nodes* (points where two or more branches come together). One node is the *reference node* (grounded node, whose voltage is zero). We number the other nodes and number and direct the branches. This we do arbitrarily. The network can now be described by a matrix $\mathbf{A} = [a_{jk}]$, where

$$a_{jk} = \begin{cases} +1 \text{ if branch } k \text{ leaves node } (j) \\ -1 \text{ if branch } k \text{ enters node } (j) \\ \ \ 0 \text{ if branch } k \text{ does not touch node } (j). \end{cases}$$

A is called the *nodal incidence matrix* of the network. Show that for the network in Fig. 155 the matrix **A** has the given form.



$$\begin{array}{l} \text{Branch} \\ \\ \text{Node } (1) \\ \text{Node } (2) \\ \text{Node } (3) \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 \end{bmatrix} \end{array}$$

Fig. 155. Network and nodal incidence matrix in Team Project 20(a)

(b) Find the nodal incidence matrices of the networks in Fig. 156.

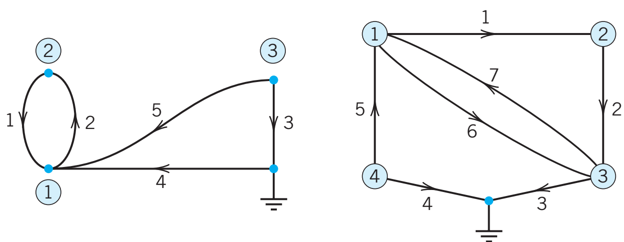


Fig. 156. Electrical networks in Team Project 20(b)

(c) Sketch the three networks corresponding to the nodal incidence matrices

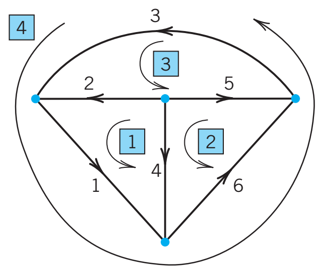
$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{bmatrix},$$

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

(d) Mesh Incidence Matrix. A network can also be characterized by the *mesh incidence matrix* $\mathbf{M} = [m_{jk}]$, where

$$m_{jk} = \begin{cases} +1 \text{ if branch } k \text{ is in mesh } \boxed{j} \\ \quad \text{and has the same orientation} \\ -1 \text{ if branch } k \text{ is in mesh } \boxed{j} \\ \quad \text{and has the opposite orientation} \\ \ \ 0 \text{ if branch } k \text{ is not in mesh } \boxed{j} \end{cases}$$

and a mesh is a loop with no branch in its interior (or in its exterior). Here, the meshes are numbered and directed (oriented) in an arbitrary fashion. Show that for the network in Fig. 157, the matrix **M** has the given form, where Row 1 corresponds to mesh 1, etc.



$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & -1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Fig. 157. Network and matrix **M** in Team Project 20(d)

7.2 Matrix Multiplication

Matrix multiplication means that one multiplies matrices by matrices. Its definition is standard but it looks artificial. *Thus you have to study matrix multiplication carefully*, multiply a few matrices together for practice until you can understand how to do it. Here then is the definition. (Motivation follows later.)

DEFINITION

Multiplication of a Matrix by a Matrix

The **product** $\mathbf{C} = \mathbf{AB}$ (in this order) of an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ times an $r \times p$ matrix $\mathbf{B} = [b_{jk}]$ is defined if and only if $r = n$ and is then the $m \times p$ matrix $\mathbf{C} = [c_{jk}]$ with entries

$$c_{jk} = \sum_{l=1}^{n} a_{jl}b_{lk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk} \qquad \begin{aligned} j &= 1, \cdots, m \\ k &= 1, \cdots, p. \end{aligned} \tag{1}$$

The condition $r = n$ means that the second factor, $\mathbf{B}$, must have as many rows as the first factor has columns, namely n. A diagram of sizes that shows when matrix multiplication is possible is as follows:

$$\begin{array}{ccccc} \mathbf{A} & \mathbf{B} & = & \mathbf{C} \\ [m \times n] & [n \times p] & = & [m \times p]. \end{array}$$

The entry c_{jk} in (1) is obtained by multiplying each entry in the jth row of $\mathbf{A}$ by the corresponding entry in the kth column of $\mathbf{B}$ and then adding these n products. For instance, $c_{21} = a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{n1}$, and so on. One calls this briefly a ***multiplication of rows into columns***. For $n = 3$, this is illustrated by

$$m = 4 \left\{ \overbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}}^{n=3} \right. \overbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}}^{p=2} = \left. \overbrace{\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{bmatrix}}^{p=2} \right\} m = 4$$

Notations in a product $\mathbf{AB} = \mathbf{C}$

where we shaded the entries that contribute to the calculation of entry c_{21} just discussed.

Matrix multiplication will be motivated by its use in *linear transformations* in this section and more fully in Sec. 7.9.

Let us illustrate the main points of matrix multiplication by some examples. Note that matrix multiplication also includes multiplying a matrix by a vector, since, after all, a vector is a special matrix.

EXAMPLE 1 **Matrix Multiplication**

$$\mathbf{AB} = \begin{bmatrix} 3 & 5 & -1 \\ 4 & 0 & 2 \\ -6 & -3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -2 & 3 & 1 \\ 5 & 0 & 7 & 8 \\ 9 & -4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 22 & -2 & 43 & 42 \\ 26 & -16 & \boxed{14} & 6 \\ -9 & 4 & -37 & -28 \end{bmatrix}$$

Here $c_{11} = 3 \cdot 2 + 5 \cdot 5 + (-1) \cdot 9 = 22$, and so on. The entry in the box is $c_{23} = 4 \cdot 3 + 0 \cdot 7 + 2 \cdot 1 = 14$. The product $\mathbf{BA}$ is not defined. ■

EXAMPLE 2 **Multiplication of a Matrix and a Vector**

$$\begin{bmatrix} 4 & 2 \\ 1 & 8 \end{bmatrix}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \cdot 3 + 2 \cdot 5 \\ 1 \cdot 3 + 8 \cdot 5 \end{bmatrix} = \begin{bmatrix} 22 \\ 43 \end{bmatrix} \quad \text{whereas} \quad \begin{bmatrix} 3 \\ 5 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 1 & 8 \end{bmatrix} \quad \text{is undefined.}$$

EXAMPLE 3 **Products of Row and Column Vectors**

$$\begin{bmatrix} 3 & 6 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = [19], \qquad \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}\begin{bmatrix} 3 & 6 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 1 \\ 6 & 12 & 2 \\ 12 & 24 & 4 \end{bmatrix}.$$

EXAMPLE 4 **CAUTION! Matrix Multiplication Is Not Commutative, AB ≠ BA in General**

This is illustrated by Examples 1 and 2, where one of the two products is not even defined, and by Example 3, where the two products have different sizes. But it also holds for square matrices. For instance,

$$\begin{bmatrix} 1 & 1 \\ 100 & 100 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 100 & 100 \end{bmatrix} = \begin{bmatrix} 99 & 99 \\ -99 & -99 \end{bmatrix}.$$

It is interesting that this also shows that $\mathbf{AB} = \mathbf{0}$ does ***not*** necessarily imply $\mathbf{BA} = \mathbf{0}$ or $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$. We shall discuss this further in Sec. 7.8, along with reasons when this happens.

Our examples show that in matrix products ***the order of factors must always be observed very carefully***. Otherwise matrix multiplication satisfies rules similar to those for numbers, namely.

$$\text{(2)} \quad \begin{aligned} &\text{(a)} \quad (k\mathbf{A})\mathbf{B} = k(\mathbf{AB}) = \mathbf{A}(k\mathbf{B}) \quad \textit{written } k\mathbf{AB} \textit{ or } \mathbf{A}k\mathbf{B} \\ &\text{(b)} \quad \mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \quad \textit{written } \mathbf{ABC} \\ &\text{(c)} \quad (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \\ &\text{(d)} \quad \mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \end{aligned}$$

provided **A**, **B**, and **C** are such that the expressions on the left are defined; here, k is any scalar. (2b) is called the **associative law**. (2c) and (2d) are called the **distributive laws**.

Since matrix multiplication is a multiplication of rows into columns, we can write the defining formula (1) more compactly as

$$\textbf{(3)} \qquad \boxed{c_{jk} = \mathbf{a}_j\mathbf{b}_k,} \qquad j = 1, \cdots, m; \quad k = 1, \cdots, p,$$

where $\mathbf{a}_j$ is the jth row vector of **A** and $\mathbf{b}_k$ is the kth column vector of **B**, so that in agreement with (1),

$$\mathbf{a}_j\mathbf{b}_k = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{jn} \end{bmatrix}\begin{bmatrix} b_{1k} \\ \cdot \\ \cdot \\ \cdot \\ b_{nk} \end{bmatrix} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk}.$$

EXAMPLE 5 Product in Terms of Row and Column Vectors

If $\mathbf{A} = [a_{jk}]$ is of size 3×3 and $\mathbf{B} = [b_{jk}]$ is of size 3×4, then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1\mathbf{b}_1 & \mathbf{a}_1\mathbf{b}_2 & \mathbf{a}_1\mathbf{b}_3 & \mathbf{a}_1\mathbf{b}_4 \\ \mathbf{a}_2\mathbf{b}_1 & \mathbf{a}_2\mathbf{b}_2 & \mathbf{a}_2\mathbf{b}_3 & \mathbf{a}_2\mathbf{b}_4 \\ \mathbf{a}_3\mathbf{b}_1 & \mathbf{a}_3\mathbf{b}_2 & \mathbf{a}_3\mathbf{b}_3 & \mathbf{a}_3\mathbf{b}_4 \end{bmatrix}. \tag{4}$$

Taking $\mathbf{a}_1 = [3 \quad 5 \quad -1]$, $\mathbf{a}_2 = [4 \quad 0 \quad 2]$, etc., verify (4) for the product in Example 1. ∎

Parallel processing of products on the computer is facilitated by a variant of (3) for computing $\mathbf{C} = \mathbf{AB}$, which is used by standard algorithms (such as in Lapack). In this method, $\mathbf{A}$ is used as given, $\mathbf{B}$ is taken in terms of its column vectors, and the product is computed columnwise; thus,

$$\mathbf{AB} = \mathbf{A}[\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_p] = [\mathbf{Ab}_1 \quad \mathbf{Ab}_2 \quad \cdots \quad \mathbf{Ab}_p]. \tag{5}$$

Columns of $\mathbf{B}$ are then assigned to different processors (individually or several to each processor), which simultaneously compute the columns of the product matrix $\mathbf{Ab}_1$, $\mathbf{Ab}_2$, etc.

EXAMPLE 6 Computing Products Columnwise by (5)

To obtain

$$\mathbf{AB} = \begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 & 7 \\ -1 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 11 & 4 & 34 \\ -17 & 8 & -23 \end{bmatrix}$$

from (5), calculate the columns

$$\begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 11 \\ -17 \end{bmatrix}, \quad \begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}, \quad \begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 34 \\ -23 \end{bmatrix}$$

of $\mathbf{AB}$ and then write them as a single matrix, as shown in the first formula on the right. ∎

Motivation of Multiplication by Linear Transformations

Let us now motivate the "unnatural" matrix multiplication by its use in **linear transformations**. For $n = 2$ variables these transformations are of the form

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \end{aligned} \tag{6*}$$

and suffice to explain the idea. (For general n they will be discussed in Sec. 7.9.) For instance, (6*) may relate an x_1x_2-coordinate system to a y_1y_2-coordinate system in the plane. In vectorial form we can write (6*) as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}. \tag{6}$$

Now suppose further that the x_1x_2-system is related to a w_1w_2-system by another linear transformation, say,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{Bw} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} b_{11}w_1 + b_{12}w_2 \\ b_{21}w_1 + b_{22}w_2 \end{bmatrix}. \tag{7}$$

Then the y_1y_2-system is related to the w_1w_2-system indirectly via the x_1x_2-system, and we wish to express this relation directly. Substitution will show that this direct relation is a linear transformation, too, say,

$$\mathbf{y} = \mathbf{Cw} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} c_{11}w_1 + c_{12}w_2 \\ c_{21}w_1 + c_{22}w_2 \end{bmatrix}. \tag{8}$$

Indeed, substituting (7) into (6), we obtain

$$\begin{aligned} y_1 &= a_{11}(b_{11}w_1 + b_{12}w_2) + a_{12}(b_{21}w_1 + b_{22}w_2) \\ &= (a_{11}b_{11} + a_{12}b_{21})w_1 + (a_{11}b_{12} + a_{12}b_{22})w_2 \\ y_2 &= a_{21}(b_{11}w_1 + b_{12}w_2) + a_{22}(b_{21}w_1 + b_{22}w_2) \\ &= (a_{21}b_{11} + a_{22}b_{21})w_1 + (a_{21}b_{12} + a_{22}b_{22})w_2. \end{aligned}$$

Comparing this with (8), we see that

$$\begin{aligned} c_{11} &= a_{11}b_{11} + a_{12}b_{21} \qquad & c_{12} &= a_{11}b_{12} + a_{12}b_{22} \\ c_{21} &= a_{21}b_{11} + a_{22}b_{21} \qquad & c_{22} &= a_{21}b_{12} + a_{22}b_{22}. \end{aligned}$$

This proves that $\mathbf{C} = \mathbf{AB}$ with the product defined as in (1). For larger matrix sizes the idea and result are exactly the same. Only the number of variables changes. We then have m variables y and n variables x and p variables w. The matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C} = \mathbf{AB}$ then have sizes $m \times n$, $n \times p$, and $m \times p$, respectively. And the requirement that $\mathbf{C}$ be the product $\mathbf{AB}$ leads to formula (1) in its general form. *This motivates matrix multiplication.*

Transposition

We obtain the transpose of a matrix by writing its rows as columns (or equivalently its columns as rows). This also applies to the transpose of vectors. Thus, a row vector becomes a column vector and vice versa. In addition, for square matrices, we can also "reflect" the elements along the main diagonal, that is, interchange entries that are symmetrically positioned with respect to the main diagonal to obtain the transpose. Hence a_{12} becomes a_{21}, a_{31} becomes a_{13}, and so forth. Example 7 illustrates these ideas. Also note that, if $\mathbf{A}$ is the given matrix, then we denote its transpose by $\mathbf{A}^\mathsf{T}$.

EXAMPLE 7 Transposition of Matrices and Vectors

If $$\mathbf{A} = \begin{bmatrix} 5 & -8 & 1 \\ 4 & 0 & 0 \end{bmatrix}, \quad \text{then} \quad \mathbf{A}^\mathsf{T} = \begin{bmatrix} 5 & 4 \\ -8 & 0 \\ 1 & 0 \end{bmatrix}.$$

A little more compactly, we can write

$$\begin{bmatrix} 5 & -8 & 1 \\ 4 & 0 & 0 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} 5 & 4 \\ -8 & 0 \\ 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 3 & 0 & 7 \\ 8 & -1 & 5 \\ 1 & -9 & 4 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} 3 & 8 & 1 \\ 0 & -1 & -9 \\ 7 & 5 & 4 \end{bmatrix},$$

Furthermore, the transpose $[6 \quad 2 \quad 3]^{\mathsf{T}}$ of the row vector $[6 \quad 2 \quad 3]$ is the column vector

$$[6 \quad 2 \quad 3]^{\mathsf{T}} = \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix}. \qquad \text{Conversely,} \qquad \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix}^{\mathsf{T}} = [6 \quad 2 \quad 3].$$

■

DEFINITION

Transposition of Matrices and Vectors

The transpose of an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ is the $n \times m$ matrix $\mathbf{A}^{\mathsf{T}}$ (read *A transpose*) that has the first *row* of $\mathbf{A}$ as its first *column*, the second *row* of $\mathbf{A}$ as its second *column*, and so on. Thus the transpose of $\mathbf{A}$ in (2) is $\mathbf{A}^{\mathsf{T}} = [a_{kj}]$, written out

$$\mathbf{A}^{\mathsf{T}} = [a_{kj}] = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdot & \cdot & \cdots & \cdot \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}. \tag{9}$$

As a special case, transposition converts row vectors to column vectors and conversely.

Transposition gives us a choice in that we can work either with the matrix or its transpose, whichever is more convenient.

Rules for transposition are

$$\textbf{(10)} \qquad \begin{aligned} &\text{(a)} & (\mathbf{A}^{\mathsf{T}})^{\mathsf{T}} &= \mathbf{A} \\ &\text{(b)} & (\mathbf{A} + \mathbf{B})^{\mathsf{T}} &= \mathbf{A}^{\mathsf{T}} + \mathbf{B}^{\mathsf{T}} \\ &\text{(c)} & (c\mathbf{A})^{\mathsf{T}} &= c\mathbf{A}^{\mathsf{T}} \\ &\text{(d)} & (\mathbf{AB})^{\mathsf{T}} &= \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}. \end{aligned}$$

CAUTION! Note that in (10d) the transposed matrices are ***in reversed order.*** We leave the proofs as an exercise in Probs. 9 and 10.

Special Matrices

Certain kinds of matrices will occur quite frequently in our work, and we now list the most important ones of them.

Symmetric and Skew-Symmetric Matrices. Transposition gives rise to two useful classes of matrices. **Symmetric** matrices are square matrices whose transpose equals the matrix itself. **Skew-symmetric** matrices are square matrices whose transpose equals ***minus*** the matrix. Both cases are defined in (11) and illustrated by Example 8.

$$\textbf{(11)} \qquad \underbrace{\mathbf{A}^{\mathsf{T}} = \mathbf{A} \quad (\text{thus } a_{kj} = a_{jk}),}_{\text{Symmetric Matrix}} \qquad \underbrace{\mathbf{A}^{\mathsf{T}} = -\mathbf{A} \quad (\text{thus } a_{kj} = -a_{jk}, \text{ hence } a_{jj} = 0).}_{\text{Skew-Symmetric Matrix}}$$

EXAMPLE 8 **Symmetric and Skew-Symmetric Matrices**

$$\mathbf{A} = \begin{bmatrix} 20 & 120 & 200 \\ 120 & 10 & 150 \\ 200 & 150 & 30 \end{bmatrix} \quad \text{is symmetric, and} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & -3 \\ -1 & 0 & -2 \\ 3 & 2 & 0 \end{bmatrix} \quad \text{is skew-symmetric.}$$

For instance, if a company has three building supply centers C_1, C_2, C_3, then $\mathbf{A}$ could show costs, say, a_{jj} for handling 1000 bags of cement at center C_j, and a_{jk} $(j \neq k)$ the cost of shipping 1000 bags from C_j to C_k. Clearly, $a_{jk} = a_{kj}$ if we assume shipping in the opposite direction will cost the same.

Symmetric matrices have several general properties which make them important. This will be seen as we proceed. ■

Triangular Matrices. **Upper triangular matrices** are square matrices that can have nonzero entries only on and ***above*** the main diagonal, whereas any entry below the diagonal must be zero. Similarly, **lower triangular matrices** can have nonzero entries only on and ***below*** the main diagonal. Any entry on the main diagonal of a triangular matrix may be zero or not.

EXAMPLE 9 **Upper and Lower Triangular Matrices**

$$\underbrace{\begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & 6 \end{bmatrix},}_{\text{Upper triangular}} \quad \underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 8 & -1 & 0 \\ 7 & 6 & 8 \end{bmatrix}, \quad \begin{bmatrix} 3 & 0 & 0 & 0 \\ 9 & -3 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 9 & 3 & 6 \end{bmatrix}.}_{\text{Lower triangular}} \qquad ■$$

Diagonal Matrices. These are square matrices that can have nonzero entries only on the main diagonal. Any entry above or below the main diagonal must be zero.

If all the diagonal entries of a diagonal matrix $\mathbf{S}$ are equal, say, c, we call $\mathbf{S}$ a **scalar matrix** because multiplication of any square matrix $\mathbf{A}$ of the same size by $\mathbf{S}$ has the same effect as the multiplication by a scalar, that is,

$$(12) \qquad \mathbf{AS} = \mathbf{SA} = c\mathbf{A}.$$

In particular, a scalar matrix, whose entries on the main diagonal are all 1, is called a **unit matrix** (or **identity matrix**) and is denoted by $\mathbf{I}_n$ or simply by $\mathbf{I}$. For $\mathbf{I}$, formula (12) becomes

$$(13) \qquad \mathbf{AI} = \mathbf{IA} = \mathbf{A}.$$

EXAMPLE 10 **Diagonal Matrix D. Scalar Matrix S. Unit Matrix I**

$$\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad ■$$

Some Applications of Matrix Multiplication

EXAMPLE 11 **Computer Production. Matrix Times Matrix**

Supercomp Ltd produces two computer models PC1086 and PC1186. The matrix **A** shows the cost per computer (in thousands of dollars) and **B** the production figures for the year 2010 (in multiples of 10,000 units.) Find a matrix **C** that shows the shareholders the cost per quarter (in millions of dollars) for raw material, labor, and miscellaneous.

$$\begin{array}{c} \\ \mathbf{A} = \end{array} \begin{array}{c} \begin{array}{cc} \text{PC1086} & \text{PC1186} \end{array} \\ \begin{bmatrix} 1.2 & 1.6 \\ 0.3 & 0.4 \\ 0.5 & 0.6 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{Raw Components} \\ \text{Labor} \\ \text{Miscellaneous} \end{array} \qquad \begin{array}{c} \text{Quarter} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \mathbf{B} = \begin{bmatrix} 3 & 8 & 6 & 9 \\ 6 & 2 & 4 & 3 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{PC1086} \\ \text{PC1186} \end{array}$$

Solution.

$$\mathbf{C} = \mathbf{AB} = \begin{array}{c} \text{Quarter} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 13.2 & 12.8 & 13.6 & 15.6 \\ 3.3 & 3.2 & 3.4 & 3.9 \\ 5.1 & 5.2 & 5.4 & 6.3 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{Raw Components} \\ \text{Labor} \\ \text{Miscellaneous} \end{array}$$

Since cost is given in multiples of \$1000 and production in multiples of 10,000 units, the entries of **C** are multiples of \$10 millions; thus $c_{11} = 13.2$ means \$132 million, etc. ■

EXAMPLE 12 **Weight Watching. Matrix Times Vector**

Suppose that in a weight-watching program, a person of 185 lb burns 350 cal/hr in walking (3 mph), 500 in bicycling (13 mph), and 950 in jogging (5.5 mph). Bill, weighing 185 lb, plans to exercise according to the matrix shown. Verify the calculations (W = Walking, B = Bicycling, J = Jogging).

$$\begin{array}{l} \\ \text{MON} \\ \text{WED} \\ \text{FRI} \\ \text{SAT} \end{array} \begin{array}{c} \begin{array}{ccc} \text{W} & \text{B} & \text{J} \end{array} \\ \begin{bmatrix} 1.0 & 0 & 0.5 \\ 1.0 & 1.0 & 0.5 \\ 1.5 & 0 & 0.5 \\ 2.0 & 1.5 & 1.0 \end{bmatrix} \end{array} \begin{bmatrix} 350 \\ 500 \\ 950 \end{bmatrix} = \begin{bmatrix} 825 \\ 1325 \\ 1000 \\ 2400 \end{bmatrix} \begin{array}{l} \text{MON} \\ \text{WED} \\ \text{FRI} \\ \text{SAT} \end{array}$$

■

EXAMPLE 13 **Markov Process. Powers of a Matrix. Stochastic Matrix**

Suppose that the 2004 state of land use in a city of 60 mi^2 of built-up area is

C: Commercially Used 25% I: Industrially Used 20% R: Residentially Used 55%.

Find the states in 2009, 2014, and 2019, assuming that the transition probabilities for 5-year intervals are given by the matrix **A** and remain practically the same over the time considered.

$$\mathbf{A} = \begin{array}{c} \begin{array}{ccc} \text{From C} & \text{From I} & \text{From R} \end{array} \\ \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0.2 & 0.9 & 0.2 \\ 0.1 & 0 & 0.8 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{To C} \\ \text{To I} \\ \text{To R} \end{array}$$

$\mathbf{A}$ is a **stochastic matrix**, that is, a square matrix with all entries nonnegative and all column sums equal to 1. Our example concerns a **Markov process**,[1] that is, a process for which the probability of entering a certain state depends only on the last state occupied (and the matrix $\mathbf{A}$), not on any earlier state.

Solution. From the matrix $\mathbf{A}$ and the 2004 state we can compute the 2009 state,

$$\begin{array}{c} \text{C} \\ \text{I} \\ \text{R} \end{array} \begin{bmatrix} 0.7 \cdot 25 + 0.1 \cdot 20 + 0 \quad \cdot 55 \\ 0.2 \cdot 25 + 0.9 \cdot 20 + 0.2 \cdot 55 \\ 0.1 \cdot 25 + \quad 0 \cdot 20 + 0.8 \cdot 55 \end{bmatrix} = \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0.2 & 0.9 & 0.2 \\ 0.1 & 0 & 0.8 \end{bmatrix} \begin{bmatrix} 25 \\ 20 \\ 55 \end{bmatrix} = \begin{bmatrix} 19.5 \\ 34.0 \\ 46.5 \end{bmatrix}.$$

To explain: The 2009 figure for C equals 25% times the probability 0.7 that C goes into C, plus 20% times the probability 0.1 that I goes into C, plus 55% times the probability 0 that R goes into C. Together,

$$25 \cdot 0.7 + 20 \cdot 0.1 + 55 \cdot 0 = 19.5\,[\%]. \qquad \text{Also} \qquad 25 \cdot 0.2 + 20 \cdot 0.9 + 55 \cdot 0.2 = 34\,[\%].$$

Similarly, the new R is 46.5%. We see that the 2009 state vector is the column vector

$$\mathbf{y} = [19.5 \quad 34.0 \quad 46.5]^{\mathsf{T}} = \mathbf{Ax} = \mathbf{A}\,[25 \quad 20 \quad 55]^{\mathsf{T}}$$

where the column vector $\mathbf{x} = [25 \quad 20 \quad 55]^{\mathsf{T}}$ is the given 2004 state vector. Note that the sum of the entries of $\mathbf{y}$ is 100 [%]. Similarly, you may verify that for 2014 and 2019 we get the state vectors

$$\mathbf{z} = \mathbf{Ay} = \mathbf{A}(\mathbf{Ax}) = \mathbf{A}^2\mathbf{x} = [17.05 \quad 43.80 \quad 39.15]^{\mathsf{T}}$$

$$\mathbf{u} = \mathbf{Az} = \mathbf{A}^2\mathbf{y} = \mathbf{A}^3\mathbf{x} = [16.315 \quad 50.660 \quad 33.025]^{\mathsf{T}}.$$

Answer. In 2009 the commercial area will be 19.5% (11.7 mi^2), the industrial 34% (20.4 mi^2), and the residential 46.5% (27.9 mi^2). For 2014 the corresponding figures are 17.05%, 43.80%, and 39.15%. For 2019 they are 16.315%, 50.660%, and 33.025%. (In Sec. 8.2 we shall see what happens in the limit, assuming that those probabilities remain the same. In the meantime, can you experiment or guess?) ■

PROBLEM SET 7.2

1–10 GENERAL QUESTIONS

1. **Multiplication.** Why is multiplication of matrices restricted by conditions on the factors?
2. **Square matrix.** What form does a 3×3 matrix have if it is symmetric as well as skew-symmetric?
3. **Product of vectors.** Can every 3×3 matrix be represented by two vectors as in Example 3?
4. **Skew-symmetric matrix.** How many different entries can a 4×4 skew-symmetric matrix have? An $n \times n$ skew-symmetric matrix?
5. Same questions as in Prob. 4 for symmetric matrices.
6. **Triangular matrix.** If $\mathbf{U}_1, \mathbf{U}_2$ are upper triangular and $\mathbf{L}_1, \mathbf{L}_2$ are lower triangular, which of the following are triangular?

   $$\mathbf{U}_1 + \mathbf{U}_2, \quad \mathbf{U}_1\mathbf{U}_2, \quad \mathbf{U}_1^2, \quad \mathbf{U}_1 + \mathbf{L}_1, \quad \mathbf{U}_1\mathbf{L}_1,$$
   $$\mathbf{L}_1 + \mathbf{L}_2$$

7. **Idempotent matrix**, defined by $\mathbf{A}^2 = \mathbf{A}$. Can you find four 2×2 idempotent matrices?
8. **Nilpotent matrix**, defined by $\mathbf{B}^m = \mathbf{0}$ for some m. Can you find three 2×2 nilpotent matrices?
9. **Transposition.** Can you prove (10a)–(10c) for 3×3 matrices? For $m \times n$ matrices?
10. **Transposition. (a)** Illustrate (10d) by simple examples. **(b)** Prove (10d).

11–20 MULTIPLICATION, ADDITION, AND TRANSPOSITION OF MATRICES AND VECTORS

Let

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

$$\mathbf{C} = \begin{bmatrix} 0 & 1 \\ 3 & 2 \\ -2 & 0 \end{bmatrix}, \qquad \mathbf{a} = [1 \quad -2 \quad 0], \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}.$$

[1] ANDREI ANDREJEVITCH MARKOV (1856–1922), Russian mathematician, known for his work in probability theory.

Showing all intermediate results, calculate the following expressions or give reasons why they are undefined:

11. $\mathbf{AB}$, $\mathbf{AB}^\mathsf{T}$, $\mathbf{BA}$, $\mathbf{B}^\mathsf{T}\mathbf{A}$

12. $\mathbf{AA}^\mathsf{T}$, $\mathbf{A}^2$, $\mathbf{BB}^\mathsf{T}$, $\mathbf{B}^2$

13. $\mathbf{CC}^\mathsf{T}$, $\mathbf{BC}$, $\mathbf{CB}$, $\mathbf{C}^\mathsf{T}\mathbf{B}$

14. $3\mathbf{A} - 2\mathbf{B}$, $(3\mathbf{A} - 2\mathbf{B})^\mathsf{T}$, $3\mathbf{A}^\mathsf{T} - 2\mathbf{B}^\mathsf{T}$, $(3\mathbf{A} - 2\mathbf{B})^\mathsf{T}\mathbf{a}^\mathsf{T}$

15. $\mathbf{Aa}$, $\mathbf{Aa}^\mathsf{T}$, $(\mathbf{Ab})^\mathsf{T}$, $\mathbf{b}^\mathsf{T}\mathbf{A}^\mathsf{T}$

16. $\mathbf{BC}$, $\mathbf{BC}^\mathsf{T}$, $\mathbf{Bb}$, $\mathbf{b}^\mathsf{T}\mathbf{B}$

17. $\mathbf{ABC}$, $\mathbf{ABa}$, $\mathbf{ABb}$, $\mathbf{Ca}^\mathsf{T}$

18. $\mathbf{ab}$, $\mathbf{ba}$, $\mathbf{aA}$, $\mathbf{Bb}$

19. $1.5\mathbf{a} + 3.0\mathbf{b}$, $1.5\mathbf{a}^\mathsf{T} + 3.0\mathbf{b}$, $(\mathbf{A} - \mathbf{B})\mathbf{b}$, $\mathbf{Ab} - \mathbf{Bb}$

20. $\mathbf{b}^\mathsf{T}\mathbf{Ab}$, $\mathbf{aBa}^\mathsf{T}$, $\mathbf{aCC}^\mathsf{T}$, $\mathbf{C}^\mathsf{T}\mathbf{ba}$

21. General rules. Prove (2) for 2×2 matrices $\mathbf{A} = [a_{jk}]$, $\mathbf{B} = [b_{jk}]$, $\mathbf{C} = [c_{jk}]$, and a general scalar.

22. Product. Write $\mathbf{AB}$ in Prob. 11 in terms of row and column vectors.

23. Product. Calculate $\mathbf{AB}$ in Prob. 11 columnwise. See Example 1.

24. Commutativity. Find all 2×2 matrices $\mathbf{A} = [a_{jk}]$ that commute with $\mathbf{B} = [b_{jk}]$, where $b_{jk} = j + k$.

25. TEAM PROJECT. Symmetric and Skew-Symmetric Matrices. These matrices occur quite frequently in applications, so it is worthwhile to study some of their most important properties.

(a) Verify the claims in (11) that $a_{kj} = a_{jk}$ for a symmetric matrix, and $a_{kj} = -a_{jk}$ for a skew-symmetric matrix. Give examples.

(b) Show that for every square matrix $\mathbf{C}$ the matrix $\mathbf{C} + \mathbf{C}^\mathsf{T}$ is symmetric and $\mathbf{C} - \mathbf{C}^\mathsf{T}$ is skew-symmetric. Write $\mathbf{C}$ in the form $\mathbf{C} = \mathbf{S} + \mathbf{T}$, where $\mathbf{S}$ is symmetric and $\mathbf{T}$ is skew-symmetric and find $\mathbf{S}$ and $\mathbf{T}$ in terms of $\mathbf{C}$. Represent $\mathbf{A}$ and $\mathbf{B}$ in Probs. 11–20 in this form.

(c) A **linear combination** of matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \cdots, \mathbf{M}$ of the same size is an expression of the form

$$a\mathbf{A} + b\mathbf{B} + c\mathbf{C} + \cdots + m\mathbf{M}, \tag{14}$$

where $a, \cdots, m$ are any scalars. Show that if these matrices are square and symmetric, so is (14); similarly, if they are skew-symmetric, so is (14).

(d) Show that $\mathbf{AB}$ with symmetric $\mathbf{A}$ and $\mathbf{B}$ is symmetric if and only if $\mathbf{A}$ and $\mathbf{B}$ **commute**, that is, $\mathbf{AB} = \mathbf{BA}$.

(e) Under what condition is the product of skew-symmetric matrices skew-symmetric?

26–30 FURTHER APPLICATIONS

26. Production. In a production process, let N mean "no trouble" and T "trouble." Let the transition probabilities from one day to the next be 0.8 for $N \to N$, hence 0.2 for $N \to T$, and 0.5 for $T \to N$, hence 0.5 for $T \to T$.

If today there is no trouble, what is the probability of N two days after today? Three days after today?

27. CAS Experiment. Markov Process. Write a program for a Markov process. Use it to calculate further steps in Example 13 of the text. Experiment with other stochastic 3×3 matrices, also using different starting values.

28. Concert subscription. In a community of 100,000 adults, subscribers to a concert series tend to renew their subscription with probability 90% and persons presently not subscribing will subscribe for the next season with probability 0.2%. If the present number of subscribers is 1200, can one predict an increase, decrease, or no change over each of the next three seasons?

29. Profit vector. Two factory outlets F_1 and F_2 in New York and Los Angeles sell sofas (S), chairs (C), and tables (T) with a profit of \$35, \$62, and \$30, respectively. Let the sales in a certain week be given by the matrix

$$\mathbf{A} = \begin{array}{c} \begin{array}{ccc} \mathrm{S} & \mathrm{C} & \mathrm{T} \end{array} \\ \begin{bmatrix} 400 & 60 & 240 \\ 100 & 120 & 500 \end{bmatrix} \end{array} \begin{array}{c} \\ F_1 \\ F_2 \end{array}$$

Introduce a "profit vector" $\mathbf{p}$ such that the components of $\mathbf{v} = \mathbf{Ap}$ give the total profits of F_1 and F_2.

30. TEAM PROJECT. Special Linear Transformations. Rotations have various applications. We show in this project how they can be handled by matrices.

(a) Rotation in the plane. Show that the linear transformation $\mathbf{y} = \mathbf{Ax}$ with

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

is a counterclockwise rotation of the Cartesian x_1x_2-coordinate system in the plane about the origin, where θ is the angle of rotation.

(b) Rotation through $n\theta$. Show that in (a)

$$\mathbf{A}^n = \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}.$$

Is this plausible? Explain this in words.

(c) Addition formulas for cosine and sine. By geometry we should have

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} = \begin{bmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{bmatrix}.$$

Derive from this the addition formulas (6) in App. A3.1.

(d) Computer graphics. To visualize a three-dimensional object with plane faces (e.g., a cube), we may store the position vectors of the vertices with respect to a suitable $x_1x_2x_3$-coordinate system (and a list of the connecting edges) and then obtain a two-dimensional image on a video screen by projecting the object onto a coordinate plane, for instance, onto the x_1x_2-plane by setting $x_3 = 0$. To change the appearance of the image, we can impose a linear transformation on the position vectors stored. Show that a diagonal matrix $\mathbf{D}$ with main diagonal entries 3, 1, $\frac{1}{2}$ gives from an $\mathbf{x} = [x_j]$ the new position vector $\mathbf{y} = \mathbf{Dx}$, where $y_1 = 3x_1$ (stretch in the x_1-direction by a factor 3), $y_2 = x_2$ (unchanged), $y_3 = \frac{1}{2}x_3$ (contraction in the x_3-direction). What effect would a scalar matrix have?

(e) Rotations in space. Explain $\mathbf{y} = \mathbf{Ax}$ geometrically when $\mathbf{A}$ is one of the three matrices

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix},$$

$$\begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix}, \quad \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

What effect would these transformations have in situations such as that described in (d)?

7.3 Linear Systems of Equations. Gauss Elimination

We now come to one of the most important use of matrices, that is, using matrices to solve systems of linear equations. We showed informally, in Example 1 of Sec. 7.1, how to represent the information contained in a system of linear equations by a matrix, called the augmented matrix. This matrix will then be used in solving the linear system of equations. Our approach to solving linear systems is called the Gauss elimination method. Since this method is so fundamental to linear algebra, the student should be alert.

