Copy this, but do no take home. Thanks.
Nando

Chapter 7

Linear Algebra: Matrices, Vectors, Determinants

Linear algebra includes the theory and application of linear systems of equations (briefly called linear systems), linear transformations, and eigenvalue problems, as they arise, for instance, from electrical networks, frameworks in mechanics, curve fitting and other optimization problems, processes in statistics, systems of differential equations, and so on.

Linear algebra makes systematic use of **vectors** and **matrices** (Sec. 7.1) and, to a lesser extent, **determinants** (Sec. 7.8); and the study of properties of matrices is by itself a central task of linear algebra.

A matrix is a rectangular array of numbers. Matrices occur in various problems, for instance, as arrays of coefficients of equations (Sec. 7.4). Matrices (and vectors) are useful because they enable us to consider an array of many numbers as a single object, denote it by a single symbol, and perform calculations with these symbols in a very compact form. The "mathematical shorthand" thus obtained is very elegant and powerful and is suitable for various practical problems. It entered applied mathematics more than 60 years ago and is of increasing importance in various fields.

This chapter has three big parts:

Calculation with matrices, Secs. 7.1–7.3
Systems of linear equations, Secs. 7.4–7.9
Eigenvalue problems, Secs. 7.10–7.14

and a (more abstract) optional section (7.15) on vector and inner product spaces and linear transformations.

Thus we first introduce matrices and vectors and related concepts (Sec. 7.1) and define the algebraic operations for matrices (Secs. 7.2, 7.3). Next we consider linear systems—solution by Gauss elimination in Sec. 7.4, existence of solutions in Sec. 7.6, determinants and Cramer's rule in Secs. 7.8 and 7.9. Then we study eigenvalue problems in general (Secs. 7.10, 7.11) and for important special real matrices (Sec. 7.12) and complex matrices (Sec. 7.13). Finally, we discuss the diagonalization of matrices and the reduction of quadratic forms to principal axes (Sec. 7.14). Other important concepts in this chapter are the rank of a matrix (Secs. 7.5, 7.9) and the inverse of a matrix (Sec. 7.7). Applications of matrices to practical problems are shown throughout the chapter.

NUMERICAL METHODS in Chap. 19 can be studied immediately after the corresponding material in the present chapter.

Prerequisite for this chapter: None.
Sections that may be omitted in a shorter course: 7.12–7.15.
References: Appendix 1, Part B.
Answers to problems: Appendix 2.

7.1 Basic Concepts

The first three sections of this chapter introduce the basic concepts and rules of matrix and vector algebra. The main application to linear systems of equations begins in Sec. 7.4.

A **matrix** is a rectangular array of numbers (or functions) enclosed in brackets. These numbers (or functions) are called *entries* or *elements* of the matrix. For example,

$$(1)\qquad \begin{bmatrix} 2 & 0.4 & 8 \\ 5 & -32 & 0 \end{bmatrix}, \quad \begin{bmatrix} 6 \\ 1 \end{bmatrix}, \quad [a_1 \quad a_2 \quad a_3], \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \begin{bmatrix} e^x & 3x \\ e^{2x} & x^2 \end{bmatrix}$$

are matrices. The first has two "*rows*" (horizontal lines) and three "*columns*" (vertical lines). The second consists of a single column, and we call it a *column vector*. The third consists of a single row, and we call it a *row vector*. The last two are *square matrices*, that is, each has as many rows as columns (two in this case).

Matrices are practical in many applications. For example, in a system of equations such as

$$\begin{aligned} 5x - 2y + z &= 0 \\ 3x + \phantom{2y +{}} 4z &= 0 \end{aligned}$$

the coefficients of the unknowns x, y, z are the entries of the *coefficient matrix*, call it $\mathbf{A}$,

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

which displays these coefficients in the pattern of the equations. Sales figures for three products I, II, III in a store on Monday (M), Tuesday (T), $\cdots$ may for each week be arranged in a matrix

$$\begin{array}{c} \\ \mathbf{A} = \end{array} \begin{array}{c} \begin{matrix} \text{M} & \text{T} & \text{W} & \text{Th} & \text{F} & \text{S} \end{matrix} \\ \begin{bmatrix} 40 & 33 & 81 & 0 & 21 & 47 \\ 0 & 12 & 78 & 50 & 50 & 96 \\ 10 & 0 & 0 & 27 & 43 & 78 \end{bmatrix} \end{array} \begin{matrix} \text{I} \\ \text{II} \\ \text{III} \end{matrix}$$

and if the company has ten stores, we can set up ten 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 for which matrices are feasible? For instance, in transportation or storage problems? Or in recording phone calls, or in listing distances in a network of roads?

General Notations and Concepts

Our discussion suggests the following. We denote matrices by capital boldface letters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \cdots$, or by writing the general entry in brackets; thus, $\mathbf{A} = [a_{jk}]$, and so on. By an **$m \times n$ matrix** (read "m by n matrix") we mean a matrix with m rows, also called **row vectors,** and n columns, also called **column vectors** of the matrix. Thus, an $m \times n$ matrix $\mathbf{A}$ is of the form

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

Hence the matrices in (1) are 2×3, 2×1, 1×3, 2×2, and 2×2.

In the **double-subscript notation** *for the entries, the first subscript always denotes the* **row** *and the second the* **column** *in which the given entry stands.* Thus a_{23} is the entry in the second row and third column.

If $m = n$, we call $\mathbf{A}$ an $n \times n$ **square matrix.** Then its diagonal containing the entries $a_{11}, a_{22}, \cdots, a_{nn}$ is called the **main diagonal** or *principal diagonal* of $\mathbf{A}$. Thus the last two matrices in (1) are square. Square matrices are particularly important, as we shall see.

A **submatrix** of an $m \times n$ matrix $\mathbf{A}$ is a matrix obtained by omitting some rows or columns (or both) from $\mathbf{A}$. For convenience, this includes $\mathbf{A}$ itself (as the matrix obtained by omitting no rows or columns of $\mathbf{A}$).

EXAMPLE 1 **Submatrices of a matrix**

The 2×3 matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

contains three 2×2 submatrices, namely,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix}, \quad \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix},$$

two 1×3 submatrices (the two row vectors), three 2×1 submatrices (the column vectors), six 1×2 submatrices, namely,

$$[a_{11} \quad a_{12}], \quad [a_{11} \quad a_{13}], \quad [a_{12} \quad a_{13}],$$

$$[a_{21} \quad a_{22}], \quad [a_{21} \quad a_{23}], \quad [a_{22} \quad a_{23}],$$

and six 1×1 submatrices, $[a_{11}]$, $[a_{12}]$, $\cdots$, $[a_{23}]$. ∎

Vectors

A **vector** is a matrix that has only one row—then we call it a **row vector**—or only one column—then we call it a **column vector.** In both cases we call its entries **components** and denote the vector by a ***lowercase*** boldface letter such as **a, b,** $\cdots$, or by its general component in brackets, $\mathbf{a} = [a_j]$, and so on. Thus

$$\mathbf{a} = [a_1 \quad a_2 \cdots \quad a_n]$$

is a row vector, and

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

is a column vector. It will depend on our purpose as to which of the two is more practical, but we often want to switch from one type of vector to the other. We can do this by "**transposition,**" which is indicated by $^{\mathsf{T}}$; thus, if

$$\mathbf{b} = \begin{bmatrix} 4 \\ 0 \\ -7 \end{bmatrix}, \quad \text{then} \quad \mathbf{b}^{\mathsf{T}} = [4 \quad 0 \quad -7];$$

Conversely, if

$$\mathbf{a} = [5 \quad 3 \quad \tfrac{1}{2}], \quad \text{then} \quad \mathbf{a}^{\mathsf{T}} = \begin{bmatrix} 5 \\ 3 \\ \tfrac{1}{2} \end{bmatrix}.$$

Transposition

It is practical to define **transposition** for any matrix. The **transpose** $\mathbf{A}^{\mathsf{T}}$ of an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ as given in (2) is the $n \times m$ matrix that has the first *row* of **A** as its first *column*, the second *row* of **A** as its second *colun*. and so on. Thus the transpose of **A** in (2) is

$$\text{(3)} \qquad \mathbf{A}^{\mathsf{T}} = [a_{kj}] = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdot & \cdot & \cdots & \cdot \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

EXAMPLE 2 **Transposition of a matrix**

If

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

■

Symmetric matrices and **skew-symmetric matrices** are square matrices whose transpose equals the matrix or minus the matrix, respectively:

$$\mathbf{A}^{\mathsf{T}} = \mathbf{A} \quad \text{(symmetric matrix)}, \qquad \mathbf{A}^{\mathsf{T}} = -\mathbf{A} \quad \text{(skew-symmetric matrix)}.$$

These matrices are quite important, and we shall use them often in this chapter.

Rules of matrix calculation follow in the next section and problems at the end of it.

7.2 Matrix Addition, Scalar Multiplication

What makes matrices and vectors really useful is the fact that we can calculate with them almost as easily as with numbers. Indeed, practical applications suggested the rules of addition and multiplication by scalars (numbers), which we now introduce. (Multiplication of matrices by matrices follows in the next section.)

We say briefly that two matrices have the **same size** if they are both $m \times n$, for instance, both 3×4. We begin by defining 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.

EXAMPLE 1 **Equality of matrices**

The definition implies that

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

A cannot be equal to, say, a 2×3 matrix. A column vector cannot be equal to a row vector, by the very definition of equality. ∎

We shall now define two algebraic operations, called *matrix addition* and *scalar multiplication*, which turn out to be practical and very useful in applications, as we shall see later in this chapter.

Definition. Addition of matrices

Addition is defined only for matrices $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ of the same and their **sum,** written $\mathbf{A} + \mathbf{B}$, is then obtained by adding the corresponding entries. 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 2 **Addition of matrices and vectors**

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

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

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), written $c\mathbf{A}$, is the $m \times n$ matrix $c\mathbf{A} = [ca_{jk}]$ obtained by multiplying each entry in $\mathbf{A}$ 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 3 **Scalar multiplication**

If

$$\mathbf{A} = \begin{bmatrix} 2.7 & -1.8 \\ 0 & 0.9 \\ 9.0 & -4.5 \end{bmatrix},$$

then

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

An $m \times n$ **zero matrix** is an $m \times n$ matrix with all entries zero. It is denoted by $\mathbf{0}$. The last matrix in Example 3 is the 3×2 zero matrix.

From the definition we see that matrix addition enjoys properties quite similar to those of the addition of real numbers; namely, for matrices of the same size we have

$$(1)\quad \begin{aligned} &(a) && \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \\ &(b) && (\mathbf{U} + \mathbf{V}) + \mathbf{W} = \mathbf{U} + (\mathbf{V} + \mathbf{W}) \qquad (\text{written } \mathbf{U} + \mathbf{V} + \mathbf{W}) \\ &(c) && \mathbf{A} + \mathbf{0} = \mathbf{A} \\ &(d) && \mathbf{A} + (-\mathbf{A}) = \mathbf{0}. \end{aligned}$$

Furthermore, from the definitions of matrix addition and scalar multiplication we also obtain

$$(2)\quad \begin{aligned} &(a) && c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B} \\ &(b) && (c + k)\mathbf{A} = c\mathbf{A} + k\mathbf{A} \\ &(c) && c(k\mathbf{A}) = (ck)\mathbf{A} \qquad (\text{written } ck\mathbf{A}) \\ &(d) && 1\mathbf{A} = \mathbf{A}. \end{aligned}$$

For the transpose (Sec. 7.1) of a sum of two $m \times n$ matrices we have

$$(3)\qquad (\mathbf{A} + \mathbf{B})^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} + \mathbf{B}^{\mathsf{T}},$$

as the reader may prove; also,

$$(4)\qquad (c\mathbf{A})^{\mathsf{T}} = c\mathbf{A}^{\mathsf{T}}.$$

One more algebraic operation, the multiplication of matrices by matrices, follows in the next section. Then we shall be ready for applications.