A shorter term for systems of linear equations is just **linear systems**. Linear systems model many applications in engineering, economics, statistics, and many other areas. Electrical networks, traffic flow, and commodity markets may serve as specific examples of applications.

Linear System, Coefficient Matrix, Augmented Matrix

A **linear system of *m* equations in *n* unknowns** $x_1, \cdots, x_n$ is a set of equations of the form

$$\textbf{(1)} \qquad \begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots&\cdots\cdots\cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

The system is called *linear* because each variable x_j appears in the first power only, just as in the equation of a straight line. $a_{11}, \cdots, a_{mn}$ are given numbers, called the **coefficients** of the system. $b_1, \cdots, b_m$ on the right are also given numbers. If all the b_j are zero, then (1) is called a **homogeneous system**. If at least one b_j is not zero, then (1) is called a **nonhomogeneous system**.

A **solution** of (1) is a set of numbers $x_1, \cdots, x_n$ that satisfies all the m equations. A **solution vector** of (1) is a vector $\mathbf{x}$ whose components form a solution of (1). If the system (1) is homogeneous, it always has at least the **trivial solution** $x_1 = 0, \cdots, x_n = 0$.

Matrix Form of the Linear System (1). From the definition of matrix multiplication we see that the m equations of (1) may be written as a single vector equation

$$\mathbf{Ax} = \mathbf{b} \tag{2}$$

where the **coefficient matrix** $\mathbf{A} = [a_{jk}]$ is the $m \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

are column vectors. We assume that the coefficients a_{jk} are not all zero, so that $\mathbf{A}$ is not a zero matrix. Note that $\mathbf{x}$ has n components, whereas $\mathbf{b}$ has m components. The matrix

$$\tilde{\mathbf{A}} = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

is called the **augmented matrix** of the system (1). The dashed vertical line could be omitted, as we shall do later. It is merely a reminder that the last column of $\tilde{\mathbf{A}}$ did not come from matrix $\mathbf{A}$ but came from vector $\mathbf{b}$. Thus, we *augmented* the matrix $\mathbf{A}$.

Note that ***the augmented matrix*** $\tilde{\mathbf{A}}$ ***determines the system* (1) *completely*** because it contains all the given numbers appearing in (1).

EXAMPLE 1 Geometric Interpretation. Existence and Uniqueness of Solutions

If $m = n = 2$, we have two equations in two unknowns x_1, x_2

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2.$$

If we interpret x_1, x_2 as coordinates in the x_1x_2-plane, then each of the two equations represents a straight line, and (x_1, x_2) is a solution if and only if the point P with coordinates x_1, x_2 lies on both lines. Hence there are three possible cases (see Fig. 158 on next page):

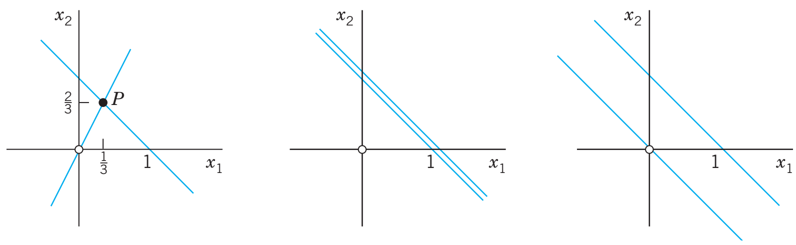
(*a*) Precisely one solution if the lines intersect

(*b*) Infinitely many solutions if the lines coincide

(*c*) No solution if the lines are parallel

For instance,

Case (a)	Case (b)	Case (c)
$x_1 + x_2 = 1$	$x_1 + x_2 = 1$	$x_1 + x_2 = 1$
$2x_1 - x_2 = 0$	$2x_1 + 2x_2 = 2$	$x_1 + x_2 = 0$



If the system is homogenous, Case (c) cannot happen, because then those two straight lines pass through the origin, whose coordinates (0, 0) constitute the trivial solution. Similarly, our present discussion can be extended from two equations in two unknowns to three equations in three unknowns. We give the geometric interpretation of three possible cases concerning solutions in Fig. 158. Instead of straight lines we have planes and the solution depends on the positioning of these planes in space relative to each other. The student may wish to come up with some specific examples. ■

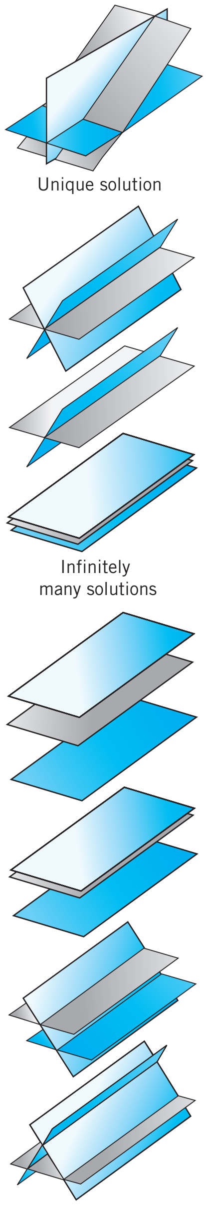


Fig. 158. Three equations in three unknowns interpreted as planes in space

Our simple example illustrated that a system (1) may have no solution. This leads to such questions as: Does a given system (1) have a solution? Under what conditions does it have precisely one solution? If it has more than one solution, how can we characterize the set of all solutions? We shall consider such questions in Sec. 7.5.

First, however, let us discuss an important systematic method for solving linear systems.

Gauss Elimination and Back Substitution

The Gauss elimination method can be motivated as follows. Consider a linear system that is in ***triangular form*** (in full, *upper* triangular form) such as

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ 13x_2 &= -26 \end{aligned}$$

(*Triangular* means that all the nonzero entries of the corresponding coefficient matrix lie above the diagonal and form an upside-down 90° triangle.) Then we can solve the system by **back substitution**, that is, we solve the last equation for the variable, $x_2 = -26/13 = -2$, and then work backward, substituting $x_2 = -2$ into the first equation and solving it for x_1, obtaining $x_1 = \frac{1}{2}(2 - 5x_2) = \frac{1}{2}(2 - 5 \cdot (-2)) = 6$. This gives us the idea of first reducing a general system to triangular form. For instance, let the given system be

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ -4x_1 + 3x_2 &= -30. \end{aligned} \qquad \text{Its augmented matrix is} \qquad \begin{bmatrix} 2 & 5 & 2 \\ -4 & 3 & -30 \end{bmatrix}.$$

We leave the first equation as it is. We eliminate x_1 from the second equation, to get a triangular system. For this we add twice the first equation to the second, and we do the same operation on the ***rows*** of the augmented matrix. This gives $-4x_1 + 4x_1 + 3x_2 + 10x_2 = -30 + 2 \cdot 2$, that is,

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ 13x_2 &= -26 \end{aligned} \qquad \text{Row 2} + 2\text{ Row 1} \qquad \begin{bmatrix} 2 & 5 & 2 \\ 0 & 13 & -26 \end{bmatrix}$$

where Row 2 + 2 Row 1 means "Add twice Row 1 to Row 2" in the original matrix. This is the **Gauss elimination** (for 2 equations in 2 unknowns) giving the triangular form, from which back substitution now yields $x_2 = -2$ and $x_1 = 6$, as before.

Since a linear system is completely determined by its augmented matrix, ***Gauss elimination can be done by merely considering the matrices***, as we have just indicated. We do this again in the next example, emphasizing the matrices by writing them first and the equations behind them, just as a help in order not to lose track.

EXAMPLE 2 Gauss Elimination. Electrical Network

Solve the linear system

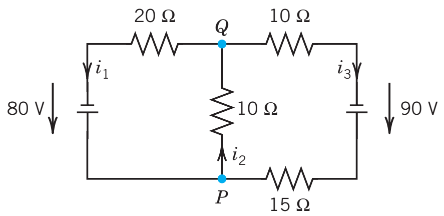
$$\begin{aligned} x_1 - x_2 + x_3 &= 0 \\ -x_1 + x_2 - x_3 &= 0 \\ 10x_2 + 25x_3 &= 90 \\ 20x_1 + 10x_2 &= 80. \end{aligned}$$

Derivation from the circuit in Fig. 159 (Optional). This is the system for the unknown currents $x_1 = i_1$, $x_2 = i_2$, $x_3 = i_3$ in the electrical network in Fig. 159. To obtain it, we label the currents as shown, choosing directions arbitrarily; if a current will come out negative, this will simply mean that the current flows against the direction of our arrow. The current entering each battery will be the same as the current leaving it. The equations for the currents result from Kirchhoff's laws:

Kirchhoff's Current Law (KCL). *At any point of a circuit, the sum of the inflowing currents equals the sum of the outflowing currents.*

Kirchhoff's Voltage Law (KVL). *In any closed loop, the sum of all voltage drops equals the impressed electromotive force.*

Node P gives the first equation, node Q the second, the right loop the third, and the left loop the fourth, as indicated in the figure.



Node P:	$i_1 - i_2 + i_3 = 0$
Node Q:	$-i_1 + i_2 - i_3 = 0$
Right loop:	$10i_2 + 25i_3 = 90$
Left loop:	$20i_1 + 10i_2 = 80$

Fig. 159. Network in Example 2 and equations relating the currents

Solution by Gauss Elimination. This system could be solved rather quickly by noticing its particular form. But this is not the point. The point is that the Gauss elimination is systematic and will work in general, also for large systems. We apply it to our system and then do back substitution. As indicated, let us write the augmented matrix of the system first and then the system itself:

Augmented Matrix $\tilde{\mathbf{A}}$

$$\begin{array}{l} \text{Pivot 1} \longrightarrow \\ \\ \text{Eliminate} \longrightarrow \\ \\ \end{array}\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0\\ \boxed{-1} & 1 & -1 & 0\\ \boxed{0} & 10 & 25 & 90\\ \boxed{20} & 10 & 0 & 80\end{array}\right]$$

Equations

$$\begin{array}{l} \text{Pivot 1} \longrightarrow \\ \\ \text{Eliminate} \longrightarrow \\ \\ \end{array}\begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0\\ -x_1 & + & x_2 & - & x_3 & = & 0\\ & & 10x_2 & + & 25x_3 & = & 90\\ 20x_1 & + & 10x_2 & & & = & 80.\end{array}$$

Step 1. Elimination of x_1

Call the first row of $\mathbf{A}$ the **pivot row** and the first equation the **pivot equation**. Call the coefficient 1 of its x_1-term the **pivot** in this step. Use this equation to eliminate x_1 (get rid of x_1) in the other equations. For this, do:

Add 1 times the pivot equation to the second equation.

Add -20 times the pivot equation to the fourth equation.

This corresponds to **row operations** on the augmented matrix as indicated in BLUE behind the ***new matrix*** in (3). So the operations are performed on the ***preceding matrix***. The result is

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0\\ 0 & 0 & 0 & 0\\ 0 & 10 & 25 & 90\\ 0 & 30 & -20 & 80\end{array}\right]\begin{array}{l} \\ \text{Row 2 + Row 1}\\ \\ \text{Row 4} - \text{20 Row 1}\end{array}\qquad \begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0\\ & & & & 0 & = & 0\\ & & 10x_2 & + & 25x_3 & = & 90\\ & & 30x_2 & - & 20x_3 & = & 80.\end{array} \tag{3}$$

Step 2. Elimination of x_2

The first equation remains as it is. We want the new second equation to serve as the next pivot equation. But since it has no x_2-term (in fact, it is $0 = 0$), we must first change the order of the equations and the corresponding rows of the new matrix. We put $0 = 0$ at the end and move the third equation and the fourth equation one place up. This is called **partial pivoting** (as opposed to the rarely used *total pivoting*, in which the order of the unknowns is also changed). It gives

$$\begin{array}{l} \\ \text{Pivot 10} \longrightarrow \\ \text{Eliminate 30} \longrightarrow \\ \\ \end{array}\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0\\ 0 & 10 & 25 & 90\\ 0 & \boxed{30} & -20 & 80\\ 0 & 0 & 0 & 0\end{array}\right]\qquad \begin{array}{l} \\ \text{Pivot 10} \longrightarrow \\ \text{Eliminate } 30x_2 \longrightarrow \\ \\ \end{array}\begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0\\ & & 10x_2 & + & 25x_3 & = & 90\\ & & \boxed{30x_2} & - & 20x_3 & = & 80\\ & & & & 0 & = & 0.\end{array}$$

To eliminate x_2, do:

Add -3 times the pivot equation to the third equation.

The result is

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0\\ 0 & 10 & 25 & 90\\ 0 & 0 & -95 & -190\\ 0 & 0 & 0 & 0\end{array}\right]\begin{array}{l} \\ \\ \text{Row 3} - \text{3 Row 2}\\ \\ \end{array}\qquad \begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0\\ & & 10x_2 & + & 25x_3 & = & 90\\ & & & - & 95x_3 & = & -190\\ & & & & 0 & = & 0.\end{array} \tag{4}$$

Back Substitution. ***Determination of x_3, x_2, x_1 (in this order)***

Working backward from the last to the first equation of this "triangular" system (4), we can now readily find x_3, then x_2, and then x_1:

$$\begin{array}{rcrcrcr} & & & - & 95x_3 & = & -190\\ & & 10x_2 & + & 25x_3 & = & 90\\ x_1 & - & x_2 & + & x_3 & = & 0\end{array}\qquad \begin{array}{r} x_3 = i_3 = 2\,[\mathrm{A}]\\ x_2 = \frac{1}{10}(90 - 25x_3) = i_2 = 4\,[\mathrm{A}]\\ x_1 = x_2 - x_3 = i_1 = 2\,[\mathrm{A}]\end{array}$$

where A stands for "amperes." This is the answer to our problem. The solution is unique. ■

Elementary Row Operations. Row-Equivalent Systems

Example 2 illustrates the operations of the Gauss elimination. These are the first two of three operations, which are called

Elementary Row Operations for Matrices:

Interchange of two rows

Addition of a constant multiple of one row to another row

*Multiplication of a row by a **nonzero** constant c*

CAUTION! These operations are for rows, ***not for columns***! They correspond to the following

Elementary Operations for Equations:

Interchange of two equations

Addition of a constant multiple of one equation to another equation

*Multiplication of an equation by a **nonzero** constant c*

Clearly, the interchange of two equations does not alter the solution set. Neither does their addition because we can undo it by a corresponding subtraction. Similarly for their multiplication, which we can undo by multiplying the new equation by $1/c$ (since $c \neq 0$), producing the original equation.

We now call a linear system S_1 **row-equivalent** to a linear system S_2 if S_1 can be obtained from S_2 by (finitely many!) row operations. This justifies Gauss elimination and establishes the following result.

THEOREM 1

Row-Equivalent Systems

Row-equivalent linear systems have the same set of solutions.

Because of this theorem, systems having the same solution sets are often called *equivalent systems*. But note well that we are dealing with ***row operations***. No column operations on the augmented matrix are permitted in this context because they would generally alter the solution set.

A linear system (1) is called **overdetermined** if it has more equations than unknowns, as in Example 2, **determined** if $m = n$, as in Example 1, and **underdetermined** if it has fewer equations than unknowns.

Furthermore, a system (1) is called **consistent** if it has at least one solution (thus, one solution or infinitely many solutions), but **inconsistent** if it has no solutions at all, as $x_1 + x_2 = 1, x_1 + x_2 = 0$ in Example 1, Case (c).

Gauss Elimination: The Three Possible Cases of Systems

We have seen, in Example 2, that Gauss elimination can solve linear systems that have a unique solution. This leaves us to apply Gauss elimination to a system with infinitely many solutions (in Example 3) and one with no solution (in Example 4).

EXAMPLE 3 Gauss Elimination if Infinitely Many Solutions Exist

Solve the following linear system of three equations in four unknowns whose augmented matrix is

$$(5)\qquad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0.6 & 1.5 & 1.5 & -5.4 & 2.7 \\ 1.2 & -0.3 & -0.3 & 2.4 & 2.1 \end{array}\right]. \qquad \text{Thus,} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ \boxed{0.6x_1} + 1.5x_2 + 1.5x_3 - 5.4x_4 &= 2.7 \\ \boxed{1.2x_1} - 0.3x_2 - 0.3x_3 + 2.4x_4 &= 2.1. \end{aligned}$$

Solution. As in the previous example, we circle pivots and box terms of equations and corresponding entries to be eliminated. We indicate the operations in terms of equations and operate on both equations and matrices.

Step 1. Elimination of x_1 from the second and third equations by adding

$$-0.6/3.0 = -0.2 \text{ times the first equation to the second equation,}$$

$$-1.2/3.0 = -0.4 \text{ times the first equation to the third equation.}$$

This gives the following, in which the pivot of the next step is circled.

$$(6)\qquad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0 & 1.1 & 1.1 & -4.4 & 1.1 \\ 0 & -1.1 & -1.1 & 4.4 & -1.1 \end{array}\right] \begin{array}{l} \\ \text{Row 2} - 0.2 \text{ Row 1} \\ \text{Row 3} - 0.4 \text{ Row 1} \end{array} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ 1.1x_2 + 1.1x_3 - 4.4x_4 &= 1.1 \\ \boxed{-1.1x_2} - 1.1x_3 + 4.4x_4 &= -1.1. \end{aligned}$$

Step 2. Elimination of x_2 from the third equation of (6) by adding

$$1.1/1.1 = 1 \text{ times the second equation to the third equation.}$$

This gives

$$(7)\qquad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0 & 1.1 & 1.1 & -4.4 & 1.1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{array}{l} \\ \\ \text{Row 3} + \text{Row 2} \end{array} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ 1.1x_2 + 1.1x_3 - 4.4x_4 &= 1.1 \\ 0 &= 0. \end{aligned}$$

Back Substitution. From the second equation, $x_2 = 1 - x_3 + 4x_4$. From this and the first equation, $x_1 = 2 - x_4$. Since x_3 and x_4 remain arbitrary, we have infinitely many solutions. If we choose a value of x_3 and a value of x_4, then the corresponding values of x_1 and x_2 are uniquely determined.

On Notation. If unknowns remain arbitrary, it is also customary to denote them by other letters $t_1, t_2, \cdots$. In this example we may thus write $x_1 = 2 - x_4 = 2 - t_2$, $x_2 = 1 - x_3 + 4x_4 = 1 - t_1 + 4t_2$, $x_3 = t_1$ (first arbitrary unknown), $x_4 = t_2$ (second arbitrary unknown). ■

EXAMPLE 4 Gauss Elimination if no Solution Exists

What will happen if we apply the Gauss elimination to a linear system that has no solution? The answer is that in this case the method will show this fact by producing a contradiction. For instance, consider

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 2 & 1 & 1 & 0 \\ 6 & 2 & 4 & 6 \end{array}\right] \qquad \begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ \boxed{2x_1} + x_2 + x_3 &= 0 \\ \boxed{6x_1} + 2x_2 + 4x_3 &= 6. \end{aligned}$$

Step 1. Elimination of x_1 from the second and third equations by adding

$$-\tfrac{2}{3} \text{ times the first equation to the second equation,}$$

$$-\tfrac{6}{3} = -2 \text{ times the first equation to the third equation.}$$

This gives

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & -2 & 2 & 0 \end{array}\right] \begin{array}{l} \\ \text{Row 2} - \frac{2}{3}\text{ Row 1} \\ \text{Row 3} - 2\text{ Row 1} \end{array} \qquad \begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ -\tfrac{1}{3}x_2 + \tfrac{1}{3}x_3 &= -2 \\ \boxed{-2x_2} + 2x_3 &= 0. \end{aligned}$$

Step 2. Elimination of x_2 from the third equation gives

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{array}\right] \begin{array}{l} \\ \\ \text{Row 3} - 6\text{ Row 2} \end{array} \qquad \begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ -\tfrac{1}{3}x_2 + \tfrac{1}{3}x_3 &= -2 \\ 0 &= 12. \end{aligned}$$

The false statement $0 = 12$ shows that the system has no solution. ■

Row Echelon Form and Information From It

At the end of the Gauss elimination the form of the coefficient matrix, the augmented matrix, and the system itself are called the **row echelon form**. In it, rows of zeros, if present, are the last rows, and, in each nonzero row, the leftmost nonzero entry is farther to the right than in the previous row. For instance, in Example 4 the coefficient matrix and its augmented in row echelon form are

$$\text{(8)} \qquad \begin{bmatrix} 3 & 2 & 1 \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{array}\right].$$

Note that we do not require that the leftmost nonzero entries be 1 since this would have no theoretic or numeric advantage. (The so-called *reduced echelon form*, in which those entries *are* 1, will be discussed in Sec. 7.8.)

The original system of m equations in n unknowns has augmented matrix $[\mathbf{A} \mid \mathbf{b}]$. This is to be row reduced to matrix $[\mathbf{R} \mid \mathbf{f}]$. The two systems $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{Rx} = \mathbf{f}$ are equivalent: if either one has a solution, so does the other, and the solutions are identical.

At the end of the Gauss elimination (before the back substitution), the row echelon form of the augmented matrix will be

$$\text{(9)} \qquad \left[\begin{array}{ccccc|c} r_{11} & r_{12} & \cdots\cdots & & r_{1n} & f_1 \\ & r_{22} & \cdots\cdots & & r_{2n} & f_2 \\ & & \ddots & & \vdots & \vdots \\ & & & r_{rr} \;\cdots & r_{rn} & f_r \\ & & & & & f_{r+1} \\ & & & & & \vdots \\ & & & & & f_m \end{array}\right].$$

Here, $r \leqq m$, $r_{11} \neq 0$, and all entries in the blue triangle and blue rectangle are zero.

The number of nonzero rows, r, in the row-reduced coefficient matrix $\mathbf{R}$ is called the **rank of R** and also the **rank of A**. Here is the method for determining whether $\mathbf{Ax} = \mathbf{b}$ has solutions and what they are:

(a) No solution. If r is less than m (meaning that $\mathbf{R}$ actually has at least one row of all 0s) *and* at least one of the numbers $f_{r+1}, f_{r+2}, \cdots, f_m$ is not zero, then the system

$\mathbf{Rx} = \mathbf{f}$ is inconsistent: No solution is possible. Therefore the system $\mathbf{Ax} = \mathbf{b}$ is inconsistent as well. See Example 4, where $r = 2 < m = 3$ and $f_{r+1} = f_3 = 12$.

If the system is consistent (either $r = m$, or $r < m$ *and* all the numbers $f_{r+1}, f_{r+2}, \cdots, f_m$ are zero), then there are solutions.

(b) Unique solution. If the system is consistent and $r = n$, there is exactly one solution, which can be found by back substitution. See Example 2, where $r = n = 3$ and $m = 4$.

(c) Infinitely many solutions. To obtain any of these solutions, choose values of $x_{r+1}, \cdots, x_n$ arbitrarily. Then solve the rth equation for x_r (in terms of those arbitrary values), then the $(r - 1)$st equation for x_{r-1}, and so on up the line. See Example 3.

Orientation. Gauss elimination is reasonable in computing time and storage demand. We shall consider those aspects in Sec. 20.1 in the chapter on numeric linear algebra. Section 7.4 develops fundamental concepts of linear algebra such as linear independence and rank of a matrix. These in turn will be used in Sec. 7.5 to fully characterize the behavior of linear systems in terms of existence and uniqueness of solutions.

PROBLEM SET 7.3

1–14 GAUSS ELIMINATION

Solve the linear system given explicitly or by its augmented matrix. Show details.

1. $$\begin{aligned} 4x - 6y &= -11 \\ -3x + 8y &= 10 \end{aligned}$$

2. $$\begin{bmatrix} 3.0 & -0.5 & 0.6 \\ 1.5 & 4.5 & 6.0 \end{bmatrix}$$

3. $$\begin{aligned} x + y - z &= 9 \\ 8y + 6z &= -6 \\ -2x + 4y - 6z &= 40 \end{aligned}$$

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

5. $$\begin{bmatrix} 13 & 12 & -6 \\ -4 & 7 & -73 \\ 11 & -13 & 157 \end{bmatrix}$$

6. $$\begin{bmatrix} 4 & -8 & 3 & 16 \\ -1 & 2 & -5 & -21 \\ 3 & -6 & 1 & 7 \end{bmatrix}$$

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

8. $$\begin{aligned} 4y + 3z &= 8 \\ 2x - z &= 2 \\ 3x + 2y &= 5 \end{aligned}$$

9. $$\begin{aligned} -2y - 2z &= -8 \\ 3x + 4y - 5z &= 13 \end{aligned}$$

10. $$\begin{bmatrix} 5 & -7 & 3 & 17 \\ -15 & 21 & -9 & 50 \end{bmatrix}$$

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

12. $$\begin{bmatrix} 2 & -2 & 4 & 0 & 0 \\ -3 & 3 & -6 & 5 & 15 \\ 1 & -1 & 2 & 0 & 0 \end{bmatrix}$$

13. $$\begin{aligned} 10x + 4y - 2z &= -4 \\ -3w - 17x + y + 2z &= 2 \\ w + x + y &= 6 \\ 8w - 34x + 16y - 10z &= 4 \end{aligned}$$

14. $$\begin{bmatrix} 2 & 3 & 1 & -11 & 1 \\ 5 & -2 & 5 & -4 & 5 \\ 1 & -1 & 3 & -3 & 3 \\ 3 & 4 & -7 & 2 & -7 \end{bmatrix}$$

15. Equivalence relation. By definition, an *equivalence relation* on a set is a relation satisfying three conditions: (named as indicated)

(i) Each element A of the set is equivalent to itself (*Reflexivity*).

(ii) If A is equivalent to B, then B is equivalent to A (*Symmetry*).

(iii) If A is equivalent to B and B is equivalent to C, then A is equivalent to C (*Transitivity*).

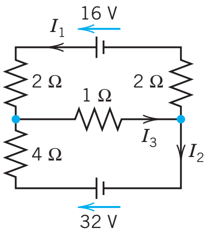
Show that row equivalence of matrices satisfies these three conditions. *Hint.* Show that for each of the three elementary row operations these conditions hold.

16. CAS PROJECT. Gauss Elimination and Back Substitution. Write a program for Gauss elimination and back substitution **(a)** that does not include pivoting and **(b)** that does include pivoting. Apply the programs to Probs. 11–14 and to some larger systems of your choice.

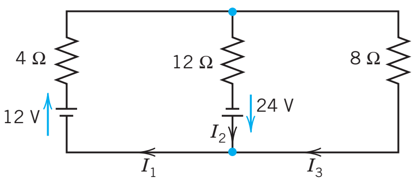
17–21 MODELS OF NETWORKS

In Probs. 17–19, using Kirchhoff's laws (see Example 2) and showing the details, find the currents:

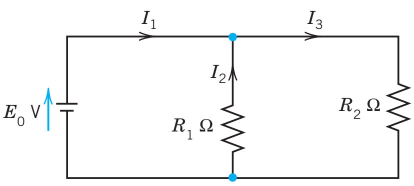
17.



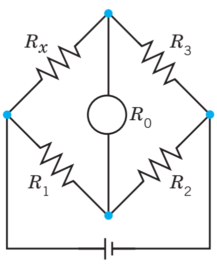
18.



19.

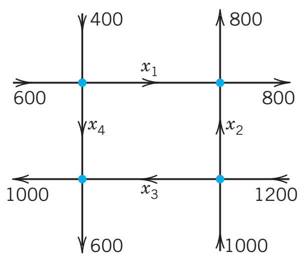


20. Wheatstone bridge. Show that if $R_x/R_3 = R_1/R_2$ in the figure, then $I = 0$. (R_0 is the resistance of the instrument by which I is measured.) This bridge is a method for determining R_x. R_1, R_2, R_3 are known. R_3 is variable. To get R_x, make $I = 0$ by varying R_3. Then calculate $R_x = R_3R_1/R_2$.



Wheatstone bridge

Problem 20



Net of one-way streets

Problem 21

21. Traffic flow. Methods of electrical circuit analysis have applications to other fields. For instance, applying the analog of Kirchhoff's Current Law, find the traffic flow (cars per hour) in the net of one-way streets (in the directions indicated by the arrows) shown in the figure. Is the solution unique?

22. Models of markets. Determine the equilibrium solution ($D_1 = S_1, D_2 = S_2$) of the two-commodity market with linear model (D, S, P = demand, supply, price; index 1 = first commodity, index 2 = second commodity)

$$D_1 = 40 - 2P_1 - P_2, \qquad S_1 = 4P_1 - P_2 + 4,$$

$$D_2 = 5P_1 - 2P_2 + 16, \qquad S_2 = 3P_2 - 4.$$

23. Balancing a chemical equation $x_1C_3H_8 + x_2O_2 \rightarrow x_3CO_2 + x_4H_2O$ means finding integer x_1, x_2, x_3, x_4 such that the numbers of atoms of carbon (C), hydrogen (H), and oxygen (O) are the same on both sides of this reaction, in which propane C_3H_8 and O_2 give carbon dioxide and water. Find the smallest positive integers $x_1, \cdots, x_4$.

24. PROJECT. Elementary Matrices. The idea is that elementary operations can be accomplished by matrix multiplication. If **A** is an $m \times n$ matrix on which we want to do an elementary operation, then there is a matrix **E** such that **EA** is the new matrix after the operation. Such an **E** is called an **elementary matrix**. This idea can be helpful, for instance, in the design of algorithms. (*Computationally*, it is generally preferable to do row operations *directly*, rather than by multiplication by **E**.)

(a) Show that the following are elementary matrices, for interchanging Rows 2 and 3, for adding -5 times the first row to the third, and for multiplying the fourth row by 8.

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{E}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -5 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{E}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}.$$

Apply $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$ to a vector and to a 4×3 matrix of your choice. Find $\mathbf{B} = \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A}$, where $\mathbf{A} = [a_{jk}]$ is the general 4×2 matrix. Is $\mathbf{B}$ equal to $\mathbf{C} = \mathbf{E}_1\mathbf{E}_2\mathbf{E}_3\mathbf{A}$?

(b) Conclude that $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$ are obtained by doing the corresponding elementary operations on the 4×4 unit matrix. Prove that *if* $\mathbf{M}$ *is obtained from* $\mathbf{A}$ *by an elementary row operation, then*

$$\mathbf{M} = \mathbf{EA},$$

where $\mathbf{E}$ *is obtained from the* $n \times n$ *unit matrix* $\mathbf{I}_n$ *by the same row operation.*

7.4 Linear Independence. Rank of a Matrix. Vector Space

Since our next goal is to fully characterize the behavior of linear systems in terms of existence and uniqueness of solutions (Sec. 7.5), we have to introduce new fundamental linear algebraic concepts that will aid us in doing so. Foremost among these are ***linear independence*** and the ***rank of a matrix***. Keep in mind that these concepts are intimately linked with the important Gauss elimination method and how it works.

Linear Independence and Dependence of Vectors

Given any set of m vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ (with the same number of components), a **linear combination** of these vectors is an expression of the form

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)}$$

where $c_1, c_2, \cdots, c_m$ are any scalars. Now consider the equation

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)} = \mathbf{0}. \tag{1}$$

Clearly, this vector equation (1) holds if we choose all c_j's zero, because then it becomes $\mathbf{0} = \mathbf{0}$. If this is the only m-tuple of scalars for which (1) holds, then our vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ are said to form a *linearly independent set* or, more briefly, we call them **linearly independent**. Otherwise, if (1) also holds with scalars not all zero, we call these vectors **linearly dependent**. This means that we can express at least one of the vectors as a linear combination of the other vectors. For instance, if (1) holds with, say, $c_1 \neq 0$, we can solve (1) for $\mathbf{a}_{(1)}$:

$$\mathbf{a}_{(1)} = k_2\mathbf{a}_{(2)} + \cdots + k_m\mathbf{a}_{(m)} \qquad \text{where } k_j = -c_j/c_1.$$

(Some k_j's may be zero. Or even all of them, namely, if $\mathbf{a}_{(1)} = \mathbf{0}$.)

Why is linear independence important? Well, if a set of vectors is linearly dependent, then we can get rid of at least one or perhaps more of the vectors until we get a linearly independent set. This set is then the smallest "truly essential" set with which we can work. Thus, we cannot express any of the vectors, of this set, linearly in terms of the others.

EXAMPLE 1 Linear Independence and Dependence

The three vectors

$$\mathbf{a}_{(1)} = [\ 3 \quad 0 \quad 2 \quad 2]$$
$$\mathbf{a}_{(2)} = [-6 \quad 42 \quad 24 \quad 54]$$
$$\mathbf{a}_{(3)} = [\ 21 \quad -21 \quad 0 \quad -15]$$

are linearly dependent because

$$6\mathbf{a}_{(1)} - \tfrac{1}{2}\mathbf{a}_{(2)} - \mathbf{a}_{(3)} = \mathbf{0}.$$

Although this is easily checked by vector arithmetic (do it!), it is not so easy to discover. However, a systematic method for finding out about linear independence and dependence follows below.

The first two of the three vectors are linearly independent because $c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} = \mathbf{0}$ implies $c_2 = 0$ (from the second components) and then $c_1 = 0$ (from any other component of $\mathbf{a}_{(1)}$. ■

Rank of a Matrix

DEFINITION

The **rank** of a matrix $\mathbf{A}$ is the maximum number of linearly independent row vectors of $\mathbf{A}$. It is denoted by rank $\mathbf{A}$.

Our further discussion will show that the rank of a matrix is an important key concept for understanding general properties of matrices and linear systems of equations.

EXAMPLE 2 Rank

The matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ -6 & 42 & 24 & 54 \\ 21 & -21 & 0 & -15 \end{bmatrix} \tag{2}$$

has rank 2, because Example 1 shows that the first two row vectors are linearly independent, whereas all three row vectors are linearly dependent.

Note further that rank $\mathbf{A} = 0$ if and only if $\mathbf{A} = \mathbf{0}$. This follows directly from the definition. ■

We call a matrix $\mathbf{A}_1$ **row-equivalent** to a matrix $\mathbf{A}_2$ if $\mathbf{A}_1$ can be obtained from $\mathbf{A}_2$ by (finitely many!) elementary row operations.

Now the maximum number of linearly independent row vectors of a matrix does not change if we change the order of rows or multiply a row by a nonzero c or take a linear combination by adding a multiple of a row to another row. This shows that rank is **invariant** under elementary row operations:

THEOREM 1

Row-Equivalent Matrices

Row-equivalent matrices have the same rank.

Hence we can determine the rank of a matrix by reducing the matrix to row-echelon form, as was done in Sec. 7.3. Once the matrix is in row-echelon form, we count the number of nonzero rows, which is precisely the rank of the matrix.

EXAMPLE 3 Determination of Rank

For the matrix in Example 2 we obtain successively

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ -6 & 42 & 24 & 54 \\ 21 & -21 & 0 & -15 \end{bmatrix} \quad \text{(given)}$$

$$\begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 42 & 28 & 58 \\ 0 & -21 & -14 & -29 \end{bmatrix} \quad \begin{matrix} \\ \text{Row 2} + 2 \text{ Row 1} \\ \text{Row 3} - 7 \text{ Row 1} \end{matrix}$$

$$\begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 42 & 28 & 58 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{matrix} \\ \\ \text{Row 3} + \tfrac{1}{2} \text{ Row 2.} \end{matrix}$$

The last matrix is in row-echelon form and has two nonzero rows. Hence rank $\mathbf{A} = 2$, as before. ■

Examples 1–3 illustrate the following useful theorem (with $p = 3$, $n = 3$, and the rank of the matrix $= 2$).

THEOREM 2 Linear Independence and Dependence of Vectors

Consider p vectors that each have n components. Then these vectors are linearly independent if the matrix formed, with these vectors as row vectors, has rank p. However, these vectors are linearly dependent if that matrix has rank less than p.

Further important properties will result from the basic

THEOREM 3 Rank in Terms of Column Vectors

The rank r of a matrix $\mathbf{A}$ *equals the maximum number of linearly independent* ***column*** *vectors of* $\mathbf{A}$.

Hence $\mathbf{A}$ *and its transpose* $\mathbf{A}^{\mathsf{T}}$ *have the same rank.*

PROOF In this proof we write simply "rows" and "columns" for row and column vectors. Let $\mathbf{A}$ be an $m \times n$ matrix of rank $\mathbf{A} = r$. Then by definition of rank, $\mathbf{A}$ has r linearly independent rows which we denote by $\mathbf{v}_{(1)}, \cdots, \mathbf{v}_{(r)}$ (regardless of their position in $\mathbf{A}$), and all the rows $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ of $\mathbf{A}$ are linear combinations of those, say,

$$\begin{aligned} \mathbf{a}_{(1)} &= c_{11}\mathbf{v}_{(1)} + c_{12}\mathbf{v}_{(2)} + \cdots + c_{1r}\mathbf{v}_{(r)} \\ \mathbf{a}_{(2)} &= c_{21}\mathbf{v}_{(1)} + c_{22}\mathbf{v}_{(2)} + \cdots + c_{2r}\mathbf{v}_{(r)} \\ &\vdots \\ \mathbf{a}_{(m)} &= c_{m1}\mathbf{v}_{(1)} + c_{m2}\mathbf{v}_{(2)} + \cdots + c_{mr}\mathbf{v}_{(r)}. \end{aligned} \tag{3}$$

These are vector equations for rows. To switch to columns, we write (3) in terms of components as n such systems, with $k = 1, \cdots, n$,

$$
\begin{aligned}
a_{1k} &= c_{11}v_{1k} + c_{12}v_{2k} + \cdots + c_{1r}v_{rk} \\
a_{2k} &= c_{21}v_{1k} + c_{22}v_{2k} + \cdots + c_{2r}v_{rk} \\
&\ \ \vdots \\
a_{mk} &= c_{m1}v_{1k} + c_{m2}v_{2k} + \cdots + c_{mr}v_{rk}
\end{aligned}
\tag{4}
$$

and collect components in columns. Indeed, we can write (4) as

$$
\begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} = v_{1k} \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix} + v_{2k} \begin{bmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{m2} \end{bmatrix} + \cdots + v_{rk} \begin{bmatrix} c_{1r} \\ c_{2r} \\ \vdots \\ c_{mr} \end{bmatrix}
\tag{5}
$$

where $k = 1, \cdots, n$. Now the vector on the left is the kth column vector of $\mathbf{A}$. We see that each of these n columns is a linear combination of the same r columns on the right. Hence $\mathbf{A}$ cannot have more linearly independent columns than rows, whose number is rank $\mathbf{A} = r$. Now rows of $\mathbf{A}$ are columns of the transpose $\mathbf{A}^\mathsf{T}$. For $\mathbf{A}^\mathsf{T}$ our conclusion is that $\mathbf{A}^\mathsf{T}$ cannot have more linearly independent columns than rows, so that $\mathbf{A}$ cannot have more linearly independent rows than columns. Together, the number of linearly independent columns of $\mathbf{A}$ must be r, the rank of $\mathbf{A}$. This completes the proof. ■

EXAMPLE 4 Illustration of Theorem 3

The matrix in (2) has rank 2. From Example 3 we see that the first two row vectors are linearly independent and by "working backward" we can verify that Row 3 $=$ 6 Row 1 $- \frac{1}{2}$ Row 2. Similarly, the first two columns are linearly independent, and by reducing the last matrix in Example 3 by columns we find that

$$\text{Column } 3 = \tfrac{2}{3}\,\text{Column } 1 + \tfrac{2}{3}\,\text{Column } 2 \qquad \text{and} \qquad \text{Column } 4 = \tfrac{2}{3}\,\text{Column } 1 + \tfrac{29}{21}\,\text{Column } 2.$$ ■

Combining Theorems 2 and 3 we obtain

THEOREM 4 Linear Dependence of Vectors

Consider p vectors each having n components. If $n < p$, then these vectors *are linearly dependent.*

PROOF The matrix $\mathbf{A}$ with those p vectors as row vectors has p rows and $n < p$ columns; hence by Theorem 3 it has rank $\mathbf{A} \leqq n < p$, which implies linear dependence by Theorem 2. ■

Vector Space

The following related concepts are of general interest in linear algebra. In the present context they provide a clarification of essential properties of matrices and their role in connection with linear systems.

Consider a nonempty set V of vectors where each vector has the same number of components. If, for any two vectors $\mathbf{a}$ and $\mathbf{b}$ in V, we have that all their linear combinations $\alpha\mathbf{a} + \beta\mathbf{b}$ (α, β any real numbers) are also elements of V, and if, furthermore, $\mathbf{a}$ and $\mathbf{b}$ satisfy the laws (3a), (3c), (3d), and (4) in Sec. 7.1, as well as any vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ in V satisfy (3b) then V is a vector space. Note that here we wrote laws (3) and (4) of Sec. 7.1 in lowercase letters $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, which is our notation for vectors. More on vector spaces in Sec. 7.9.

The maximum number of linearly independent vectors in V is called the **dimension** of V and is denoted by dim V. Here we assume the dimension to be finite; infinite dimension will be defined in Sec. 7.9.

A linearly independent set in V consisting of a maximum possible number of vectors in V is called a **basis** for V. In other words, any largest possible set of independent vectors in V forms basis for V. That means, if we add one or more vector to that set, the set will be linearly dependent. (See also the beginning of Sec. 7.4 on linear independence and dependence of vectors.) Thus, the number of vectors of a basis for V equals dim V.

The set of all linear combinations of given vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(p)}$ with the same number of components is called the **span** of these vectors. Obviously, a span is a vector space. If in addition, the given vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(p)}$ are linearly independent, then they form a basis for that vector space.

This then leads to another equivalent definition of basis. A set of vectors is a **basis** for a vector space V if (1) the vectors in the set are linearly independent, and if (2) any vector in V can be expressed as a linear combination of the vectors in the set. If (2) holds, we also say that the set of vectors **spans** the vector space V.

By a **subspace** of a vector space V we mean a nonempty subset of V (including V itself) that forms a vector space with respect to the two algebraic operations (addition and scalar multiplication) defined for the vectors of V.

EXAMPLE 5 Vector Space, Dimension, Basis

The span of the three vectors in Example 1 is a vector space of dimension 2. A basis of this vector space consists of any two of those three vectors, for instance, $\mathbf{a}_{(1)}$, $\mathbf{a}_{(2)}$, or $\mathbf{a}_{(1)}$, $\mathbf{a}_{(3)}$, etc. ■

We further note the simple

THEOREM 5 Vector Space R^n

The vector space R^n consisting of all vectors with n components (n real numbers) has dimension n.

PROOF A basis of n vectors is $\mathbf{a}_{(1)} = [1 \quad 0 \quad \cdots \quad 0]$, $\mathbf{a}_{(2)} = [0 \quad 1 \quad 0 \quad \cdots \quad 0]$, $\cdots$, $\mathbf{a}_{(n)} = [0 \quad \cdots \quad 0 \quad 1]$. ■

For a matrix $\mathbf{A}$, we call the span of the row vectors the **row space** of $\mathbf{A}$. Similarly, the span of the column vectors of $\mathbf{A}$ is called the **column space** of $\mathbf{A}$.

Now, Theorem 3 shows that a matrix $\mathbf{A}$ has as many linearly independent rows as columns. By the definition of dimension, their number is the dimension of the row space or the column space of $\mathbf{A}$. This proves

THEOREM 6 Row Space and Column Space

The row space and the column space of a matrix $\mathbf{A}$ have the same dimension, equal to rank $\mathbf{A}$.

Finally, for a given matrix **A** the solution set of the homogeneous system $\mathbf{Ax} = 0$ is a vector space, called the **null space** of **A**, and its dimension is called the **nullity** of **A**. In the next section we motivate and prove the basic relation

$$\text{rank } \mathbf{A} + \text{nullity } \mathbf{A} = \text{Number of columns of } \mathbf{A}. \tag{6}$$

PROBLEM SET 7.4

1–10 RANK, ROW SPACE, COLUMN SPACE

Find the rank. Find a basis for the row space. Find a basis for the column space. *Hint.* Row-reduce the matrix and its transpose. (You may omit obvious factors from the vectors of these bases.)

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

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

3. $\begin{bmatrix} 0 & 3 & 5 \\ 3 & 5 & 0 \\ 5 & 0 & 10 \end{bmatrix}$

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

5. $\begin{bmatrix} 0.2 & -0.1 & 0.4 \\ 0 & 1.1 & -0.3 \\ 0.1 & 0 & -2.1 \end{bmatrix}$

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

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

8. $\begin{bmatrix} 2 & 4 & 8 & 16 \\ 16 & 8 & 4 & 2 \\ 4 & 8 & 16 & 2 \\ 2 & 16 & 8 & 4 \end{bmatrix}$

9. $\begin{bmatrix} 9 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

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

11. CAS Experiment. Rank. (a) Show experimentally that the $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ with $a_{jk} = j + k - 1$ has rank 2 for any n. (Problem 20 shows $n = 4$.) Try to prove it.

(b) Do the same when $a_{jk} = j + k + c$, where c is any positive integer.

(c) What is rank **A** if $a_{jk} = 2^{j+k-2}$? Try to find other large matrices of low rank independent of n.

12–16 GENERAL PROPERTIES OF RANK

Show the following:

12. rank $\mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$ = rank **AB**. (Note the order!)

13. rank **A** = rank **B** does *not* imply rank $\mathbf{A}^2$ = rank $\mathbf{B}^2$. (Give a counterexample.)

14. If **A** is not square, either the row vectors or the column vectors of **A** are linearly dependent.

15. If the row vectors of a square matrix are linearly independent, so are the column vectors, and conversely.

16. Give examples showing that the rank of a product of matrices cannot exceed the rank of either factor.

17–25 LINEAR INDEPENDENCE

Are the following sets of vectors linearly independent? Show the details of your work.

17. $[3\ \ 4\ \ 0\ \ 2]$, $[2\ \ -1\ \ 3\ \ 7]$, $[1\ \ 16\ \ -12\ \ -22]$

18. $[1\ \ \tfrac{1}{2}\ \ \tfrac{1}{3}\ \ \tfrac{1}{4}]$, $[\tfrac{1}{2}\ \ \tfrac{1}{3}\ \ \tfrac{1}{4}\ \ \tfrac{1}{5}]$, $[\tfrac{1}{3}\ \ \tfrac{1}{4}\ \ \tfrac{1}{5}\ \ \tfrac{1}{6}]$, $[\tfrac{1}{4}\ \ \tfrac{1}{5}\ \ \tfrac{1}{6}\ \ \tfrac{1}{7}]$

19. $[0\ \ 1\ \ 1]$, $[1\ \ 1\ \ 1]$, $[0\ \ 0\ \ 1]$

20. $[1\ \ 2\ \ 3\ \ 4]$, $[2\ \ 3\ \ 4\ \ 5]$, $[3\ \ 4\ \ 5\ \ 6]$, $[4\ \ 5\ \ 6\ \ 7]$

21. $[2\ \ 0\ \ 0\ \ 7]$, $[2\ \ 0\ \ 0\ \ 8]$, $[2\ \ 0\ \ 0\ \ 9]$, $[2\ \ 0\ \ 1\ \ 0]$

22. $[0.4\ \ -0.2\ \ 0.2]$, $[0\ \ 0\ \ 0]$, $[3.0\ \ -0.6\ \ 1.5]$

23. $[9\ \ 8\ \ 7\ \ 6\ \ 5]$, $[9\ \ 7\ \ 5\ \ 3\ \ 1]$

24. $[4\ \ -1\ \ 3]$, $[0\ \ 8\ \ 1]$, $[1\ \ 3\ \ -5]$, $[2\ \ 6\ \ 1]$

25. $[6\ \ 0\ \ -1\ \ 3]$, $[2\ \ 2\ \ 5\ \ 0]$, $[-4\ \ -4\ \ -4\ \ -4]$

26. Linearly independent subset. Beginning with the last of the vectors $[3\ \ 0\ \ 1\ \ 2]$, $[6\ \ 1\ \ 0\ \ 0]$, $[12\ \ 1\ \ 2\ \ 4]$, $[6\ \ 0\ \ 2\ \ 4]$, and $[9\ \ 0\ \ 1\ \ 2]$, omit one after another until you get a linearly independent set.

27–35 VECTOR SPACE

Is the given set of vectors a vector space? Give reasons. If your answer is yes, determine the dimension and find a basis. ($v_1, v_2, \cdots$ denote components.)

27. All vectors in R^3 with $v_1 - v_2 + 2v_3 = 0$

28. All vectors in R^3 with $3v_2 + v_3 = k$

29. All vectors in R^2 with $v_1 \geqq v_2$

30. All vectors in R^n with the first $n - 2$ components zero

31. All vectors in R^5 with positive components

32. All vectors in R^3 with $3v_1 - 2v_2 + v_3 = 0$, $4v_1 + 5v_2 = 0$

33. All vectors in R^3 with $3v_1 - v_3 = 0$, $2v_1 + 3v_2 - 4v_3 = 0$

34. All vectors in R^n with $|v_j| = 1$ for $j = 1, \cdots, n$

35. All vectors in R^4 with $v_1 = 2v_2 = 3v_3 = 4v_4$

7.5 Solutions of Linear Systems: Existence, Uniqueness

Rank, as just defined, gives complete information about existence, uniqueness, and general structure of the solution set of linear systems as follows.

A linear system of equations in n unknowns has a unique solution if the coefficient matrix and the augmented matrix have the same rank n, and infinitely many solutions if that common rank is less than n. The system has no solution if those two matrices have different rank.

To state this precisely and prove it, we shall use the generally important concept of a **submatrix** of **A**. By this we mean any matrix obtained from **A** by omitting some rows or columns (or both). By definition this includes **A** itself (as the matrix obtained by omitting no rows or columns); this is practical.

THEOREM 1

Fundamental Theorem for Linear Systems

(a) Existence. *A linear system of m equations in n unknowns* $x_1, \cdots, x_n$

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

is **consistent**, *that is, has solutions, if and only if the coefficient matrix* **A** *and the augmented matrix* $\tilde{\mathbf{A}}$ *have the same rank. Here,*

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{A}} = \left[\begin{array}{ccc:c} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

(b) Uniqueness. *The system* (1) *has precisely one solution if and only if this common rank r of* **A** *and* $\tilde{\mathbf{A}}$ *equals n.*

(c) Infinitely many solutions. *If this common rank r is less than n, the system* (1) *has infinitely many solutions. All of these solutions are obtained by determining r suitable unknowns* (*whose submatrix of coefficients must have rank r*) *in terms of the remaining $n - r$ unknowns, to which arbitrary values can be assigned.* (See Example 3 in Sec. 7.3.)

(d) Gauss elimination (Sec. 7.3). *If solutions exist, they can all be obtained by the Gauss elimination.* (This method will automatically reveal whether or not solutions exist; see Sec. 7.3.)

PROOF **(a)** We can write the system (1) in vector form $\mathbf{Ax} = \mathbf{b}$ or in terms of column vectors $\mathbf{c}_{(1)}, \cdots, \mathbf{c}_{(n)}$ of $\mathbf{A}$:

$$\mathbf{c}_{(1)}x_1 + \mathbf{c}_{(2)}x_2 + \cdots + \mathbf{c}_{(n)}x_n = \mathbf{b}. \tag{2}$$

$\widetilde{\mathbf{A}}$ is obtained by augmenting $\mathbf{A}$ by a single column $\mathbf{b}$. Hence, by Theorem 3 in Sec. 7.4, rank $\widetilde{\mathbf{A}}$ equals rank $\mathbf{A}$ or rank $\mathbf{A} + 1$. Now if (1) has a solution $\mathbf{x}$, then (2) shows that $\mathbf{b}$ must be a linear combination of those column vectors, so that $\widetilde{\mathbf{A}}$ and $\mathbf{A}$ have the same maximum number of linearly independent column vectors and thus the same rank.

Conversely, if rank $\widetilde{\mathbf{A}}$ = rank $\mathbf{A}$, then $\mathbf{b}$ must be a linear combination of the column vectors of $\mathbf{A}$, say,

$$\mathbf{b} = \alpha_1\mathbf{c}_{(1)} + \cdots + \alpha_n\mathbf{c}_{(n)} \tag{2*}$$

since otherwise rank $\widetilde{\mathbf{A}}$ = rank $\mathbf{A} + 1$. But (2*) means that (1) has a solution, namely, $x_1 = \alpha_1, \cdots, x_n = \alpha_n$, as can be seen by comparing (2*) and (2).

(b) If rank $\mathbf{A} = n$, the n column vectors in (2) are linearly independent by Theorem 3 in Sec. 7.4. We claim that then the representation (2) of $\mathbf{b}$ is unique because otherwise

$$\mathbf{c}_{(1)}x_1 + \cdots + \mathbf{c}_{(n)}x_n = \mathbf{c}_{(1)}\tilde{x}_1 + \cdots + \mathbf{c}_{(n)}\tilde{x}_n.$$

This would imply (take all terms to the left, with a minus sign)

$$(x_1 - \tilde{x}_1)\mathbf{c}_{(1)} + \cdots + (x_n - \tilde{x}_n)\mathbf{c}_{(n)} = \mathbf{0}$$

and $x_1 - \tilde{x}_1 = 0, \cdots, x_n - \tilde{x}_n = 0$ by linear independence. But this means that the scalars $x_1, \cdots, x_n$ in (2) are uniquely determined, that is, the solution of (1) is unique.

(c) If rank $\mathbf{A}$ = rank $\widetilde{\mathbf{A}} = r < n$, then by Theorem 3 in Sec. 7.4 there is a linearly independent set K of r column vectors of $\mathbf{A}$ such that the other $n - r$ column vectors of $\mathbf{A}$ are linear combinations of those vectors. We renumber the columns and unknowns, denoting the renumbered quantities by $\hat{}$, so that $\{\hat{\mathbf{c}}_{(1)}, \cdots, \hat{\mathbf{c}}_{(r)}\}$ is that linearly independent set K. Then (2) becomes

$$\hat{\mathbf{c}}_{(1)}\hat{x}_1 + \cdots + \hat{\mathbf{c}}_{(r)}\hat{x}_r + \hat{\mathbf{c}}_{(r+1)}\hat{x}_{r+1} + \cdots + \hat{\mathbf{c}}_{(n)}\hat{x}_n = \mathbf{b},$$

$\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{\mathbf{c}}_{(n)}$ are linear combinations of the vectors of K, and so are the vectors $\hat{x}_{r+1}\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{x}_n\hat{\mathbf{c}}_{(n)}$. Expressing these vectors in terms of the vectors of K and collecting terms, we can thus write the system in the form

$$\hat{\mathbf{c}}_{(1)}y_1 + \cdots + \hat{\mathbf{c}}_{(r)}y_r = \mathbf{b} \tag{3}$$

with $y_j = \hat{x}_j + \beta_j$, where β_j results from the $n - r$ terms $\hat{\mathbf{c}}_{(r+1)}\hat{x}_{r+1}, \cdots, \hat{\mathbf{c}}_{(n)}\hat{x}_n$; here, $j = 1, \cdots, r$. Since the system has a solution, there are $y_1, \cdots, y_r$ satisfying (3). These scalars are unique since K is linearly independent. Choosing $\hat{x}_{r+1}, \cdots, \hat{x}_n$ fixes the β_j and corresponding $\hat{x}_j = y_j - \beta_j$, where $j = 1, \cdots, r$.

(d) This was discussed in Sec. 7.3 and is restated here as a reminder. ■

The theorem is illustrated in Sec. 7.3. In Example 2 there is a unique solution since rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A} = n = 3$ (as can be seen from the last matrix in the example). In Example 3 we have rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A} = 2 < n = 4$ and can choose x_3 and x_4 arbitrarily. In Example 4 there is no solution because rank $\mathbf{A} = 2 <$ rank $\tilde{\mathbf{A}} = 3$.

Homogeneous Linear System

Recall from Sec. 7.3 that a linear system (1) is called **homogeneous** if all the b_j's are zero, and **nonhomogeneous** if one or several b_j's are not zero. For the homogeneous system we obtain from the Fundamental Theorem the following results.

THEOREM 2

Homogeneous Linear System

A homogeneous linear system

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\
\cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0
\end{aligned}
\tag{4}
$$

always has the **trivial solution** $x_1 = 0, \cdots, x_n = 0$. *Nontrivial solutions exist if and only if rank* $\mathbf{A} < n$. *If* rank $\mathbf{A} = r < n$, *these solutions, together with* $\mathbf{x} = \mathbf{0}$, *form a vector space* (see Sec. 7.4) *of dimension* $n - r$ called the **solution space** of (4).

In particular, if $\mathbf{x}_{(1)}$ *and* $\mathbf{x}_{(2)}$ *are solution vectors of* (4), *then* $\mathbf{x} = c_1\mathbf{x}_{(1)} + c_2\mathbf{x}_{(2)}$ *with any scalars* c_1 *and* c_2 *is a solution vector of* (4). (This ***does not hold*** for nonhomogeneous systems. Also, the term *solution space is used for homogeneous* systems only.)

PROOF The first proposition can be seen directly from the system. It agrees with the fact that $\mathbf{b} = \mathbf{0}$ implies that rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A}$, so that a homogeneous system is always ***consistent***. If rank $\mathbf{A} = n$, the trivial solution is the unique solution according to (b) in Theorem 1. If rank $\mathbf{A} < n$, there are nontrivial solutions according to (c) in Theorem 1. The solutions form a vector space because if $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$ are any of them, then $\mathbf{A}\mathbf{x}_{(1)} = \mathbf{0}$, $\mathbf{A}\mathbf{x}_{(2)} = \mathbf{0}$, and this implies $\mathbf{A}(\mathbf{x}_{(1)} + \mathbf{x}_{(2)}) = \mathbf{A}\mathbf{x}_{(1)} + \mathbf{A}\mathbf{x}_{(2)} = \mathbf{0}$ as well as $\mathbf{A}(c\mathbf{x}_{(1)}) = c\mathbf{A}\mathbf{x}_{(1)} = \mathbf{0}$, where c is arbitrary. If rank $\mathbf{A} = r < n$, Theorem 1 (c) implies that we can choose $n - r$ suitable unknowns, call them $x_{r+1}, \cdots, x_n$, in an arbitrary fashion, and every solution is obtained in this way. Hence a basis for the solution space, briefly called a **basis of solutions** of (4), is $\mathbf{y}_{(1)}, \cdots, \mathbf{y}_{(n-r)}$, where the basis vector $\mathbf{y}_{(j)}$ is obtained by choosing $x_{r+j} = 1$ and the other $x_{r+1}, \cdots, x_n$ zero; the corresponding first r components of this solution vector are then determined. Thus the solution space of (4) has dimension $n - r$. This proves Theorem 2. ■

The solution space of (4) is also called the **null space** of $\mathbf{A}$ because $\mathbf{Ax} = \mathbf{0}$ for every $\mathbf{x}$ in the solution space of (4). Its dimension is called the **nullity** of $\mathbf{A}$. Hence Theorem 2 states that

$$\text{rank } \mathbf{A} + \text{nullity } \mathbf{A} = n \tag{5}$$

where n is the number of unknowns (number of columns of $\mathbf{A}$).

Furthermore, by the definition of rank we have rank $\mathbf{A} \leqq m$ in (4). Hence if $m < n$, then rank $\mathbf{A} < n$. By Theorem 2 this gives the practically important

THEOREM 3

Homogeneous Linear System with Fewer Equations Than Unknowns

A homogeneous linear system with fewer equations than unknowns always has nontrivial solutions.

Nonhomogeneous Linear Systems

The characterization of all solutions of the linear system (1) is now quite simple, as follows.

THEOREM 4

Nonhomogeneous Linear System

If a nonhomogeneous linear system (1) *is consistent, then all of its solutions are obtained as*

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_h \tag{6}$$

where $\mathbf{x}_0$ *is any* (fixed) *solution of* (1) *and* $\mathbf{x}_h$ *runs through all the solutions of the corresponding homogeneous system* (4).

PROOF The difference $\mathbf{x}_h = \mathbf{x} - \mathbf{x}_0$ of any two solutions of (1) is a solution of (4) because $\mathbf{Ax}_h = \mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{Ax} - \mathbf{Ax}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Since $\mathbf{x}$ is any solution of (1), we get all the solutions of (1) if in (6) we take any solution $\mathbf{x}_0$ of (1) and let $\mathbf{x}_h$ vary throughout the solution space of (4). ■

This covers a main part of our discussion of characterizing the solutions of systems of linear equations. Our next main topic is determinants and their role in linear equations.

7.6 For Reference: Second- and Third-Order Determinants

We created this section as a quick general reference section on second- and third-order determinants. It is completely independent of the theory in Sec. 7.7 and suffices as a reference for many of our examples and problems. Since this section is for reference, ***go on to the next section, consulting this material only when needed***.

A **determinant of second order** is denoted and defined by

$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{1}$$

So here we have ***bars*** (whereas a matrix has ***brackets***).

Cramer's rule for solving linear systems of two equations in two unknowns

$$\text{(2)} \qquad \begin{aligned} &\text{(a)} \quad a_{11}x_1 + a_{12}x_2 = b_1 \\ &\text{(b)} \quad a_{21}x_1 + a_{22}x_2 = b_2 \end{aligned}$$

is

$$\text{(3)} \qquad \begin{aligned} x_1 &= \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{D} = \frac{b_1a_{22} - a_{12}b_2}{D}, \\ x_2 &= \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{D} = \frac{a_{11}b_2 - b_1a_{21}}{D} \end{aligned}$$

with D as in (1), provided

$$D \neq 0.$$

The value $D = 0$ appears for homogeneous systems with nontrivial solutions.

PROOF We prove (3). To eliminate x_2 multiply (2a) by a_{22} and (2b) by $-a_{12}$ and add,

$$(a_{11}a_{22} - a_{12}a_{21})x_1 = b_1a_{22} - a_{12}b_2.$$

Similarly, to eliminate x_1 multiply (2a) by $-a_{21}$ and (2b) by a_{11} and add,

$$(a_{11}a_{22} - a_{12}a_{21})x_2 = a_{11}b_2 - b_1a_{21}.$$

Assuming that $D = a_{11}a_{22} - a_{12}a_{21} \neq 0$, dividing, and writing the right sides of these two equations as determinants, we obtain (3). ■

EXAMPLE 1 Cramer's Rule for Two Equations

If $\begin{aligned} 4x_1 + 3x_2 &= 12 \\ 2x_1 + 5x_2 &= -8 \end{aligned}$ then $x_1 = \dfrac{\begin{vmatrix} 12 & 3 \\ -8 & 5 \end{vmatrix}}{\begin{vmatrix} 4 & 3 \\ 2 & 5 \end{vmatrix}} = \dfrac{84}{14} = 6, \quad x_2 = \dfrac{\begin{vmatrix} 4 & 12 \\ 2 & -8 \end{vmatrix}}{\begin{vmatrix} 4 & 3 \\ 2 & 5 \end{vmatrix}} = \dfrac{-56}{14} = -4.$ ■

Third-Order Determinants

A **determinant of third order** can be defined by

$$\text{(4)} \qquad D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

Note the following. The signs on the right are $+ - +$. Each of the three terms on the right is an entry in the first column of D times its **minor**, that is, the second-order determinant obtained from D by deleting the row and column of that entry; thus, for a_{11} delete the first row and first column, and so on.

If we write out the minors in (4), we obtain

$$D = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{21}a_{13}a_{32} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{13}a_{22}. \tag{4*}$$

Cramer's Rule for Linear Systems of Three Equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{5}$$

is

$$x_1 = \frac{D_1}{D}, \quad x_2 = \frac{D_2}{D}, \quad x_3 = \frac{D_3}{D} \qquad (D \neq 0) \tag{6}$$

with the *determinant D of the system* given by (4) and

$$D_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}, \quad D_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}, \quad D_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}.$$

Note that D_1, D_2, D_3 are obtained by replacing Columns 1, 2, 3, respectively, by the column of the right sides of (5).

Cramer's rule (6) can be derived by eliminations similar to those for (3), but it also follows from the general case (Theorem 4) in the next section.

7.7 Determinants. Cramer's Rule

Determinants were originally introduced for solving linear systems. Although ***impractical in computations***, they have important engineering applications in eigenvalue problems (Sec. 8.1), differential equations, vector algebra (Sec. 9.3), and in other areas. They can be introduced in several equivalent ways. Our definition is particularly for dealing with linear systems.

A **determinant of order** n is a scalar associated with an $n \times n$ (hence ***square!***) matrix $\mathbf{A} = [a_{jk}]$, and is denoted by

$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{1}$$

For $n = 1$, this determinant is defined by

$$D = a_{11}. \tag{2}$$

For $n \geqq 2$ by

$$D = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn} \qquad (j = 1, 2, \cdots, \text{or } n) \tag{3a}$$

or

$$D = a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk} \quad (k = 1, 2, \cdots, \text{or } n). \tag{3b}$$

Here,

$$C_{jk} = (-1)^{j+k} M_{jk}$$

and M_{jk} is a determinant of order $n - 1$, namely, the determinant of the submatrix of $\mathbf{A}$ obtained from $\mathbf{A}$ by omitting the row and column of the entry a_{jk}, that is, the jth row and the kth column.

In this way, D is defined in terms of n determinants of order $n - 1$, each of which is, in turn, defined in terms of $n - 1$ determinants of order $n - 2$, and so on—until we finally arrive at second-order determinants, in which those submatrices consist of single entries whose determinant is defined to be the entry itself.

From the definition it follows that *we may **expand** D by any row or column,* that is, choose in (3) the entries in any row or column, similarly when expanding the C_{jk}'s in (3), and so on.

This definition is unambiguous, that is, it yields the same value for D no matter which columns or rows we choose in expanding. A proof is given in App. 4.

Terms used in connection with determinants are taken from matrices. In D we have n^2 **entries** a_{jk}, also n **rows** and n **columns**, and a **main diagonal** on which $a_{11}, a_{22}, \cdots, a_{nn}$ stand. Two terms are new:

M_{jk} is called the **minor** *of a_{jk} in D,* and C_{jk} the **cofactor** *of a_{jk} in D.*

For later use we note that (3) may also be written in terms of minors

$$D = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (j = 1, 2, \cdots, \text{or } n) \tag{4a}$$

$$D = \sum_{j=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (k = 1, 2, \cdots, \text{or } n). \tag{4b}$$

EXAMPLE 1 Minors and Cofactors of a Third-Order Determinant

In (4) of the previous section the minors and cofactors of the entries in the first column can be seen directly. For the entries in the second row the minors are

$$M_{21} = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}, \qquad M_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \qquad M_{23} = \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

and the cofactors are $C_{21} = -M_{21}$, $C_{22} = +M_{22}$, and $C_{23} = -M_{23}$. Similarly for the third row—write these down yourself. And verify that the signs in C_{jk} form a **checkerboard pattern**

$$\begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix}$$

■

EXAMPLE 2 **Expansions of a Third-Order Determinant**

$$D = \begin{vmatrix} 1 & 3 & 0 \\ 2 & 6 & 4 \\ -1 & 0 & 2 \end{vmatrix} = 1\begin{vmatrix} 6 & 4 \\ 0 & 2 \end{vmatrix} - 3\begin{vmatrix} 2 & 4 \\ -1 & 2 \end{vmatrix} + 0\begin{vmatrix} 2 & 6 \\ -1 & 0 \end{vmatrix}$$

$$= 1(12 - 0) - 3(4 + 4) + 0(0 + 6) = -12.$$

This is the expansion by the first row. The expansion by the third column is

$$D = 0\begin{vmatrix} 2 & 6 \\ -1 & 0 \end{vmatrix} - 4\begin{vmatrix} 1 & 3 \\ -1 & 0 \end{vmatrix} + 2\begin{vmatrix} 1 & 3 \\ 2 & 6 \end{vmatrix} = 0 - 12 + 0 = -12.$$

Verify that the other four expansions also give the value -12. ■

EXAMPLE 3 **Determinant of a Triangular Matrix**

$$\begin{vmatrix} -3 & 0 & 0 \\ 6 & 4 & 0 \\ -1 & 2 & 5 \end{vmatrix} = -3\begin{vmatrix} 4 & 0 \\ 2 & 5 \end{vmatrix} = -3 \cdot 4 \cdot 5 = -60.$$

Inspired by this, can you formulate a little theorem on determinants of triangular matrices? Of diagonal matrices? ■

General Properties of Determinants

There is an attractive way of finding determinants (1) that consists of applying elementary row operations to (1). By doing so we obtain an "upper triangular" determinant (see Sec. 7.1, for definition with "matrix" replaced by "determinant") whose value is then very easy to compute, being just the product of its diagonal entries. This approach is ***similar*** (***but not the same!***) to what we did to matrices in Sec. 7.3. *In particular, be aware that interchanging two rows in a determinant introduces a multiplicative factor of* -1 *to the value of the determinant*! Details are as follows.

THEOREM 1 **Behavior of an *n*th-Order Determinant under Elementary Row Operations**

(a) *Interchange of two rows multiplies the value of the determinant by* -1.

(b) *Addition of a multiple of a row to another row does not alter the value of the determinant.*

(c) *Multiplication of a row by a nonzero constant c multiplies the value of the determinant by c.* (This holds also when $c = 0$, but no longer gives an elementary row operation.)

PROOF **(a)** By induction. The statement holds for $n = 2$ because

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \quad \text{but} \quad \begin{vmatrix} c & d \\ a & b \end{vmatrix} = bc - ad.$$

We now make the induction hypothesis that (a) holds for determinants of order $n - 1 \geqq 2$ and show that it then holds for determinants of order n. Let D be of order n. Let E be obtained from D by the interchange of two rows. Expand D and E by a row that is ***not*** one of those interchanged, call it the jth row. Then by (4a),

$$D = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk}, \qquad E = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} N_{jk} \tag{5}$$

where N_{jk} is obtained from the minor M_{jk} of a_{jk} in D by the interchange of those two rows which have been interchanged in D (and which N_{jk} must both contain because we expand by another row!). Now these minors are of order $n - 1$. Hence the induction hypothesis applies and gives $N_{jk} = -M_{jk}$. Thus $E = -D$ by (5).

(b) Add c times Row i to Row j. Let $\tilde{D}$ be the new determinant. Its entries in Row j are $a_{jk} + ca_{ik}$. If we expand $\tilde{D}$ by this Row j, we see that we can write it as $\tilde{D} = D_1 + cD_2$, where $D_1 = D$ has in Row j the a_{jk}, whereas D_2 has in that Row j the a_{jk} from the addition. Hence D_2 has a_{jk} in both Row i and Row j. Interchanging these two rows gives D_2 back, but on the other hand it gives $-D_2$ by (a). Together $D_2 = -D_2 = 0$, so that $\tilde{D} = D_1 = D$.

(c) Expand the determinant by the row that has been multiplied. ■

CAUTION! $\det(c\mathbf{A}) = c^n \det \mathbf{A}$ (not $c \det \mathbf{A}$). Explain why. ■

EXAMPLE 4 **Evaluation of Determinants by Reduction to Triangular Form**

Because of Theorem 1 we may evaluate determinants by reduction to triangular form, as in the Gauss elimination for a matrix. For instance (with the blue explanations always referring to the ***preceding determinant***)

$$D = \begin{vmatrix} 2 & 0 & -4 & 6 \\ 4 & 5 & 1 & 0 \\ 0 & 2 & 6 & -1 \\ -3 & 8 & 9 & 1 \end{vmatrix}$$

$$= \begin{vmatrix} 2 & 0 & -4 & 6 \\ 0 & 5 & 9 & -12 \\ 0 & 2 & 6 & -1 \\ 0 & 8 & 3 & 10 \end{vmatrix} \begin{matrix} \\ \text{Row 2} - 2 \text{ Row 1} \\ \\ \text{Row 4} + 1.5 \text{ Row 1} \end{matrix}$$

$$= \begin{vmatrix} 2 & 0 & -4 & 6 \\ 0 & 5 & 9 & -12 \\ 0 & 0 & 2.4 & 3.8 \\ 0 & 0 & -11.4 & 29.2 \end{vmatrix} \begin{matrix} \\ \\ \text{Row 3} - 0.4 \text{ Row 2} \\ \text{Row 4} - 1.6 \text{ Row 2} \end{matrix}$$

$$= \begin{vmatrix} 2 & 0 & -4 & 6 \\ 0 & 5 & 9 & -12 \\ 0 & 0 & 2.4 & 3.8 \\ 0 & 0 & -0 & 47.25 \end{vmatrix} \begin{matrix} \\ \\ \\ \text{Row 4} + 4.75 \text{ Row 3} \end{matrix}$$

$$= 2 \cdot 5 \cdot 2.4 \cdot 47.25 = 1134.$$ ■

THEOREM 2

Further Properties of nth-Order Determinants

(a)–(c) *in Theorem* 1 *hold also for columns.*

(d) ***Transposition*** *leaves the value of a determinant unaltered.*

(e) ***A zero row or column*** *renders the value of a determinant zero.*

(f) ***Proportional rows or columns*** *render the value of a determinant zero. In particular, a determinant with two identical rows or columns has the value zero.*

PROOF **(a)–(e)** follow directly from the fact that a determinant can be expanded by any row column. In **(d)**, transposition is defined as for matrices, that is, the jth row becomes the jth column of the transpose.

(f) If Row $j = c$ times Row i, then $D = cD_1$, where D_1 has Row j = Row i. Hence an interchange of these rows reproduces D_1, but it also gives $-D_1$ by Theorem 1(a). Hence $D_1 = 0$ and $D = cD_1 = 0$. Similarly for columns. ■

It is quite remarkable that the important concept of the rank of a matrix **A**, which is the maximum number of linearly independent row or column vectors of **A** (see Sec. 7.4), can be related to determinants. Here we may assume that rank $\mathbf{A} > 0$ because the only matrices with rank 0 are the zero matrices (see Sec. 7.4).