Problem Set 7.1–7.2

Let $\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & -2 \\ 4 & 5 \end{bmatrix}$, $\mathbf{C} = \begin{bmatrix} 3 & 0 & 2 \\ 4 & 0 & 1 \end{bmatrix}$, $\mathbf{D} = \begin{bmatrix} 6 & 1 & -5 \\ 5 & -2 & 13 \end{bmatrix}$.

Find the following expressions or give reasons why they are undefined.

1. $\mathbf{A} + \mathbf{B}$, $\mathbf{B} + \mathbf{A}$
2. $4\mathbf{A}$, $-3\mathbf{C}$, $3\mathbf{A} - 3\mathbf{B}$, $3(\mathbf{A} - \mathbf{B})$
3. $2\mathbf{C} + 2\mathbf{D}$, $2(\mathbf{C} + \mathbf{D})$
4. $\mathbf{A} + \mathbf{B} + \mathbf{C}$, $\mathbf{C} - \mathbf{D}$
5. $\mathbf{A} - \mathbf{C}$, $\mathbf{A} + 0\mathbf{C}$, $\mathbf{C} + 0\mathbf{A}$
6. $\mathbf{A} + \mathbf{A}^{\mathsf{T}}$, $(\mathbf{A} + \mathbf{B})^{\mathsf{T}}$, $\mathbf{A}^{\mathsf{T}} + \mathbf{B}^{\mathsf{T}}$, $(\mathbf{A}^{\mathsf{T}})^{\mathsf{T}}$
7. $4\mathbf{B} + 8\mathbf{B}^{\mathsf{T}}$, $4(\mathbf{B} + 2\mathbf{B}^{\mathsf{T}})$
8. $(2\mathbf{C})^{\mathsf{T}}$, $2\mathbf{C}^{\mathsf{T}}$, $\mathbf{C} + \mathbf{C}^{\mathsf{T}}$, $\mathbf{C}^{\mathsf{T}} - 2\mathbf{D}^{\mathsf{T}}$

Let $\mathbf{K} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 2 \\ 0 & 2 & 5 \end{bmatrix}$, $\mathbf{L} = \begin{bmatrix} 0 & 2 & -8 \\ -2 & 0 & 6 \\ 8 & -6 & 0 \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 9 \\ 0 \\ 5 \end{bmatrix}$.

Find the following expressions or give reasons why they are undefined.

9. $\mathbf{K} + \mathbf{L}$, $\mathbf{K} - \mathbf{L}$
10. $3(\mathbf{a} - 4\mathbf{b})$, $3\mathbf{a} - 12\mathbf{b}$, $\mathbf{K} + \mathbf{a}$, $\mathbf{a} + \mathbf{a}^{\mathsf{T}}$
11. $\mathbf{K} - \mathbf{K}^{\mathsf{T}}$, $\mathbf{L} + \mathbf{L}^{\mathsf{T}}$, $\mathbf{a}^{\mathsf{T}} + \mathbf{b}^{\mathsf{T}}$
12. $3\mathbf{K} + 4\mathbf{L}$, $6\mathbf{K} + 8\mathbf{L}$
13. $\mathbf{K} + \mathbf{K}^{\mathsf{T}} + \mathbf{L} - \mathbf{L}^{\mathsf{T}}$
14. $6\mathbf{a}^{\mathsf{T}} - 9\mathbf{b}^{\mathsf{T}}$, $3(2\mathbf{a} - 3\mathbf{b})^{\mathsf{T}}$, $3(3\mathbf{b}^{\mathsf{T}} - 2\mathbf{a}^{\mathsf{T}})$

Symmetric and skew-symmetric matrices

15. Show that **K** is symmetric and **L** is skew-symmetric.

16. Show that for a symmetric matrix $\mathbf{A} = [a_{jk}]$ we have $a_{jk} = a_{kj}$.

17. Show that if $\mathbf{A} = [a_{jk}]$ is skew-symmetric, then $a_{jk} = -a_{kj}$, in particular, $a_{jj} = 0$.

18. Write **A** (in Probs. 1–8) as the sum of a symmetric and a skew-symmetric matrix.

19. Write **B** as the sum of a symmetric and a skew-symmetric matrix.

20. Show that if **A** is any square matrix, then $\mathbf{S} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^\mathsf{T})$ is symmetric, $\mathbf{T} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^\mathsf{T})$ is skew-symmetric, and $\mathbf{A} = \mathbf{S} + \mathbf{T}$.

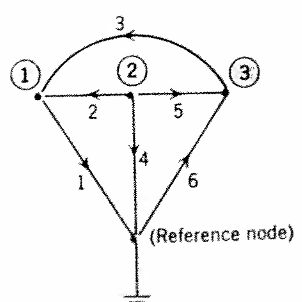
21. Prove (3) and (4) in Sec. 7.2 as well as $(\mathbf{A}^\mathsf{T})^\mathsf{T} = \mathbf{A}$.

Use of matrices in modeling networks. Matrices have various engineering applications, as we shall see. For instance, they may be used to characterize connections (in electrical networks, in nets of roads connecting cities, in production processes, etc.), as follows.

22. **(Nodal incidence matrix)** Figure 131 shows an electrical network having 6 *branches* (connections) and 4 *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

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

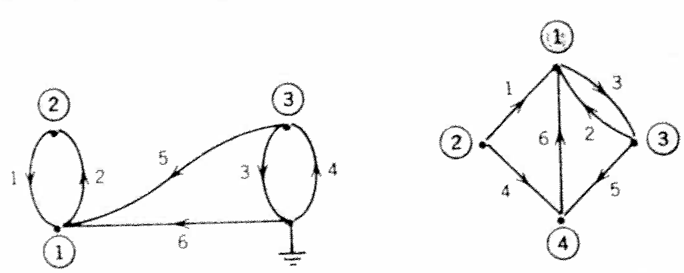
A is called the *nodal incidence matrix* of the network. Show that for the network in Fig. 131, **A** has the given form.



Branch	1	2	3	4	5	6
Node (1)	1	−1	−1	0	0	0
Node (2)	0	1	0	1	1	0
Node (3)	0	0	1	0	−1	−1

Fig. 131. Network and nodal incidence matrix in Prob. 22

23. Find the nodal incidence matrix of the electrical network in Fig. 132A.



(A) Problem 23 **(B) Problem 24**

Fig. 132. Electrical network and net of one-way streets

24. Methods of electrical network analysis have applications in other fields, too. Determine the analog of the nodal incidence matrix for the net of one-way streets (directions as indicated by the arrows) shown in Fig. 132B.

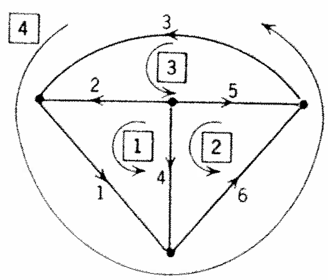
Sketch the network whose nodal incidence matrix is

25. $\begin{bmatrix} 0 & 0 & -1 \\ 1 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix}$ **26.** $\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ **27.** $\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$

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

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

where 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. 133, the matrix **M** has the given form.



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

Fig. 133. Network and matrix **M** in Problem 28

7.3 Matrix Multiplication

As the last algebraic operation we shall now define the multiplication of matrices by matrices. This definition will at first look somewhat artificial, but afterward it will be fully motivated by the use of matrices in linear transformations, by which this multiplication is suggested.

Definition. Multiplication of a matrix by a matrix

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

Number of rows of 2nd factor **B** = Number of columns of 1st factor **A**,

and is then defined as the $m \times p$ matrix $\mathbf{C} = [c_{jk}]$ with entries

$$c_{jk} = \sum_{l=1}^{n} a_{jl}b_{lk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk} \tag{1}$$

(where $j = 1, \cdots, m$ and $k = 1, \cdots, p$); that is, multiply each entry in the jth *row* of **A** by the corresponding entry in the kth *column* of **B** and then add these n products. One says briefly that this is a ***"multiplication of rows into columns."*** Figure 134 illustrates this.

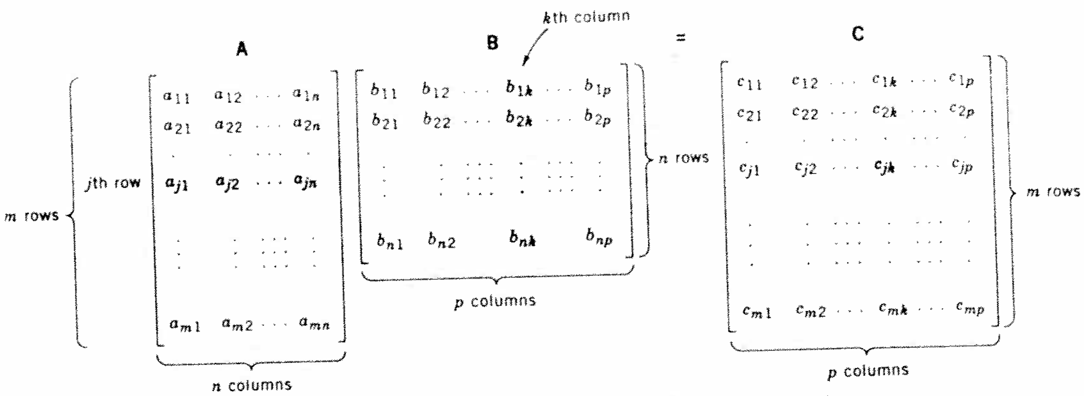


Fig. 134. Matrix multiplication **AB** = **C**

Examples. Properties of Matrix Multiplication

EXAMPLE 1 **Matrix multiplication**

$$\mathbf{AB} = \begin{bmatrix} 4 & 3 \\ 7 & 2 \\ 9 & 0 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 4 \cdot 2 + 3 \cdot 1 & 4 \cdot 5 + 3 \cdot 6 \\ 7 \cdot 2 + 2 \cdot 1 & 7 \cdot 5 + 2 \cdot 6 \\ 9 \cdot 2 + 0 \cdot 1 & 9 \cdot 5 + 0 \cdot 6 \end{bmatrix} = \begin{bmatrix} 11 & 38 \\ 16 & 47 \\ 18 & 45 \end{bmatrix}.$$

Here **A** is 3×2 and **B** is 2×2, so that **AB** comes out 3×2, whereas **BA** is not defined. ∎

EXAMPLE 2 **Multiplication of a matrix and a vector**

$$\begin{bmatrix} 4 & 2 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 12 + 10 \\ 3 + 40 \end{bmatrix} = \begin{bmatrix} 22 \\ 43 \end{bmatrix} \quad \text{whereas} \quad \begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} 4 & 2 \\ 1 & 8 \end{bmatrix} \quad \text{is undefined.}$$ ∎