THEOREM 3

Rank in Terms of Determinants

Consider an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$:

(1) *$\mathbf{A}$ has rank $r \geqq 1$ if and only if $\mathbf{A}$ has an $r \times r$ submatrix with a nonzero determinant.*

(2) *The determinant of any square submatrix with more than r rows, contained in $\mathbf{A}$ (if such a matrix exists!) has a value equal to zero.*

Furthermore, if $m = n$, we have:

(3) *An $n \times n$ square matrix $\mathbf{A}$ has rank n if and only if*

$$\det \mathbf{A} \neq 0.$$

PROOF The key idea is that elementary row operations (Sec. 7.3) alter neither rank (by Theorem 1 in Sec. 7.4) nor the property of a determinant being nonzero (by Theorem 1 in this section). The echelon form $\hat{\mathbf{A}}$ of **A** (see Sec. 7.3) has r nonzero row vectors (which are the first r row vectors) if and only if rank $\mathbf{A} = r$. Without loss of generality, we can assume that $r \geqq 1$. Let $\hat{\mathbf{R}}$ be the $r \times r$ submatrix in the left upper corner of $\hat{\mathbf{A}}$ (so that the entries of $\hat{\mathbf{R}}$ are in both the first r rows and r columns of $\hat{\mathbf{A}}$). Now $\hat{\mathbf{R}}$ is triangular, with all diagonal entries r_{jj} nonzero. Thus, $\det \hat{\mathbf{R}} = r_{11} \cdots r_{rr} \neq 0$. Also $\det \mathbf{R} \neq 0$ for the corresponding $r \times r$ submatrix **R** of **A** because $\hat{\mathbf{R}}$ results from **R** by elementary row operations. This proves part (1).

Similarly, $\det \mathbf{S} = 0$ for any square submatrix **S** of $r + 1$ or more rows perhaps contained in **A** because the corresponding submatrix $\hat{\mathbf{S}}$ of $\hat{\mathbf{A}}$ must contain a row of zeros (otherwise we would have rank $\mathbf{A} \geqq r + 1$), so that $\det \hat{\mathbf{S}} = 0$ by Theorem 2. This proves part (2). Furthermore, we have proven the theorem for an $m \times n$ matrix.

For an $n \times n$ square matrix $\mathbf{A}$ we proceed as follows. To prove (3), we apply part (1) (already proven!). This gives us that rank $\mathbf{A} = n \geqq 1$ if and only if $\mathbf{A}$ contains an $n \times n$ submatrix with nonzero determinant. But the only such submatrix contained in our square matrix $\mathbf{A}$, is $\mathbf{A}$ itself, hence $\det \mathbf{A} \neq 0$. This proves part (3). ■

Cramer's Rule

Theorem 3 opens the way to the classical solution formula for linear systems known as **Cramer's rule**,[2] which gives solutions as quotients of determinants. ***Cramer's rule is not practical in computations*** for which the methods in Secs. 7.3 and 20.1–20.3 are suitable. However, Cramer's rule is of ***theoretical interest*** in differential equations (Secs. 2.10 and 3.3) and in other theoretical work that has engineering applications.

THEOREM 4

Cramer's Theorem (Solution of Linear Systems by Determinants)

(a) *If a linear system of n equations in the same number of unknowns* $x_1, \cdots, x_n$

(6)
$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad &\cdot \quad \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

has a nonzero coefficient determinant $D = \det \mathbf{A}$, *the system has precisely one solution. This solution is given by the formulas*

(7)
$$x_1 = \frac{D_1}{D}, \quad x_2 = \frac{D_2}{D}, \cdots, \quad x_n = \frac{D_n}{D} \qquad \textbf{(Cramer's rule)}$$

where D_k *is the determinant obtained from D by replacing in D the kth column by the column with the entries* $b_1, \cdots, b_n$.

(b) *Hence if the system (6) is* ***homogeneous*** *and* $D \neq 0$, *it has only the trivial solution* $x_1 = 0, x_2 = 0, \cdots, x_n = 0$. *If* $D = 0$, *the homogeneous system also has nontrivial solutions.*

PROOF The augmented matrix $\widetilde{\mathbf{A}}$ of the system (6) is of size $n \times (n + 1)$. Hence its rank can be at most n. Now if

(8)
$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \neq 0,$$

[2]GABRIEL CRAMER (1704–1752), Swiss mathematician.

then rank $\mathbf{A} = n$ by Theorem 3. Thus rank $\tilde{\mathbf{A}} = \text{rank}\,\mathbf{A}$. Hence, by the Fundamental Theorem in Sec. 7.5, the system (6) has a unique solution.

Let us now prove (7). Expanding D by its kth column, we obtain

$$D = a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk}, \tag{9}$$

where C_{ik} is the cofactor of entry a_{ik} in D. If we replace the entries in the kth column of D by any other numbers, we obtain a new determinant, say, $\hat{D}$. Clearly, its expansion by the kth column will be of the form (9), with $a_{1k}, \cdots, a_{nk}$ replaced by those new numbers and the cofactors C_{ik} as before. In particular, if we choose as new numbers the entries $a_{1l}, \cdots, a_{nl}$ of the lth column of D (where $l \neq k$), we have a new determinant $\hat{D}$ which has the column $[a_{1l} \quad \cdots \quad a_{nl}]^{\mathsf{T}}$ twice, once as its lth column, and once as its kth because of the replacement. Hence $\hat{D} = 0$ by Theorem 2(f). If we now expand $\hat{D}$ by the column that has been replaced (the kth column), we thus obtain

$$a_{1l}C_{1k} + a_{2l}C_{2k} + \cdots + a_{nl}C_{nk} = 0 \qquad (l \neq k). \tag{10}$$

We now multiply the first equation in (6) by C_{1k} on both sides, the second by $C_{2k}, \cdots,$ the last by C_{nk}, and add the resulting equations. This gives

$$\begin{aligned} C_{1k}(a_{11}x_1 + \cdots + a_{1n}x_n) + \cdots + C_{nk}(a_{n1}x_1 + \cdots + a_{nn}x_n) \\ = b_1C_{1k} + \cdots + b_nC_{nk}. \end{aligned} \tag{11}$$

Collecting terms with the same x_j, we can write the left side as

$$x_1(a_{11}C_{1k} + a_{21}C_{2k} + \cdots + a_{n1}C_{nk}) + \cdots + x_n(a_{1n}C_{1k} + a_{2n}C_{2k} + \cdots + a_{nn}C_{nk}).$$

From this we see that x_k is multiplied by

$$a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk}.$$

Equation (9) shows that this equals D. Similarly, x_1 is multiplied by

$$a_{1l}C_{1k} + a_{2l}C_{2k} + \cdots + a_{nl}C_{nk}.$$

Equation (10) shows that this is zero when $l \neq k$. Accordingly, the left side of (11) equals simply x_kD, so that (11) becomes

$$x_kD = b_1C_{1k} + b_2C_{2k} + \cdots + b_nC_{nk}.$$

Now the right side of this is D_k as defined in the theorem, expanded by its kth column, so that division by D gives (7). This proves Cramer's rule.

If (6) is homogeneous and $D \neq 0$, then each D_k has a column of zeros, so that $D_k = 0$ by Theorem 2(e), and (7) gives the trivial solution.

Finally, if (6) is homogeneous and $D = 0$, then rank $\mathbf{A} < n$ by Theorem 3, so that nontrivial solutions exist by Theorem 2 in Sec. 7.5. ■

EXAMPLE 5 Illustration of Cramer's Rule (Theorem 4)

For $n = 2$, see Example 1 of Sec. 7.6. Also, at the end of that section, we give Cramer's rule for a general linear system of three equations. ■

Finally, an important application for Cramer's rule dealing with inverse matrices will be given in the next section.

PROBLEM SET 7.7

1–6 GENERAL PROBLEMS

1. **General Properties of Determinants.** Illustrate each statement in Theorems 1 and 2 with an example of your choice.

2. **Second-Order Determinant.** Expand a general second-order determinant in four possible ways and show that the results agree.

3. **Third-Order Determinant.** Do the task indicated in Theorem 2. Also evaluate D by reduction to triangular form.

4. **Expansion Numerically Impractical.** Show that the computation of an nth-order determinant by expansion involves $n!$ multiplications, which if a multiplication takes 10^{-9} sec would take these times:

n	10	15	20	25
Time	0.004 sec	22 min	77 years	$0.5 \cdot 10^9$ years

5. **Multiplication by Scalar.** Show that $\det(k\mathbf{A}) = k^n \det \mathbf{A}$ (not $k \det \mathbf{A}$). Give an example.

6. **Minors, cofactors.** Complete the list in Example 1.

7–15 EVALUATION OF DETERMINANTS

Showing the details, evaluate:

7. $\begin{vmatrix} \cos\alpha & \sin\alpha \\ \sin\beta & \cos\beta \end{vmatrix}$

8. $\begin{vmatrix} 0.4 & 4.9 \\ 1.5 & -1.3 \end{vmatrix}$

9. $\begin{vmatrix} \cos n\theta & \sin n\theta \\ -\sin n\theta & \cos n\theta \end{vmatrix}$

10. $\begin{vmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{vmatrix}$

11. $\begin{vmatrix} 4 & -1 & 8 \\ 0 & 2 & 3 \\ 0 & 0 & 5 \end{vmatrix}$

12. $\begin{vmatrix} a & b & c \\ c & a & b \\ b & c & a \end{vmatrix}$

13. $\begin{vmatrix} 0 & 4 & -1 & 5 \\ -4 & 0 & 3 & -2 \\ 1 & -3 & 0 & 1 \\ -5 & 2 & -1 & 0 \end{vmatrix}$

14. $\begin{vmatrix} 4 & 7 & 0 & 0 \\ 2 & 8 & 0 & 0 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & -2 & 2 \end{vmatrix}$

15. $\begin{vmatrix} 1 & 2 & 0 & 0 \\ 2 & 4 & 2 & 0 \\ 0 & 2 & 9 & 2 \\ 0 & 0 & 2 & 16 \end{vmatrix}$

16. **CAS EXPERIMENT. Determinant of Zeros and Ones.** Find the value of the determinant of the $n \times n$ matrix A_n with main diagonal entries all 0 and all others 1. Try to find a formula for this. Try to prove it by induction. Interpret A_3 and A_4 as *incidence matrices* (as in Problem Set 7.1 but without the minuses) of a triangle and a tetrahedron, respectively; similarly for an *n-simplex*, having n vertices and $n(n-1)/2$ edges (and spanning R^{n-1}, $n = 5, 6, \cdots$).

17–19 RANK BY DETERMINANTS

Find the rank by Theorem 3 (which is not very practical) and check by row reduction. Show details.

17. $\begin{bmatrix} 4 & 9 \\ -8 & -6 \\ 16 & 12 \end{bmatrix}$

18. $\begin{bmatrix} 0 & 4 & -6 \\ 4 & 0 & 10 \\ -6 & 10 & 0 \end{bmatrix}$

19. $\begin{bmatrix} 1 & 5 & 2 & 2 \\ 1 & 3 & 2 & 6 \\ 4 & 0 & 8 & 48 \end{bmatrix}$

20. **TEAM PROJECT. Geometric Applications: Curves and Surfaces Through Given Points.** The idea is to get an equation from the vanishing of the determinant of a homogeneous linear system as the condition for a nontrivial solution in Cramer's theorem. We explain the trick for obtaining such a system for the case of a line L through two given points P_1: (x_1, y_1) and P_2: (x_2, y_2). The unknown line is $ax + by = -c$, say. We write it as $ax + by + c \cdot 1 = 0$. To get a nontrivial solution a, b, c, the determinant of the "coefficients" $x, y, 1$ must be zero. The system is

$$\begin{aligned} &ax + by + c \cdot 1 = 0 && (\text{Line } L) \\ (12)\quad &ax_1 + by_1 + c \cdot 1 = 0 && (P_1 \text{ on } L) \\ &ax_2 + by_2 + c \cdot 1 = 0 && (P_2 \text{ on } L). \end{aligned}$$

(a) Line through two points. Derive from $D = 0$ in (12) the familiar formula

$$\frac{x - x_1}{x_1 - x_2} = \frac{y - y_1}{y_1 - y_2}.$$

(b) Plane. Find the analog of (12) for a plane through three given points. Apply it when the points are (1, 1, 1), (3, 2, 6), (5, 0, 5).

(c) Circle. Find a similar formula for a circle in the plane through three given points. Find and sketch the circle through (2, 6), (6, 4), (7, 1).

(d) Sphere. Find the analog of the formula in (c) for a sphere through four given points. Find the sphere through (0, 0, 5), (4, 0, 1), (0, 4, 1), (0, 0, −3) by this formula or by inspection.

(e) General conic section. Find a formula for a general conic section (the vanishing of a determinant of 6th order). Try it out for a quadratic parabola and for a more general conic section of your own choice.

21–25 CRAMER'S RULE

Solve by Cramer's rule. Check by Gauss elimination and back substitution. Show details.

21. $\begin{aligned} 3x - 5y &= 15.5 \\ 6x + 16y &= 5.0 \end{aligned}$

22. $\begin{aligned} 2x - 4y &= -24 \\ 5x + 2y &= 0 \end{aligned}$

23. $\begin{aligned} 3y - 4z &= 16 \\ 2x - 5y + 7z &= -27 \\ -x \qquad - 9z &= 9 \end{aligned}$

24. $\begin{aligned} 3x - 2y + z &= 13 \\ -2x + y + 4z &= 11 \\ x + 4y - 5z &= -31 \end{aligned}$

25. $\begin{aligned} -4w + x + y \qquad &= -10 \\ w - 4x \qquad + z &= 1 \\ w \qquad - 4y + z &= -7 \\ x + y - 4z &= 10 \end{aligned}$

7.8 Inverse of a Matrix. Gauss–Jordan Elimination

***In this section we consider* square *matrices exclusively*.**

The **inverse** of an $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ is denoted by $\mathbf{A}^{-1}$ and is an $n \times n$ matrix such that

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{1}$$

where $\mathbf{I}$ is the $n \times n$ unit matrix (see Sec. 7.2).

If $\mathbf{A}$ has an inverse, then $\mathbf{A}$ is called a **nonsingular matrix**. If $\mathbf{A}$ has no inverse, then $\mathbf{A}$ is called a **singular matrix**.

If $\mathbf{A}$ *has an inverse, the inverse is unique.*

Indeed, if both $\mathbf{B}$ and $\mathbf{C}$ are inverses of $\mathbf{A}$, then $\mathbf{AB} = \mathbf{I}$ and $\mathbf{CA} = \mathbf{I}$, so that we obtain the uniqueness from

$$\mathbf{B} = \mathbf{IB} = (\mathbf{CA})\mathbf{B} = \mathbf{C}(\mathbf{AB}) = \mathbf{CI} = \mathbf{C}.$$

We prove next that $\mathbf{A}$ has an inverse (is nonsingular) if and only if it has maximum possible rank n. The proof will also show that $\mathbf{Ax} = \mathbf{b}$ implies $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ provided $\mathbf{A}^{-1}$ exists, and will thus give a motivation for the inverse as well as a relation to linear systems. (But this will ***not*** give a good method of solving $\mathbf{Ax} = \mathbf{b}$ ***numerically*** because the Gauss elimination in Sec. 7.3 requires fewer computations.)

THEOREM 1

Existence of the Inverse

The inverse $\mathbf{A}^{-1}$ *of an* $n \times n$ *matrix* $\mathbf{A}$ *exists if and only if* rank $\mathbf{A} = n$, *thus* (by Theorem 3, Sec. 7.7) *if and only if* det $\mathbf{A} \neq 0$. *Hence* $\mathbf{A}$ *is nonsingular if* rank $\mathbf{A} = n$, *and is singular if* rank $\mathbf{A} < n$.

PROOF Let $\mathbf{A}$ be a given $n \times n$ matrix and consider the linear system

$$\mathbf{Ax} = \mathbf{b}. \tag{2}$$

If the inverse $\mathbf{A}^{-1}$ exists, then multiplication from the left on both sides and use of (1) gives

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

This shows that (2) has a solution $\mathbf{x}$, which is unique because, for another solution $\mathbf{u}$, we have $\mathbf{Au} = \mathbf{b}$, so that $\mathbf{u} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{x}$. Hence $\mathbf{A}$ must have rank n by the Fundamental Theorem in Sec. 7.5.

Conversely, let rank $\mathbf{A} = n$. Then by the same theorem, the system (2) has a unique solution $\mathbf{x}$ for any $\mathbf{b}$. Now the back substitution following the Gauss elimination (Sec. 7.3) shows that the components x_j of $\mathbf{x}$ are linear combinations of those of $\mathbf{b}$. Hence we can write

$$\mathbf{x} = \mathbf{Bb} \tag{3}$$

with $\mathbf{B}$ to be determined. Substitution into (2) gives

$$\mathbf{Ax} = \mathbf{A}(\mathbf{Bb}) = (\mathbf{AB})\mathbf{b} = \mathbf{Cb} = \mathbf{b} \qquad (\mathbf{C} = \mathbf{AB})$$

for any $\mathbf{b}$. Hence $\mathbf{C} = \mathbf{AB} = \mathbf{I}$, the unit matrix. Similarly, if we substitute (2) into (3) we get

$$\mathbf{x} = \mathbf{Bb} = \mathbf{B}(\mathbf{Ax}) = (\mathbf{BA})\mathbf{x}$$

for any $\mathbf{x}$ (and $\mathbf{b} = \mathbf{Ax}$). Hence $\mathbf{BA} = \mathbf{I}$. Together, $\mathbf{B} = \mathbf{A}^{-1}$ exists. ■

Determination of the Inverse by the Gauss–Jordan Method

To actually determine the inverse $\mathbf{A}^{-1}$ of a nonsingular $n \times n$ matrix $\mathbf{A}$, we can use a variant of the Gauss elimination (Sec. 7.3), called the **Gauss–Jordan elimination**.[3] The idea of the method is as follows.

Using $\mathbf{A}$, we form n linear systems

$$\mathbf{Ax}_{(1)} = \mathbf{e}_{(1)}, \quad \cdots, \quad \mathbf{Ax}_{(n)} = \mathbf{e}_{(n)}$$

where the vectors $\mathbf{e}_{(1)}, \cdots, \mathbf{e}_{(n)}$ are the columns of the $n \times n$ unit matrix $\mathbf{I}$; thus, $\mathbf{e}_{(1)} = [1 \quad 0 \quad \cdots \quad 0]^\mathsf{T}$, $\mathbf{e}_{(2)} = [0 \quad 1 \quad 0 \quad \cdots \quad 0]^\mathsf{T}$, etc. These are n vector equations in the unknown vectors $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$. We combine them into a single matrix equation

[3] WILHELM JORDAN (1842–1899), German geodesist and mathematician. He did important geodesic work in Africa, where he surveyed oases. [See Althoen, S.C. and R. McLaughlin, Gauss–Jordan reduction: A brief history. *American Mathematical Monthly*, Vol. **94**, No. 2 (1987), pp. 130–142.]

We do ***not recommend*** it as a method for solving systems of linear equations, since the number of operations in addition to those of the Gauss elimination is larger than that for back substitution, which the Gauss–Jordan elimination avoids. See also Sec. 20.1.

$\mathbf{AX} = \mathbf{I}$, with the unknown matrix $\mathbf{X}$ having the columns $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$. Correspondingly, we combine the n augmented matrices $[\mathbf{A} \quad \mathbf{e}_{(1)}], \cdots, [\mathbf{A} \quad \mathbf{e}_{(n)}]$ into one wide $n \times 2n$ "augmented matrix" $\widetilde{\mathbf{A}} = [\mathbf{A} \quad \mathbf{I}]$. Now multiplication of $\mathbf{AX} = \mathbf{I}$ by $\mathbf{A}^{-1}$ from the left gives $\mathbf{X} = \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}^{-1}$. Hence, to solve $\mathbf{AX} = \mathbf{I}$ for $\mathbf{X}$, we can apply the Gauss elimination to $\widetilde{\mathbf{A}} = [\mathbf{A} \quad \mathbf{I}]$. This gives a matrix of the form $[\mathbf{U} \quad \mathbf{H}]$ with upper triangular $\mathbf{U}$ because the Gauss elimination triangularizes systems. The Gauss–Jordan method reduces $\mathbf{U}$ by further elementary row operations to diagonal form, in fact to the unit matrix $\mathbf{I}$. This is done by eliminating the entries of $\mathbf{U}$ above the main diagonal and making the diagonal entries all 1 by multiplication (see Example 1). Of course, the method operates on the entire matrix $[\mathbf{U} \quad \mathbf{H}]$, transforming $\mathbf{H}$ into some matrix $\mathbf{K}$, hence the entire $[\mathbf{U} \quad \mathbf{H}]$ to $[\mathbf{I} \quad \mathbf{K}]$. This is the "augmented matrix" of $\mathbf{IX} = \mathbf{K}$. Now $\mathbf{IX} = \mathbf{X} = \mathbf{A}^{-1}$, as shown before. By comparison, $\mathbf{K} = \mathbf{A}^{-1}$, so that we can read $\mathbf{A}^{-1}$ directly from $[\mathbf{I} \quad \mathbf{K}]$.

The following example illustrates the practical details of the method.

EXAMPLE 1 Finding the Inverse of a Matrix by Gauss–Jordan Elimination

Determine the inverse $\mathbf{A}^{-1}$ of

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

Solution. We apply the Gauss elimination (Sec. 7.3) to the following $n \times 2n = 3 \times 6$ matrix, where BLUE always refers to the previous matrix.

$$[\mathbf{A} \quad \mathbf{I}] = \left[\begin{array}{ccc|ccc} -1 & 1 & 2 & 1 & 0 & 0 \\ 3 & -1 & 1 & 0 & 1 & 0 \\ -1 & 3 & 4 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|ccc} -1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 7 & 3 & 1 & 0 \\ 0 & 2 & 2 & -1 & 0 & 1 \end{array}\right] \begin{array}{l} \\ \text{Row 2} + 3\text{ Row 1} \\ \text{Row 3} - \text{Row 1} \end{array}$$

$$\left[\begin{array}{ccc|ccc} -1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 7 & 3 & 1 & 0 \\ 0 & 0 & -5 & -4 & -1 & 1 \end{array}\right] \begin{array}{l} \\ \\ \text{Row 3} - \text{Row 2} \end{array}$$

This is $[\mathbf{U} \quad \mathbf{H}]$ as produced by the Gauss elimination. Now follow the additional Gauss–Jordan steps, reducing $\mathbf{U}$ to $\mathbf{I}$, that is, to diagonal form with entries 1 on the main diagonal.

$$\left[\begin{array}{ccc|ccc} 1 & -1 & -2 & -1 & 0 & 0 \\ 0 & 1 & 3.5 & 1.5 & 0.5 & 0 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} -\text{Row 1} \\ 0.5\text{ Row 2} \\ -0.2\text{ Row 3} \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 0.6 & 0.4 & -0.4 \\ 0 & 1 & 0 & -1.3 & -0.2 & 0.7 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} \text{Row 1} + 2\text{ Row 3} \\ \text{Row 2} - 3.5\text{ Row 3} \\ \\ \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -0.7 & 0.2 & 0.3 \\ 0 & 1 & 0 & -1.3 & -0.2 & 0.7 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} \text{Row 1} + \text{Row 2} \\ \\ \\ \end{array}$$

The last three columns constitute $\mathbf{A}^{-1}$. Check:

$$\begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix} \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Similarly, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. ■

Formulas for Inverses

Since finding the inverse of a matrix is really a problem of solving a system of linear equations, it is not surprising that Cramer's rule (Theorem 4, Sec. 7.7) might come into play. And similarly, as Cramer's rule was useful for theoretical study but not for computation, so too is the explicit formula (4) in the following theorem useful for theoretical considerations but not recommended for actually determining inverse matrices, except for the frequently occurring 2×2 case as given in (4*).

THEOREM 2

Inverse of a Matrix by Determinants

The inverse of a nonsingular $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ is given by

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}[C_{jk}]^{\mathsf{T}} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \cdot & \cdot & \cdots & \cdot \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}, \tag{4}$$

where C_{jk} is the cofactor of a_{jk} in det $\mathbf{A}$ (see Sec. 7.7). (CAUTION! Note well that in $\mathbf{A}^{-1}$, the cofactor C_{jk} occupies the same place as a_{kj} (not a_{jk}) does in $\mathbf{A}$.)

In particular, the inverse of

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \textit{is} \quad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \tag{4*}$$

PROOF We denote the right side of (4) by $\mathbf{B}$ and show that $\mathbf{BA} = \mathbf{I}$. We first write

$$\mathbf{BA} = \mathbf{G} = [g_{kl}] \tag{5}$$

and then show that $\mathbf{G} = \mathbf{I}$. Now by the definition of matrix multiplication and because of the form of $\mathbf{B}$ in (4), we obtain (CAUTION! C_{sk}, not C_{ks})

$$g_{kl} = \sum_{s=1}^{n} \frac{C_{sk}}{\det \mathbf{A}} a_{sl} = \frac{1}{\det \mathbf{A}}(a_{1l}C_{1k} + \cdots + a_{nl}C_{nk}). \tag{6}$$

Now (9) and (10) in Sec. 7.7 show that the sum $(\cdots)$ on the right is $D = \det \mathbf{A}$ when $l = k$, and is zero when $l \neq k$. Hence

$$g_{kk} = \frac{1}{\det \mathbf{A}} \det \mathbf{A} = 1,$$

$$g_{kl} = 0 \quad (l \neq k).$$

In particular, for $n = 2$ we have in (4), in the first row, $C_{11} = a_{22}$, $C_{21} = -a_{12}$ and, in the second row, $C_{12} = -a_{21}$, $C_{22} = a_{11}$. This gives (4*). ■

The special case $n = 2$ occurs quite frequently in geometric and other applications. You may perhaps want to memorize formula (4*). Example 2 gives an illustration of (4*).

EXAMPLE 2 Inverse of a 2 × 2 Matrix by Determinants

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}, \quad \mathbf{A}^{-1} = \frac{1}{10}\begin{bmatrix} 4 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 0.4 & -0.1 \\ -0.2 & 0.3 \end{bmatrix}$$ ■

EXAMPLE 3 Further Illustration of Theorem 2

Using (4), find the inverse of

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

Solution. We obtain $\det \mathbf{A} = -1(-7) - 1 \cdot 13 + 2 \cdot 8 = 10$, and in (4),

$$C_{11} = \begin{vmatrix} -1 & 1 \\ 3 & 4 \end{vmatrix} = -7, \qquad C_{21} = -\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 2, \qquad C_{31} = \begin{vmatrix} 1 & 2 \\ -1 & 1 \end{vmatrix} = 3,$$

$$C_{12} = -\begin{vmatrix} 3 & 1 \\ -1 & 4 \end{vmatrix} = -13, \qquad C_{22} = \begin{vmatrix} -1 & 2 \\ -1 & 4 \end{vmatrix} = -2, \qquad C_{32} = -\begin{vmatrix} -1 & 2 \\ 3 & 1 \end{vmatrix} = 7,$$

$$C_{13} = \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} = 8, \qquad C_{23} = -\begin{vmatrix} -1 & 1 \\ -1 & 3 \end{vmatrix} = 2, \qquad C_{33} = \begin{vmatrix} -1 & 1 \\ 3 & -1 \end{vmatrix} = -2,$$

so that by (4), in agreement with Example 1,

$$\mathbf{A}^{-1} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix}.$$ ■

Diagonal matrices $\mathbf{A} = [a_{jk}]$, $a_{jk} = 0$ when $j \neq k$, have an inverse if and only if all $a_{jj} \neq 0$. Then $\mathbf{A}^{-1}$ is diagonal, too, with entries $1/a_{11}, \cdots, 1/a_{nn}$.

PROOF For a diagonal matrix we have in (4)

$$\frac{C_{11}}{D} = \frac{a_{22} \cdots a_{nn}}{a_{11}a_{22} \cdots a_{nn}} = \frac{1}{a_{11}}, \qquad \text{etc.}$$ ■

EXAMPLE 4 Inverse of a Diagonal Matrix

Let

$$\mathbf{A} = \begin{bmatrix} -0.5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then we obtain the inverse $\mathbf{A}^{-1}$ by inverting each individual diagonal element of $\mathbf{A}$, that is, by taking $1/(-0.5)$, $\frac{1}{4}$, and $\frac{1}{1}$ as the diagonal entries of $\mathbf{A}^{-1}$, that is,

$$\mathbf{A}^{-1} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Products can be inverted by taking the inverse of each factor and multiplying these inverses ***in reverse order***,

$$(\mathbf{AC})^{-1} = \mathbf{C}^{-1}\mathbf{A}^{-1}. \tag{7}$$

Hence for more than two factors,

$$(\mathbf{AC} \cdots \mathbf{PQ})^{-1} = \mathbf{Q}^{-1}\mathbf{P}^{-1} \cdots \mathbf{C}^{-1}\mathbf{A}^{-1}. \tag{8}$$

PROOF The idea is to start from (1) for $\mathbf{AC}$ instead of $\mathbf{A}$, that is, $\mathbf{AC}(\mathbf{AC})^{-1} = \mathbf{I}$, and multiply it on both sides from the left, first by $\mathbf{A}^{-1}$, which because of $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ gives

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{AC}(\mathbf{AC})^{-1} &= \mathbf{C}(\mathbf{AC})^{-1} \\ &= \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}^{-1}, \end{aligned}$$

and then multiplying this on both sides from the left, this time by $\mathbf{C}^{-1}$ and by using $\mathbf{C}^{-1}\mathbf{C} = \mathbf{I}$,

$$\mathbf{C}^{-1}\mathbf{C}(\mathbf{AC})^{-1} = (\mathbf{AC})^{-1} = \mathbf{C}^{-1}\mathbf{A}^{-1}.$$

This proves (7), and from it, (8) follows by induction.

We also note that *the inverse of the inverse is the given matrix*, as you may prove,

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}. \tag{9}$$

Unusual Properties of Matrix Multiplication. Cancellation Laws

Section 7.2 contains warnings that some properties of matrix multiplication deviate from those for numbers, and we are now able to explain the restricted validity of the so-called **cancellation laws** [2] and [3] below, using rank and inverse, concepts that were not yet available in Sec. 7.2. The deviations from the usual are of great practical importance and must be carefully observed. They are as follows.

[1] Matrix multiplication is not commutative, that is, in general we have

$$\mathbf{AB} \neq \mathbf{BA}.$$

[2] $\mathbf{AB} = \mathbf{0}$ does not generally imply $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$ (or $\mathbf{BA} = \mathbf{0}$); for example,

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

[3] $\mathbf{AC} = \mathbf{AD}$ does not generally imply $\mathbf{C} = \mathbf{D}$ (even when $\mathbf{A} \neq \mathbf{0}$).

Complete answers to [2] and [3] are contained in the following theorem.

THEOREM 3

Cancellation Laws

Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *be* $n \times n$ *matrices. Then:*

(a) *If* rank $\mathbf{A} = n$ *and* $\mathbf{AB} = \mathbf{AC}$, *then* $\mathbf{B} = \mathbf{C}$.

(b) *If* rank $\mathbf{A} = n$, *then* $\mathbf{AB} = \mathbf{0}$ *implies* $\mathbf{B} = \mathbf{0}$. *Hence if* $\mathbf{AB} = \mathbf{0}$, *but* $\mathbf{A} \neq \mathbf{0}$ *as well as* $\mathbf{B} \neq \mathbf{0}$, *then* rank $\mathbf{A} < n$ *and* rank $\mathbf{B} < n$.

(c) *If* $\mathbf{A}$ *is singular, so are* $\mathbf{BA}$ *and* $\mathbf{AB}$.

PROOF **(a)** The inverse of $\mathbf{A}$ exists by Theorem 1. Multiplication by $\mathbf{A}^{-1}$ from the left gives $\mathbf{A}^{-1}\mathbf{AB} = \mathbf{A}^{-1}\mathbf{AC}$, hence $\mathbf{B} = \mathbf{C}$.

(b) Let rank $\mathbf{A} = n$. Then $\mathbf{A}^{-1}$ exists, and $\mathbf{AB} = \mathbf{0}$ implies $\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B} = \mathbf{0}$. Similarly when rank $\mathbf{B} = n$. This implies the second statement in (b).

($\mathbf{c}_1$) Rank $\mathbf{A} < n$ by Theorem 1. Hence $\mathbf{Ax} = \mathbf{0}$ has nontrivial solutions by Theorem 2 in Sec. 7.5. Multiplication by $\mathbf{B}$ shows that these solutions are also solutions of $\mathbf{BAx} = \mathbf{0}$, so that rank ($\mathbf{BA}$) $< n$ by Theorem 2 in Sec. 7.5 and $\mathbf{BA}$ is singular by Theorem 1.

($\mathbf{c}_2$) $\mathbf{A}^\mathsf{T}$ is singular by Theorem 2(d) in Sec. 7.7. Hence $\mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$ is singular by part (c_1), and is equal to $(\mathbf{AB})^\mathsf{T}$ by (10d) in Sec. 7.2. Hence $\mathbf{AB}$ is singular by Theorem 2(d) in Sec. 7.7. ■

Determinants of Matrix Products

The determinant of a matrix product $\mathbf{AB}$ or $\mathbf{BA}$ can be written as the product of the determinants of the factors, and it is interesting that det $\mathbf{AB}$ = det $\mathbf{BA}$, although $\mathbf{AB} \neq \mathbf{BA}$ in general. The corresponding formula (10) is needed occasionally and can be obtained by Gauss–Jordan elimination (see Example 1) and from the theorem just proved.

THEOREM 4

Determinant of a Product of Matrices

For any $n \times n$ *matrices* $\mathbf{A}$ *and* $\mathbf{B}$,

$$\det(\mathbf{AB}) = \det(\mathbf{BA}) = \det \mathbf{A} \det \mathbf{B}. \tag{10}$$

PROOF If **A** or **B** is singular, so are **AB** and **BA** by Theorem 3(c), and (10) reduces to $0 = 0$ by Theorem 3 in Sec. 7.7.

Now let **A** and **B** be nonsingular. Then we can reduce **A** to a diagonal matrix $\hat{\mathbf{A}} = [a_{jk}]$ by Gauss–Jordan steps. Under these operations, det **A** retains its value, by Theorem 1 in Sec. 7.7, (a) and (b) [not (c)] except perhaps for a sign reversal in row interchanging when pivoting. But the same operations reduce **AB** to $\hat{\mathbf{A}}\mathbf{B}$ with the same effect on det (**AB**). Hence it remains to prove (10) for $\hat{\mathbf{A}}\mathbf{B}$; written out,

$$\hat{\mathbf{A}}\mathbf{B} = \begin{bmatrix} \hat{a}_{11} & 0 & \cdots & 0 \\ 0 & \hat{a}_{22} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \hat{a}_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ & & \vdots & \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} \hat{a}_{11}b_{11} & \hat{a}_{11}b_{12} & \cdots & \hat{a}_{11}b_{1n} \\ \hat{a}_{22}b_{21} & \hat{a}_{22}b_{22} & \cdots & \hat{a}_{22}b_{2n} \\ & & \vdots & \\ \hat{a}_{nn}b_{n1} & \hat{a}_{nn}b_{n2} & \cdots & \hat{a}_{nn}b_{nn} \end{bmatrix}.$$

We now take the determinant det ($\hat{\mathbf{A}}\mathbf{B}$). On the right we can take out a factor $\hat{a}_{11}$ from the first row, $\hat{a}_{22}$ from the second, $\cdots$, $\hat{a}_{nn}$ from the nth. But this product $\hat{a}_{11}\hat{a}_{22}\cdots\hat{a}_{nn}$ equals det $\hat{\mathbf{A}}$ because $\hat{\mathbf{A}}$ is diagonal. The remaining determinant is det **B**. This proves (10) for det (**AB**), and the proof for det (**BA**) follows by the same idea. ■

This completes our discussion of linear systems (Secs. 7.3–7.8). Section 7.9 on vector spaces and linear transformations is optional. ***Numeric methods*** are discussed in Secs. 20.1–20.4, which are independent of other sections on numerics.

PROBLEM SET 7.8

1–10 INVERSE

Find the inverse by Gauss–Jordan (or by (4*) if $n = 2$). Check by using (1).

1. $\begin{bmatrix} 1.80 & -2.32 \\ -0.25 & 0.60 \end{bmatrix}$

2. $\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{bmatrix}$

3. $\begin{bmatrix} 0.3 & -0.1 & 0.5 \\ 2 & 6 & 4 \\ 5 & 0 & 9 \end{bmatrix}$

4. $\begin{bmatrix} 0 & 0 & 0.1 \\ 0 & -0.4 & 0 \\ 2.5 & 0 & 0 \end{bmatrix}$

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

6. $\begin{bmatrix} -4 & 0 & 0 \\ 0 & 8 & 13 \\ 0 & 3 & 5 \end{bmatrix}$

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

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

9. $\begin{bmatrix} 0 & 8 & 0 \\ 0 & 0 & 4 \\ 2 & 0 & 0 \end{bmatrix}$

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

11–18 SOME GENERAL FORMULAS

11. Inverse of the square. Verify $(\mathbf{A}^2)^{-1} = (\mathbf{A}^{-1})^2$ for **A** in Prob. 1.

12. Prove the formula in Prob. 11.

13. Inverse of the transpose. Verify $(\mathbf{A}^{\mathsf{T}})^{-1} = (\mathbf{A}^{-1})^{\mathsf{T}}$ for $\mathbf{A}$ in Prob. 1.