EXAMPLE 3 **Products of row and column vectors**

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

EXAMPLE 4 **CAUTION! Matrix multiplication is not commutative, AB ≠ BA in general**

This is illustrated by Examples 2 and 3, but also holds for square matrices; for instance,

$$\begin{bmatrix} 9 & 3 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 9 \cdot 1 + 3 \cdot 2 & 9 \cdot (-4) + 3 \cdot 5 \\ -2 \cdot 1 + 0 \cdot 2 & (-2) \cdot (-4) + 0 \cdot 5 \end{bmatrix} = \begin{bmatrix} 15 & -21 \\ -2 & 8 \end{bmatrix}$$

whereas

$$\begin{bmatrix} 1 & -4 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 9 & 3 \\ -2 & 0 \end{bmatrix} = \begin{bmatrix} 1 \cdot 9 + (-4) \cdot (-2) & 1 \cdot 3 + (-4) \cdot 0 \\ 2 \cdot 9 + 5 \cdot (-2) & 2 \cdot 3 + 5 \cdot 0 \end{bmatrix} = \begin{bmatrix} 17 & 3 \\ 8 & 6 \end{bmatrix}.$$ ∎

EXAMPLE 5 **AB = 0 does not necessarily imply A = 0 or B = 0 or BA = 0**

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

We have thus discovered the two properties

$$\text{(2a)} \qquad \mathbf{AB} \neq \mathbf{BA} \qquad \text{in general}$$

and

$$\text{(2b)} \quad \mathbf{AB} = \mathbf{0} \text{ does not necessarily imply } \mathbf{A} = \mathbf{0} \text{ or } \mathbf{B} = \mathbf{0} \text{ or } \mathbf{BA} = \mathbf{0},$$

by which matrix multiplication differs from the multiplication of numbers. Hence, always observe the order of factors very carefully! To emphasize this, we say that in **AB**, the matrix **B** is ***premultiplied,*** or ***multiplied from the left,*** by **A**, and **A** is ***postmultiplied,*** or ***multiplied from the right,*** by **B**. More about (2b) will be said in Sec. 7.7. The other properties of matrix multiplication are similar to those of the multiplication of numbers, namely,

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

provided **A**, **B**, and **C** are such that the expressions on the left are defined; here, k is any scalar.

Special Matrices

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

Triangular matrices

A square matrix whose entries above the main diagonal are all zero is called a *lower triangular matrix*. Similarly, an *upper triangular matrix* is a square matrix whose entries below the main diagonal are all zero. For instance,

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 3 & 0 \\ 5 & 0 & 2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{T}_2 = \begin{bmatrix} 1 & 6 & -1 \\ 0 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}$$

are lower and upper triangular, respectively. An entry *on* the main diagonal of a triangular matrix may be zero or not.

Diagonal matrices

A square matrix $\mathbf{A} = [a_{jk}]$ whose entries above *and* below the main diagonal are all zero, that is, $a_{jk} = 0$ for all $j \neq k$, is called a **diagonal matrix.** For example,

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

are diagonal matrices.

A diagonal matrix whose entries on the main diagonal are all equal is called a **scalar matrix.** Thus a scalar matrix is of the form

$$\mathbf{S} = \begin{bmatrix} c & 0 & \cdots & 0 \\ 0 & c & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & c \end{bmatrix}$$

where c is any number. The name comes from the fact that an $n \times n$ scalar matrix $\mathbf{S}$ commutes with any $n \times n$ matrix $\mathbf{A}$, and the multiplication by $\mathbf{S}$ has the same effect as the multiplication by a scalar,

$$\mathbf{AS} = \mathbf{SA} = c\mathbf{A}. \tag{3}$$

In particular, a scalar matrix whose entries on the main diagonal are all 1 is called a **unit matrix** and is denoted by $\mathbf{I}_n$ or simply by $\mathbf{I}$. For $\mathbf{I}$, formula (3) becomes

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}. \tag{4}$$

For example, the 3×3 unit matrix is

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

Transpose of a Product

The transpose (see Sec. 7.1) *of a product equals the product of the transposed factors, taken in* ***reverse*** *order,*

$$\boxed{(\mathbf{AB})^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}.} \tag{5}$$

The proof of the useful formula (5) follows from the definition of matrix multiplication and is left to the student.

EXAMPLE 6 **Transposition of a product**

Formula (5) is illustrated by

$$(\mathbf{AB})^{\mathsf{T}} = \left(\begin{bmatrix} 4 & 9 \\ 0 & 2 \\ 1 & 6 \end{bmatrix}\begin{bmatrix} 3 & 7 \\ 2 & 8 \end{bmatrix}\right)^{\mathsf{T}} = \begin{bmatrix} 30 & 100 \\ 4 & 16 \\ 15 & 55 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} 30 & 4 & 15 \\ 100 & 16 & 55 \end{bmatrix}.$$

$$\mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}} = \begin{bmatrix} 3 & 2 \\ 7 & 8 \end{bmatrix}\begin{bmatrix} 4 & 0 & 1 \\ 9 & 2 & 6 \end{bmatrix} = \begin{bmatrix} 30 & 4 & 15 \\ 100 & 16 & 55 \end{bmatrix}.$$ ∎

Inner Product of Vectors

This is just a special case of our definition of matrix multiplication, which occurs frequently, so that it pays to give it a special name and notation, as follows.

If **a** and **b** are column vectors with n components, then $\mathbf{a}^{\mathsf{T}}$ is a row vector, and matrix multiplication of these vectors gives a 1×1 matrix, thus a real number, which is called the **inner product** or **dot product** of **a** and **b** and is denoted by $\mathbf{a} \cdot \mathbf{b}$; thus

$$\text{(6)} \quad \mathbf{a} \cdot \mathbf{b} = \mathbf{a}^{\mathsf{T}}\mathbf{b} = [a_1 \cdots a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{l=1}^{n} a_l b_l = a_1 b_1 + \cdots + a_n b_n.$$

Inner products have interesting applications in mechanics and geometry, as we shall see in Sec. 8.2. At present we shall use them to express matrix products in a condensed form, which is often quite useful.

Product in Terms of Row and Column Vectors

Matrix multiplication is a multiplication of rows into columns, as we know, and we can thus write (1) in terms of inner products. Indeed, every entry of $\mathbf{C} = \mathbf{AB}$ is an inner product,

$$c_{11} = \mathbf{a}_1 \cdot \mathbf{b}_1 = (\text{first row of } \mathbf{A}) \cdot (\text{first column of } \mathbf{B})$$

$$c_{12} = \mathbf{a}_1 \cdot \mathbf{b}_2 = (\text{first row of } \mathbf{A}) \cdot (\text{second column of } \mathbf{B})$$

and so on, the general term being

$$\text{(7)} \quad \boxed{c_{jk} = \mathbf{a}_j \cdot \mathbf{b}_k = (j\text{th row of } \mathbf{A}) \cdot (k\text{th column of } \mathbf{B}).}$$

Accordingly, if we write $\mathbf{A}$ in terms of its row vectors,

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}, \qquad \text{where} \qquad \begin{array}{l} \mathbf{a}_1 = [a_{11} \quad a_{12} \quad \cdots \quad a_{1n}] \\ \mathbf{a}_2 = [a_{21} \quad a_{22} \quad \cdots \quad a_{2n}] \\ \qquad \vdots \\ \mathbf{a}_n = [a_{m1} \quad a_{m2} \quad \cdots \quad a_{mn}] \end{array}$$

and $\mathbf{B}$ in terms of its column vectors, $\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_p]$, where

$$\mathbf{b}_1 = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{n2} \end{bmatrix}, \quad \cdots, \quad \mathbf{b}_p = \begin{bmatrix} b_{1p} \\ b_{2p} \\ \vdots \\ b_{np} \end{bmatrix}$$

we see from (1) or (7) that the product $\mathbf{C} = \mathbf{AB}$ can be written

$$\text{(8)} \qquad \mathbf{C} = \mathbf{AB} = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \mathbf{a}_1 \cdot \mathbf{b}_2 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_p \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \mathbf{a}_2 \cdot \mathbf{b}_2 & \cdots & \mathbf{a}_2 \cdot \mathbf{b}_p \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{a}_m \cdot \mathbf{b}_1 & \mathbf{a}_m \cdot \mathbf{b}_2 & \cdots & \mathbf{a}_m \cdot \mathbf{b}_p \end{bmatrix}.$$

This idea sometimes helps in applications to see more clearly what is going on.

Furthermore, $\mathbf{Ab}_1$ is a column vector

$$\mathbf{Ab}_1 = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} \\ \cdot \\ \cdot \\ \cdot \\ b_{n1} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_m \cdot \mathbf{b}_1 \end{bmatrix}$$

and (8) shows that this is the first column of $\mathbf{AB}$. Similarly for the other columns of $\mathbf{AB}$, so that we can write

$$\text{(9)} \qquad \mathbf{AB} = [\mathbf{Ab}_1 \quad \mathbf{Ab}_2 \quad \cdots \quad \mathbf{Ab}_p],$$

a formula that is often useful (for instance, in Sec. 7.14).

EXAMPLE 7 **Product in terms of row and column vectors**

Writing a 2×2 matrix $\mathbf{A}$ in terms of row vectors, say,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} \quad \text{where} \quad \begin{aligned} \mathbf{a}_1 &= [a_{11} \quad a_{12}] \\ \mathbf{a}_2 &= [a_{21} \quad a_{22}] \end{aligned}$$

and a 2×2 matrix $\mathbf{B}$ in terms of column vectors, say,

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = [\mathbf{b}_1 \quad \mathbf{b}_2] \quad \text{where} \quad \mathbf{b}_1 = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} b_{12} \\ b_{22} \end{bmatrix}.$$

we see that (8) takes the form

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} [\mathbf{b}_1 \quad \mathbf{b}_2] = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \mathbf{a}_1 \cdot \mathbf{b}_2 \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \mathbf{a}_2 \cdot \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

Also, $\mathbf{AB} = [\mathbf{Ab}_1 \quad \mathbf{Ab}_2]$ by (9). ∎

Motivation of Matrix Multiplication

Matrix multiplication may look somewhat strange at first sight, but there is a good reason for such an "unnatural" definition, which comes from the use of matrices in connection with "linear transformations." To see this, we consider three coordinate systems in the plane, which we denote as the w_1w_2-system, the x_1x_2-system, and the y_1y_2-system, and we assume that these systems are related by transformations

$$(10) \qquad \begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \end{aligned}$$

and

$$(11) \qquad \begin{aligned} x_1 &= b_{11}w_1 + b_{12}w_2 \\ x_2 &= b_{21}w_1 + b_{22}w_2 \end{aligned}$$

which are (special) *linear transformations*. By substituting (11) into (10) we see that the y_1y_2-coordinates can be obtained directly from the w_1w_2-coordinates by a single linear transformation of the form

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

Now this substitution gives

$$y_1 = a_{11}(b_{11}w_1 + b_{12}w_2) + a_{12}(b_{21}w_1 + b_{22}w_2)$$
$$y_2 = a_{21}(b_{11}w_1 + b_{12}w_2) + a_{22}(b_{21}w_1 + b_{22}w_2).$$

Comparing this with (12), we see that we must have

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} \qquad c_{12} = a_{11}b_{12} + a_{12}b_{22}$$
$$c_{21} = a_{21}b_{11} + a_{22}b_{21} \qquad c_{22} = a_{21}b_{12} + a_{22}b_{22}$$

or briefly

$$(13) \qquad c_{jk} = a_{j1}b_{1k} + a_{j2}b_{2k} = \sum_{i=1}^{2} a_{ji}b_{ik} \qquad j, k = 1, 2.$$

This is (1) with $m = n = p = 2$.

What does our calculation show? Essentially two things. First, matrix multiplication is defined in such a way that linear transformations can be written in compact form, using matrices; in our case, (10) becomes

$$(10^*) \quad \mathbf{y} = \mathbf{Ax} \quad \text{where} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and (11) becomes

$$(11^*) \quad \mathbf{x} = \mathbf{Bw} \quad \text{where} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Second, if we substitute linear transformations into each other, we can obtain the coefficient matrix **C** of the composite transformation (the transformation obtained by the substitution) simply by multiplying the coefficient matrices **A** and **B** of the given transformations, in the right order suggested by the substitution; from (10*), (11*), and (12) we get

$$\mathbf{y} = \mathbf{Ax} = \mathbf{A}(\mathbf{Bw}) = \mathbf{ABw} = \mathbf{Cw}, \qquad \text{where} \qquad \mathbf{C} = \mathbf{AB}.$$

For higher dimensions the idea and the result are exactly the same; only the number of variables changes. We then have m variables $y_1, \cdots, y_m$ and n variables $x_1, \cdots, x_n$ and p variables $w_1, \cdots, w_p$. The matrix **A** is $m \times n$, the matrix **B** is $n \times p$, and **C** is $m \times p$, as in Fig. 134. And the requirement that **C** be the product **AB** leads to formula (1) in its general form. This completely motivates the definition of matrix multiplication.

We shall say more about (general) linear transformations and related matrices in Sec. 7.15, after we have gained more experience with matrices by considering linear systems of equations, beginning in the next section.

An Application of Matrix Multiplication

EXAMPLE 8 **Stochastic matrix. Markov process**

Suppose that the 1993 state of land use in a city of 50 square miles of (nonvacant) area is

I	(Residentially used)	30%
II	(Commercially used)	20%
III	(Industrially used)	50%.

Find the states in 1998 and 2003, assuming that the transition probabilities for 5-year intervals are given by the following matrix $\mathbf{A} = [a_{jk}]$.

$$\begin{array}{c} \\ \text{From I} \\ \text{From II} \\ \text{From III} \end{array} \begin{array}{c} \begin{array}{ccc} \text{To I} & \text{To II} & \text{To III} \end{array} \\ \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0 & 0.1 & 0.9 \end{bmatrix} \end{array}$$

Remark. A square matrix with nonnegative entries and row sums all equal to 1 is called a **stochastic matrix.** Thus **A** is a stochastic matrix. A stochastic process for which the probability of entering a certain state depends only on the *last* state occupied (and on the matrix governing the process) is called a **Markov process.**[1] Thus our example concerns a Markov process.

Solution. From the matrix **A** and the 1993 state we can compute the 1998 state

I	(Residential)	$0.8 \cdot 30 + 0.1 \cdot 20 + 0 \cdot 50 = 26$ [%]
II	(Commercial)	$0.1 \cdot 30 + 0.7 \cdot 20 + 0.1 \cdot 50 = 22$ [%]
III	(Industrial)	$0.1 \cdot 30 + 0.2 \cdot 20 + 0.9 \cdot 50 = 52$ [%].

The sum is 100%, as it should be. We write this in matrix form. Let the column vector **x** denote the 1993 state; thus, $\mathbf{x}^{\mathsf{T}} = [30 \quad 20 \quad 50]$. Let **y** denote the 1998 state. Then

$$\mathbf{y}^{\mathsf{T}} = \mathbf{x}^{\mathsf{T}}\mathbf{A} = [30 \quad 20 \quad 50] \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0 & 0.1 & 0.9 \end{bmatrix} = [26 \quad 22 \quad 52].$$

Similarly, for the vector **z** of the 2003 state we get, as the reader may verify,

$$\mathbf{z}^{\mathsf{T}} = \mathbf{y}^{\mathsf{T}}\mathbf{A} = (\mathbf{x}^{\mathsf{T}}\mathbf{A})\mathbf{A} = \mathbf{x}^{\mathsf{T}}\mathbf{A}^2 = [23 \quad 23.2 \quad 53.8].$$

Answer. In 1998, the residential area will be 26% (13 square miles), the commercial 22% (11 square miles) and the industrial 52% (26 square miles). For 2003, the corresponding figures are 23%, 23.2%, 53.8%. ∎

This is the end of the first portion of Chap. 7, in which we have defined the rules of matrix and vector algebra. We are now ready for applications, beginning in the next section.

---

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

Problem Set 7.3

Let $\mathbf{a} = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 0 & -2 \\ 2 & 3 \end{bmatrix}$, $\mathbf{C} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 0 \\ 0 & 3 & 4 \end{bmatrix}$, $\mathbf{d} = [1 \quad 0 \quad 2]$.

Find those of the following expressions that are defined.

1. $\mathbf{CB}, \mathbf{B}^T\mathbf{C}^T, \mathbf{BC}^T$

2. $\mathbf{C}^2, \mathbf{C}^3, \mathbf{CC}^T, \mathbf{C}^T\mathbf{C}$

3. $\mathbf{Ca}, \mathbf{Cd}^T, \mathbf{C}^T\mathbf{d}^T$

4. $\mathbf{B}^T\mathbf{a}, \mathbf{Bd}, \mathbf{dB}, \mathbf{ad}$

5. $\mathbf{B}^T\mathbf{C}, \mathbf{B}^T\mathbf{B}$

6. $\mathbf{BB}^T, \mathbf{BB}^T\mathbf{C}, \mathbf{BB}^T\mathbf{a}$

7. $\mathbf{a}^T\mathbf{a}, \mathbf{a}^T\mathbf{Ca}, \mathbf{dCd}^T$

8. $\mathbf{dd}^T, \mathbf{d}^T\mathbf{d}, \mathbf{adB}, \mathbf{adBB}^T$

9. Prove (5).

10. Find real 2×2 matrices (as many as you can) whose square is $\mathbf{I}$, the unit matrix.

11. Find a 2×2 matrix $\mathbf{A} \neq \mathbf{0}$ such that $\mathbf{A}^2 = \mathbf{0}$.

12. Find two 2×2 matrices $\mathbf{A}, \mathbf{B}$ such that $(\mathbf{A} + \mathbf{B})^2 \neq \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2$.

13. **(Idempotent matrix)** A matrix $\mathbf{A}$ is said to be *idempotent* if $\mathbf{A}^2 = \mathbf{A}$. Give examples of idempotent matrices, different from the zero or unit matrix.

14. Show that $\mathbf{AA}^T$ is symmetric.

15. Find all real square matrices that are both symmetric and skew-symmetric.

16. Show that the product of symmetric matrices $\mathbf{A}, \mathbf{B}$ is symmetric if and only if $\mathbf{A}$ and $\mathbf{B}$ commute, $\mathbf{AB} = \mathbf{BA}$.

Special linear transformations were used in the text to motivate matrix multiplication, and we add some problems of practical interest. (Linear transformations in general follow in Sec. 7.15.)

17. **(Rotation)** Show that the linear transformation $\mathbf{y} = \mathbf{Ax}$ with matrix

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

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.

18. Show that in Prob. 17,

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

What does this result mean geometrically?

19. **(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?

20. **(Rotations in space in computer graphics)** What effect would the following matrices have in the situation described in Prob. 19?

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

21. **(Assignment problem)** Contractors C_1, C_2, C_3 bid for jobs J_1, J_2, J_3 as the cost matrix (in 100 000-dollar units) shows. What assignment minimizes the total cost (a) under no condition? (b) Under the condition that each contractor be assigned to only one job?

$$\begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \end{array} \begin{array}{c} \begin{array}{ccc} J_1 & J_2 & J_3 \end{array} \\ \begin{bmatrix} 24 & 4 & 10 \\ 18 & 6 & 12 \\ 16 & 8 & 8 \end{bmatrix} \end{array}$$

Cost matrix for Problem 21

$$\mathbf{A} = [a_{jk}] = \begin{bmatrix} 40 & 28 & 20 \\ 34 & 26 & 14 \\ 36 & 30 & 20 \end{bmatrix}$$

Matrix A for Problem 22

22. If worker W_j can do job J_k in a_{jk} hours, as shown by the matrix $\mathbf{A}$, and each worker should do one job only, which assignment would minimize the total time?

23. **(Markov process)** For the Markov process with transition matrix $\mathbf{A} = [a_{jk}]$, whose entries are $a_{11} = a_{12} = 0.5$, $a_{21} = 0.2$, $a_{22} = 0.8$, and initial state $[0.7 \quad 0.7]^{\mathsf{T}}$, compute the next 3 states.

24. 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 trouble 2 days after today? 3 days after today?

7.4 Linear Systems of Equations. Gauss Elimination

The most important practical use of matrices is in the solution of linear systems of equations, which appear frequently as models of various problems, for instance, in frameworks, electrical networks, traffic flow, production and consumption, assignment of jobs to workers, population growth, statistics, numerical methods for differential equations (Chap. 20), and many others. We begin in this section with an important solution method, the Gauss elimination, and discuss general properties of solutions in the next sections.

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

$$
\text{(1)} \qquad \boxed{\begin{array}{l} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{array}}
$$

Thus, a system of two equations in three unknown is

$$
\begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2, \end{array} \quad \text{for example,} \quad \begin{array}{r} 5x_1 + 2x_2 - x_3 = 4 \\ x_1 - 4x_2 + 3x_3 = 6. \end{array}
$$

The a_{jk} are given numbers, which are called the **coefficients** of the system. The b_i are also given numbers. If the b_i are all zero, then (1) is called a **homogeneous system.** If at least one b_i is not zero, then (1) is called a **nonhomogenous system.**

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

Coefficient Matrix, Augmented Matrix

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

$$
\text{(2)} \qquad \boxed{\mathbf{Ax} = \mathbf{b}}
$$

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

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

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

$$\tilde{\mathbf{A}} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix}$$

is called the **augmented matrix** of the system (1). We see that $\tilde{\mathbf{A}}$ is obtained by augmenting $\mathbf{A}$ by the column $\mathbf{b}$. The matrix $\tilde{\mathbf{A}}$ determines the system (1) completely, because it contains all the given numbers appearing in (1).

EXAMPLE 1 **Geometric interpretation. Existence of solutions**

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

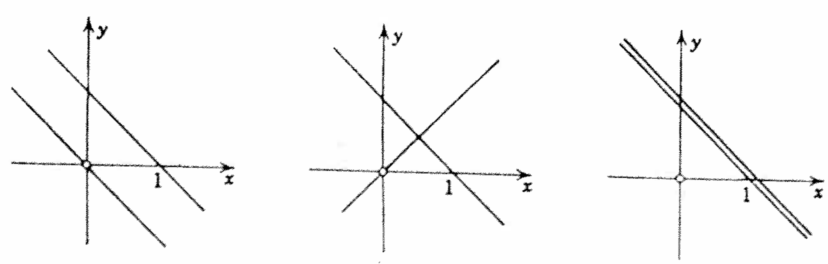
$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2.$$

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:

(a) No solution if the lines are parallel.
(b) Precisely one solution if they intersect.
(c) Infinitely many solutions if they coincide.