14. Prove the formula in Prob. 13.

15. Inverse of the inverse. Prove that $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

16. Rotation. Give an application of the matrix in Prob. 2 that makes the form of the inverse obvious.

17. Triangular matrix. Is the inverse of a triangular matrix always triangular (as in Prob. 5)? Give reason.

18. Row interchange. Same task as in Prob. 16 for the matrix in Prob. 7.

19–20 FORMULA (4)

Formula (4) is occasionally needed in theory. To understand it, apply it and check the result by Gauss–Jordan:

19. In Prob. 3

20. In Prob. 6

7.9 Vector Spaces, Inner Product Spaces, Linear Transformations *Optional*

We have captured the essence of vector spaces in Sec. 7.4. There we dealt with ***special vector spaces*** that arose quite naturally in the context of matrices and linear systems. The elements of these vector spaces, called *vectors*, satisfied rules (3) and (4) of Sec. 7.1 (which were similar to those for numbers). These special vector spaces were generated by *spans,* that is, linear combination of finitely many vectors. Furthermore, each such vector had n real numbers as *components*. Review this material before going on.

We can generalize this idea by taking *all* vectors with n real numbers as components and obtain the very important ***real n-dimensional vector space*** R^n. The vectors are known as "real vectors." Thus, each vector in R^n is an ordered n-tuple of real numbers.

Now we can consider special values for n. For $n = 2$, we obtain R^2, the vector space of all ordered pairs, which correspond to the **vectors in the plane**. For $n = 3$, we obtain R^3, the vector space of all ordered triples, which are the **vectors in 3-space**. These vectors have wide applications in mechanics, geometry, and calculus and are basic to the engineer and physicist.

Similarly, if we take all ordered n-tuples of *complex numbers* as vectors and complex numbers as scalars, we obtain the **complex vector space** C^n, which we shall consider in Sec. 8.5.

Furthermore, there are other sets of practical interest consisting of matrices, functions, transformations, or others for which addition and scalar multiplication can be defined in an almost natural way so that they too form vector spaces.

It is perhaps not too great an intellectual jump to create, from the ***concrete model*** R^n, the ***abstract concept*** of a ***real vector space*** V by taking the basic properties (3) and (4) in Sec. 7.1 as axioms. In this way, the definition of a real vector space arises.

DEFINITION

Real Vector Space

A nonempty set V of elements $\mathbf{a}, \mathbf{b}, \cdots$ is called a **real vector space** (or *real linear space*), and these elements are called **vectors** (regardless of their nature, which will come out from the context or will be left arbitrary) if, in V, there are defined two algebraic operations (called *vector addition* and *scalar multiplication*) as follows.

I. Vector addition associates with every pair of vectors $\mathbf{a}$ and $\mathbf{b}$ of V a unique vector of V, called the *sum* of $\mathbf{a}$ and $\mathbf{b}$ and denoted by $\mathbf{a} + \mathbf{b}$, such that the following axioms are satisfied.

I.1 *Commutativity.* For any two vectors $\mathbf{a}$ and $\mathbf{b}$ of V,

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}.$$

I.2 *Associativity.* For any three vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ of V,

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}) \qquad \text{(written } \mathbf{a} + \mathbf{b} + \mathbf{c}\text{).}$$

I.3 There is a unique vector in V, called the *zero vector* and denoted by $\mathbf{0}$, such that for every $\mathbf{a}$ in V,

$$\mathbf{a} + \mathbf{0} = \mathbf{a}.$$

I.4 For every $\mathbf{a}$ in V there is a unique vector in V that is denoted by $-\mathbf{a}$ and is such that

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

II. Scalar multiplication. The real numbers are called **scalars**. Scalar multiplication associates with every $\mathbf{a}$ in V and every scalar c a unique vector of V, called the *product* of c and $\mathbf{a}$ and denoted by $c\mathbf{a}$ (or $\mathbf{a}c$) such that the following axioms are satisfied.

II.1 *Distributivity.* For every scalar c and vectors $\mathbf{a}$ and $\mathbf{b}$ in V,

$$c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b}.$$

II.2 *Distributivity.* For all scalars c and k and every $\mathbf{a}$ in V,

$$(c + k)\mathbf{a} = c\mathbf{a} + k\mathbf{a}.$$

II.3 *Associativity.* For all scalars c and k and every $\mathbf{a}$ in V,

$$c(k\mathbf{a}) = (ck)\mathbf{a} \qquad \text{(written } ck\mathbf{a}\text{).}$$

II.4 For every $\mathbf{a}$ in V,

$$1\mathbf{a} = \mathbf{a}.$$

If, in the above definition, we take complex numbers as scalars instead of real numbers, we obtain the axiomatic definition of a **complex vector space**.

Take a look at the axioms in the above definition. Each axiom stands on its own: It is concise, useful, and it expresses a simple property of V. There are as few axioms as possible and together they express ***all*** the desired properties of V. Selecting good axioms is a process of trial and error that often extends over a long period of time. But once agreed upon, axioms become *standard* such as the ones in the definition of a real vector space.

The following concepts related to a vector space are exactly defined as those given in Sec. 7.4. Indeed, a **linear combination** of vectors $a_{(1)}, \cdots, a_{(m)}$ in a vector space V is an expression

$$c_1\mathbf{a}_{(1)} + \cdots + c_m\mathbf{a}_m \qquad (c_1, \cdots, c_m \text{ any scalars}).$$

These vectors form a **linearly independent set** (briefly, they are called **linearly independent**) if

$$c_1\mathbf{a}_{(1)} + \cdots + c_m\mathbf{a}_{(m)} = \mathbf{0} \tag{1}$$

implies that $c_1 = 0, \cdots, c_m = 0$. Otherwise, if (1) also holds with scalars not all zero, the vectors are called **linearly dependent**.

Note that (1) with $m = 1$ is $c\mathbf{a} = \mathbf{0}$ and shows that a single vector $\mathbf{a}$ is linearly independent if and only if $\mathbf{a} \neq \mathbf{0}$.

V has **dimension n**, or is **n-dimensional**, if it contains a linearly independent set of n vectors, whereas any set of more than n vectors in V is linearly dependent. That set of n linearly independent vectors is called a **basis** for V. Then every vector in V can be written as a linear combination of the basis vectors. Furthermore, for a given basis, this representation is unique (see Prob. 2).

EXAMPLE 1 **Vector Space of Matrices**

The real 2×2 matrices form a four-dimensional real vector space. A basis is

$$\mathbf{B}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

because any 2×2 matrix $\mathbf{A} = [a_{jk}]$ has a unique representation $\mathbf{A} = a_{11}\mathbf{B}_{11} + a_{12}\mathbf{B}_{12} + a_{21}\mathbf{B}_{21} + a_{22}\mathbf{B}_{22}$. Similarly, the real $m \times n$ matrices with fixed m and n form an mn-dimensional vector space. What is the dimension of the vector space of all 3×3 skew-symmetric matrices? Can you find a basis? ■

EXAMPLE 2 **Vector Space of Polynomials**

The set of all constant, linear, and quadratic polynomials in x together is a vector space of dimension 3 with basis $\{1, x, x^2\}$ under the usual addition and multiplication by real numbers because these two operations give polynomials not exceeding degree 2. What is the dimension of the vector space of all polynomials of degree not exceeding a given fixed n? Can you find a basis? ■

If a vector space V contains a linearly independent set of n vectors for every n, no matter how large, then V is called **infinite dimensional**, as opposed to a *finite dimensional* (n-dimensional) vector space just defined. An example of an infinite dimensional vector space is the space of all continuous functions on some interval $[a, b]$ of the x-axis, as we mention without proof.

Inner Product Spaces

If $\mathbf{a}$ and $\mathbf{b}$ are vectors in R^n, regarded as column vectors, we can form the product $\mathbf{a}^\mathsf{T}\mathbf{b}$. This is a 1×1 matrix, which we can identify with its single entry, that is, with a number.

This product is called the **inner product** or **dot product** of **a** and **b**. Other notations for it are (**a**, **b**) and **a** • **b**. Thus

$$\mathbf{a}^{\mathsf{T}}\mathbf{b} = (\mathbf{a}, \mathbf{b}) = \mathbf{a} \bullet \mathbf{b} = [a_1 \cdots a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_l b_l = a_1 b_1 + \cdots + a_n b_n.$$

We now extend this concept to general real vector spaces by taking basic properties of (**a**, **b**) as axioms for an "abstract inner product" (**a**, **b**) as follows.

DEFINITION

Real Inner Product Space

A real vector space V is called a **real inner product space** (or *real pre-Hilbert*[4] *space*) if it has the following property. With every pair of vectors **a** and **b** in V there is associated a real number, which is denoted by (**a**, **b**) and is called the **inner product** of **a** and **b**, such that the following axioms are satisfied.

I. For all scalars q_1 and q_2 and all vectors **a**, **b**, **c** in V,

$$(q_1\mathbf{a} + q_2\mathbf{b}, \mathbf{c}) = q_1(\mathbf{a}, \mathbf{c}) + q_2(\mathbf{b}, \mathbf{c}) \qquad (\textit{Linearity}).$$

II. For all vectors **a** and **b** in V,

$$(\mathbf{a}, \mathbf{b}) = (\mathbf{b}, \mathbf{a}) \qquad (\textit{Symmetry}).$$

III. For every **a** in V,

$$\left.\begin{aligned} (\mathbf{a}, \mathbf{a}) &\geqq 0, \\ (\mathbf{a}, \mathbf{a}) = 0 \quad &\text{if and only if} \quad \mathbf{a} = \mathbf{0} \end{aligned}\right\} \qquad (\textit{Positive-definiteness}).$$

Vectors whose inner product is zero are called **orthogonal**.
The *length* or **norm** of a vector in V is defined by

$$\|\mathbf{a}\| = \sqrt{(\mathbf{a}, \mathbf{a})} \quad (\geqq 0). \tag{2}$$

A vector of norm 1 is called a **unit vector**.

---

[4]DAVID HILBERT (1862–1943), great German mathematician, taught at Königsberg and Göttingen and was the creator of the famous Göttingen mathematical school. He is known for his basic work in algebra, the calculus of variations, integral equations, functional analysis, and mathematical logic. His "Foundations of Geometry" helped the axiomatic method to gain general recognition. His famous 23 problems (presented in 1900 at the International Congress of Mathematicians in Paris) considerably influenced the development of modern mathematics.

If V is finite dimensional, it is actually a so-called *Hilbert space*; see [GenRef7], p. 128, listed in App. 1.

From these axioms and from (2) one can derive the basic inequality

$$|(\mathbf{a}, \mathbf{b})| \leqq \|\mathbf{a}\| \, \|\mathbf{b}\| \qquad \text{(3)} \qquad (\textit{Cauchy–Schwarz}^{5} \textit{ inequality}).$$

From this follows

$$\|\mathbf{a} + \mathbf{b}\| \leqq \|\mathbf{a}\| + \|\mathbf{b}\| \qquad \text{(4)} \qquad (\textit{Triangle inequality}).$$

A simple direct calculation gives

$$\|\mathbf{a} + \mathbf{b}\|^2 + \|\mathbf{a} - \mathbf{b}\|^2 = 2(\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2) \qquad \text{(5)} \qquad (\textit{Parallelogram equality}).$$

EXAMPLE 3 *n*-Dimensional Euclidean Space

R^n with the inner product

$$(\mathbf{a}, \mathbf{b}) = \mathbf{a}^{\mathsf{T}}\mathbf{b} = a_1 b_1 + \cdots + a_n b_n \qquad \text{(6)}$$

(where both **a** and **b** are *column* vectors) is called the ***n*-dimensional Euclidean space** and is denoted by E^n or again simply by R^n. Axioms I–III hold, as direct calculation shows. Equation (2) gives the "**Euclidean norm**"

$$\|\mathbf{a}\| = \sqrt{(\mathbf{a}, \mathbf{a})} = \sqrt{\mathbf{a}^{\mathsf{T}}\mathbf{a}} = \sqrt{a_1^2 + \cdots + a_n^2}. \qquad \text{(7)}$$ ■

EXAMPLE 4 An Inner Product for Functions. Function Space

The set of all real-valued continuous functions $f(x), g(x), \cdots$ on a given interval $\alpha \leqq x \leqq \beta$ is a real vector space under the usual addition of functions and multiplication by scalars (real numbers). On this "**function space**" we can define an inner product by the integral

$$(f, g) = \int_{\alpha}^{\beta} f(x)\, g(x)\, dx. \qquad \text{(8)}$$

Axioms I–III can be verified by direct calculation. Equation (2) gives the norm

$$\|f\| = \sqrt{(f, f)} = \sqrt{\int_{\alpha}^{\beta} f(x)^2\, dx}. \qquad \text{(9)}$$ ■

Our examples give a first impression of the great generality of the abstract concepts of vector spaces and inner product spaces. Further details belong to more advanced courses (on functional analysis, meaning abstract modern analysis; see [GenRef7] listed in App. 1) and cannot be discussed here. Instead we now take up a related topic where matrices play a central role.

Linear Transformations

Let X and Y be any vector spaces. To each vector $\mathbf{x}$ in X we assign a unique vector $\mathbf{y}$ in Y. Then we say that a **mapping** (or **transformation** or **operator**) of X into Y is given. Such a mapping is denoted by a capital letter, say F. The vector $\mathbf{y}$ in Y assigned to a vector $\mathbf{x}$ in X is called the **image** of $\mathbf{x}$ under F and is denoted by $F(x)$ [or $F\mathbf{x}$, without parentheses].

[5] HERMANN AMANDUS SCHWARZ (1843–1921). German mathematician, known by his work in complex analysis (conformal mapping) and differential geometry. For Cauchy see Sec. 2.5.

F is called a **linear mapping** or **linear transformation** if, for all vectors $\mathbf{v}$ and $\mathbf{x}$ in X and scalars c,

$$\begin{aligned} F(\mathbf{v} + \mathbf{x}) &= F(\mathbf{v}) + F(\mathbf{x}) \\ F(c\mathbf{x}) &= cF(\mathbf{x}). \end{aligned} \tag{10}$$

Linear Transformation of Space R^n into Space R^m

From now on we let $X = R^n$ and $Y = R^m$. Then any real $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ gives a transformation of R^n into R^m,

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{11}$$

Since $\mathbf{A}(\mathbf{u} + \mathbf{x}) = \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{x}$ and $\mathbf{A}(c\mathbf{x}) = c\mathbf{A}\mathbf{x}$, this transformation is linear.

We show that, conversely, every linear transformation F of R^n into R^m can be given in terms of an $m \times n$ matrix $\mathbf{A}$, after a basis for R^n and a basis for R^m have been chosen. This can be proved as follows.

Let $\mathbf{e}_{(1)}, \cdots, \mathbf{e}_{(n)}$ be any basis for R^n. Then every $\mathbf{x}$ in R^n has a unique representation

$$\mathbf{x} = x_1\mathbf{e}_{(1)} + \cdots + x_n\mathbf{e}_{(n)}.$$

Since F is linear, this representation implies for the image $F(\mathbf{x})$:

$$F(\mathbf{x}) = F(x_1\mathbf{e}_{(1)} + \cdots + x_n\mathbf{e}_{(n)}) = x_1F(\mathbf{e}_{(1)}) + \cdots + x_nF(\mathbf{e}_{(n)}).$$

Hence F is uniquely determined by the images of the vectors of a basis for R^n. We now choose for R^n the "**standard basis**"

$$\mathbf{e}_{(1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_{(2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad \mathbf{e}_{(n)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{12}$$

where $\mathbf{e}_{(j)}$ has its jth component equal to 1 and all others 0. We show that we can now determine an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ such that for every $\mathbf{x}$ in R^n and image $\mathbf{y} = F(\mathbf{x})$ in R^m,

$$\mathbf{y} = F(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

Indeed, from the image $\mathbf{y}^{(1)} = F(\mathbf{e}_{(1)})$ of $\mathbf{e}_{(1)}$ we get the condition

$$\mathbf{y}^{(1)} = \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_m^{(1)} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

from which we can determine the first column of $\mathbf{A}$, namely $a_{11} = y_1^{(1)}$, $a_{21} = y_2^{(1)}$, $\cdots$, $a_{m1} = y_m^{(1)}$. Similarly, from the image of $\mathbf{e}_{(2)}$ we get the second column of $\mathbf{A}$, and so on. This completes the proof. ■

We say that **A represents** F, or *is a representation of* F, with respect to the bases for R^n and R^m. Quite generally, the purpose of a "**representation**" is the replacement of one object of study by another object whose properties are more readily apparent.

In three-dimensional Euclidean space E^3 the standard basis is usually written $\mathbf{e}_{(1)} = \mathbf{i}$, $\mathbf{e}_{(2)} = \mathbf{j}$, $\mathbf{e}_{(3)} = \mathbf{k}$. Thus,

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{13}$$

These are the three unit vectors in the positive directions of the axes of the **Cartesian coordinate system in space**, that is, the usual coordinate system with the same scale of measurement on the three mutually perpendicular coordinate axes.

EXAMPLE 5 Linear Transformations

Interpreted as transformations of Cartesian coordinates in the plane, the matrices

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

represent a reflection in the line $x_2 = x_1$, a reflection in the x_1-axis, a reflection in the origin, and a stretch (when $a > 1$, or a contraction when $0 < a < 1$) in the x_1-direction, respectively. ■

EXAMPLE 6 Linear Transformations

Our discussion preceding Example 5 is simpler than it may look at first sight. To see this, find $\mathbf{A}$ representing the linear transformation that maps (x_1, x_2) onto $(2x_1 - 5x_2, 3x_1 + 4x_2)$.

Solution. Obviously, the transformation is

$$y_1 = 2x_1 - 5x_2$$
$$y_2 = 3x_1 + 4x_2.$$

From this we can directly see that the matrix is

$$\mathbf{A} = \begin{bmatrix} 2 & -5 \\ 3 & 4 \end{bmatrix}. \qquad \text{Check:} \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & -5 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 - 5x_2 \\ 3x_1 + 4x_2 \end{bmatrix}.$$ ■

If $\mathbf{A}$ in (11) is square, $n \times n$, then (11) maps R^n into R^n. If this $\mathbf{A}$ is nonsingular, so that $\mathbf{A}^{-1}$ exists (see Sec. 7.8), then multiplication of (11) by $\mathbf{A}^{-1}$ from the left and use of $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ gives the **inverse transformation**

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}. \tag{14}$$

It maps every $\mathbf{y} = \mathbf{y}_0$ onto that $\mathbf{x}$, which by (11) is mapped onto $\mathbf{y}_0$. *The inverse of a linear transformation is itself linear,* because it is given by a matrix, as (14) shows.

Composition of Linear Transformations

We want to give you a flavor of how linear transformations in general vector spaces work. You will notice, if you read carefully, that definitions and verifications (Example 7) strictly follow the given rules and you can think your way through the material by going in a slow systematic fashion.

The last operation we want to discuss is composition of linear transformations. Let X, Y, W be general vector spaces. As before, let F be a linear transformation from X to Y. Let G be a linear transformation from W to X. Then we denote, by H, the **composition** of F and G, that is,

$$H = F \circ G = FG = F(G),$$

which means *we take transformation G and then apply transformation F to it* (***in that order!***, *i.e. you go from left to right*).

Now, to give this a more concrete meaning, if we let $\mathbf{w}$ be a vector in W, then $G(\mathbf{w})$ is a vector in X and $F(G(\mathbf{w}))$ is a vector in Y. Thus, H maps W to Y, and we can write

$$H(\mathbf{w}) = (F \circ G)(\mathbf{w}) = (FG)(\mathbf{w}) = F(G(\mathbf{w})), \tag{15}$$

which completes the definition of composition in a general vector space setting. But is composition really linear? To check this we have to verify that H, as defined in (15), obeys the two equations of (10).

EXAMPLE 7 The Composition of Linear Transformations Is Linear

To show that H is indeed linear we must show that (10) holds. We have, for two vectors $\mathbf{w}_1$, $\mathbf{w}_2$ in W,

$$\begin{aligned} H(\mathbf{w}_1 + \mathbf{w}_2) &= (F \circ G)(\mathbf{w}_1 + \mathbf{w}_2) \\ &= F(G(\mathbf{w}_1 + \mathbf{w}_2)) \\ &= F(G(\mathbf{w}_1) + G(\mathbf{w}_2)) && \text{(by linearity of } G) \\ &= F(G(\mathbf{w}_1)) + F(G(\mathbf{w}_2)) && \text{(by linearity of } F) \\ &= (F \circ G)(\mathbf{w}_1) + (F \circ G)(\mathbf{w}_2) && \text{(by (15))} \\ &= H(\mathbf{w}_1) + H(\mathbf{w}_2) && \text{(by definition of } H). \end{aligned}$$

Similarly, $H(c\mathbf{w}_2) = (F \circ G)(c\mathbf{w}_2) = F(G(c\mathbf{w}_2)) = F(c(G(\mathbf{w}_2)))$

$$= cF(G(\mathbf{w}_2)) = c(F \circ G)(\mathbf{w}_2) = cH(\mathbf{w}_2).$$ ■

We defined composition as a linear transformation in a general vector space setting and showed that the composition of linear transformations is indeed linear.

Next we want to relate composition of linear transformations to matrix multiplication.

To do so we let $X = R^n$, $Y = R^m$, and $W = R^p$. This choice of particular vector spaces allows us to represent the linear transformations as matrices and form matrix equations, as was done in (11). Thus F can be represented by a general real $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ and G by an $n \times p$ matrix $\mathbf{B} = [b_{jk}]$. Then we can write for F, with column vectors $\mathbf{x}$ with n entries, and resulting vector $\mathbf{y}$, with m entries

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{16}$$

and similarly for G, with column vector $\mathbf{w}$ with p entries,

$$\mathbf{x} = \mathbf{B}\mathbf{w}. \tag{17}$$

Substituting (17) into (16) gives

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{B}\mathbf{w}) = (\mathbf{A}\mathbf{B})\mathbf{w} = \mathbf{A}\mathbf{B}\mathbf{w} = \mathbf{C}\mathbf{w} \qquad \text{where } \mathbf{C} = \mathbf{A}\mathbf{B}. \tag{18}$$

This is (15) in a matrix setting, this is, ***we can define the composition of linear transformations in the Euclidean spaces as multiplication by matrices***. Hence, the real $m \times p$ matrix $\mathbf{C}$ represents a linear transformation H which maps R^p to R^n with vector $\mathbf{w}$, a column vector with p entries.

Remarks. Our discussion is similar to the one in Sec. 7.2, where we motivated the "unnatural" matrix multiplication of matrices. Look back and see that our current, more general, discussion is written out there for the case of dimension $m = 2$, $n = 2$, and $p = 2$. (You may want to write out our development by picking small *distinct* dimensions, such as $m = 2$, $n = 3$, and $p = 4$, and writing down the matrices and vectors. This is a trick of the trade of mathematicians in that we like to develop and test theories on smaller examples to see that they work.)

EXAMPLE 8 Linear Transformations. Composition

In Example 5 of Sec. 7.9, let $\mathbf{A}$ be the first matrix and $\mathbf{B}$ be the fourth matrix with $a > 1$. Then, applying $\mathbf{B}$ to a vector $\mathbf{w} = [w_1 \quad w_2]^\mathsf{T}$, stretches the element w_1 by a in the x_1 direction. Next, when we apply $\mathbf{A}$ to the "stretched" vector, we reflect the vector along the line $x_1 = x_2$, resulting in a vector $\mathbf{y} = [w_2 \quad aw_1]^\mathsf{T}$. But this represents, precisely, a geometric description for the composition H of two linear transformations F and G represented by matrices $\mathbf{A}$ and $\mathbf{B}$. We now show that, for this example, our result can be obtained by straightforward matrix multiplication, that is,

$$\mathbf{AB} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ a & 0 \end{bmatrix}$$

and as in (18) calculate

$$\mathbf{ABw} = \begin{bmatrix} 0 & 1 \\ a & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} w_2 \\ aw_1 \end{bmatrix},$$

which is the same as before. This shows that indeed $\mathbf{AB} = \mathbf{C}$, and we see the composition of linear transformations can be represented by a linear transformation. It also shows that the order of matrix multiplication is important (!). You may want to try applying $\mathbf{A}$ first and then $\mathbf{B}$, resulting in $\mathbf{BA}$. What do you see? Does it make geometric sense? Is it the same result as $\mathbf{AB}$? ■

We have learned several abstract concepts such as vector space, inner product space, and linear transformation. *The introduction of such concepts allows engineers and scientists to communicate in a concise and common language.* For example, the concept of a vector space encapsulated a lot of ideas in a very concise manner. For the student, learning such concepts provides a foundation for more advanced studies in engineering.

This concludes Chapter 7. The central theme was *the Gaussian elimination of* Sec. 7.3 from which most of the other concepts and theory flowed. The next chapter again has a central theme, that is, *eigenvalue problems,* an area very rich in applications such as in engineering, modern physics, and other areas.

PROBLEM SET 7.9

1. **Basis.** Find three bases of R^2.
2. **Uniqueness.** Show that the representation $\mathbf{v} = c_1\mathbf{a}_{(1)} + \cdots + c_n\mathbf{a}_{(n)}$ of any given vector in an n-dimensional vector space V in terms of a given basis $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(n)}$ for V is unique. ***Hint***. Take two representations and consider the difference.

3–10 VECTOR SPACE

(More problems in Problem Set 9.4.) Is the given set, taken with the usual addition and scalar multiplication, a vector space? Give reason. If your answer is yes, find the dimension and a basis.

3. All vectors in R^3 satisfying $-v_1 + 2v_2 + 3v_3 = 0$, $-4v_1 + v_2 + v_3 = 0$.
4. All skew-symmetric 3×3 matrices.
5. All polynomials in x of degree 4 or less with nonnegative coefficients.
6. All functions $y(x) = a \cos 2x + b \sin 2x$ with arbitrary constants a and b.
7. All functions $y(x) = (ax + b)e^{-x}$ with any constant a and b.
8. All $n \times n$ matrices $\mathbf{A}$ with fixed n and $\det \mathbf{A} = 0$.
9. All 2×2 matrices $[a_{jk}]$ with $a_{11} + a_{22} = 0$.
10. All 3×2 matrices $[a_{jk}]$ with first column any multiple of $[3 \quad 0 \quad -5]^\mathsf{T}$.

11–14 LINEAR TRANSFORMATIONS

Find the inverse transformation. Show the details.

11. $y_1 = 0.5x_1 - 0.5x_2$
$y_2 = 1.5x_1 - 2.5x_2$

12. $y_1 = 3x_1 + 2x_2$
$y_2 = 4x_1 + x_2$

13. $y_1 = 5x_1 + 3x_2 - 3x_3$
$y_2 = 3x_1 + 2x_2 - 2x_3$
$y_3 = 2x_1 - x_2 + 2x_3$

14. $y_1 = 0.2x_1 - 0.1x_2$
$y_2 = -0.2x_2 + 0.1x_3$
$y_3 = 0.1x_1 + 0.1x_3$

15–20 EUCLIDEAN NORM

Find the Euclidean norm of the vectors:

15. $[3 \quad 1 \quad -4]^\mathsf{T}$
16. $[\frac{1}{2} \quad \frac{1}{3} \quad -\frac{1}{2} \quad -\frac{1}{3}]^\mathsf{T}$
17. $[1 \quad 0 \quad 0 \quad 1 \quad -1 \quad 0 \quad -1 \quad 1]^\mathsf{T}$
18. $[-4 \quad 8 \quad -1]^\mathsf{T}$
19. $[\frac{2}{3} \quad \frac{2}{3} \quad \frac{1}{3} \quad 0]^\mathsf{T}$
20. $[\frac{1}{2} \quad -\frac{1}{2} \quad -\frac{1}{2} \quad \frac{1}{2}]^\mathsf{T}$

21–25 INNER PRODUCT. ORTHOGONALITY

21. **Orthogonality.** For what value(s) of k are the vectors $[2 \quad \frac{1}{2} \quad -4 \quad 0]^\mathsf{T}$ and $[5 \quad k \quad 0 \quad \frac{1}{4}]^\mathsf{T}$ orthogonal?
22. **Orthogonality.** Find all vectors in R^3 orthogonal to $[2 \quad 0 \quad 1]$. Do they form a vector space?
23. **Triangle inequality.** Verify (4) for the vectors in Probs. 15 and 18.
24. **Cauchy–Schwarz inequality.** Verify (3) for the vectors in Probs. 16 and 19.
25. **Parallelogram equality.** Verify (5) for the first two column vectors of the coefficient matrix in Prob. 13.

CHAPTER 7 REVIEW QUESTIONS AND PROBLEMS

1. What properties of matrix multiplication differ from those of the multiplication of numbers?
2. Let $\mathbf{A}$ be a 100×100 matrix and $\mathbf{B}$ a 100×50 matrix. Are the following expressions defined or not? $\mathbf{A} + \mathbf{B}$, $\mathbf{A}^2$, $\mathbf{B}^2$, $\mathbf{AB}$, $\mathbf{BA}$, $\mathbf{AA}^\mathsf{T}$, $\mathbf{B}^\mathsf{T}\mathbf{A}$, $\mathbf{B}^\mathsf{T}\mathbf{B}$, $\mathbf{BB}^\mathsf{T}$, $\mathbf{B}^\mathsf{T}\mathbf{AB}$. Give reasons.
3. Are there any linear systems without solutions? With one solution? With more than one solution? Give simple examples.
4. Let $\mathbf{C}$ be 10×10 matrix and $\mathbf{a}$ a column vector with 10 components. Are the following expressions defined or not? $\mathbf{Ca}$, $\mathbf{C}^\mathsf{T}\mathbf{a}$, $\mathbf{Ca}^\mathsf{T}$, $\mathbf{aC}$, $\mathbf{a}^\mathsf{T}\mathbf{C}$, $(\mathbf{Ca}^\mathsf{T})^\mathsf{T}$.
5. Motivate the definition of matrix multiplication.
6. Explain the use of matrices in linear transformations.
7. How can you give the rank of a matrix in terms of row vectors? Of column vectors? Of determinants?
8. What is the role of rank in connection with solving linear systems?
9. What is the idea of Gauss elimination and back substitution?
10. What is the inverse of a matrix? When does it exist? How would you determine it?

11–20 MATRIX AND VECTOR CALCULATIONS

Showing the details, calculate the following expressions or give reason why they are not defined, when

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

$$\mathbf{u} = \begin{bmatrix} 2 \\ 0 \\ -5 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 7 \\ -3 \\ 3 \end{bmatrix}$$

11. $\mathbf{AB}$, $\mathbf{BA}$

12. $\mathbf{A}^{\mathsf{T}}$, $\mathbf{B}^{\mathsf{T}}$

13. $\mathbf{Au}$, $\mathbf{u}^{\mathsf{T}}\mathbf{A}$

14. $\mathbf{u}^{\mathsf{T}}\mathbf{v}$, $\mathbf{uv}^{\mathsf{T}}$

15. $\mathbf{u}^{\mathsf{T}}\mathbf{Au}$, $\mathbf{v}^{\mathsf{T}}\mathbf{Bv}$

16. $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$

17. $\det \mathbf{A}$, $\det \mathbf{A}^2$, $(\det \mathbf{A})^2$, $\det \mathbf{B}$

18. $(\mathbf{A}^2)^{-1}$, $(\mathbf{A}^{-1})^2$

19. $\mathbf{AB} - \mathbf{BA}$

20. $(\mathbf{A} + \mathbf{A}^{\mathsf{T}})(\mathbf{B} - \mathbf{B}^{\mathsf{T}})$

21–28 LINEAR SYSTEMS

Showing the details, find all solutions or indicate that no solution exists.

21.
$$\begin{aligned} 4y + z &= 0 \\ 12x - 5y - 3z &= 34 \\ -6x \qquad + 4z &= 8 \end{aligned}$$

22.
$$\begin{aligned} 5x - 3y + z &= 7 \\ 2x + 3y - z &= 0 \\ 8x + 9y - 3z &= 2 \end{aligned}$$

23.
$$\begin{aligned} 9x + 3y - 6z &= 60 \\ 2x - 4y + 8z &= 4 \end{aligned}$$

24.
$$\begin{aligned} -6x + 39y - 9z &= -12 \\ 2x - 13y + 3z &= 4 \end{aligned}$$

25.
$$\begin{aligned} 0.3x - 0.7y + 1.3z &= 3.24 \\ 0.9y - 0.8z &= -2.53 \\ 0.7z &= 1.19 \end{aligned}$$

26.
$$\begin{aligned} 2x + 3y - 7z &= 3 \\ -4x - 6y + 14z &= 7 \end{aligned}$$

27.
$$\begin{aligned} x + 2y &= 6 \\ 3x + 5y &= 20 \\ -4x + y &= -42 \end{aligned}$$

28.
$$\begin{aligned} -8x \qquad + 2z &= 1 \\ 6y + 4z &= 3 \\ 12x + 2y \qquad &= 2 \end{aligned}$$

29–32 RANK

Determine the ranks of the coefficient matrix and the augmented matrix and state how many solutions the linear system will have.

29. In Prob. 23

30. In Prob. 24

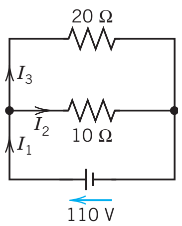
31. In Prob. 27

32. In Prob. 26

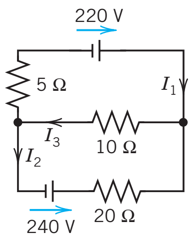
33–35 NETWORKS

Find the currents.

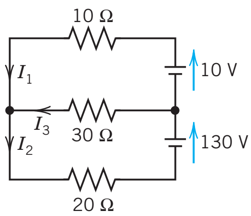
33.



34.



35.



SUMMARY OF CHAPTER 7

Linear Algebra: Matrices, Vectors, Determinants. Linear Systems

An $m \times n$ **matrix** $\mathbf{A} = [a_{jk}]$ is a rectangular array of numbers or functions ("entries," "elements") arranged in m horizontal **rows** and n vertical **columns**. If $m = n$, the matrix is called **square**. A $1 \times n$ matrix is called a **row vector** and an $m \times 1$ matrix a **column vector** (Sec. 7.1).

The **sum** $\mathbf{A} + \mathbf{B}$ of matrices of the same **size** (i.e., both $m \times n$) is obtained by adding corresponding entries. The **product** of $\mathbf{A}$ by a scalar c is obtained by multiplying each a_{jk} by c (Sec. 7.1).

The **product** $\mathbf{C} = \mathbf{AB}$ of an $m \times n$ matrix $\mathbf{A}$ by an $r \times p$ matrix $\mathbf{B} = [b_{jk}]$ is defined only when $r = n$, and is the $m \times p$ matrix $\mathbf{C} = [c_{jk}]$ with entries

$$c_{jk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk} \qquad \text{(row } j \text{ of } \mathbf{A} \text{ times column } k \text{ of } \mathbf{B}\text{).} \tag{1}$$

This multiplication is motivated by the composition of **linear transformations** (Secs. 7.2, 7.9). It is associative, but is ***not commutative:*** if $\mathbf{AB}$ is defined, $\mathbf{BA}$ may not be defined, but even if $\mathbf{BA}$ is defined, $\mathbf{AB} \neq \mathbf{BA}$ in general. Also $\mathbf{AB} = \mathbf{0}$ may not imply $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$ or $\mathbf{BA} = \mathbf{0}$ (Secs. 7.2, 7.8). Illustrations:

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

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

$$[1 \quad 2]\begin{bmatrix} 3 \\ 4 \end{bmatrix} = [11], \qquad \begin{bmatrix} 3 \\ 4 \end{bmatrix}[1 \quad 2] = \begin{bmatrix} 3 & 6 \\ 4 & 8 \end{bmatrix}.$$

The **transpose** $\mathbf{A}^\mathsf{T}$ of a matrix $\mathbf{A} = [a_{jk}]$ is $\mathbf{A}^\mathsf{T} = [a_{kj}]$; rows become columns and conversely (Sec. 7.2). Here, $\mathbf{A}$ need not be square. If it is and $\mathbf{A} = \mathbf{A}^\mathsf{T}$, then $\mathbf{A}$ is called **symmetric;** if $\mathbf{A} = -\mathbf{A}^\mathsf{T}$, it is called **skew-symmetric**. For a product, $(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$ (Sec. 7.2).

A main application of matrices concerns **linear systems of equations**

$$\mathbf{Ax} = \mathbf{b} \qquad \text{(Sec. 7.3)} \tag{2}$$

(m equations in n unknowns $x_1, \cdots, x_n$; $\mathbf{A}$ and $\mathbf{b}$ given). The most important method of solution is the **Gauss elimination** (Sec. 7.3), which reduces the system to "triangular" form by *elementary row operations,* which leave the set of solutions unchanged. (Numeric aspects and variants, such as *Doolittle's* and *Cholesky's methods,* are discussed in Secs. 20.1 and 20.2.)

Cramer's rule (Secs. 7.6, 7.7) represents the unknowns in a system (2) of n equations in n unknowns as quotients of determinants; for numeric work it is impractical. **Determinants** (Sec. 7.7) have decreased in importance, but will retain their place in eigenvalue problems, elementary geometry, etc.

The **inverse** $\mathbf{A}^{-1}$ of a square matrix satisfies $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. It exists if and only if $\det \mathbf{A} \neq 0$. It can be computed by the *Gauss–Jordan elimination* (Sec. 7.8).

The **rank** r of a matrix $\mathbf{A}$ is the maximum number of linearly independent rows or columns of $\mathbf{A}$ or, equivalently, the number of rows of the largest square submatrix of $\mathbf{A}$ with nonzero determinant (Secs. 7.4, 7.7).

The system (2) has solutions if and only if rank $\mathbf{A}$ = rank $[\mathbf{A} \quad \mathbf{b}]$, where $[\mathbf{A} \quad \mathbf{b}]$ is the **augmented matrix** (Fundamental Theorem, Sec. 7.5).

The **homogeneous system**

$$\mathbf{A}\mathbf{x} = \mathbf{0} \tag{3}$$

has solutions $\mathbf{x} \neq \mathbf{0}$ ("nontrivial solutions") if and only if rank $\mathbf{A} < n$, in the case $m = n$ equivalently if and only if $\det \mathbf{A} = \mathbf{0}$ (Secs. 7.6, 7.7).

Vector spaces, inner product spaces, and linear transformations are discussed in Sec. 7.9. See also Sec. 7.4.

CHAPTER 8

Linear Algebra: Matrix Eigenvalue Problems

A matrix eigenvalue problem considers the vector equation

$$\mathbf{Ax} = \lambda\mathbf{x}. \tag{1}$$

Here $\mathbf{A}$ is a given square matrix, λ an unknown scalar, and $\mathbf{x}$ an unknown vector. In a matrix eigenvalue problem, the task is to determine λ's and $\mathbf{x}$'s that satisfy (1). Since $\mathbf{x} = \mathbf{0}$ is always a solution for any λ and thus not interesting, we only admit solutions with $\mathbf{x} \neq \mathbf{0}$.

The solutions to (1) are given the following names: The λ's that satisfy (1) are called **eigenvalues of A** and the corresponding nonzero $\mathbf{x}$'s that also satisfy (1) are called **eigenvectors of A**.

From this rather innocent looking vector equation flows an amazing amount of relevant theory and an incredible richness of applications. Indeed, eigenvalue problems come up all the time in engineering, physics, geometry, numerics, theoretical mathematics, biology, environmental science, urban planning, economics, psychology, and other areas. Thus, in your career you are likely to encounter eigenvalue problems.

We start with a basic and thorough introduction to eigenvalue problems in Sec. 8.1 and explain (1) with several simple matrices. This is followed by a section devoted entirely to applications ranging from mass–spring systems of physics to population control models of environmental science. We show you these diverse examples to train your skills in modeling and solving eigenvalue problems. Eigenvalue problems for real symmetric, skew-symmetric, and orthogonal matrices are discussed in Sec. 8.3 and their complex counterparts (which are important in modern physics) in Sec. 8.5. In Sec. 8.4 we show how by diagonalizing a matrix, we obtain its eigenvalues.

COMMENT. ***Numerics for eigenvalues* (Secs. 20.6–20.9) *can be studied immediately after this chapter*.**

Prerequisite: Chap. 7.
Sections that may be omitted in a shorter course: 8.4, 8.5.
References and Answers to Problems: App. 1 Part B, App. 2.

The following chart identifies where different types of eigenvalue problems appear in the book.

Topic	Where to find it
Matrix Eigenvalue Problem (algebraic eigenvalue problem)	Chap. 8
Eigenvalue Problems in Numerics	Secs. 20.6–20.9
Eigenvalue Problem for ODEs (Sturm–Liouville problems)	Secs. 11.5, 11.6
Eigenvalue Problems for Systems of ODEs	Chap. 4
Eigenvalue Problems for PDEs	Secs. 12.3–12.11

8.1 The Matrix Eigenvalue Problem. Determining Eigenvalues and Eigenvectors

Consider multiplying nonzero vectors by a given square matrix, such as

$$\begin{bmatrix} 6 & 3 \\ 4 & 7 \end{bmatrix}\begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 33 \\ 27 \end{bmatrix}, \qquad \begin{bmatrix} 6 & 3 \\ 4 & 7 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 30 \\ 40 \end{bmatrix}.$$

We want to see what influence the multiplication of the given matrix has on the vectors. In the first case, we get a totally new vector with a different direction and different length when compared to the original vector. This is what usually happens and is of no interest here. In the second case something interesting happens. The multiplication produces a vector $[30 \quad 40]^\mathsf{T} = 10\,[3 \quad 4]^\mathsf{T}$, which means the new vector has the same direction as the original vector. The scale constant, which we denote by λ is 10. *The problem of systematically finding such λ's and nonzero vectors for a given square matrix will be the theme of this chapter*. It is called the *matrix eigenvalue* problem or, more commonly, the *eigenvalue* problem.

We formalize our observation. Let $\mathbf{A} = [a_{jk}]$ be a given nonzero square matrix of dimension $n \times n$. Consider the following vector equation:

$$\mathbf{Ax} = \lambda\mathbf{x}. \tag{1}$$

The problem of finding nonzero $\mathbf{x}$'s and λ's that satisfy equation (1) is called an eigenvalue problem.

Remark. So $\mathbf{A}$ is a given square (!) matrix, $\mathbf{x}$ is an unknown vector, and λ is an unknown scalar. Our task is to find λ's and nonzero $\mathbf{x}$'s that satisfy (1). Geometrically, we are looking for vectors, $\mathbf{x}$, for which the multiplication by $\mathbf{A}$ has the same effect as the multiplication by a scalar λ; in other words, $\mathbf{Ax}$ should be proportional to $\mathbf{x}$. Thus, the multiplication has the effect of producing, from the original vector $\mathbf{x}$, a new vector $\lambda\mathbf{x}$ that has the same or opposite (minus sign) direction as the original vector. (This was all demonstrated in our intuitive opening example. Can you see that the second equation in that example satisfies (1) with $\lambda = 10$ and $\mathbf{x} = [3 \quad 4]^\mathsf{T}$, and $\mathbf{A}$ the given 2×2 matrix? Write it out.) Now why do we require $\mathbf{x}$ to be nonzero? The reason is that $\mathbf{x} = \mathbf{0}$ is always a solution of (1) for any value of λ, because $\mathbf{A0} = \mathbf{0}$. This is of no interest.

We introduce more terminology. A value of λ, for which (1) has a solution $\mathbf{x} \neq \mathbf{0}$, is called an **eigenvalue** or *characteristic value* of the matrix $\mathbf{A}$. Another term for λ is a *latent* root. ("Eigen" is German and means "proper" or "characteristic."). The corresponding solutions $\mathbf{x} \neq \mathbf{0}$ of (1) are called the **eigenvectors** or *characteristic vectors* of $\mathbf{A}$ corresponding to that eigenvalue λ. The set of all the eigenvalues of $\mathbf{A}$ is called the **spectrum** of $\mathbf{A}$. We shall see that the spectrum consists of at least one eigenvalue and at most of n numerically different eigenvalues. The largest of the absolute values of the eigenvalues of $\mathbf{A}$ is called the *spectral radius* of $\mathbf{A}$, a name to be motivated later.

How to Find Eigenvalues and Eigenvectors

Now, with the new terminology for (1), we can just say that the problem of determining the eigenvalues and eigenvectors of a matrix is called an eigenvalue problem. (However, more precisely, we are considering an algebraic eigenvalue problem, as opposed to an eigenvalue problem involving an ODE or PDE, as considered in Secs. 11.5 and 12.3, or an integral equation.)

Eigenvalues have a very large number of applications in diverse fields such as in engineering, geometry, physics, mathematics, biology, environmental science, economics, psychology, and other areas. You will encounter applications for elastic membranes, Markov processes, population models, and others in this chapter.

Since, from the viewpoint of engineering applications, eigenvalue problems are the most important problems in connection with matrices, the student should carefully follow our discussion.

Example 1 demonstrates how to systematically solve a simple eigenvalue problem.

EXAMPLE 1 Determination of Eigenvalues and Eigenvectors

We illustrate all the steps in terms of the matrix

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

Solution. **(a)** ***Eigenvalues.*** These must be determined ***first***. Equation (1) is

$$\mathbf{Ax} = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \qquad \text{in components,} \qquad \begin{aligned} -5x_1 + 2x_2 &= \lambda x_1 \\ 2x_1 - 2x_2 &= \lambda x_2. \end{aligned}$$

Transferring the terms on the right to the left, we get

$$\begin{aligned} (-5 - \lambda)x_1 + \quad 2x_2 \quad &= 0 \\ 2x_1 \quad + (-2 - \lambda)x_2 &= 0. \end{aligned} \tag{2*}$$

This can be written in matrix notation

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \tag{3*}$$

because (1) is $\mathbf{Ax} - \lambda\mathbf{x} = \mathbf{Ax} - \lambda\mathbf{Ix} = (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$, which gives (3*). We see that this is a ***homogeneous*** linear system. By Cramer's theorem in Sec. 7.7 it has a nontrivial solution $\mathbf{x} \neq \mathbf{0}$ (an eigenvector of $\mathbf{A}$ we are looking for) if and only if its coefficient determinant is zero, that is,

$$D(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} -5 - \lambda & 2 \\ 2 & -2 - \lambda \end{vmatrix} = (-5 - \lambda)(-2 - \lambda) - 4 = \lambda^2 + 7\lambda + 6 = 0. \tag{4*}$$

We call $D(\lambda)$ the **characteristic determinant** or, if expanded, the **characteristic polynomial**, and $D(\lambda) = 0$ the **characteristic equation** of $\mathbf{A}$. The solutions of this quadratic equation are $\lambda_1 = -1$ and $\lambda_2 = -6$. These are the eigenvalues of $\mathbf{A}$.

(b_1) ***Eigenvector of* A *corresponding to*** λ_1. This vector is obtained from (2*) with $\lambda = \lambda_1 = -1$, that is,

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

A solution is $x_2 = 2x_1$, as we see from either of the two equations, so that we need only one of them. This determines an eigenvector corresponding to $\lambda_1 = -1$ up to a scalar multiple. If we choose $x_1 = 1$, we obtain the eigenvector

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \text{Check:} \qquad \mathbf{A}\mathbf{x}_1 = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix} = (-1)\mathbf{x}_1 = \lambda_1\mathbf{x}_1.$$

(b_2) ***Eigenvector of* A *corresponding to*** λ_2. For $\lambda = \lambda_2 = -6$, equation (2*) becomes

$$\begin{aligned} x_1 + 2x_2 &= 0 \\ 2x_1 + 4x_2 &= 0. \end{aligned}$$

A solution is $x_2 = -x_1/2$ with arbitrary x_1. If we choose $x_1 = 2$, we get $x_2 = -1$. Thus an eigenvector of $\mathbf{A}$ corresponding to $\lambda_2 = -6$ is

$$\mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \qquad \text{Check:} \qquad \mathbf{A}\mathbf{x}_2 = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -12 \\ 6 \end{bmatrix} = (-6)\mathbf{x}_2 = \lambda_2\mathbf{x}_2.$$

For the matrix in the intuitive opening example at the start of Sec. 8.1, the characteristic equation is $\lambda^2 - 13\lambda + 30 = (\lambda - 10)(\lambda - 3) = 0$. The eigenvalues are $\{10, \ 3\}$. Corresponding eigenvectors are $[3 \quad 4]^T$ and $[-1 \quad 1]^T$, respectively. The reader may want to verify this. ■

This example illustrates the general case as follows. Equation (1) written in components is

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= \lambda x_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= \lambda x_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots& \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= \lambda x_n. \end{aligned}$$

Transferring the terms on the right side to the left side, we have

$$\text{(2)} \qquad \begin{array}{ccccccc} (a_{11} - \lambda)x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & (a_{22} - \lambda)x_2 & + \cdots + & a_{2n}x_n & = & 0 \\ \cdots\cdots & & \cdots\cdots & & \cdots\cdots & & \\ a_{n1}x_1 & + & a_{n2}x_2 & + \cdots + & (a_{nn} - \lambda)x_n & = & 0. \end{array}$$

In matrix notation,

$$\textbf{(3)} \qquad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}.$$

By Cramer's theorem in Sec. 7.7, this homogeneous linear system of equations has a nontrivial solution if and only if the corresponding determinant of the coefficients is zero:

$$D(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{4}$$

$\mathbf{A} - \lambda\mathbf{I}$ is called the **characteristic matrix** and $D(\lambda)$ the **characteristic determinant** of $\mathbf{A}$. Equation (4) is called the **characteristic equation** of $\mathbf{A}$. By developing $D(\lambda)$ we obtain a polynomial of nth degree in λ. This is called the **characteristic polynomial** of $\mathbf{A}$.

This proves the following important theorem.

THEOREM 1

Eigenvalues

The eigenvalues of a square matrix $\mathbf{A}$ *are the roots of the characteristic equation* (4) *of* $\mathbf{A}$.

Hence an $n \times n$ *matrix has at least one eigenvalue and at most* n *numerically different eigenvalues.*

For larger n, the actual computation of eigenvalues will, in general, require the use of Newton's method (Sec. 19.2) or another numeric approximation method in Secs. 20.7–20.9.

The eigenvalues must be determined first. Once these are known, corresponding eigen*vectors* are obtained from the system (2), for instance, by the Gauss elimination, where λ is the eigenvalue for which an eigenvector is wanted. This is what we did in Example 1 and shall do again in the examples below. (To prevent misunderstandings: *numeric approximation methods*, such as in Sec. 20.8, may determine eigen*vectors first*.)

Eigenvectors have the following properties.

THEOREM 2

Eigenvectors, Eigenspace

If $\mathbf{w}$ *and* $\mathbf{x}$ *are eigenvectors of a matrix* $\mathbf{A}$ *corresponding to* ***the same*** *eigenvalue* λ, *so are* $\mathbf{w} + \mathbf{x}$ (provided $\mathbf{x} \neq -\mathbf{w}$) *and* $k\mathbf{x}$ *for any* $k \neq 0$.

Hence the eigenvectors corresponding to one and the same eigenvalue λ *of* $\mathbf{A}$, *together with* $\mathbf{0}$, *form a vector space* (cf. Sec. 7.4), *called the* **eigenspace** *of* $\mathbf{A}$ *corresponding to that* λ.

PROOF $\mathbf{Aw} = \lambda\mathbf{w}$ and $\mathbf{Ax} = \lambda\mathbf{x}$ *imply* $\mathbf{A}(\mathbf{w} + \mathbf{x}) = \mathbf{Aw} + \mathbf{Ax} = \lambda\mathbf{w} + \lambda\mathbf{x} = \lambda(\mathbf{w} + \mathbf{x})$ and $\mathbf{A}(k\mathbf{w}) = k(\mathbf{Aw}) = k(\lambda\mathbf{w}) = \lambda(k\mathbf{w})$; hence $\mathbf{A}(k\mathbf{w} + \ell\mathbf{x}) = \lambda(k\mathbf{w} + \ell\mathbf{x})$. ■

In particular, *an eigenvector* $\mathbf{x}$ *is determined only up to a constant factor.* Hence we can **normalize** $\mathbf{x}$, that is, multiply it by a scalar to get a unit vector (see Sec. 7.9). For instance, $\mathbf{x}_1 = [1 \quad 2]^\mathsf{T}$ in Example 1 has the length $\|\mathbf{x}_1\| = \sqrt{1^2 + 2^2} = \sqrt{5}$; hence $[1/\sqrt{5} \quad 2/\sqrt{5}]^\mathsf{T}$ is a normalized eigenvector (a unit eigenvector).

Examples 2 and 3 will illustrate that an $n \times n$ matrix may have n linearly independent eigenvectors, or it may have fewer than n. In Example 4 we shall see that a *real* matrix may have *complex* eigenvalues and eigenvectors.

EXAMPLE 2 **Multiple Eigenvalues**

Find the eigenvalues and eigenvectors of

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

Solution. For our matrix, the characteristic determinant gives the characteristic equation

$$-\lambda^3 - \lambda^2 + 21\lambda + 45 = 0.$$

The roots (eigenvalues of $\mathbf{A}$) are $\lambda_1 = 5$, $\lambda_2 = \lambda_3 = -3$. (If you have trouble finding roots, you may want to use a root finding algorithm such as Newton's method (Sec. 19.2). Your CAS or scientific calculator can find roots. However, to really learn and remember this material, you have to do some exercises with paper and pencil.) To find eigenvectors, we apply the Gauss elimination (Sec. 7.3) to the system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$, first with $\lambda = 5$ and then with $\lambda = -3$. For $\lambda = 5$ the characteristic matrix is

$$\mathbf{A} - \lambda\mathbf{I} = \mathbf{A} - 5\mathbf{I} = \begin{bmatrix} -7 & 2 & -3 \\ 2 & -4 & -6 \\ -1 & -2 & -5 \end{bmatrix}. \qquad \text{It row-reduces to} \qquad \begin{bmatrix} -7 & 2 & -3 \\ 0 & -\frac{24}{7} & -\frac{48}{7} \\ 0 & 0 & 0 \end{bmatrix}.$$

Hence it has rank 2. Choosing $x_3 = -1$ we have $x_2 = 2$ from $-\frac{24}{7}x_2 - \frac{48}{7}x_3 = 0$ and then $x_1 = 1$ from $-7x_1 + 2x_2 - 3x_3 = 0$. Hence an eigenvector of $\mathbf{A}$ corresponding to $\lambda = 5$ is $\mathbf{x}_1 = [1 \quad 2 \quad -1]^\mathsf{T}$.

For $\lambda = -3$ the characteristic matrix

$$\mathbf{A} - \lambda\mathbf{I} = \mathbf{A} + 3\mathbf{I} = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{bmatrix} \qquad \text{row-reduces to} \qquad \begin{bmatrix} 1 & 2 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Hence it has rank 1. From $x_1 + 2x_2 - 3x_3 = 0$ we have $x_1 = -2x_2 + 3x_3$. Choosing $x_2 = 1, x_3 = 0$ and $x_2 = 0, x_3 = 1$, we obtain two linearly independent eigenvectors of $\mathbf{A}$ corresponding to $\lambda = -3$ [as they must exist by (5), Sec. 7.5, with rank $= 1$ and $n = 3$],

$$\mathbf{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

and

$$\mathbf{x}_3 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}. \qquad \blacksquare$$

The order M_λ of an eigenvalue λ as a root of the characteristic polynomial is called the **algebraic multiplicity** of λ. The number m_λ of linearly independent eigenvectors corresponding to λ is called the **geometric multiplicity** of λ. Thus m_λ is the dimension of the eigenspace corresponding to this λ.

Since the characteristic polynomial has degree n, the sum of all the algebraic multiplicities must equal n. In Example 2 for $\lambda = -3$ we have $m_\lambda = M_\lambda = 2$. In general, $m_\lambda \leqq M_\lambda$, as can be shown. The difference $\Delta_\lambda = M_\lambda - m_\lambda$ is called the **defect** of λ. Thus $\Delta_{-3} = 0$ in Example 2, but positive defects Δ_λ can easily occur:

EXAMPLE 3 Algebraic Multiplicity, Geometric Multiplicity. Positive Defect

The characteristic equation of the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{is} \qquad \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & 1 \\ 0 & -\lambda \end{vmatrix} = \lambda^2 = 0.$$

Hence $\lambda = 0$ is an eigenvalue of algebraic multiplicity $M_0 = 2$. But its geometric multiplicity is only $m_0 = 1$, since eigenvectors result from $-0x_1 + x_2 = 0$, hence $x_2 = 0$, in the form $[x_1 \quad 0]^\mathsf{T}$. Hence for $\lambda = 0$ the defect is $\Delta_0 = 1$.

Similarly, the characteristic equation of the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 0 & 3 \end{bmatrix} \qquad \text{is} \qquad \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 3-\lambda & 2 \\ 0 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 = 0.$$

Hence $\lambda = 3$ is an eigenvalue of algebraic multiplicity $M_3 = 2$, but its geometric multiplicity is only $m_3 = 1$, since eigenvectors result from $0x_1 + 2x_2 = 0$ in the form $[x_1 \quad 0]^\mathsf{T}$. ■

EXAMPLE 4 Real Matrices with Complex Eigenvalues and Eigenvectors

Since real polynomials may have complex roots (which then occur in conjugate pairs), a real matrix may have complex eigenvalues and eigenvectors. For instance, the characteristic equation of the skew-symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad \text{is} \qquad \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0.$$

It gives the eigenvalues $\lambda_1 = i\ (= \sqrt{-1})$, $\lambda_2 = -i$. Eigenvectors are obtained from $-ix_1 + x_2 = 0$ and $ix_1 + x_2 = 0$, respectively, and we can choose $x_1 = 1$ to get

$$\begin{bmatrix} 1 \\ i \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$ ■

In the next section we shall need the following simple theorem.

THEOREM 3 Eigenvalues of the Transpose

The transpose $\mathbf{A}^\mathsf{T}$ *of a square matrix* $\mathbf{A}$ *has the same eigenvalues as* $\mathbf{A}$.

PROOF Transposition does not change the value of the characteristic determinant, as follows from Theorem 2d in Sec. 7.7. ■

Having gained a first impression of matrix eigenvalue problems, we shall illustrate their importance with some typical applications in Sec. 8.2.

PROBLEM SET 8.1

1–16 EIGENVALUES, EIGENVECTORS

Find the eigenvalues. Find the corresponding eigenvectors. Use the given λ or factor in Probs. 11 and 15.

1. $\begin{bmatrix} 3.0 & 0 \\ 0 & -0.6 \end{bmatrix}$

2. $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

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

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

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

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

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

8. $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$

9. $\begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}$

10. $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

11. $\begin{bmatrix} 6 & 2 & -2 \\ 2 & 5 & 0 \\ -2 & 0 & 7 \end{bmatrix}, \quad \lambda = 3$

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

13. $\begin{bmatrix} 13 & 5 & 2 \\ 2 & 7 & -8 \\ 5 & 4 & 7 \end{bmatrix}$

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

15. $\begin{bmatrix} -1 & 0 & 12 & 0 \\ 0 & -1 & 0 & 12 \\ 0 & 0 & -1 & -4 \\ 0 & 0 & -4 & -1 \end{bmatrix}, \quad (\lambda + 1)^2$

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

17–20 LINEAR TRANSFORMATIONS AND EIGENVALUES

Find the matrix $\mathbf{A}$ in the linear transformation $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{x} = [x_1 \;\; x_2]^\mathsf{T}$ ($\mathbf{x} = [x_1 \;\; x_2 \;\; x_3]^\mathsf{T}$) are Cartesian coordinates. Find the eigenvalues and eigenvectors and explain their geometric meaning.

17. Counterclockwise rotation through the angle $\pi/2$ about the origin in R^2.

18. Reflection about the x_1-axis in R^2.

19. Orthogonal projection (perpendicular projection) of R^2 onto the x_2-axis.

20. Orthogonal projection of R^3 onto the plane $x_2 = x_1$.

21–25 GENERAL PROBLEMS

21. Nonzero defect. Find further 2×2 and 3×3 matrices with positive defect. See Example 3.

22. Multiple eigenvalues. Find further 2×2 and 3×3 matrices with multiple eigenvalues. See Example 2.

23. Complex eigenvalues. Show that the eigenvalues of a real matrix are real or complex conjugate in pairs.

24. Inverse matrix. Show that $\mathbf{A}^{-1}$ exists if and only if the eigenvalues $\lambda_1, \cdots, \lambda_n$ are all nonzero, and then $\mathbf{A}^{-1}$ has the eigenvalues $1/\lambda_1, \cdots, 1/\lambda_n$.

25. Transpose. Illustrate Theorem 3 with examples of your own.

8.2 Some Applications of Eigenvalue Problems

We have selected some typical examples from the wide range of applications of matrix eigenvalue problems. The last example, that is, Example 4, shows an application involving vibrating springs and ODEs. It falls into the domain of Chapter 4, which covers matrix eigenvalue problems related to ODE's modeling mechanical systems and electrical networks. Example 4 is included to keep our discussion independent of Chapter 4. (However, the reader not interested in ODEs may want to skip Example 4 without loss of continuity.)

EXAMPLE 1 **Stretching of an Elastic Membrane**

An elastic membrane in the x_1x_2-plane with boundary circle $x_1^2 + x_2^2 = 1$ (Fig. 160) is stretched so that a point P: (x_1, x_2) goes over into the point Q: (y_1, y_2) given by

$$(1) \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{Ax} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \qquad \text{in components,} \qquad \begin{aligned} y_1 &= 5x_1 + 3x_2 \\ y_2 &= 3x_1 + 5x_2. \end{aligned}$$

Find the **principal directions**, that is, the directions of the position vector $\mathbf{x}$ of P for which the direction of the position vector $\mathbf{y}$ of Q is the same or exactly opposite. What shape does the boundary circle take under this deformation?

Solution. We are looking for vectors $\mathbf{x}$ such that $\mathbf{y} = \lambda\mathbf{x}$. Since $\mathbf{y} = \mathbf{Ax}$, this gives $\mathbf{Ax} = \lambda\mathbf{x}$, the equation of an eigenvalue problem. In components, $\mathbf{Ax} = \lambda\mathbf{x}$ is

$$(2) \qquad \begin{aligned} 5x_1 + 3x_2 &= \lambda x_1 \\ 3x_1 + 5x_2 &= \lambda x_2 \end{aligned} \qquad \text{or} \qquad \begin{aligned} (5-\lambda)x_1 + \quad 3x_2 \quad &= 0 \\ 3x_1 \quad + (5-\lambda)x_2 &= 0. \end{aligned}$$

The characteristic equation is

$$(3) \qquad \begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} = (5-\lambda)^2 - 9 = 0.$$

Its solutions are $\lambda_1 = 8$ and $\lambda_2 = 2$. These are the eigenvalues of our problem. For $\lambda = \lambda_1 = 8$, our system (2) becomes

$$\begin{aligned} -3x_1 + 3x_2 &= 0, \\ 3x_1 - 3x_2 &= 0. \end{aligned} \quad \left| \quad \begin{aligned} &\text{Solution } x_2 = x_1,\ x_1 \text{ arbitrary,} \\ &\text{for instance, } x_1 = x_2 = 1. \end{aligned} \right.$$

For $\lambda_2 = 2$, our system (2) becomes

$$\begin{aligned} 3x_1 + 3x_2 &= 0, \\ 3x_1 + 3x_2 &= 0. \end{aligned} \quad \left| \quad \begin{aligned} &\text{Solution } x_2 = -x_1,\ x_1 \text{ arbitrary,} \\ &\text{for instance, } x_1 = 1, x_2 = -1. \end{aligned} \right.$$

We thus obtain as eigenvectors of $\mathbf{A}$, for instance, $[1 \quad 1]^\mathsf{T}$ corresponding to λ_1 and $[1 \quad -1]^\mathsf{T}$ corresponding to λ_2 (or a nonzero scalar multiple of these). These vectors make 45° and 135° angles with the positive x_1-direction. They give the principal directions, the answer to our problem. The eigenvalues show that in the principal directions the membrane is stretched by factors 8 and 2, respectively; see Fig. 160.

Accordingly, if we choose the principal directions as directions of a new Cartesian u_1u_2-coordinate system, say, with the positive u_1-semi-axis in the first quadrant and the positive u_2-semi-axis in the second quadrant of the x_1x_2-system, and if we set $u_1 = r\cos\phi$, $u_2 = r\sin\phi$, then a boundary point of the unstretched circular membrane has coordinates $\cos\phi$, $\sin\phi$. Hence, after the stretch we have

$$z_1 = 8\cos\phi, \qquad z_2 = 2\sin\phi.$$

Since $\cos^2\phi + \sin^2\phi = 1$, this shows that the deformed boundary is an ellipse (Fig. 160)

$$(4) \qquad \frac{z_1^2}{8^2} + \frac{z_2^2}{2^2} = 1. \qquad \blacksquare$$

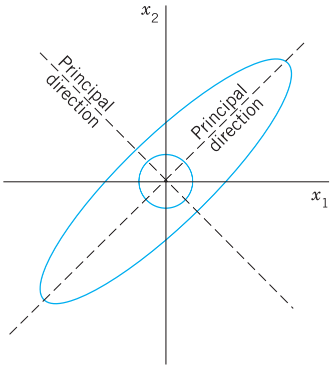


Fig. 160. Undeformed and deformed membrane in Example 1

EXAMPLE 2 Eigenvalue Problems Arising from Markov Processes

Markov processes as considered in Example 13 of Sec. 7.2 lead to eigenvalue problems if we ask for the limit state of the process in which the state vector $\mathbf{x}$ is reproduced under the multiplication by the stochastic matrix $\mathbf{A}$ governing the process, that is, $\mathbf{Ax} = \mathbf{x}$. Hence $\mathbf{A}$ should have the eigenvalue 1, and $\mathbf{x}$ should be a corresponding eigenvector. This is of practical interest because it shows the long-term tendency of the development modeled by the process.

In that example,

$$\mathbf{A} = \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0.2 & 0.9 & 0.2 \\ 0.1 & 0 & 0.8 \end{bmatrix}. \qquad \text{For the transpose,} \qquad \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.9 & 0 \\ 0 & 0.2 & 0.8 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Hence $\mathbf{A}^\mathsf{T}$ has the eigenvalue 1, and the same is true for $\mathbf{A}$ by Theorem 3 in Sec. 8.1. An eigenvector $\mathbf{x}$ of $\mathbf{A}$ for $\lambda = 1$ is obtained from

$$\mathbf{A} - \mathbf{I} = \begin{bmatrix} -0.3 & 0.1 & 0 \\ 0.2 & -0.1 & 0.2 \\ 0.1 & 0 & -0.2 \end{bmatrix}, \qquad \text{row-reduced to} \qquad \begin{bmatrix} -\frac{3}{10} & \frac{1}{10} & 0 \\ 0 & -\frac{1}{30} & \frac{1}{5} \\ 0 & 0 & 0 \end{bmatrix}.$$

Taking $x_3 = 1$, we get $x_2 = 6$ from $-x_2/30 + x_3/5 = 0$ and then $x_1 = 2$ from $-3x_1/10 + x_2/10 = 0$. This gives $\mathbf{x} = [2 \quad 6 \quad 1]^\mathsf{T}$. It means that in the long run, the ratio Commercial:Industrial:Residential will approach 2:6:1, provided that the probabilities given by $\mathbf{A}$ remain (about) the same. (We switched to ordinary fractions to avoid rounding errors.) ■

EXAMPLE 3 Eigenvalue Problems Arising from Population Models. Leslie Model

The Leslie model describes age-specified population growth, as follows. Let the oldest age attained by the females in some animal population be 9 years. Divide the population into three age classes of 3 years each. Let the "*Leslie matrix*" be

$$\mathbf{L} = [l_{jk}] = \begin{bmatrix} 0 & 2.3 & 0.4 \\ 0.6 & 0 & 0 \\ 0 & 0.3 & 0 \end{bmatrix} \tag{5}$$

where l_{1k} is the average number of daughters born to a single female during the time she is in age class k, and $l_{j,j-1}(j = 2, 3)$ is the fraction of females in age class $j - 1$ that will survive and pass into class j. (a) What is the number of females in each class after 3, 6, 9 years if each class initially consists of 400 females? (b) For what initial distribution will the number of females in each class change by the same proportion? What is this rate of change?

Solution. **(a)** Initially, $\mathbf{x}_{(0)}^{\mathsf{T}} = [400 \quad 400 \quad 400]$. After 3 years,

$$\mathbf{x}_{(3)} = \mathbf{L}\mathbf{x}_{(0)} = \begin{bmatrix} 0 & 2.3 & 0.4 \\ 0.6 & 0 & 0 \\ 0 & 0.3 & 0 \end{bmatrix} \begin{bmatrix} 400 \\ 400 \\ 400 \end{bmatrix} = \begin{bmatrix} 1080 \\ 240 \\ 120 \end{bmatrix}.$$

Similarly, after 6 years the number of females in each class is given by $\mathbf{x}_{(6)}^{\mathsf{T}} = (\mathbf{L}\mathbf{x}_{(3)})^{\mathsf{T}} = [600 \quad 648 \quad 72]$, and after 9 years we have $\mathbf{x}_{(9)}^{\mathsf{T}} = (\mathbf{L}\mathbf{x}_{(6)})^{\mathsf{T}} = [1519.2 \quad 360 \quad 194.4]$.

(b) Proportional change means that we are looking for a distribution vector $\mathbf{x}$ such that $\mathbf{L}\mathbf{x} = \lambda\mathbf{x}$, where λ is the rate of change (growth if $\lambda > 1$, decrease if $\lambda < 1$). The characteristic equation is (develop the characteristic determinant by the first column)

$$\det(\mathbf{L} - \lambda\mathbf{I}) = -\lambda^3 - 0.6(-2.3\lambda - 0.3 \cdot 0.4) = -\lambda^3 + 1.38\lambda + 0.072 = 0.$$

A positive root is found to be (for instance, by Newton's method, Sec. 19.2) $\lambda = 1.2$. A corresponding eigenvector $\mathbf{x}$ can be determined from the characteristic matrix

$$\mathbf{A} - 1.2\mathbf{I} = \begin{bmatrix} -1.2 & 2.3 & 0.4 \\ 0.6 & -1.2 & 0 \\ 0 & 0.3 & -1.2 \end{bmatrix}, \qquad \text{say,} \qquad \mathbf{x} = \begin{bmatrix} 1 \\ 0.5 \\ 0.125 \end{bmatrix}$$

where $x_3 = 0.125$ is chosen, $x_2 = 0.5$ then follows from $0.3x_2 - 1.2x_3 = 0$, and $x_1 = 1$ from $-1.2x_1 + 2.3x_2 + 0.4x_3 = 0$. To get an initial population of 1200 as before, we multiply $\mathbf{x}$ by $1200/(1 + 0.5 + 0.125) = 738$. *Answer:* Proportional growth of the numbers of females in the three classes will occur if the initial values are 738, 369, 92 in classes 1, 2, 3, respectively. The growth rate will be 1.2 per 3 years. ■

EXAMPLE 4 Vibrating System of Two Masses on Two Springs (Fig. 161)

Mass–spring systems involving several masses and springs can be treated as eigenvalue problems. For instance, the mechanical system in Fig. 161 is governed by the system of ODEs

$$\begin{aligned} y_1'' &= -3y_1 - 2(y_1 - y_2) = -5y_1 + 2y_2 \\ y_2'' &= -2(y_2 - y_1) = 2y_1 - 2y_2 \end{aligned} \tag{6}$$

where y_1 and y_2 are the displacements of the masses from rest, as shown in the figure, and primes denote derivatives with respect to time t. In vector form, this becomes

$$\mathbf{y}'' = \begin{bmatrix} y_1'' \\ y_2'' \end{bmatrix} = \mathbf{A}\mathbf{y} = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}. \tag{7}$$

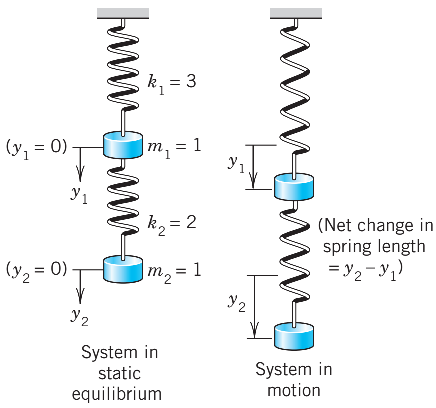


Fig. 161. Masses on springs in Example 4

We try a vector solution of the form

$$\mathbf{y} = \mathbf{x}e^{\omega t}. \tag{8}$$

This is suggested by a mechanical system of a single mass on a spring (Sec. 2.4), whose motion is given by exponential functions (and sines and cosines). Substitution into (7) gives

$$\omega^2 \mathbf{x} e^{\omega t} = \mathbf{A}\mathbf{x}e^{\omega t}.$$

Dividing by $e^{\omega t}$ and writing $\omega^2 = \lambda$, we see that our mechanical system leads to the eigenvalue problem

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x} \qquad \text{where } \lambda = \omega^2. \tag{9}$$

From Example 1 in Sec. 8.1 we see that $\mathbf{A}$ has the eigenvalues $\lambda_1 = -1$ and $\lambda_2 = -6$. Consequently, $\omega = \pm\sqrt{-1} = \pm i$ and $\sqrt{-6} = \pm i\sqrt{6}$, respectively. Corresponding eigenvectors are

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}. \tag{10}$$

From (8) we thus obtain the four complex solutions [see (10), Sec. 2.2]

$$\mathbf{x}_1 e^{\pm it} = \mathbf{x}_1(\cos t \pm i \sin t),$$
$$\mathbf{x}_2 e^{\pm i\sqrt{6}t} = \mathbf{x}_2(\cos \sqrt{6}\, t \pm i \sin \sqrt{6}\, t).$$

By addition and subtraction (see Sec. 2.2) we get the four real solutions

$$\mathbf{x}_1 \cos t, \qquad \mathbf{x}_1 \sin t, \qquad \mathbf{x}_2 \cos \sqrt{6}\, t, \qquad \mathbf{x}_2 \sin \sqrt{6}\, t.$$

A general solution is obtained by taking a linear combination of these,

$$\mathbf{y} = \mathbf{x}_1(a_1 \cos t + b_1 \sin t) + \mathbf{x}_2(a_2 \cos \sqrt{6}\, t + b_2 \sin \sqrt{6}\, t)$$

with arbitrary constants a_1, b_1, a_2, b_2 (to which values can be assigned by prescribing initial displacement and initial velocity of each of the two masses). By (10), the components of $\mathbf{y}$ are

$$y_1 = a_1 \cos t + b_1 \sin t + 2a_2 \cos \sqrt{6}\, t + 2b_2 \sin \sqrt{6}\, t$$
$$y_2 = 2a_1 \cos t + 2b_1 \sin t - a_2 \cos \sqrt{6}\, t - b_2 \sin \sqrt{6}\, t.$$

These functions describe harmonic oscillations of the two masses. Physically, this had to be expected because we have neglected damping. ■

PROBLEM SET 8.2

1–6 ELASTIC DEFORMATIONS

Given $\mathbf{A}$ in a deformation $\mathbf{y} = \mathbf{A}\mathbf{x}$, find the principal directions and corresponding factors of extension or contraction. Show the details.