For instance,

Case (a)	Case (b)	Case (c)
$x + y = 1$	$x + y = 1$	$x + y = 1$
$x + y = 0$	$x - y = 0$	$2x + 2y = 2$



If the system is homogeneous, Case (a) cannot happen, because then those two straight lines pass through the origin, whose coordinates 0, 0 constitute the trivial solution. The reader may consider three equations in three unknowns as representations of three planes in space and discuss the various possible cases in a similar fashion. ∎

Our simple example illustrates that a system (1) may not always have a solution, and relevant problems are as follows. 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? How can we obtain the solutions? We discuss the last question first and the others in Sec. 7.6.

Gauss Elimination

The Gauss elimination is a standard method for solving linear systems. This is a systematic process of elimination, a method of great importance that works in practice and is reasonable with respect to computing time and storage demand (two aspects we shall consider in Sec. 19.1 on numerical methods). We first explain the method by some typical examples. Since a linear system is completely determined by its augmented matrix, the process of elimination can be performed by merely considering the matrices. To see this correspondence, we shall write systems of equations and augmented matrices side by side.

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. 135 (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. 135. 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:

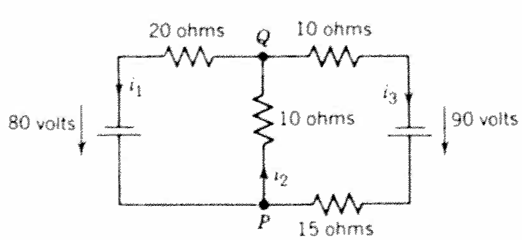
Kirchhoff's current law (KCL). *At any point of a circuit, the sum of the inflowing currents equals the sum of the outflowing currents.*

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.

Solution by Gauss's method. This system is so simple that we could almost solve it by inspection. This is not the point. The point is to perform a systematic elimination—the Gauss elimination—which will work in general, also for large systems. It is a reduction to "*triangular form*" from which we shall then readily obtain the values of the unknowns by "*back substitution.*"

We write the system and its augmented matrix side by side:



$$\begin{aligned} &\text{Node } P: & i_1 - i_2 + i_3 &= 0 \\ &\text{Node } Q: & -i_1 + i_2 - i_3 &= 0 \\ &\text{Right loop:} & 10i_2 + 25i_3 &= 90 \\ &\text{Left loop:} & 20i_1 + 10i_2 + \quad &= 80 \end{aligned}$$

Fig. 135. Network in Example 2 and equations for the currents

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

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

First Step. Elimination of x_1

Call the first equation the **pivot equation** and its x_1-term the **pivot** in this step, and use this equation to eliminate x_1 (get rid of x_1) in the other equations. For this, do these operations:

Subtract -1 times the pivot equation from the second equation.[2]

Subtract 20 times the pivot equation from the fourth equation.

This corresponds to row operations on the augmented matrix, which we indicate behind the *new* matrix in (3). The result is

$$
(3) \qquad
\begin{array}{rcl}
x_1 - x_2 + x_3 &=& 0 \\
0 &=& 0 \\
10x_2 + 25x_3 &=& 90 \\
30x_2 - 20x_3 &=& 80
\end{array}
\qquad
\begin{bmatrix}
1 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 \\
0 & 10 & 25 & 90 \\
0 & 30 & -20 & 80
\end{bmatrix}
\begin{array}{l}
\\
\text{Row 2 + Row 1} \\
\\
\text{Row 4} - 20\ \text{Row}
\end{array}
$$

Second Step. Elimination of x_2

The first equation, which has just served as the pivot equation, remains untouched. We want to take the (new!) second equation as the next pivot equation. Since it contains no x_2-term (needed as the next pivot)—in fact, it is $0 = 0$—first we have to change the order of equations (and corresponding rows of the new matrix) to get a nonzero pivot. We put the second equation ($0 = 0$) at the end and move the third and fourth equations one place up; this is called **partial pivoting.**[3] We get

$$
\begin{array}{l}
\\
\text{Pivot} \longrightarrow \\
\text{Eliminate} \longrightarrow \\
\\
\end{array}
\begin{array}{rcl}
x_1 - x_2 + x_3 &=& 0 \\
\boxed{10x_2} + 25x_3 &=& 90 \\
\boxed{30x_2} - 20x_3 &=& 80 \\
0 &=& 0
\end{array}
\qquad
\begin{bmatrix}
1 & -1 & 1 & 0 \\
0 & 10 & 25 & 90 \\
0 & 30 & -20 & 80 \\
0 & 0 & 0 & 0
\end{bmatrix}
$$

To eliminate x_2, do:

Subtract 3 times the pivot equation from the third equation.

The result is

$$
(4) \qquad
\begin{array}{rcl}
x_1 - x_2 + x_3 &=& 0 \\
10x_2 + 25x_3 &=& 90 \\
-95x_3 &=& -190 \\
0 &=& 0
\end{array}
\qquad
\begin{bmatrix}
1 & -1 & 1 & 0 \\
0 & 10 & 25 & 90 \\
0 & 0 & -95 & -190 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{array}{l}
\\
\\
\text{Row 3} - 3\ \text{Row} \\
\\
\end{array}
$$

---

[2] To call all the operations "subtractions" rather than "subtractions" and "additions" is preferable from the viewpoint of uniformity of numerical algorithms. See also Sec. 19.1.

[3] As opposed to **total pivoting,** in which also the order of the unknowns is changed. Total pivoting is hardly used in practice.

Back Substitution. Determination of x_3, x_2, x_1
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:

$$\begin{aligned} -95x_3 &= -190, & x_3 &= i_3 = 2 \text{ [amperes]},\\ 10x_2 + 25x_3 &= 90, & x_2 &= \tfrac{1}{10}(90 - 25x_3) = i_2 = 4 \text{ [amperes]},\\ x_1 - x_2 + x_3 &= 0, & x_1 &= x_2 - x_3 = i_1 = 2 \text{ [amperes]}.\end{aligned}$$

This is the answer to our problem. The solution is unique. ∎

A 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 (1) has fewer equations than unknowns. An underdetermined system always has solutions, whereas in the other two cases, solutions may or may not exist. (Details follow in Sec. 7.6.) We want to illustrate next that the Gauss elimination applies to any system, no matter whether it has many solutions, a unique solution, or no solutions.

EXAMPLE 3 Gauss elimination for an underdetermined system

Solve the linear system of three equations in four unknowns

$$(5)\qquad \begin{aligned} \boxed{3.0x_1} + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0\\ \boxed{0.6x_1} + 1.5x_2 + 1.5x_3 - 5.4x_4 &= 2.7\\ \boxed{1.2x_1} - 0.3x_2 - 0.3x_3 + 2.4x_4 &= 2.1\end{aligned} \qquad \begin{bmatrix} 3.0 & 2.0 & 2.0 & -5.0 & 8.0\\ 0.6 & 1.5 & 1.5 & -5.4 & 2.7\\ 1.2 & -0.3 & -0.3 & 2.4 & 2.1\end{bmatrix}$$

Solution. As in the previous example, we circle pivots and box terms to be eliminated.

First Step. Elimination of x_1 from the second and third equations by subtracting

$0.6/3.0 = 0.2$ times the first equation from the second equation,

$1.2/3.0 = 0.4$ times the first equation from the third equation.

This gives a new system of equations

$$(6)\qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0\\ \boxed{1.1x_2} + 1.1x_3 - 4.4x_4 &= 1.1\\ \boxed{-1.1x_2} - 1.1x_3 + 4.4x_4 &= -1.1\end{aligned} \qquad \begin{bmatrix} 3.0 & 2.0 & 2.0 & -5.0 & 8.0\\ 0 & 1.1 & 1.1 & -4.4 & 1.1\\ 0 & -1.1 & -1.1 & 4.4 & -1.1\end{bmatrix}$$

and we circle the pivot to be used in the next step.

Second Step. Elimination of x_2 from the third equation of (6) by subtracting

$-1.1/1.1 = -1$ times the second equation from the third equation.

This gives

$$(7)\qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0\\ 1.1x_2 + 1.1x_3 - 4.4x_4 &= 1.1\\ 0 &= 0\end{aligned} \qquad \begin{bmatrix} 3.0 & 2.0 & 2.0 & -5.0 & 8.0\\ 0 & 1.1 & 1.1 & -4.4 & 1.1\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

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. ∎

EXAMPLE 4 **Gauss elimination if a unique solution exists**

Solve the system

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

First Step. Elimination of x_1 from the second and third equations gives

$$\begin{aligned} -x_1 + x_2 + 2x_3 &= 2 \\ 2x_2 + 7x_3 &= 12 \\ 2x_2 + 2x_3 &= 2 \end{aligned} \qquad \begin{bmatrix} -1 & 1 & 2 & 2 \\ 0 & 2 & 7 & 12 \\ 0 & 2 & 2 & 2 \end{bmatrix} \begin{matrix} \\ \text{Row 2} + 3\ \text{Row 1} \\ \text{Row 3} - \text{Row 1} \end{matrix}$$

Second Step. Elimination of x_2 from the third equation gives

$$\begin{aligned} -x_1 + x_2 + 2x_3 &= 2 \\ 2x_2 + 7x_3 &= 12 \\ -5x_3 &= -10 \end{aligned} \qquad \begin{bmatrix} -1 & 1 & 2 & 2 \\ 0 & 2 & 7 & 12 \\ 0 & 0 & -5 & -10 \end{bmatrix} \begin{matrix} \\ \\ \text{Row 3} - \text{Row 2} \end{matrix}$$

Back Substitution. Beginning with the last equation, we obtain successively $x_3 = 2$, $x_2 = -1$, $x_1 = 1$. We see that the system has a unique solution. ∎

EXAMPLE 5 **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

$$\begin{aligned} \boxed{3x_1} + 2x_2 + x_3 &= 3 \\ \boxed{2x_1} + x_2 + x_3 &= 0 \\ \boxed{6x_1} + 2x_2 + 4x_3 &= 6 \end{aligned} \qquad \begin{bmatrix} 3 & 2 & 1 & 3 \\ 2 & 1 & 1 & 0 \\ 6 & 2 & 4 & 6 \end{bmatrix}$$

First Step. Elimination of x_1 from the second and third equations by subtracting

2/3 times the first equation from the second equation,

6/3 = 2 times the first equation from the third equation.

This gives

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ \boxed{-\tfrac{1}{3}x_2} + \tfrac{1}{3}x_3 &= -2 \\ \boxed{-2x_2} + 2x_3 &= 0 \end{aligned} \qquad \begin{bmatrix} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & -2 & 2 & 0 \end{bmatrix}$$

Second Step. Elimination of x_2 from the third equation gives

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ -\tfrac{1}{3}x_2 + \tfrac{1}{3}x_3 &= -2 \\ 0 &= 12 \end{aligned} \qquad \begin{bmatrix} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{bmatrix}$$

This shows that the system has no solution. ∎

The form of the system and of the matrix in the last step of the Gauss elimination is called the **echelon form.** Thus in Example 5 the echelon forms of the coefficient matrix and the augmented matrix are

$$\begin{bmatrix} 3 & 2 & 1 \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{bmatrix}.$$

At the end of the Gauss elimination (before the back substitution) the reduced system will have the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots\cdots + a_{1n}x_n &= b_1 \\ c_{22}x_2 + \cdots\cdots + c_{2n}x_n &= b_2^* \\ &\vdots \\ k_{rr}x_r + \cdots + k_{rn}x_n &= \tilde{b}_r \\ 0 &= \tilde{b}_{r+1} \\ &\vdots \\ 0 &= \tilde{b}_m \end{aligned} \tag{8}$$

where $r \leqq m$ (and $a_{11} \neq 0$, $c_{22} \neq 0, \cdots, k_{rr} \neq 0$). From this we see that with respect to solutions of this system (8), there are three possible cases:

(a) No solution if $r < m$ and one of the numbers $\tilde{b}_{r+1}, \cdots, \tilde{b}_m$ is not zero. This is illustrated by Example 5, where $r = 2 < m = 3$ and $\tilde{b}_{r+1} = \tilde{b}_3 = 12$.

(b) Precisely one solution if $r = n$ and $\tilde{b}_{r+1}, \cdots, \tilde{b}_m$, if present, are zero. This solution is obtained by solving the nth equation of (8) for x_n, then the $(n-1)$th equation for x_{n-1}, and so on up the line. See Example 2, where $r = n = 3$ and $m = 4$.

(c) Infinitely many solutions if $r < n$ and $\tilde{b}_{r+1}, \cdots, \tilde{b}_m$, if present, are zero. Then any of these solutions is obtained by choosing values at pleasure for the unknowns $x_{r+1}, \cdots, x_n$, solving the rth equation for x_r, then the $(r-1)$th equation for x_{r-1}, and so on up the line. Example 3 illustrates this case.

Elementary Row Operations

To justify the Gauss elimination as a method of solving linear systems, we first introduce two related concepts.

Elementary operations for equations

Interchange of two equations
Multiplication of an equation by a nonzero constant
Addition of a constant multiple of one equation to another equation.

To these correspond the following

Elementary row operations for matrices

Interchange of two rows
Multiplication of a row by a nonzero constant
Addition of a constant multiple of one row to another row.

The Gauss elimination consists of the use of the third of these operations[4] (for getting zeros) and of the first (in pivoting).

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!) elementary row operations. Clearly, the system produced by the Gauss elimination at the end is row-equivalent to the original system to be solved. Hence the desired justification of the Gauss elimination as a solution method now follows from the subsequent theorem, which implies that the Gauss elimination gives all solutions of the *original* system.

Theorem 1 (Row-equivalent systems)

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

Proof. The interchange of two equations does not alter the solution set. Neither does the multiplication of an equation by a (nonzero!) constant c, because multiplication of the new equation by $1/c$ produces the original equation. Similarly for the addition of an equation E_1 to an equation E_2, since by adding $-E_1$ (the equation obtained from E_1 by multiplying E_1 by -1) to the equation resulting from the addition we get back the original equation. ∎

This justifies the Gauss elimination. Numerical aspects of it are discussed in Sec. 19.1 (which is independent of other sections on numerical methods) and popular variants of it (Doolittle's, Crout's, and Cholesky's methods) in Sec. 19.2.

Problem Set 7.4

Solve the following linear systems by the Gauss elimination.

1. $2x + 3y = 4$, $3x + 2y = -4$
2. $3x + 2y = -17$, $10x + y = 0$
3. $-x + 2y = 4$, $3x + 4y = 38$

[4] In the Gauss elimination we said "subtraction of a constant multiple" (rather than "addition"), as being more suggestive in getting zeros; of course, this is a mere matter of language.

4.
$$\begin{aligned} x + 2y - 8z &= 0 \\ 2x - 3y + 5z &= 0 \\ 3x + 2y - 12z &= 0 \end{aligned}$$

5.
$$\begin{aligned} 3x - y + z &= -2 \\ x + 5y + 2z &= 6 \\ 2x + 3y + z &= 0 \end{aligned}$$

6.
$$\begin{aligned} 7x - y - 2z &= 0 \\ 9x - y - 3z &= 0 \\ 2x + 4y - 7z &= 0 \end{aligned}$$

7.
$$\begin{aligned} x + y + z &= -1 \\ 4y + 6z &= 6 \\ y + z &= 1 \end{aligned}$$

8.
$$\begin{aligned} 5x + 3y &= 22 \\ -4x + 7y &= 20 \\ 9x - 2y &= 15 \end{aligned}$$

9.
$$\begin{aligned} 4y + 3z &= 13 \\ x - 2y + z &= 3 \\ 3x + 5y &= 11 \end{aligned}$$

10.
$$\begin{aligned} 7x - 4y - 2z &= -6 \\ 16x + 2y + z &= 3 \end{aligned}$$

11.
$$\begin{aligned} x - 3y + 2z &= 2 \\ 5x - 15y + 7z &= 10 \end{aligned}$$

12.
$$\begin{aligned} 3x - 3y - 7z &= -4 \\ x - y + 2z &= 3 \end{aligned}$$

13.
$$\begin{aligned} 3w - 6x - y - z &= 0 \\ w - 2x + 5y - 3z &= 0 \\ 2w - 4x + 3y - z &= 3 \end{aligned}$$

14.
$$\begin{aligned} 4w + 3x - 9y + z &= 1 \\ -w + 2x - 13y + 3z &= 3 \\ 3w - x + 8y - 2z &= -2 \end{aligned}$$

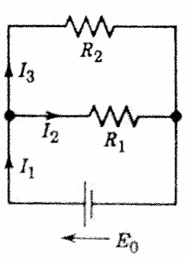
15.
$$\begin{aligned} w + x + y &= 6 \\ -3w - 17x + y + 2z &= 2 \\ 4w - 17x + 8y - 5z &= 2 \\ -5x - 2y + z &= 2 \end{aligned}$$

16.
$$\begin{aligned} w - x + 3y - 3z &= 3 \\ -5w + 2x - 5y + 4z &= -5 \\ -3w - 4x + 7y - 2z &= 7 \\ 2w + 3x + y - 11z &= 1 \end{aligned}$$

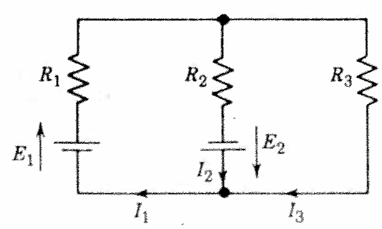
Models of electrical networks

Using Kirchhoff's laws (see Example 2), find the currents in the following networks.

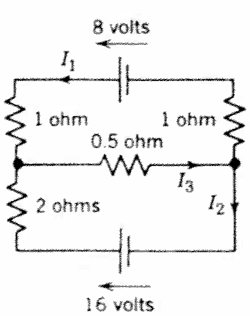
17.



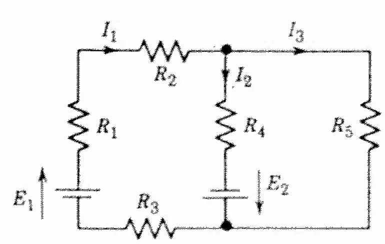
18.

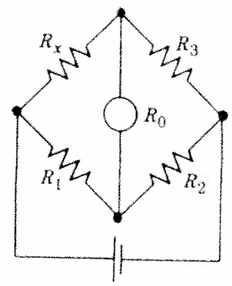


19.

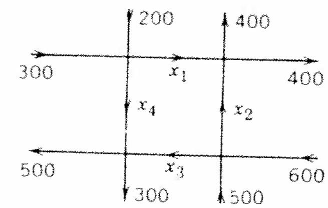


20.





Problem 21
Wheatstone bridge



Problem 22
Net of one-way streets

21. **(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.)
22. **(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?
23. **(Models of markets)** Determine the equilibrium solution ($D_1 = S_1$, $D_2 = S_2$) of the two-commodity market with linear model

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

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

where D, S, P mean demand, supply, price, and the subscripts 1 and 2 refer to the first and second commodity, respectively.

24. **(Equivalence relation)** By definition, an *equivalence relation* on a set is a relation satisfying three conditions:
    (1) Each element A of the set is equivalent to itself.
    (2) If A is equivalent to B, then B is equivalent to A.
    (3) If A is equivalent to B and B is equivalent to C, then A is equivalent to C.

    For instance, equality is an equivalence relation on the set of real numbers. Show that row equivalence satisfies these three conditions.

7.5 Linear Independence. Vector Space. Rank of a Matrix

In the last section we explained the most important practical method for solving linear systems of equations, the Gauss elimination. We have also seen that a system may have no solutions or a single solution or more than just one solution (and then infinitely many solutions). So we ask whether we can make general statements about these *problems of existence and uniqueness*. The answer is yes, and we shall do so in the next section. For this we shall need the concepts of linear independence and rank, which we now introduce; these are of great general importance, far beyond our present discussion.

Linear Independence and Dependence of Vectors

Given any set of m vectors[5] $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ (with the same number of components), a **linear combination** of these vectors is an expression of the form

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

where $c_1, \cdots, c_m$ are any scalars.[6] 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 holds if we choose all c'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,** because then we can express (at least) one of them as a linear combination of the others; for instance, if (1) holds with, say, $c_1 \neq 0$, we can solve (1) for $\mathbf{a}_{(1)}$:

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

(and some or even all k's may be zero).

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 (do it!), it is not so easy to discover; however, a 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 pair of components). ∎

Vector Space, Dimension, Basis

Given m vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ with n components each, as before, we can form the set V of all linear combinations of these vectors. V is called the **span** of these m vectors.

V is a **vector space.**[7] By definition, this means that V is a set of vectors with the two algebraic operations of *addition* and *scalar multiplication* defined for these vectors such that the following holds.

---

[5] Write simply $\mathbf{a}_1, \cdots, \mathbf{a}_m$ if you wish, but keep in mind that these are *vectors*, not vector *components*.

[6] In this section, scalars will be *real* numbers.

[7] Here, we give just what we need in the next section. General vector spaces follow in Sec. 7.15.

1. The sum $\mathbf{a} + \mathbf{b}$ of any vectors $\mathbf{a}$ and $\mathbf{b}$ in V is also in V and the product $k\mathbf{a}$ of any vector $\mathbf{a}$ in V and scalar k is also in V.

2. For all vectors and scalars we have the familiar rules

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

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

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

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

as well as

$$k(\mathbf{a} + \mathbf{b}) = k\mathbf{a} + k\mathbf{b}$$

$$(k + \ell)\mathbf{a} = k\mathbf{a} + \ell\mathbf{a}$$

$$k(\ell\mathbf{a}) = (k\ell)\mathbf{a} \qquad (\text{written } k\ell\mathbf{a})$$

$$1\mathbf{a} = \mathbf{a}.$$

(For our vectors, these rules follow from (1) and (2) in Sec. 7.2—after all, vectors are special matrices!)

The maximum number of linearly independent vectors in V is called the **dimension** of V and is denoted by dim V.

Clearly, if those given m vectors are linearly independent, then dim $V = m$; and if they are linearly dependent, then dim $V < m$.

A linearly independent set in V consisting of a maximum possible number of vectors in V is called a **basis** for V. Thus the number of vectors of a basis for V equals dim V.