1. $\begin{bmatrix} 3.0 & 1.5 \\ 1.5 & 3.0 \end{bmatrix}$

2. $\begin{bmatrix} 2.0 & 0.4 \\ 0.4 & 2.0 \end{bmatrix}$

3. $\begin{bmatrix} 7 & \sqrt{6} \\ \sqrt{6} & 2 \end{bmatrix}$

4. $\begin{bmatrix} 5 & 2 \\ 2 & 13 \end{bmatrix}$

5. $\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$

6. $\begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$

7–9 MARKOV PROCESSES

Find the limit state of the Markov process modeled by the given matrix. Show the details.

7. $\begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}$

8. $\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.6 & 0.1 \\ 0.3 & 0.1 & 0.6 \end{bmatrix}$

9. $\begin{bmatrix} 0.6 & 0.1 & 0.2 \\ 0.4 & 0.1 & 0.4 \\ 0 & 0.8 & 0.4 \end{bmatrix}$

10–12 AGE-SPECIFIC POPULATION

Find the growth rate in the Leslie model (see Example 3) with the matrix as given. Show the details.

10. $\begin{bmatrix} 0 & 9.0 & 5.0 \\ 0.4 & 0 & 0 \\ 0 & 0.4 & 0 \end{bmatrix}$ **11.** $\begin{bmatrix} 0 & 3.45 & 0.60 \\ 0.90 & 0 & 0 \\ 0 & 0.45 & 0 \end{bmatrix}$

12. $\begin{bmatrix} 0 & 3.0 & 2.0 & 2.0 \\ 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{bmatrix}$

13–15 LEONTIEF MODELS[1]

13. Leontief input–output model. Suppose that three industries are interrelated so that their outputs are used as inputs by themselves, according to the 3×3 **consumption matrix**

$$\mathbf{A} = [a_{jk}] = \begin{bmatrix} 0.1 & 0.5 & 0 \\ 0.8 & 0 & 0.4 \\ 0.1 & 0.5 & 0.6 \end{bmatrix}$$

where a_{jk} is the fraction of the output of industry k consumed (purchased) by industry j. Let p_j be the price charged by industry j for its total output. A problem is to find prices so that for each industry, total expenditures equal total income. Show that this leads to $\mathbf{Ap} = \mathbf{p}$, where $\mathbf{p} = [p_1 \quad p_2 \quad p_3]^\mathsf{T}$, and find a solution $\mathbf{p}$ with nonnegative p_1, p_2, p_3.

14. Show that a consumption matrix as considered in Prob. 13 must have column sums 1 and always has the eigenvalue 1.

15. Open Leontief input–output model. If not the whole output but only a portion of it is consumed by the industries themselves, then instead of $\mathbf{Ax} = \mathbf{x}$ (as in Prob. 13), we have $\mathbf{x} - \mathbf{Ax} = \mathbf{y}$, where $\mathbf{x} = [x_1 \quad x_2 \quad x_3]^\mathsf{T}$ is produced, $\mathbf{Ax}$ is consumed by the industries, and, thus, $\mathbf{y}$ is the net production available for other consumers. Find for what production $\mathbf{x}$ a given **demand vector** $\mathbf{y} = [0.1 \quad 0.3 \quad 0.1]^\mathsf{T}$ can be achieved if the consumption matrix is

$$\mathbf{A} = \begin{bmatrix} 0.1 & 0.4 & 0.2 \\ 0.5 & 0 & 0.1 \\ 0.1 & 0.4 & 0.4 \end{bmatrix}.$$

16–20 GENERAL PROPERTIES OF EIGENVALUE PROBLEMS

Let $\mathbf{A} = [a_{jk}]$ be an $n \times n$ matrix with (not necessarily distinct) eigenvalues $\lambda_1, \cdots, \lambda_n$. Show.

16. Trace. The sum of the main diagonal entries, called the *trace* of $\mathbf{A}$, equals the sum of the eigenvalues of $\mathbf{A}$.

17. "Spectral shift." $\mathbf{A} - k\mathbf{I}$ has the eigenvalues $\lambda_1 - k, \cdots, \lambda_n - k$ and the same eigenvectors as $\mathbf{A}$.

18. Scalar multiples, powers. $k\mathbf{A}$ has the eigenvalues $k\lambda_1, \cdots, k\lambda_n$. $\mathbf{A}^m (m = 1, 2, \cdots)$ has the eigenvalues $\lambda_1^m, \cdots, \lambda_n^m$. The eigenvectors are those of $\mathbf{A}$.

19. Spectral mapping theorem. The "**polynomial matrix**"

$$p(\mathbf{A}) = k_m\mathbf{A}^m + k_{m-1}\mathbf{A}^{m-1} + \cdots + k_1\mathbf{A} + k_0\mathbf{I}$$

has the eigenvalues

$$p(\lambda_j) = k_m\lambda_j^m + k_{m-1}\lambda_j^{m-1} + \cdots + k_1\lambda_j + k_0$$

where $j = 1, \cdots, n$, and the same eigenvectors as $\mathbf{A}$.

20. Perron's theorem. A Leslie matrix $\mathbf{L}$ with positive $l_{12}, l_{13}, l_{21}, l_{32}$ has a positive eigenvalue. (This is a special case of the Perron–Frobenius theorem in Sec. 20.7, which is difficult to prove in its general form.)

8.3 Symmetric, Skew-Symmetric, and Orthogonal Matrices

We consider three classes of real square matrices that, because of their remarkable properties, occur quite frequently in applications. The first two matrices have already been mentioned in Sec. 7.2. The goal of Sec. 8.3 is to show their remarkable properties.

[1]WASSILY LEONTIEF (1906–1999). American economist at New York University. For his input–output analysis he was awarded the Nobel Prize in 1973.

DEFINITIONS

Symmetric, Skew-Symmetric, and Orthogonal Matrices

A ***real*** square matrix $\mathbf{A} = [a_{jk}]$ is called

symmetric if transposition leaves it unchanged,

$$\mathbf{A}^\mathsf{T} = \mathbf{A}, \qquad \text{thus} \qquad a_{kj} = a_{jk}, \tag{1}$$

skew-symmetric if transposition gives the negative of **A**,

$$\mathbf{A}^\mathsf{T} = -\mathbf{A}, \qquad \text{thus} \qquad a_{kj} = -a_{jk}, \tag{2}$$

orthogonal if transposition gives the inverse of **A**,

$$\mathbf{A}^\mathsf{T} = \mathbf{A}^{-1}. \tag{3}$$

EXAMPLE 1 **Symmetric, Skew-Symmetric, and Orthogonal Matrices**

The matrices

$$\begin{bmatrix} -3 & 1 & 5 \\ 1 & 0 & -2 \\ 5 & -2 & 4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 9 & -12 \\ -9 & 0 & 20 \\ 12 & -20 & 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{bmatrix}$$

are symmetric, skew-symmetric, and orthogonal, respectively, as you should verify. Every skew-symmetric matrix has all main diagonal entries zero. (Can you prove this?) ■

Any real square matrix **A** may be written as the sum of a symmetric matrix **R** and a skew-symmetric matrix **S**, where

$$\mathbf{R} = \tfrac{1}{2}(\mathbf{A} + \mathbf{A}^\mathsf{T}) \qquad \text{and} \qquad \mathbf{S} = \tfrac{1}{2}(\mathbf{A} - \mathbf{A}^\mathsf{T}). \tag{4}$$

EXAMPLE 2 **Illustration of Formula (4)**

$$\mathbf{A} = \begin{bmatrix} 9 & 5 & 2 \\ 2 & 3 & -8 \\ 5 & 4 & 3 \end{bmatrix} = \mathbf{R} + \mathbf{S} = \begin{bmatrix} 9.0 & 3.5 & 3.5 \\ 3.5 & 3.0 & -2.0 \\ 3.5 & -2.0 & 3.0 \end{bmatrix} + \begin{bmatrix} 0 & 1.5 & -1.5 \\ -1.5 & 0 & -6.0 \\ 1.5 & 6.0 & 0 \end{bmatrix}$$ ■

THEOREM 1 **Eigenvalues of Symmetric and Skew-Symmetric Matrices**

(a) *The eigenvalues of a symmetric matrix are real.*

(b) *The eigenvalues of a skew-symmetric matrix are pure imaginary or zero.*

This basic theorem (and an extension of it) will be proved in Sec. 8.5.

EXAMPLE 3 Eigenvalues of Symmetric and Skew-Symmetric Matrices

The matrices in (1) and (7) of Sec. 8.2 are symmetric and have real eigenvalues. The skew-symmetric matrix in Example 1 has the eigenvalues 0, $-25i$, and $25i$. (Verify this.) The following matrix has the real eigenvalues 1 and 5 but is not symmetric. Does this contradict Theorem 1?

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

■

Orthogonal Transformations and Orthogonal Matrices

Orthogonal transformations are transformations

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad \text{where } \mathbf{A} \text{ is an orthogonal matrix.} \tag{5}$$

With each vector $\mathbf{x}$ in R^n such a transformation assigns a vector $\mathbf{y}$ in R^n. For instance, the plane rotation through an angle θ

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{6}$$

is an orthogonal transformation. It can be shown that any orthogonal transformation in the plane or in three-dimensional space is a **rotation** (possibly combined with a reflection in a straight line or a plane, respectively).

The main reason for the importance of orthogonal matrices is as follows.

THEOREM 2 Invariance of Inner Product

An orthogonal transformation preserves the value of the **inner product** *of vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *in* R^n, *defined by*

$$\mathbf{a} \bullet \mathbf{b} = \mathbf{a}^\mathsf{T}\mathbf{b} = [a_1 \quad \cdots \quad a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}. \tag{7}$$

That is, for any $\mathbf{a}$ *and* $\mathbf{b}$ *in* R^n, *orthogonal* $n \times n$ *matrix* $\mathbf{A}$, *and* $\mathbf{u} = \mathbf{A}\mathbf{a}, \mathbf{v} = \mathbf{A}\mathbf{b}$ *we have* $\mathbf{u} \bullet \mathbf{v} = \mathbf{a} \bullet \mathbf{b}$.

Hence the transformation also preserves the **length** *or* **norm** *of any vector* $\mathbf{a}$ *in* R^n *given by*

$$\|\mathbf{a}\| = \sqrt{\mathbf{a} \bullet \mathbf{a}} = \sqrt{\mathbf{a}^\mathsf{T}\mathbf{a}}. \tag{8}$$

PROOF Let $\mathbf{A}$ be orthogonal. Let $\mathbf{u} = \mathbf{A}\mathbf{a}$ and $\mathbf{v} = \mathbf{A}\mathbf{b}$. We must show that $\mathbf{u} \bullet \mathbf{v} = \mathbf{a} \bullet \mathbf{b}$. Now $(\mathbf{A}\mathbf{a})^\mathsf{T} = \mathbf{a}^\mathsf{T}\mathbf{A}^\mathsf{T}$ by (10d) in Sec. 7.2 and $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ by (3). Hence

$$\mathbf{u} \bullet \mathbf{v} = \mathbf{u}^\mathsf{T}\mathbf{v} = (\mathbf{A}\mathbf{a})^\mathsf{T}\mathbf{A}\mathbf{b} = \mathbf{a}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{b} = \mathbf{a}^\mathsf{T}\mathbf{I}\mathbf{b} = \mathbf{a}^\mathsf{T}\mathbf{b} = \mathbf{a} \bullet \mathbf{b}. \tag{9}$$

From this the invariance of $\|\mathbf{a}\|$ follows if we set $\mathbf{b} = \mathbf{a}$. ■

Orthogonal matrices have further interesting properties as follows.

THEOREM 3

Orthonormality of Column and Row Vectors

A real square matrix is orthogonal if and only if its column vectors $\mathbf{a}_1, \cdots, \mathbf{a}_n$ *(and also its row vectors) form an* **orthonormal system**, *that is,*

$$\mathbf{a}_j \cdot \mathbf{a}_k = \mathbf{a}_j^{\mathsf{T}}\mathbf{a}_k = \begin{cases} 0 & \text{if} \quad j \neq k \\ 1 & \text{if} \quad j = k. \end{cases} \tag{10}$$

PROOF **(a)** Let $\mathbf{A}$ be orthogonal. Then $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}^{\mathsf{T}}\mathbf{A} = \mathbf{I}$. In terms of column vectors $\mathbf{a}_1, \cdots, \mathbf{a}_n$,

$$\mathbf{I} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{A}^{\mathsf{T}}\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{a}_n^{\mathsf{T}} \end{bmatrix} [\mathbf{a}_1 \cdots \mathbf{a}_n] = \begin{bmatrix} \mathbf{a}_1^{\mathsf{T}}\mathbf{a}_1 & \mathbf{a}_1^{\mathsf{T}}\mathbf{a}_2 & \cdots & \mathbf{a}_1^{\mathsf{T}}\mathbf{a}_n \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{a}_n^{\mathsf{T}}\mathbf{a}_1 & \mathbf{a}_n^{\mathsf{T}}\mathbf{a}_2 & \cdots & \mathbf{a}_n^{\mathsf{T}}\mathbf{a}_n \end{bmatrix}. \tag{11}$$

The last equality implies (10), by the definition of the $n \times n$ unit matrix $\mathbf{I}$. From (3) it follows that the inverse of an orthogonal matrix is orthogonal (see CAS Experiment 12). Now the column vectors of $\mathbf{A}^{-1}(=\mathbf{A}^{\mathsf{T}})$ are the row vectors of $\mathbf{A}$. Hence the row vectors of $\mathbf{A}$ also form an orthonormal system.
(b) Conversely, if the column vectors of $\mathbf{A}$ satisfy (10), the off-diagonal entries in (11) must be 0 and the diagonal entries 1. Hence $\mathbf{A}^{\mathsf{T}}\mathbf{A} = \mathbf{I}$, as (11) shows. Similarly, $\mathbf{A}\mathbf{A}^{\mathsf{T}} = \mathbf{I}$. This implies $\mathbf{A}^{\mathsf{T}} = \mathbf{A}^{-1}$ because also $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ and the inverse is unique. Hence $\mathbf{A}$ is orthogonal. Similarly when the row vectors of $\mathbf{A}$ form an orthonormal system, by what has been said at the end of part (**a**). ■

THEOREM 4

Determinant of an Orthogonal Matrix

The determinant of an orthogonal matrix has the value $+1$ *or* -1.

PROOF From $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$ (Sec. 7.8, Theorem 4) and $\det \mathbf{A}^{\mathsf{T}} = \det \mathbf{A}$ (Sec. 7.7, Theorem 2d), we get for an orthogonal matrix

$$1 = \det \mathbf{I} = \det(\mathbf{A}\mathbf{A}^{-1}) = \det(\mathbf{A}\mathbf{A}^{\mathsf{T}}) = \det \mathbf{A} \det \mathbf{A}^{\mathsf{T}} = (\det \mathbf{A})^2. \quad ■$$

EXAMPLE 4 **Illustration of Theorems 3 and 4**

The last matrix in Example 1 and the matrix in (6) illustrate Theorems 3 and 4 because their determinants are -1 and $+1$, as you should verify. ■

THEOREM 5

Eigenvalues of an Orthogonal Matrix

The eigenvalues of an orthogonal matrix $\mathbf{A}$ *are real or complex conjugates in pairs and have absolute value* 1.

PROOF The first part of the statement holds for any real matrix $\mathbf{A}$ because its characteristic polynomial has real coefficients, so that its zeros (the eigenvalues of $\mathbf{A}$) must be as indicated. The claim that $|\lambda| = 1$ will be proved in Sec. 8.5. ■

EXAMPLE 5 **Eigenvalues of an Orthogonal Matrix**

The orthogonal matrix in Example 1 has the characteristic equation

$$-\lambda^3 + \tfrac{2}{3}\lambda^2 + \tfrac{2}{3}\lambda - 1 = 0.$$

Now one of the eigenvalues must be real (why?), hence $+1$ or -1. Trying, we find -1. Division by $\lambda + 1$ gives $-(\lambda^2 - 5\lambda/3 + 1) = 0$ and the two eigenvalues $(5 + i\sqrt{11})/6$ and $(5 - i\sqrt{11})/6$, which have absolute value 1. Verify all of this. ■

Looking back at this section, you will find that the numerous basic results it contains have relatively short, straightforward proofs. This is typical of large portions of matrix eigenvalue theory.

PROBLEM SET 8.3

1–10 SPECTRUM

Are the following matrices symmetric, skew-symmetric, or orthogonal? Find the spectrum of each, thereby illustrating Theorems 1 and 5. Show your work in detail.

1. $\begin{bmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{bmatrix}$ **2.** $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$

3. $\begin{bmatrix} 2 & 8 \\ -8 & 2 \end{bmatrix}$ **4.** $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

5. $\begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 5 \end{bmatrix}$ **6.** $\begin{bmatrix} a & k & k \\ k & a & k \\ k & k & a \end{bmatrix}$

7. $\begin{bmatrix} 0 & 9 & -12 \\ -9 & 0 & 20 \\ 12 & -20 & 0 \end{bmatrix}$ **8.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$

9. $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$ **10.** $\begin{bmatrix} \frac{4}{9} & \frac{8}{9} & \frac{1}{9} \\ -\frac{7}{9} & \frac{4}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{1}{9} & \frac{8}{9} \end{bmatrix}$

11. WRITING PROJECT. Section Summary. Summarize the main concepts and facts in this section, giving illustrative examples of your own.

12. CAS EXPERIMENT. Orthogonal Matrices.

(a) Products. Inverse. Prove that the product of two orthogonal matrices is orthogonal, and so is the inverse of an orthogonal matrix. What does this mean in terms of rotations?

(b) Rotation. Show that (6) is an orthogonal transformation. Verify that it satisfies Theorem 3. Find the inverse transformation.

(c) Powers. Write a program for computing powers $\mathbf{A}^m$ $(m = 1, 2, \cdots)$ of a 2×2 matrix $\mathbf{A}$ and their spectra. Apply it to the matrix in Prob. 1 (call it $\mathbf{A}$). To what rotation does $\mathbf{A}$ correspond? Do the eigenvalues of $\mathbf{A}^m$ have a limit as $m \to \infty$?

(d) Compute the eigenvalues of $(0.9\mathbf{A})^m$, where $\mathbf{A}$ is the matrix in Prob. 1. Plot them as points. What is their limit? Along what kind of curve do these points approach the limit?

(e) Find $\mathbf{A}$ such that $\mathbf{y} = \mathbf{Ax}$ is a counterclockwise rotation through 30° in the plane.

13–20 GENERAL PROPERTIES

13. Verification. Verify the statements in Example 1.

14. Verify the statements in Examples 3 and 4.

15. Sum. Are the eigenvalues of $\mathbf{A} + \mathbf{B}$ sums of the eigenvalues of $\mathbf{A}$ and of $\mathbf{B}$?

16. Orthogonality. Prove that eigenvectors of a symmetric matrix corresponding to different eigenvalues are orthogonal. Give examples.

17. Skew-symmetric matrix. Show that the inverse of a skew-symmetric matrix is skew-symmetric.

18. Do there exist nonsingular skew-symmetric $n \times n$ matrices with odd n?

19. Orthogonal matrix. Do there exist skew-symmetric orthogonal 3×3 matrices?

20. Symmetric matrix. Do there exist nondiagonal symmetric 3×3 matrices that are orthogonal?

8.4 Eigenbases. Diagonalization. Quadratic Forms

So far we have emphasized properties of eigen*values*. We now turn to general properties of eigen*vectors*. Eigenvectors of an $n \times n$ matrix $\mathbf{A}$ may (or may not!) form a basis for R^n. If we are interested in a transformation $\mathbf{y} = \mathbf{A}\mathbf{x}$, such an "**eigenbasis**" (basis of eigenvectors)—if it exists—is of great advantage because then we can represent any $\mathbf{x}$ in R^n uniquely as a linear combination of the eigenvectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$, say,

$$\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n.$$

And, denoting the corresponding (not necessarily distinct) eigenvalues of the matrix $\mathbf{A}$ by $\lambda_1, \cdots, \lambda_n$, we have $\mathbf{A}\mathbf{x}_j = \lambda_j\mathbf{x}_j$, so that we simply obtain

$$\begin{aligned} \mathbf{y} = \mathbf{A}\mathbf{x} &= \mathbf{A}(c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n) \\ &= c_1\mathbf{A}\mathbf{x}_1 + \cdots + c_n\mathbf{A}\mathbf{x}_n \\ &= c_1\lambda_1\mathbf{x}_1 + \cdots + c_n\lambda_n\mathbf{x}_n. \end{aligned} \tag{1}$$

This shows that we have decomposed the complicated action of $\mathbf{A}$ on an arbitrary vector $\mathbf{x}$ into a sum of simple actions (multiplication by scalars) on the eigenvectors of $\mathbf{A}$. This is the point of an eigenbasis.

Now if the n eigenvalues are all different, we do obtain a basis:

THEOREM 1

Basis of Eigenvectors

*If an $n \times n$ matrix $\mathbf{A}$ has n **distinct** eigenvalues, then $\mathbf{A}$ has a basis of eigenvectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$ for R^n.*

PROOF All we have to show is that $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are linearly independent. Suppose they are not. Let r be the largest integer such that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is a linearly independent set. Then $r < n$ and the set $\{\mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{x}_{r+1}\}$ is linearly dependent. Thus there are scalars $c_1, \cdots, c_{r+1}$, not all zero, such that

$$c_1\mathbf{x}_1 + \cdots + c_{r+1}\mathbf{x}_{r+1} = \mathbf{0} \tag{2}$$

(see Sec. 7.4). Multiplying both sides by $\mathbf{A}$ and using $\mathbf{A}\mathbf{x}_j = \lambda_j\mathbf{x}_j$, we obtain

$$\mathbf{A}(c_1\mathbf{x}_1 + \cdots + c_{r+1}\mathbf{x}_{r+1}) = c_1\lambda_1\mathbf{x}_1 + \cdots + c_{r+1}\lambda_{r+1}\mathbf{x}_{r+1} = \mathbf{A}\mathbf{0} = \mathbf{0}. \tag{3}$$

To get rid of the last term, we subtract λ_{r+1} times (2) from this, obtaining

$$c_1(\lambda_1 - \lambda_{r+1})\mathbf{x}_1 + \cdots + c_r(\lambda_r - \lambda_{r+1})\mathbf{x}_r = \mathbf{0}.$$

Here $c_1(\lambda_1 - \lambda_{r+1}) = 0, \cdots, c_r(\lambda_r - \lambda_{r+1}) = 0$ since $\{x_1, \cdots, x_r\}$ is linearly independent. Hence $c_1 = \cdots = c_r = 0$, since all the eigenvalues are distinct. But with this, (2) reduces to $c_{r+1}\mathbf{x}_{r+1} = \mathbf{0}$, hence $c_{r+1} = 0$, since $\mathbf{x}_{r+1} \neq \mathbf{0}$ (an eigenvector!). This contradicts the fact that not all scalars in (2) are zero. Hence the conclusion of the theorem must hold. ■

EXAMPLE 1 **Eigenbasis. Nondistinct Eigenvalues. Nonexistence**

The matrix $\mathbf{A} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$ has a basis of eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ corresponding to the eigenvalues $\lambda_1 = 8$, $\lambda_2 = 2$. (See Example 1 in Sec. 8.2.)

Even if not all n eigenvalues are different, a matrix $\mathbf{A}$ may still provide an eigenbasis for R^n. See Example 2 in Sec. 8.1, where $n = 3$.

On the other hand, $\mathbf{A}$ ***may not have enough linearly independent eigenvectors to make up a basis***. For instance, $\mathbf{A}$ in Example 3 of Sec. 8.1 is

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and has only one eigenvector} \quad \begin{bmatrix} k \\ 0 \end{bmatrix} \quad (k \neq 0, \text{ arbitrary}).$$

Actually, eigenbases exist under much more general conditions than those in Theorem 1. An important case is the following.

THEOREM 2 **Symmetric Matrices**

A symmetric matrix has an orthonormal basis of eigenvectors for R^n.

For a proof (which is involved) see Ref. [B3], vol. 1, pp. 270–272.

EXAMPLE 2 **Orthonormal Basis of Eigenvectors**

The first matrix in Example 1 is symmetric, and an orthonormal basis of eigenvectors is $[1/\sqrt{2} \quad 1/\sqrt{2}]^\mathsf{T}$, $[1/\sqrt{2} \quad -1/\sqrt{2}]^\mathsf{T}$.

Similarity of Matrices. Diagonalization

Eigenbases also play a role in reducing a matrix $\mathbf{A}$ to a diagonal matrix whose entries are the eigenvalues of $\mathbf{A}$. This is done by a "similarity transformation," which is defined as follows (and will have various applications in numerics in Chap. 20).

DEFINITION **Similar Matrices. Similarity Transformation**

An $n \times n$ matrix $\hat{\mathbf{A}}$ is called **similar** to an $n \times n$ matrix $\mathbf{A}$ if

$$\hat{\mathbf{A}} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} \tag{4}$$

for some (nonsingular!) $n \times n$ matrix $\mathbf{P}$. This transformation, which gives $\hat{\mathbf{A}}$ from $\mathbf{A}$, is called a **similarity transformation**.

The key property of this transformation is that it preserves the eigenvalues of $\mathbf{A}$:

THEOREM 3 **Eigenvalues and Eigenvectors of Similar Matrices**

If $\hat{\mathbf{A}}$ is similar to $\mathbf{A}$, then $\hat{\mathbf{A}}$ has the same eigenvalues as $\mathbf{A}$.

Furthermore, if $\mathbf{x}$ is an eigenvector of $\mathbf{A}$, then $\mathbf{y} = \mathbf{P}^{-1}\mathbf{x}$ is an eigenvector of $\hat{\mathbf{A}}$ corresponding to the same eigenvalue.

PROOF From $\mathbf{Ax} = \lambda\mathbf{x}$ (λ an eigenvalue, $\mathbf{x} \neq \mathbf{0}$) we get $\mathbf{P}^{-1}\mathbf{Ax} = \lambda\mathbf{P}^{-1}\mathbf{x}$. Now $\mathbf{I} = \mathbf{PP}^{-1}$. By this ***identity trick*** the equation $\mathbf{P}^{-1}\mathbf{Ax} = \lambda\mathbf{P}^{-1}\mathbf{x}$ gives

$$\mathbf{P}^{-1}\mathbf{Ax} = \mathbf{P}^{-1}\mathbf{AIx} = \mathbf{P}^{-1}\mathbf{APP}^{-1}\mathbf{x} = (\mathbf{P}^{-1}\mathbf{AP})\mathbf{P}^{-1}\mathbf{x} = \hat{\mathbf{A}}(\mathbf{P}^{-1}\mathbf{x}) = \lambda\mathbf{P}^{-1}\mathbf{x}.$$

Hence λ is an eigenvalue of $\hat{\mathbf{A}}$ and $\mathbf{P}^{-1}\mathbf{x}$ a corresponding eigenvector. Indeed, $\mathbf{P}^{-1}\mathbf{x} \neq \mathbf{0}$ because $\mathbf{P}^{-1}\mathbf{x} = \mathbf{0}$ would give $\mathbf{x} = \mathbf{Ix} = \mathbf{PP}^{-1}\mathbf{x} = \mathbf{P0} = \mathbf{0}$, contradicting $\mathbf{x} \neq \mathbf{0}$. ■

EXAMPLE 3 **Eigenvalues and Vectors of Similar Matrices**

Let,
$$\mathbf{A} = \begin{bmatrix} 6 & -3 \\ 4 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}.$$

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

Here $\mathbf{P}^{-1}$ was obtained from (4*) in Sec. 7.8 with $\det \mathbf{P} = 1$. We see that $\hat{\mathbf{A}}$ has the eigenvalues $\lambda_1 = 3$, $\lambda_2 = 2$. The characteristic equation of $\mathbf{A}$ is $(6 - \lambda)(-1 - \lambda) + 12 = \lambda^2 - 5\lambda + 6 = 0$. It has the roots (the eigenvalues of $\mathbf{A}$) $\lambda_1 = 3$, $\lambda_2 = 2$, confirming the first part of Theorem 3.

We confirm the second part. From the first component of $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ we have $(6 - \lambda)x_1 - 3x_2 = 0$. For $\lambda = 3$ this gives $3x_1 - 3x_2 = 0$, say, $\mathbf{x}_1 = [1 \quad 1]^\mathsf{T}$. For $\lambda = 2$ it gives $4x_1 - 3x_2 = 0$, say, $\mathbf{x}_2 = [3 \quad 4]^\mathsf{T}$. In Theorem 3 we thus have

$$\mathbf{y}_1 = \mathbf{P}^{-1}\mathbf{x}_1 = \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathbf{y}_2 = \mathbf{P}^{-1}\mathbf{x}_2 = \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Indeed, these are eigenvectors of the diagonal matrix $\hat{\mathbf{A}}$.

Perhaps we see that $\mathbf{x}_1$ and $\mathbf{x}_2$ are the columns of $\mathbf{P}$. This suggests the general method of transforming a matrix $\mathbf{A}$ to diagonal form $\mathbf{D}$ by using $\mathbf{P} = \mathbf{X}$, the matrix with eigenvectors as columns. ■

By a suitable similarity transformation we can now transform a matrix $\mathbf{A}$ to a diagonal matrix $\mathbf{D}$ whose diagonal entries are the eigenvalues of $\mathbf{A}$:

THEOREM 4 **Diagonalization of a Matrix**

If an $n \times n$ matrix $\mathbf{A}$ has a basis of eigenvectors, then

$$\mathbf{D} = \mathbf{X}^{-1}\mathbf{AX} \tag{5}$$

is diagonal, with the eigenvalues of $\mathbf{A}$ as the entries on the main diagonal. Here $\mathbf{X}$ is the matrix with these eigenvectors as column vectors. Also,

$$\mathbf{D}^m = \mathbf{X}^{-1}\mathbf{A}^m\mathbf{X} \qquad (m = 2, 3, \cdots). \tag{5*}$$

PROOF Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be a basis of eigenvectors of $\mathbf{A}$ for R^n. Let the corresponding eigenvalues of $\mathbf{A}$ be $\lambda_1, \cdots, \lambda_n$, respectively, so that $\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_1, \cdots, \mathbf{A}\mathbf{x}_n = \lambda_n\mathbf{x}_n$. Then $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]$ has rank n, by Theorem 3 in Sec. 7.4. Hence $\mathbf{X}^{-1}$ exists by Theorem 1 in Sec. 7.8. We claim that

$$\mathbf{A}\mathbf{X} = \mathbf{A}[\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n] = [\mathbf{A}\mathbf{x}_1 \quad \cdots \quad \mathbf{A}\mathbf{x}_n] = [\lambda_1\mathbf{x}_1 \quad \cdots \quad \lambda_n\mathbf{x}_n] = \mathbf{X}\mathbf{D} \tag{6}$$

where $\mathbf{D}$ is the diagonal matrix as in (5). The fourth equality in (6) follows by direct calculation. (Try it for $n = 2$ and then for general n.) The third equality uses $\mathbf{A}\mathbf{x}_k = \lambda_k\mathbf{x}_k$. The second equality results if we note that the first column of $\mathbf{AX}$ is $\mathbf{A}$ times the first column of $\mathbf{X}$, which is $\mathbf{x}_1$, and so on. For instance, when $n = 2$ and we write $\mathbf{x}_1 = [x_{11} \quad x_{21}]$, $\mathbf{x}_2 = [x_{12} \quad x_{22}]$, we have

$$\begin{aligned}
\mathbf{A}\mathbf{X} = \mathbf{A}[\mathbf{x}_1 \quad \mathbf{x}_2] &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \\
&= \begin{array}{c} \begin{bmatrix} a_{11}x_{11} + a_{12}x_{21} & a_{11}x_{12} + a_{12}x_{22} \\ a_{21}x_{11} + a_{22}x_{21} & a_{21}x_{12} + a_{22}x_{22} \end{bmatrix} \\ \text{Column 1} \qquad\qquad\qquad \text{Column 2} \end{array} = [\mathbf{A}\mathbf{x}_1 \quad \mathbf{A}\mathbf{x}_2].
\end{aligned}$$

If we multiply (6) by $\mathbf{X}^{-1}$ from the left, we obtain (5). Since (5) is a similarity transformation, Theorem 3 implies that $\mathbf{D}$ has the same eigenvalues as $\mathbf{A}$. Equation (5*) follows if we note that

$$\mathbf{D}^2 = \mathbf{D}\mathbf{D} = (\mathbf{X}^{-1}\mathbf{A}\mathbf{X})(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}) = \mathbf{X}^{-1}\mathbf{A}(\mathbf{X}\mathbf{X}^{-1})\mathbf{A}\mathbf{X} = \mathbf{X}^{-1}\mathbf{A}\mathbf{A}\mathbf{X} = \mathbf{X}^{-1}\mathbf{A}^2\mathbf{X}, \quad \text{etc.} \quad \blacksquare$$

EXAMPLE 4 **Diagonalization**

Diagonalize

$$\mathbf{A} = \begin{bmatrix} 7.3 & 0.2 & -3.7 \\ -11.5 & 1.0 & 5.5 \\ 17.7 & 1.8 & -9.3 \end{bmatrix}.$$

Solution. The characteristic determinant gives the characteristic equation $-\lambda^3 - \lambda^2 + 12\lambda = 0$. The roots (eigenvalues of $\mathbf{A}$) are $\lambda_1 = 3$, $\lambda_2 = -4$, $\lambda_3 = 0$. By the Gauss elimination applied to $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ with $\lambda = \lambda_1, \lambda_2, \lambda_3$ we find eigenvectors and then $\mathbf{X}^{-1}$ by the Gauss–Jordan elimination (Sec. 7.8, Example 1). The results are

$$\begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}, \qquad \mathbf{X}^{-1} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix}.$$

Calculating $\mathbf{AX}$ and multiplying by $\mathbf{X}^{-1}$ from the left, we thus obtain

$$\mathbf{D} = \mathbf{X}^{-1}\mathbf{A}\mathbf{X} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix} \begin{bmatrix} -3 & -4 & 0 \\ 9 & 4 & 0 \\ -3 & -12 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \qquad \blacksquare$$

Quadratic Forms. Transformation to Principal Axes

By definition, a **quadratic form** Q in the components $x_1, \cdots, x_n$ of a vector $\mathbf{x}$ is a sum of n^2 terms, namely,

$$
\textbf{(7)} \qquad
\begin{aligned}
Q = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_j x_k \\
&= \quad a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n \\
&\quad + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\
&\quad + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\quad + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2.
\end{aligned}
$$

$\mathbf{A} = [a_{jk}]$ is called the **coefficient matrix** of the form. We may assume that $\mathbf{A}$ is ***symmetric***, because we can take off-diagonal terms together in pairs and write the result as a sum of two equal terms; see the following example.

EXAMPLE 5 Quadratic Form. Symmetric Coefficient Matrix

Let

$$\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} = [x_1 \quad x_2]\begin{bmatrix} 3 & 4 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 4x_1x_2 + 6x_2x_1 + 2x_2^2 = 3x_1^2 + 10x_1x_2 + 2x_2^2.$$

Here $4 + 6 = 10 = 5 + 5$. From the corresponding ***symmetric*** matrix $\mathbf{C} = [c_{jk}]$, where $c_{jk} = \frac{1}{2}(a_{jk} + a_{kj})$, thus $c_{11} = 3$, $c_{12} = c_{21} = 5$, $c_{22} = 2$, we get the same result; indeed,

$$\mathbf{x}^\mathsf{T}\mathbf{C}\mathbf{x} = [x_1 \quad x_2]\begin{bmatrix} 3 & 5 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 5x_1x_2 + 5x_2x_1 + 2x_2^2 = 3x_1^2 + 10x_1x_2 + 2x_2^2. \qquad \blacksquare$$

Quadratic forms occur in physics and geometry, for instance, in connection with conic sections (ellipses $x_1^2/a^2 + x_2^2/b^2 = 1$, etc.) and quadratic surfaces (cones, etc.). Their transformation to principal axes is an important practical task related to the diagonalization of matrices, as follows.