EXAMPLE 2 Vector space, Dimension, Basis

The span of the three vectors in Example 1 is a vector space of dimension 2, and a basis is $\mathbf{a}_{(1)}$, $\mathbf{a}_{(2)}$, for instance, or $\mathbf{a}_{(1)}$, $\mathbf{a}_{(3)}$, etc. ∎

By the **real *n*-dimensional vector space R^n** we mean the space of all vectors with n *real* numbers as components ("*real vectors*") and real numbers as scalars. This is a standard name and notation. Hence each such vector is an ordered n-tuple of real numbers, as we know.

Thus for $n = 3$ we get R^3 consisting of ordered triples ("vectors in 3-space"), and for $n = 2$ we get R^2 consisting of ordered pairs ("vectors in the plane"). In Chaps. 8 and 9 we shall see that these special cases provide wide areas for applications of geometry, mechanics, and calculus which are of basic importance to the engineer and physicist.

Rank of a Matrix

The maximum number of linearly independent row vectors of a matrix $\mathbf{A} = [a_{jk}]$ is called the **rank** of $\mathbf{A}$ and is denoted by

$$\text{rank } \mathbf{A}.$$

EXAMPLE 3 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.

In our proposed discussion of the existence and uniqueness of solutions of systems of linear equations we shall need the following very important

Theorem 1 (Rank in terms of column vectors)

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

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

Proof. Let $\mathbf{A} = [a_{jk}]$ and let rank $\mathbf{A} = r$. Then, by definition, $\mathbf{A}$ has a linearly independent set of r row vectors, call them $\mathbf{v}_{(1)}, \cdots, \mathbf{v}_{(r)}$, and all row vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ of $\mathbf{A}$ are linear combinations of those independent ones, say,

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

These are vector equations. Each of them is equivalent to n equations for corresponding components. Denoting the components of $\mathbf{v}_{(1)}$ by $v_{11}, \cdots, v_{1n}$, the components of $\mathbf{v}_{(2)}$ by $v_{21}, \cdots, v_{2n}$, etc., and similarly for the vectors on the left, we thus have

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

where $k = 1, \cdots, n$. This can be written

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

where $k = 1, \cdots, n$. The vector on the left is the kth column vector of $\mathbf{A}$. Hence the equation shows that each column vector of $\mathbf{A}$ is a linear combination of the r vectors on the right. Hence the maximum number of linearly independent column vectors of $\mathbf{A}$ cannot exceed r, which is the maximum number of linearly independent row vectors of $\mathbf{A}$, by the definition of rank.

Now the same conclusion applies to the transpose $\mathbf{A}^T$ of $\mathbf{A}$. Since the row vectors of $\mathbf{A}^T$ are the column vectors of $\mathbf{A}$, and the column vectors of $\mathbf{A}^T$ are the row vectors of $\mathbf{A}$, that conclusion means that the maximum number of linearly independent row vectors of $\mathbf{A}$ (which is r) cannot exceed the maximum number of linearly independent column vectors of $\mathbf{A}$. Hence that number must equal r, and the proof is complete. ∎

EXAMPLE 4 **Illustration of Theorem 1**

What does Theorem 1 mean with respect to our matrix $\mathbf{A}$ in (2)? Since we have rank $\mathbf{A} = 2$, the column vectors should contain two linearly independent ones, and the other two should be linear combinations of them. Indeed, the first two column vectors are linearly independent, and

$$\begin{bmatrix} 2 \\ 24 \\ 0 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} 3 \\ -6 \\ 21 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} 0 \\ 42 \\ -21 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ 54 \\ -15 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} 3 \\ -6 \\ 21 \end{bmatrix} + \frac{29}{21}\begin{bmatrix} 0 \\ 42 \\ -21 \end{bmatrix}.$$

This is easy to verify but not so easy to see. Imagining that $\mathbf{A}^T$ were given, we realize that the determination of a rank by a direct application of the definition is not the proper way, unless a matrix is sufficiently simple. This suggests asking whether we can "simplify" (transform) a matrix without altering its rank. The answer is yes, as we show next. ∎

The span of the row vectors of a matrix $\mathbf{A}$ is called the **row space** of $\mathbf{A}$ and the span of the columns the **column space** of $\mathbf{A}$. From this and Theorem 1 we have

Theorem 2 **(Row space and column space)**

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

Invariance of Rank Under Elementary Row Operations

We claim that *elementary row operations* (Sec. 7.4) *do not alter the rank of a matrix* $\mathbf{A}$.

For the first operation (interchange of two row vectors) this is clear. The second operation (multiplication of a row vector by a nonzero constant) does not alter the rank either, since it does not alter the maximum number of linearly independent row vectors. Finally, the third operation is the addition of c times a row vector $\mathbf{a}_{(j)}$, say, to another row vector, say, $\mathbf{a}_{(i)}$. This produces a matrix that differs from $\mathbf{A}$ only in the ith row vector, which is of the form $\mathbf{a}_{(i)} + c\mathbf{a}_{(j)}$, a linear combination of the row vectors $\mathbf{a}_{(i)}$ and $\mathbf{a}_{(j)}$, so that the number of linearly independent row vectors remains the same. Hence the new matrix has the same rank as $\mathbf{A}$. Remembering from the previous section that *row-equivalent matrices* are those that can be obtained from each other by finitely many elementary row operations, our result is

Theorem 3 (Row-equivalent matrices)
Row-equivalent matrices have the same rank.

This theorem tells us what we can do to determine the rank of a matrix **A**, namely, we can reduce **A** to echelon form (Sec. 7.4), using the technique of the Gauss elimination, because this leaves the rank unchanged, by Theorem 2, and from the echelon form we can recognize the rank directly.

EXAMPLE 5 Determination of rank

For the matrix in Example 3 we obtain successively

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

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

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

The last matrix is in echelon form. From the row vectors and Theorem 3 we see immediately that rank $\mathbf{A} \leqq 2$, and rank $\mathbf{A} = 2$ by Theorem 1, since the first two column vectors are certainly linearly independent. ∎

This method of determining rank has practical applications in connection with the determination of linear dependence and independence of vectors. The key to this is the following theorem, which results immediately from the definition of rank.

Theorem 4 (Linear dependence and independence)
p vectors $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(p)}$ (with n components each) are linearly independent if the matrix with row vectors $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(p)}$ has rank p; they are linearly dependent if that rank is less than p.

Since each of those p vectors has n components, that matrix, call it **A**, has p rows and n columns; and if $n < p$, then by Theorem 1 we must have rank $\mathbf{A} \leqq n < p$, so that Theorem 4 yields the following result, which one should keep in mind.

Theorem 5 *p vectors with $n < p$ components are always linearly dependent.*

For instance, three or more vectors in the plane are linearly dependent. Similarly, four or more vectors in space are linearly dependent.

By the definition of dimension, we also have

Theorem 6 *The vector space R^n consisting of all vectors with n components has dimension n.*

Basic applications of rank follow in the next section.

Problem Set 7.5

Vector spaces. Is the given set of vectors a vector space? (Give a reason.) If your answer is yes, determine the dimension and find a basis.

1. All vectors $[v_1 \ \ v_2 \ \ v_3]^T$ in R^3 such that $v_1 + 2v_2 = 0$.
2. All vectors in R^4 such that $v_1 + v_2 = 0$, $v_3 + v_4 = 0$.
3. All vectors in R^3 satisfying $v_1 + v_2 + v_3 = 1$.
4. All vectors in R^2 such that $v_1 + v_2 = k$ (= *const*).
5. All real numbers.
6. All vectors in R^5 such that $v_1 + v_2 = 0$, $v_1 + 2v_2 + v_3 - v_4 = 0$.
7. All ordered quintuples of positive real numbers.
8. All vectors in R^4 such that $v_1 = 0$, $v_2 + v_3 + v_4 \geqq 0$.
9. All vectors in R^n with the first $n - 1$ components zero.

10. **(Subspace)** A nonempty subset W of a vector space V is called a *subspace* of V if W is itself a vector space with respect to the algebraic operations defined in V. Give examples of one- and two-dimensional subspaces of R^3.

Rank. Find the rank by inspection or by the method in Example 5.

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

12. $\begin{bmatrix} 3 & 2 & -9 \\ -6 & -4 & 18 \\ 12 & 8 & -36 \end{bmatrix}$

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

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

15. $\begin{bmatrix} a & b & c \\ b & a & c \end{bmatrix}$, $a \neq \pm b$

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

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

18. $\begin{bmatrix} 6 & 0 & 2 \\ 0 & 8 & 3 \\ 2 & 7 & 5 \\ 5 & 5 & 0 \end{bmatrix}$

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

20. Show by an example that rank $\mathbf{A}$ = rank $\mathbf{B}$ does *not* imply rank $\mathbf{A}^2$ = rank $\mathbf{B}^2$.
21. Show that rank $\mathbf{B}^T\mathbf{A}^T$ = rank $\mathbf{AB}$.
22. Prove that if the row vectors of an $n \times n$ matrix $\mathbf{A}$ are linearly independent, then so are the column vectors of $\mathbf{A}$ (and vice versa).
23. Prove that if $\mathbf{A}$ is not square, then either the row vectors or the column vectors of $\mathbf{A}$ are linearly dependent.
24. Prove that row-equivalent matrices have the same row space.
25. Find a basis of the row space and of the column space of the matrix in Prob. 11.
26. Do the same task as in Prob. 25 for the matrix in Prob. 17.

Linear independence. State whether the given vectors are linearly independent or dependent.

27. [1 5 3], [2 4 6], [3 9 11]

28. [4 3 9], [0 0 0], [1 $\frac{3}{2}$ $\frac{2}{3}$]

29. [1 0], [1 2], [3 4]

30. [1 1 0], [0 1 1], [1 0 1]

31. [1 2 3], [4 5 6], [7 8 9]

32. [3 −1 4], [6 7 5], [9 6 9]

33. [9 0 9], [0 6 6], [3 3 0]

34. [2 1 0 6], [1 9 9 0]

35. [3 0 2 4 5], [7 2 6 1 0], [1 2 2 −7 −10]

7.6 Linear Systems: General Properties of Solutions

Using the concept of the rank of a matrix, as defined in the last section, we can now settle the issue of existence and uniqueness of solutions of linear systems. The central theorem (which the student should memorize!) is as follows. (For illustrative examples, see Sec. 7.4.)

Theorem 1 **Fundamental Theorem for linear systems**

(a) *A linear system of m equations*

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

in n unknowns $x_1, \cdots, x_n$ has solutions if and only if the coefficient matrix **A** *and the augmented matrix* $\tilde{\mathbf{A}}$, *that is,*

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \textit{and} \quad \tilde{\mathbf{A}} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix},$$

have the same rank.

(b) *If this rank r equals n, the system* (1) *has precisely one solution.*

(c) *If $r < n$, the system* (1) *has infinitely many solutions, all of which 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.*

(d) *If solutions exist, they can all be obtained by the Gauss elimination* (see Sec. 7.4). (This elimination may be started without first looking at the ranks of **A** and $\tilde{\mathbf{A}}$, since it will automatically reveal whether or not solutions exist; see, for instance, Example 5 in Sec. 7.4.)

Proof. (a) We can write the system (1) in the form

$$\boxed{\mathbf{Ax} = \mathbf{b}} \tag{1}$$

or in terms of the column vectors $\mathbf{c}_{(1)}, \cdots, \mathbf{c}_{(n)}$ of $\mathbf{A}$:

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

Since $\tilde{\mathbf{A}}$ is obtained by attaching to $\mathbf{A}$ the additional column $\mathbf{b}$, Theorem 1 in Sec. 7.5 implies that rank $\tilde{\mathbf{A}}$ 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. Hence rank $\tilde{\mathbf{A}}$ cannot exceed rank $\mathbf{A}$, so that we must have rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A}$.