By Theorem 2, the *symmetric* coefficient matrix $\mathbf{A}$ of (7) has an orthonormal basis of eigenvectors. Hence if we take these as column vectors, we obtain a matrix $\mathbf{X}$ that is orthogonal, so that $\mathbf{X}^{-1} = \mathbf{X}^\mathsf{T}$. From (5) we thus have $\mathbf{A} = \mathbf{X}\mathbf{D}\mathbf{X}^{-1} = \mathbf{X}\mathbf{D}\mathbf{X}^\mathsf{T}$. Substitution into (7) gives

$$\text{(8)} \qquad Q = \mathbf{x}^\mathsf{T}\mathbf{X}\mathbf{D}\mathbf{X}^\mathsf{T}\mathbf{x}.$$

If we set $\mathbf{X}^\mathsf{T}\mathbf{x} = \mathbf{y}$, then, since $\mathbf{X}^\mathsf{T} = \mathbf{X}^{-1}$, we have $\mathbf{X}^{-1}\mathbf{x} = \mathbf{y}$ and thus obtain

$$\text{(9)} \qquad \mathbf{x} = \mathbf{X}\mathbf{y}.$$

Furthermore, in (8) we have $\mathbf{x}^\mathsf{T}\mathbf{X} = (\mathbf{X}^\mathsf{T}\mathbf{x})^\mathsf{T} = \mathbf{y}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{x} = \mathbf{y}$, so that Q becomes simply

$$\textbf{(10)} \qquad Q = \mathbf{y}^\mathsf{T}\mathbf{D}\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.$$

This proves the following basic theorem.

THEOREM 5

Principal Axes Theorem

The substitution (9) *transforms a quadratic form*

$$Q = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_j x_k \qquad (a_{kj} = a_{jk})$$

to the principal axes form or **canonical form** (10), *where* $\lambda_1, \cdots, \lambda_n$ *are the* (*not necessarily distinct*) *eigenvalues of the* (*symmetric!*) *matrix* $\mathbf{A}$, *and* $\mathbf{X}$ *is an orthogonal matrix with corresponding eigenvectors* $\mathbf{x}_1, \cdots, \mathbf{x}_n$, *respectively, as column vectors.*

EXAMPLE 6 **Transformation to Principal Axes. Conic Sections**

Find out what type of conic section the following quadratic form represents and transform it to principal axes:

$$Q = 17x_1^2 - 30x_1x_2 + 17x_2^2 = 128.$$

Solution. We have $Q = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}$, where

$$\mathbf{A} = \begin{bmatrix} 17 & -15 \\ -15 & 17 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

This gives the characteristic equation $(17 - \lambda)^2 - 15^2 = 0$. It has the roots $\lambda_1 = 2$, $\lambda_2 = 32$. Hence (10) becomes

$$Q = 2y_1^2 + 32y_2^2.$$

We see that $Q = 128$ represents the ellipse $2y_1^2 + 32y_2^2 = 128$, that is,

$$\frac{y_1^2}{8^2} + \frac{y_2^2}{2^2} = 1.$$

If we want to know the direction of the principal axes in the x_1x_2-coordinates, we have to determine normalized eigenvectors from $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ with $\lambda = \lambda_1 = 2$ and $\lambda = \lambda_2 = 32$ and then use (9). We get

$$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix},$$

hence

$$\mathbf{x} = \mathbf{X}\mathbf{y} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \qquad \begin{aligned} x_1 &= y_1/\sqrt{2} - y_2/\sqrt{2} \\ x_2 &= y_1/\sqrt{2} + y_2/\sqrt{2}. \end{aligned}$$

This is a 45° rotation. Our results agree with those in Sec. 8.2, Example 1, except for the notations. See also Fig. 160 in that example. ■

PROBLEM SET 8.4

1–5 SIMILAR MATRICES HAVE EQUAL EIGENVALUES

Verify this for $\mathbf{A}$ and $\mathbf{A} = \mathbf{P}^{-1}\mathbf{AP}$. If $\mathbf{y}$ is an eigenvector of $\mathbf{P}$, show that $\mathbf{x} = \mathbf{Py}$ are eigenvectors of $\mathbf{A}$. Show the details of your work.

1. $\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} -4 & 2 \\ 3 & -1 \end{bmatrix}$

2. $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 7 & -5 \\ 10 & -7 \end{bmatrix}$

3. $\mathbf{A} = \begin{bmatrix} 8 & -4 \\ 2 & 2 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 0.28 & 0.96 \\ -0.96 & 0.28 \end{bmatrix}$

4. $\mathbf{A} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 3 & 2 \\ 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 1 & 0 \\ 3 & 0 & 5 \end{bmatrix},$
$\lambda_1 = 3$

5. $\mathbf{A} = \begin{bmatrix} -5 & 0 & 15 \\ 3 & 4 & -9 \\ -5 & 0 & 15 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

6. **PROJECT. Similarity of Matrices.** Similarity is basic, for instance, in designing numeric methods.

   (a) Trace. By definition, the **trace** of an $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ is the sum of the diagonal entries,

   $$\text{trace } \mathbf{A} = a_{11} + a_{22} + \cdots + a_{nn}.$$

   Show that the trace equals the sum of the eigenvalues, each counted as often as its algebraic multiplicity indicates. Illustrate this with the matrices $\mathbf{A}$ in Probs. 1, 3, and 5.

   (b) Trace of product. Let $\mathbf{B} = [b_{jk}]$ be $n \times n$. Show that similar matrices have equal traces, by first proving

   $$\text{trace } \mathbf{AB} = \sum_{i=1}^{n} \sum_{l=1}^{n} a_{il}b_{li} = \text{trace } \mathbf{BA}.$$

   (c) Find a relationship between $\hat{\mathbf{A}}$ in (4) and $\hat{\mathbf{A}} = \mathbf{PAP}^{-1}$.

   (d) Diagonalization. What can you do in (5) if you want to change the order of the eigenvalues in $\mathbf{D}$, for instance, interchange $d_{11} = \lambda_1$ and $d_{22} = \lambda_2$?

7. **No basis.** Find further 2×2 and 3×3 matrices without eigenbasis.

8. **Orthonormal basis.** Illustrate Theorem 2 with further examples.

9–16 DIAGONALIZATION OF MATRICES

Find an eigenbasis (a basis of eigenvectors) and diagonalize. Show the details.

9. $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

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

11. $\begin{bmatrix} -19 & 7 \\ -42 & 16 \end{bmatrix}$

12. $\begin{bmatrix} -4.3 & 7.7 \\ 1.3 & 9.3 \end{bmatrix}$

13. $\begin{bmatrix} 4 & 0 & 0 \\ 12 & -2 & 0 \\ 21 & -6 & 1 \end{bmatrix}$

14. $\begin{bmatrix} -5 & -6 & 6 \\ -9 & -8 & 12 \\ -12 & -12 & 16 \end{bmatrix}, \quad \lambda_1 = -2$

15. $\begin{bmatrix} 4 & 3 & 3 \\ 3 & 6 & 1 \\ 3 & 1 & 6 \end{bmatrix}, \quad \lambda_1 = 10$

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

17–23 PRINCIPAL AXES. CONIC SECTIONS

What kind of conic section (or pair of straight lines) is given by the quadratic form? Transform it to principal axes. Express $\mathbf{x}^\mathsf{T} = [x_1 \quad x_2]$ in terms of the new coordinate vector $\mathbf{y}^\mathsf{T} = [y_1 \quad y_2]$, as in Example 6.

17. $7x_1^2 + 6x_1x_2 + 7x_2^2 = 200$

18. $3x_1^2 + 8x_1x_2 - 3x_2^2 = 10$

19. $3x_1^2 + 22x_1x_2 + 3x_2^2 = 0$

20. $9x_1^2 + 6x_1x_2 + x_2^2 = 10$

21. $x_1^2 - 12x_1x_2 + x_2^2 = 70$

22. $4x_1^2 + 12x_1x_2 + 13x_2^2 = 16$

23. $-11x_1^2 + 84x_1x_2 + 24x_2^2 = 156$

24. Definiteness. A quadratic form $Q(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}$ and its (symmetric!) matrix $\mathbf{A}$ are called **(a) positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$, **(b) negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$, **(c) indefinite** if $Q(\mathbf{x})$ takes both positive and negative values. (See Fig. 162.) [$Q(\mathbf{x})$ and $\mathbf{A}$ are called *positive semidefinite* (*negative semidefinite*) if $Q(\mathbf{x}) \geqq 0$ ($Q(\mathbf{x}) \leqq 0$) for all $\mathbf{x}$.] Show that a necessary and sufficient condition for (a), (b), and (c) is that the eigenvalues of $\mathbf{A}$ are (a) all positive, (b) all negative, and (c) both positive and negative.

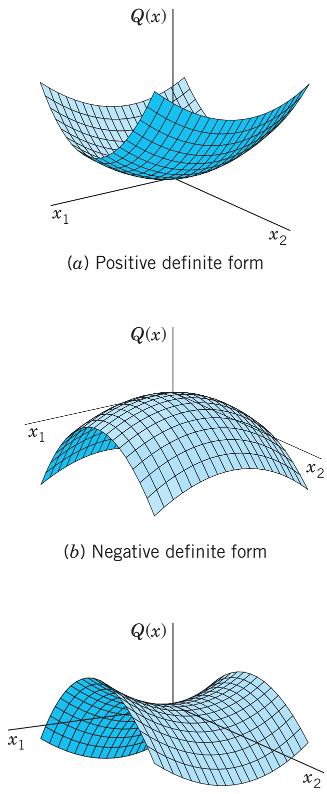
Hint. Use Theorem 5.

25. Definiteness. A necessary and sufficient condition for positive definiteness of a quadratic form $Q(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}$ with *symmetric* matrix $\mathbf{A}$ is that all the **principal minors** are positive (see Ref. [B3], vol. 1, p. 306), that is,

$$a_{11} > 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} > 0,$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} > 0, \qquad \cdots, \qquad \det \mathbf{A} > 0.$$

Show that the form in Prob. 22 is positive definite, whereas that in Prob. 23 is indefinite.



(*a*) Positive definite form

(*b*) Negative definite form

(*c*) Indefinite form

Fig. 162. Quadratic forms in two variables (Problem 24)

8.5 Complex Matrices and Forms. *Optional*

The three classes of matrices in Sec. 8.3 have complex counterparts which are of practical interest in certain applications, for instance, in quantum mechanics. This is mainly because of their spectra as shown in Theorem 1 in this section. The second topic is about extending quadratic forms of Sec. 8.4 to complex numbers. (The reader who wants to brush up on complex numbers may want to consult Sec. 13.1.)

Notations

$\overline{\mathbf{A}} = [\overline{a}_{jk}]$ is obtained from $\mathbf{A} = [a_{jk}]$ by replacing each entry $a_{jk} = \alpha + i\beta$ (α, β real) with its complex conjugate $\overline{a}_{jk} = \alpha - i\beta$. Also, $\overline{\mathbf{A}}^\mathsf{T} = [\overline{a}_{kj}]$ is the transpose of $\overline{\mathbf{A}}$, hence the conjugate transpose of $\mathbf{A}$.

EXAMPLE 1 Notations

If $\mathbf{A} = \begin{bmatrix} 3+4i & 1-i \\ 6 & 2-5i \end{bmatrix}$, then $\overline{\mathbf{A}} = \begin{bmatrix} 3-4i & 1+i \\ 6 & 2+5i \end{bmatrix}$ and $\overline{\mathbf{A}}^\mathsf{T} = \begin{bmatrix} 3-4i & 6 \\ 1+i & 2+5i \end{bmatrix}$. ■

DEFINITION

Hermitian, Skew-Hermitian, and Unitary Matrices

A square matrix $\mathbf{A} = [a_{kj}]$ is called

Hermitian	if $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}$,	that is,	$\bar{a}_{kj} = a_{jk}$
skew-Hermitian	if $\overline{\mathbf{A}}^{\mathsf{T}} = -\mathbf{A}$,	that is,	$\bar{a}_{kj} = -a_{jk}$
unitary	if $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{-1}$.		

The first two classes are named after Hermite (see footnote 13 in Problem Set 5.8).

From the definitions we see the following. If $\mathbf{A}$ is Hermitian, the entries on the main diagonal must satisfy $\bar{a}_{jj} = a_{jj}$; that is, they are real. Similarly, if $\mathbf{A}$ is skew-Hermitian, then $\bar{a}_{jj} = -a_{jj}$. If we set $a_{jj} = \alpha + i\beta$, this becomes $\alpha - i\beta = -(\alpha + i\beta)$. Hence $\alpha = 0$, so that a_{jj} must be pure imaginary or 0.

EXAMPLE 2 Hermitian, Skew-Hermitian, and Unitary Matrices

$$\mathbf{A} = \begin{bmatrix} 4 & 1 - 3i \\ 1 + 3i & 7 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 3i & 2 + i \\ -2 + i & -i \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} \frac{1}{2}i & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2}i \end{bmatrix}$$

are Hermitian, skew-Hermitian, and unitary matrices, respectively, as you may verify by using the definitions. ■

If a Hermitian matrix is real, then $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} = \mathbf{A}$. Hence a real Hermitian matrix is a symmetric matrix (Sec. 8.3).

Similarly, if a skew-Hermitian matrix is real, then $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} = -\mathbf{A}$. Hence a real skew-Hermitian matrix is a skew-symmetric matrix.

Finally, if a unitary matrix is real, then $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} = \mathbf{A}^{-1}$. Hence a real unitary matrix is an orthogonal matrix.

This shows that *Hermitian, skew-Hermitian, and unitary matrices generalize symmetric, skew-symmetric, and orthogonal matrices, respectively.*

Eigenvalues

It is quite remarkable that the matrices under consideration have spectra (sets of eigenvalues; see Sec. 8.1) that can be characterized in a general way as follows (see Fig. 163).

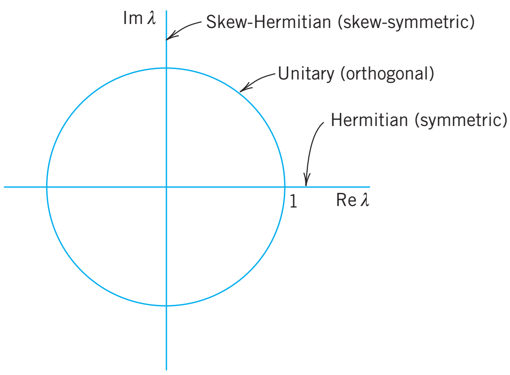


Fig. 163. Location of the eigenvalues of Hermitian, skew-Hermitian, and unitary matrices in the complex λ-plane

THEOREM 1

Eigenvalues

(a) *The eigenvalues of a Hermitian matrix (and thus of a symmetric matrix) are real.*

(b) *The eigenvalues of a skew-Hermitian matrix (and thus of a skew-symmetric matrix) are pure imaginary or zero.*

(c) *The eigenvalues of a unitary matrix (and thus of an orthogonal matrix) have absolute value* 1.

EXAMPLE 3

Illustration of Theorem 1

For the matrices in Example 2 we find by direct calculation

Matrix		Characteristic Equation	Eigenvalues
A	Hermitian	$\lambda^2 - 11\lambda + 18 = 0$	$9, \quad 2$
B	Skew-Hermitian	$\lambda^2 - 2i\lambda + 8 = 0$	$4i, \quad -2i$
C	Unitary	$\lambda^2 - i\lambda - 1 = 0$	$\frac{1}{2}\sqrt{3} + \frac{1}{2}i, \quad -\frac{1}{2}\sqrt{3} + \frac{1}{2}i$

and $|\pm\frac{1}{2}\sqrt{3} + \frac{1}{2}i|^2 = \frac{3}{4} + \frac{1}{4} = 1.$ ■

PROOF We prove Theorem 1. Let λ be an eigenvalue and $\mathbf{x}$ an eigenvector of $\mathbf{A}$. Multiply $\mathbf{Ax} = \lambda\mathbf{x}$ from the left by $\bar{\mathbf{x}}^{\mathsf{T}}$, thus $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax} = \lambda\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}$, and divide by $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x} = \bar{x}_1 x_1 + \cdots + \bar{x}_n x_n = |x_1|^2 + \cdots + |x_n|^2$, which is real and not 0 because $\mathbf{x} \neq \mathbf{0}$. This gives

$$\lambda = \frac{\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax}}{\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}}. \tag{1}$$

(a) If $\mathbf{A}$ is Hermitian, $\bar{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}$ or $\mathbf{A}^{\mathsf{T}} = \bar{\mathbf{A}}$ and we show that then the numerator in (1) is real, which makes λ real. $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax}$ is a scalar; hence taking the transpose has no effect. Thus

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax} = (\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax})^{\mathsf{T}} = \mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\bar{\mathbf{x}} = \mathbf{x}^{\mathsf{T}}\bar{\mathbf{A}}\bar{\mathbf{x}} = (\overline{\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax}}). \tag{2}$$

Hence, $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax}$ equals its complex conjugate, so that it must be real. ($a + ib = a - ib$ implies $b = 0$.)

(b) If $\mathbf{A}$ is skew-Hermitian, $\mathbf{A}^{\mathsf{T}} = -\bar{\mathbf{A}}$ and instead of (2) we obtain

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax} = -(\overline{\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax}}) \tag{3}$$

so that $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Ax}$ equals minus its complex conjugate and is pure imaginary or 0. ($a + ib = -(a - ib)$ implies $a = 0$.)

(c) Let $\mathbf{A}$ be unitary. We take $\mathbf{Ax} = \lambda\mathbf{x}$ and its conjugate transpose

$$(\overline{\mathbf{A}\mathbf{x}})^{\mathsf{T}} = (\overline{\lambda\mathbf{x}})^{\mathsf{T}} = \bar{\lambda}\bar{\mathbf{x}}^{\mathsf{T}}$$

and multiply the two left sides and the two right sides,

$$(\overline{\mathbf{A}\mathbf{x}})^{\mathsf{T}}\mathbf{Ax} = \bar{\lambda}\lambda\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x} = |\lambda|^2\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}.$$

But $\mathbf{A}$ is unitary, $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{-1}$, so that on the left we obtain

$$(\overline{\mathbf{A}}\overline{\mathbf{x}})^{\mathsf{T}}\mathbf{A}\mathbf{x} = \overline{\mathbf{x}}^{\mathsf{T}}\overline{\mathbf{A}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \overline{\mathbf{x}}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \overline{\mathbf{x}}^{\mathsf{T}}\mathbf{I}\mathbf{x} = \overline{\mathbf{x}}^{\mathsf{T}}\mathbf{x}.$$

Together, $\overline{\mathbf{x}}^{\mathsf{T}}\mathbf{x} = |\lambda|^2\overline{\mathbf{x}}^{\mathsf{T}}\mathbf{x}$. We now divide by $\overline{\mathbf{x}}^{\mathsf{T}}\mathbf{x}$ $(\neq 0)$ to get $|\lambda|^2 = 1$. Hence $|\lambda| = 1$. This proves Theorem 1 as well as Theorems 1 and 5 in Sec. 8.3. ■

Key properties of orthogonal matrices (invariance of the inner product, orthonormality of rows and columns; see Sec. 8.3) generalize to unitary matrices in a remarkable way.

To see this, instead of R^n we now use the **complex vector space** C^n of all complex vectors with n complex numbers as components, and complex numbers as scalars. For such complex vectors the **inner product** is defined by (note the overbar for the complex conjugate)

$$\mathbf{a} \bullet \mathbf{b} = \overline{\mathbf{a}}^{\mathsf{T}}\mathbf{b}. \tag{4}$$

The **length** or **norm** of such a complex vector is a *real* number defined by

$$\|\mathbf{a}\| = \sqrt{\mathbf{a} \bullet \mathbf{a}} = \sqrt{\overline{\mathbf{a}}_j^{\mathsf{T}}\mathbf{a}} = \sqrt{\overline{a}_1 a_1 + \cdots + \overline{a}_n a_n} = \sqrt{|a_1|^2 + \cdots + |a_n|^2}. \tag{5}$$

THEOREM 2

Invariance of Inner Product

A **unitary transformation**, *that is,* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *with a unitary matrix* $\mathbf{A}$, *preserves the value of the inner product* (4), *hence also the norm* (5).

PROOF The proof is the same as that of Theorem 2 in Sec. 8.3, which the theorem generalizes. In the analog of (9), Sec. 8.3, we now have bars,

$$\mathbf{u} \bullet \mathbf{v} = \overline{\mathbf{u}}^{\mathsf{T}}\mathbf{v} = (\overline{\mathbf{A}}\overline{\mathbf{a}})^{\mathsf{T}}\mathbf{A}\mathbf{b} = \overline{\mathbf{a}}^{\mathsf{T}}\overline{\mathbf{A}}^{\mathsf{T}}\mathbf{A}\mathbf{b} = \overline{\mathbf{a}}^{\mathsf{T}}\mathbf{I}\mathbf{b} = \overline{\mathbf{a}}^{\mathsf{T}}\mathbf{b} = \mathbf{a} \bullet \mathbf{b}.$$

The complex analog of an orthonormal system of real vectors (see Sec. 8.3) is defined as follows.

DEFINITION

Unitary System

A *unitary system* is a set of complex vectors satisfying the relationships

$$\mathbf{a}_j \bullet \mathbf{a}_k = \overline{\mathbf{a}}_j^{\mathsf{T}}\mathbf{a}_k = \begin{cases} 0 & \text{if} \quad j \neq k \\ 1 & \text{if} \quad j = k. \end{cases} \tag{6}$$

Theorem 3 in Sec. 8.3 extends to complex as follows.

THEOREM 3

Unitary Systems of Column and Row Vectors

A complex square matrix is unitary if and only if its column vectors (and also its row vectors) form a unitary system.

PROOF The proof is the same as that of Theorem 3 in Sec. 8.3, except for the bars required in $\bar{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{-1}$ and in (4) and (6) of the present section.

THEOREM 4

Determinant of a Unitary Matrix

Let $\mathbf{A}$ *be a unitary matrix. Then its determinant has absolute value one, that is,* $|\det \mathbf{A}| = 1$.

PROOF Similarly, as in Sec. 8.3, we obtain

$$\begin{aligned} 1 &= \det(\mathbf{A}\mathbf{A}^{-1}) = \det(\mathbf{A}\bar{\mathbf{A}}^{\mathsf{T}}) = \det \mathbf{A} \det \bar{\mathbf{A}}^{\mathsf{T}} = \det \mathbf{A} \det \bar{\mathbf{A}} \\ &= \det \mathbf{A}\, \overline{\det \mathbf{A}} = |\det \mathbf{A}|^2. \end{aligned}$$

Hence $|\det \mathbf{A}| = 1$ (where $\det \mathbf{A}$ may now be complex). ■

EXAMPLE 4

Unitary Matrix Illustrating Theorems 1c and 2–4

For the vectors $\mathbf{a}^{\mathsf{T}} = [2 \quad -i]$ and $\mathbf{b}^{\mathsf{T}} = [1+i \quad 4i]$ we get $\bar{\mathbf{a}}^{\mathsf{T}} = [2 \quad i]^{\mathsf{T}}$ and $\bar{\mathbf{a}}^{\mathsf{T}}\mathbf{b} = 2(1+i) - 4 = -2 + 2i$ and with

$$\mathbf{A} = \begin{bmatrix} 0.8i & 0.6 \\ 0.6 & 0.8i \end{bmatrix} \quad \text{also} \quad \mathbf{A}\mathbf{a} = \begin{bmatrix} i \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{A}\mathbf{b} = \begin{bmatrix} -0.8 + 3.2i \\ -2.6 + 0.6i \end{bmatrix},$$

as one can readily verify. This gives $(\overline{\mathbf{A}\mathbf{a}})^{\mathsf{T}}\mathbf{A}\mathbf{b} = -2 + 2i$, illustrating Theorem 2. The matrix is unitary. Its columns form a unitary system,

$$\bar{\mathbf{a}}_1^{\mathsf{T}}\mathbf{a}_1 = -0.8i \cdot 0.8i + 0.6^2 = 1, \qquad \bar{\mathbf{a}}_1^{\mathsf{T}}\mathbf{a}_2 = -0.8i \cdot 0.6 + 0.6 \cdot 0.8i = 0,$$
$$\bar{\mathbf{a}}_2^{\mathsf{T}}\mathbf{a}_2 = 0.6^2 + (-0.8i)0.8i = 1$$

and so do its rows. Also, $\det \mathbf{A} = -1$. The eigenvalues are $0.6 + 0.8i$ and $-0.6 + 0.8i$, with eigenvectors $[1 \quad 1]^{\mathsf{T}}$ and $[1 \quad -1]^{\mathsf{T}}$, respectively. ■

Theorem 2 in Sec. 8.4 on the existence of an eigenbasis extends to complex matrices as follows.

THEOREM 5

Basis of Eigenvectors

A Hermitian, skew-Hermitian, or unitary matrix has a basis of eigenvectors for C^n *that is a unitary system.*

For a proof see Ref. [B3], vol. 1, pp. 270–272 and p. 244 (Definition 2).

EXAMPLE 5

Unitary Eigenbases

The matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ in Example 2 have the following unitary systems of eigenvectors, as you should verify.

$$\mathbf{A}: \quad \frac{1}{\sqrt{35}}[1-3i \quad 5]^{\mathsf{T}} \quad (\lambda = 9), \qquad \frac{1}{\sqrt{14}}[1-3i \quad -2]^{\mathsf{T}} \quad (\lambda = 2)$$

$$\mathbf{B}: \quad \frac{1}{\sqrt{30}}[1-2i \quad -5]^{\mathsf{T}} \quad (\lambda = -2i), \qquad \frac{1}{\sqrt{30}}[5 \quad 1+2i]^{\mathsf{T}} \quad (\lambda = 4i)$$

$$\mathbf{C}: \quad \frac{1}{\sqrt{2}}[1 \quad 1]^{\mathsf{T}} \quad (\lambda = \tfrac{1}{2}(i + \sqrt{3})), \qquad \frac{1}{\sqrt{2}}[1 \quad -1]^{\mathsf{T}} \quad (\lambda = \tfrac{1}{2}(i - \sqrt{3})).$$ ■

Hermitian and Skew-Hermitian Forms

The concept of a quadratic form (Sec. 8.4) can be extended to complex. We call the numerator $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x}$ in (1) a **form** in the components $x_1, \cdots, x_n$ of $\mathbf{x}$, which may now be complex. This form is again a sum of n^2 terms

$$
\begin{aligned}
\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}\bar{x}_j x_k \\
&= \quad a_{11}\bar{x}_1x_1 + \cdots + a_{1n}\bar{x}_1x_n \\
&\quad + a_{21}\bar{x}_2x_1 + \cdots + a_{2n}\bar{x}_2x_n \\
&\quad + \cdots\cdots\cdots\cdots\cdots\cdots \\
&\quad + a_{n1}\bar{x}_nx_1 + \cdots + a_{nn}\bar{x}_nx_n.
\end{aligned} \tag{7}
$$

$\mathbf{A}$ is called its **coefficient matrix**. The form is called a **Hermitian** or **skew-Hermitian form** if $\mathbf{A}$ is Hermitian or skew-Hermitian, respectively. *The value of a Hermitian form is real, and that of a skew-Hermitian form is pure imaginary or zero.* This can be seen directly from (2) and (3) and accounts for the importance of these forms in physics. Note that (2) and (3) are valid for any vectors because, in the proof of (2) and (3), we did not use that $\mathbf{x}$ is an eigenvector but only that $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}$ is real and not 0.

EXAMPLE 6 Hermitian Form

For $\mathbf{A}$ in Example 2 and, say, $\mathbf{x} = [1+i \quad 5i]^{\mathsf{T}}$ we get

$$
\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = [1-i \quad -5i]\begin{bmatrix} 4 & 1-3i \\ 1+3i & 7 \end{bmatrix}\begin{bmatrix} 1+i \\ 5i \end{bmatrix} = [1-i \quad -5i]\begin{bmatrix} 4(1+i)+(1-3i)\cdot 5i \\ (1+3i)(1+i)+7\cdot 5i \end{bmatrix} = 223. \quad \blacksquare
$$

Clearly, if $\mathbf{A}$ and $\mathbf{x}$ in (4) are real, then (7) reduces to a quadratic form, as discussed in the last section.

PROBLEM SET 8.5

1–6 EIGENVALUES AND VECTORS

Is the given matrix Hermitian? Skew-Hermitian? Unitary? Find its eigenvalues and eigenvectors.

1. $\begin{bmatrix} 6 & i \\ -i & 6 \end{bmatrix}$ **2.** $\begin{bmatrix} i & 1+i \\ -1+i & 0 \end{bmatrix}$

3. $\begin{bmatrix} \frac{1}{2} & i\sqrt{\frac{3}{4}} \\ i\sqrt{\frac{3}{4}} & \frac{1}{2} \end{bmatrix}$ **4.** $\begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$

5. $\begin{bmatrix} i & 0 & 0 \\ 0 & 0 & i \\ 0 & i & 0 \end{bmatrix}$ **6.** $\begin{bmatrix} 0 & 2+2i & 0 \\ 2-2i & 0 & 2+2i \\ 0 & 2-2i & 0 \end{bmatrix}$

7. Pauli spin matrices. Find the eigenvalues and eigenvectors of the so-called *Pauli spin matrices* and show that $\mathbf{S}_x\mathbf{S}_y = i\mathbf{S}_z$, $\mathbf{S}_y\mathbf{S}_x = -i\mathbf{S}_z$, $\mathbf{S}_x^2 = \mathbf{S}_y^2 = \mathbf{S}_z^2 = \mathbf{I}$, where

$$
\mathbf{S}_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{S}_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},
$$

$$
\mathbf{S}_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.
$$

8. Eigenvectors. Find eigenvectors of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ in Examples 2 and 3.

9–12 COMPLEX FORMS

Is the matrix **A** Hermitian or skew-Hermitian? Find $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x}$. Show the details.

9. $\mathbf{A} = \begin{bmatrix} 4 & 3-2i \\ 3+2i & -4 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -4i \\ 2+2i \end{bmatrix}$

10. $\mathbf{A} = \begin{bmatrix} i & -2+3i \\ 2+3i & 0 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 2i \\ 8 \end{bmatrix}$

11. $\mathbf{A} = \begin{bmatrix} i & 1 & 2+i \\ -1 & 0 & 3i \\ -2+i & 3i & i \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ i \\ -i \end{bmatrix}$

12. $\mathbf{A} = \begin{bmatrix} 1 & i & 4 \\ -i & 3 & 0 \\ 4 & 0 & 2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ i \\ -i \end{bmatrix}$

13–20 GENERAL PROBLEMS

13. Product. Show that $\overline{(\mathbf{ABC})}^{\mathsf{T}} = -\mathbf{C}^{-1}\mathbf{BA}$ for any $n \times n$ Hermitian **A**, skew-Hermitian **B**, and unitary **C**.

14. Product. Show $\overline{(\mathbf{BA})}^{\mathsf{T}} = -\mathbf{AB}$ for **A** and **B** in Example 2. For any $n \times n$ Hermitian **A** and skew-Hermitian **B**.

15. Decomposition. Show that any square matrix may be written as the sum of a Hermitian and a skew-Hermitian matrix. Give examples.

16. Unitary matrices. Prove that the product of two unitary $n \times n$ matrices and the inverse of a unitary matrix are unitary. Give examples.

17. Powers of unitary matrices in applications may sometimes be very simple. Show that $\mathbf{C}^{12} = \mathbf{I}$ in Example 2. Find further examples.

18. Normal matrix. This important concept denotes a matrix that commutes with its conjugate transpose, $\mathbf{A}\bar{\mathbf{A}}^{\mathsf{T}} = \bar{\mathbf{A}}^{\mathsf{T}}\mathbf{A}$. Prove that Hermitian, skew-Hermitian, and unitary matrices are normal. Give corresponding examples of your own.

19. Normality criterion. Prove that **A** is normal if and only if the Hermitian and skew-Hermitian matrices in Prob. 18 commute.

20. Find a simple matrix that is not normal. Find a normal matrix that is not Hermitian, skew-Hermitian, or unitary.

CHAPTER 8 REVIEW QUESTIONS AND PROBLEMS

1. In solving an eigenvalue problem, what is given and what is sought?
2. Give a few typical applications of eigenvalue problems.
3. Do there exist square matrices without eigenvalues?
4. Can a real matrix have complex eigenvalues? Can a complex matrix have real eigenvalues?
5. Does a 5×5 matrix always have a real eigenvalue?
6. What is algebraic multiplicity of an eigenvalue? Defect?
7. What is an eigenbasis? When does it exist? Why is it important?
8. When can we expect orthogonal eigenvectors?
9. State the definitions and main properties of the three classes of real matrices and of complex matrices that we have discussed.
10. What is diagonalization? Transformation to principal axes?

11–15 SPECTRUM

Find the eigenvalues. Find the eigenvectors.

11. $\begin{bmatrix} 2.5 & 0.5 \\ 0.5 & 2.5 \end{bmatrix}$ **12.** $\begin{bmatrix} -7 & 4 \\ -12 & 7 \end{bmatrix}$

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

14. $\begin{bmatrix} 7 & 2 & -1 \\ 2 & 7 & 1 \\ -1 & 1 & 8.5 \end{bmatrix}$

15. $\begin{bmatrix} 0 & -3 & -6 \\ 3 & 0 & -6 \\ 6 & 6 & 0 \end{bmatrix}$

16–17 SIMILARITY

Verify that **A** and $\hat{\mathbf{A}} = \mathbf{p}^{-1}\mathbf{AP}$ have the same spectrum.

16. $\mathbf{A} = \begin{bmatrix} 19 & 12 \\ 12 & 1 \end{bmatrix}$, $\mathbf{P} = \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}$

17. $\mathbf{A} = \begin{bmatrix} 7 & -4 \\ 12 & -7 \end{bmatrix}$, $\mathbf{P} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

18. $\mathbf{A} = \begin{bmatrix} -4 & 6 & 6 \\ 0 & 2 & 0 \\ -1 & 1 & 1 \end{bmatrix}$, $\mathbf{P} = \begin{bmatrix} 1 & 8 & -7 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$

19–21 **DIAGONALIZATION**

Find an eigenbasis and diagonalize.

9. $\begin{bmatrix} -1.4 & 1.0 \\ -1.0 & 1.1 \end{bmatrix}$

20. $\begin{bmatrix} 72 & -56 \\ -56 & 513 \end{bmatrix}$

21. $\begin{bmatrix} -12 & 22 & 6 \\ 8 & 2 & 6 \\ -8 & 20 & 16 \end{bmatrix}$

22–25 **CONIC SECTIONS. PRINCIPAL AXES**

Transform to canonical form (to principal axes). Express $[x_1 \quad x_2]^\mathsf{T}$ in terms of the new variables $[y_1 \quad y_2]^\mathsf{T}$.

22. $9x_1^2 - 6x_1x_2 + 17x_2^2 = 36$

23. $4x_1^2 + 24x_1x_2 - 14x_2^2 = 20$

24. $5x_1^2 + 24x_1x_2 - 5x_2^2 = 0$

25. $3.7x_1^2 + 3.2x_1x_2 + 1.3x_2^2 = 4.5$

SUMMARY OF CHAPTER 8

Linear Algebra: Matrix Eigenvalue Problems

The practical importance of matrix eigenvalue problems can hardly be overrated. The problems are defined by the vector equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \tag{1}$$

$\mathbf{A}$ is a given square matrix. All matrices in this chapter are ***square***. λ is a scalar. To *solve* the problem (1) means to determine values of λ, called **eigenvalues** (or **characteristic values**) of $\mathbf{A}$, such that (1) has a nontrivial solution $\mathbf{x}$ (that is, $\mathbf{x} \neq \mathbf{0}$), called an **eigenvector** of $\mathbf{A}$ corresponding to that λ. An $n \times n$ matrix has at least one and at most n numerically different eigenvalues. These are the solutions of the **characteristic equation** (Sec. 8.1)

$$D(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{2}$$

$D(\lambda)$ is called the **characteristic determinant** of $\mathbf{A}$. By expanding it we get the **characteristic polynomial** of $\mathbf{A}$, which is of degree n in λ. Some typical applications are shown in Sec. 8.2.

Section 8.3 is devoted to eigenvalue problems for **symmetric** ($\mathbf{A}^\mathsf{T} = \mathbf{A}$), **skew-symmetric** ($\mathbf{A}^\mathsf{T} = -\mathbf{A}$), and **orthogonal matrices** ($\mathbf{A}^\mathsf{T} = \mathbf{A}^{-1}$). Section 8.4 concerns the diagonalization of matrices and the transformation of quadratic forms to principal axes and its relation to eigenvalues.

Section 8.5 extends Sec. 8.3 to the complex analogs of those real matrices, called **Hermitian** ($\bar{\mathbf{A}}^\mathsf{T} = \mathbf{A}$), **skew-Hermitian** ($\bar{\mathbf{A}}^\mathsf{T} = -\mathbf{A}$), and **unitary matrices** ($\bar{\mathbf{A}}^\mathsf{T} = \mathbf{A}^{-1}$). All the eigenvalues of a Hermitian matrix (and a symmetric one) are real. For a skew-Hermitian (and a skew-symmetric) matrix they are pure imaginary or zero. For a unitary (and an orthogonal) matrix they have absolute value 1.