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

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

since otherwise rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A} + 1$. But this means that (1) has a solution, namely, $x_1 = \alpha_1, \cdots, x_n = \alpha_n$.

(b) If rank $\mathbf{A} = r = n$, then the set $C = \{\mathbf{c}_{(1)}, \cdots, \mathbf{c}_{(n)}\}$ is linearly independent, by Theorem 1 in Sec. 7.5. It follows that then the representation (2) of $\mathbf{b}$ is unique because

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

would imply

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

and $x_1 - \tilde{x}_1 = 0, \cdots, x_n - \tilde{x}_n = 0$ by the linear independence. Hence 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 $\tilde{\mathbf{A}} = r < n$, by Theorem 1, Sec. 7.5, 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

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

$\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{\mathbf{c}}_{(n)}$ 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 terms $\hat{x}_{r+1}\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{x}_n\hat{\mathbf{c}}_{(n)}$; 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 proved in Sec. 7.5 and is restated here as a reminder. ∎

The theorem is illustrated by the examples in Sec. 7.4: in Example 3 we have rank $\mathbf{A}$ = rank $\tilde{\mathbf{A}} = 2 < n = 4$ and can choose x_3 and x_4 arbitrarily; in Example 4 there is a unique solution since rank $\mathbf{A}$ = rank $\tilde{\mathbf{A}} = n = 3$; and in Example 5 there is no solution, since rank $\mathbf{A} = 2 <$ rank $\tilde{\mathbf{A}} = 3$.

The Homogeneous System

The system (1) is called **homogeneous** if all the b_j's on the right side are zero. Otherwise it is called **nonhomogeneous.** (See also Sec. 7.4.) From the Fundamental Theorem we readily obtain the following results.

Theorem 2 (Homogeneous system)

A homogeneous linear system

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

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 of dimension* $n - r$ *(see Sec. 7.5). In particular, if* $\mathbf{x}_{(1)}$ *and* $\mathbf{x}_{(2)}$ *are solution vectors of* (4), *then* $\mathbf{x} = c_1\mathbf{x}_{(1)} + c_2\mathbf{x}_{(2)}$, *where* c_1 *and* c_2 *are any scalars, is a solution vector of* (4). *(This* ***does not hold*** *for nonhomogeneous systems.)*

Proof. The first proposition is obvious and is in agreement with the fact that for a homogeneous system the matrix of the coefficients and the augmented matrix have the same rank. The solution vectors 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 $\mathbf{A} = r < n$, the Fundamental Theorem 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. It follows that a **basis of solutions** (that is, a basis for the vector space of these solutions) is $\mathbf{y}_{(1)}, \cdots, \mathbf{y}_{(n-r)}$, where the solution vector $\mathbf{y}_{(j)}$, $j = 1, \cdots, n - r$, is obtained by choosing $x_{r+j} = 1$ and the other $x_{r+1}, \cdots, x_n$ zero; the corresponding $x_1, \cdots, x_r$ are then determined. This proves that the vector space of all solutions has dimension $n - r$ and completes the proof. ∎

We mention that the vector space of all solutions of (4) is called the **null space** of the coefficient matrix **A**, because if we multiply any **x** in this null space by **A** we get **0**. The dimension of the null space is called the **nullity** of **A**. In terms of these concepts, 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 **A**). If we have rank $\mathbf{A} = n$, then nullity $\mathbf{A} = 0$, so that the system has only the trivial solution. If rank $\mathbf{A} = r < n$, then nullity $\mathbf{A} = n - r > 0$, so that we have nontrivial solutions which, together with **0**, form a vector space of dimension $n - r > 0$.

Note that rank $\mathbf{A} \leqq m$ in (4), by the definition, so that rank $\mathbf{A} < n$ when $m < n$. By Theorem 2 this proves the following theorem, which is of considerable practical importance.

Theorem 3 **(System with fewer equations than unknowns)**
A homogeneous system of linear equations with fewer equations than unknowns always has nontrivial solutions.

The Nonhomogeneous System

If a nonhomogeneous system of linear equations has solutions, their totality can be characterized as follows.

Theorem 4 **(Nonhomogeneous system)**
If a nonhomogeneous linear system of equations of the form (1) *has solutions, then all these solutions are of the form*

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

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. Let **x** be any given solution of (1) and $\mathbf{x}_0$ an arbitrarily chosen solution of (1). Then $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$ and, therefore,

$$\mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}_0 = \mathbf{0}.$$

This shows that the difference $\mathbf{x} - \mathbf{x}_0$ of any solution **x** of (1) and any fixed solution $\mathbf{x}_0$ of (1) is a solution of (4), say, $\mathbf{x}_h$. Hence all solutions of (1) are obtained by letting $\mathbf{x}_h$ run through all the solutions of the homogeneous system (4), and the proof is complete. ∎

7.7 Inverse of a Matrix

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

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

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

where $\mathbf{I}$ is the $n \times n$ unit matrix (see Sec. 7.3).

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 thus give a motivation for the inverse as well as a relation to linear systems.[8]

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 $\mathbf{A} = n$. *Hence* $\mathbf{A}$ *is nonsingular if* rank $\mathbf{A} = n$, *and is singular if* rank $\mathbf{A} < n$.

Proof. Consider the linear system

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

with the given matrix $\mathbf{A}$ as coefficient matrix. If the inverse exists, then multiplication from the left on both sides gives by (1)

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

This shows that (2) has a unique solution $\mathbf{x}$, so that $\mathbf{A}$ must have rank n by the Fundamental Theorem in the last section.

Conversely, let rank $\mathbf{A} = n$. Then by the same theorem, the system (2) has a unique solution $\mathbf{x}$ for any $\mathbf{b}$, and the Gauss elimination (in Sec. 7.4) shows that its components x_j are linear combinations of those of $\mathbf{b}$, so that we can write

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

---

[8] But *not* a method of solving $\mathbf{Ax} = \mathbf{b}$ *numerically*, because the Gauss elimination (Sec. 7.4) requires fewer computations.

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

Determination of the Inverse

We want to show that for practically determining the inverse $\mathbf{A}^{-1}$ of a nonsingular $n \times n$ matrix **A** we can use the Gauss elimination (Sec. 7.4), actually, a variant of it, called the **Gauss–Jordan elimination.**[9] Our idea is as follows. Using **A**, we form the n systems $\mathbf{Ax}_{(1)} = \mathbf{e}_{(1)}, \cdots, \mathbf{Ax}_{(n)} = \mathbf{e}_{(n)}$, where $\mathbf{e}_{(j)}$ has the jth component 1 and the other components 0. Introducing the $n \times n$ matrices $\mathbf{X} = [\mathbf{x}_{(1)} \cdots \mathbf{x}_{(n)}]$ and $\mathbf{I} = [\mathbf{e}_{(1)} \cdots \mathbf{e}_{(n)}]$, we combine the n systems into a single matrix equation $\mathbf{AX} = \mathbf{I}$ and the n augmented matrices $[\mathbf{A} \quad \mathbf{e}_{(1)}], \cdots, [\mathbf{A} \quad \mathbf{e}_{(n)}]$ into a single augmented matrix $\tilde{\mathbf{A}} = [\mathbf{A} \quad \mathbf{I}]$. Now $\mathbf{AX} = \mathbf{I}$ implies $\mathbf{X} = \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}^{-1}$, and to solve $\mathbf{AX} = \mathbf{I}$ for **X** we can apply the Gauss elimination to $\tilde{\mathbf{A}} = [\mathbf{A} \quad \mathbf{I}]$ to get $[\mathbf{U} \quad \mathbf{H}]$, where **U** is upper triangular, since the Gauss elimination triangularizes systems. The Gauss–Jordan elimination now operates on $[\mathbf{U} \quad \mathbf{H}]$ and, by eliminating the entries in **U** above the main diagonal, reduces it to $[\mathbf{I} \quad \mathbf{K}]$, the augmented matrix of $\mathbf{IX} = \mathbf{A}^{-1}$. Hence we must have $\mathbf{K} = \mathbf{A}^{-1}$ and can thus read off $\mathbf{A}^{-1}$ at the end.

(A formula for the entries of $\mathbf{A}^{-1}$ in terms of those of **A** follows in Sec. 7.9, in connection with determinants.)

EXAMPLE 1 **Inverse of a matrix. Gauss–Jordan elimination**

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

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

[9] WILHELM JORDAN (1842—1899), German mathematician and geodesist. [See *American Mathematical Monthly* **94** (1987), 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. 19.1.

Solution. We apply the Gauss elimination (Sec. 7.4) to

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

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

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

This is [U H] as produced by the Gauss elimination, and U agrees with Example 4 in Sec. 7.4. Now follow the additional Gauss–Jordan steps, reducing U to I, that is, to diagonal form with entries 1 on the main diagonal.

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

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

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

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, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. ∎

Some Useful Formulas for Inverses

For a nonsingular 2×2 matrix we obtain

$$(4) \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix},$$

where $\det \mathbf{A} = a_{11}a_{22} - a_{12}a_{21}$ and will be discussed in the next section. Indeed, one can readily verify that (1) holds.

Similarly, for a nonsingular diagonal matrix we simply have

$$(5) \qquad \mathbf{A} = \begin{bmatrix} a_{11} & \cdots & 0 \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ 0 & \cdots & a_{nn} \end{bmatrix}, \qquad \mathbf{A}^{-1} = \begin{bmatrix} 1/a_{11} & \cdots & 0 \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ 0 & \cdots & 1/a_{nn} \end{bmatrix};$$

the entries of $\mathbf{A}^{-1}$ on the main diagonal are the reciprocals of those of $\mathbf{A}$.

EXAMPLE 2 **Inverse of a 2 × 2 matrix**

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

EXAMPLE 3 **Inverse of a diagonal matrix**

$$\mathbf{A} = \begin{bmatrix} -0.5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{A}^{-1} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ ∎

The inverse of the inverse is the given matrix **A**:

$$(6) \qquad (\mathbf{A}^{-1})^{-1} = \mathbf{A}.$$

The simple proof is left to the reader (Prob. 16).

Inverse of a product. The inverse of a product **AC** can be calculated by inverting each factor separately and multiplying the results *in reverse order:*

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

To prove (7), we start from (1), with **A** replaced by **AC**, that is,

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

Multiplying this by $\mathbf{A}^{-1}$ from the left and using $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, we obtain

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

If we multiply this by $\mathbf{C}^{-1}$ from the left, the result follows.

Of course, (7) may be generalized to products of more than two matrices; by induction we obtain

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

Vanishing of Matrix Products. Cancellation Law

We can now obtain more information about the strange fact that for matrix multiplication, the cancellation law is not true, in general; that is, $\mathbf{AB} = \mathbf{0}$ does not necessarily imply that $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$ (as for numbers), and it does also not imply that $\mathbf{BA} = \mathbf{0}$. These facts were stated in Sec. 7.3 and illustrated with

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

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

Each of these two matrices has rank less than $n = 2$. This is typical, and the situation changes when $n \times n$ matrices have rank n:

Theorem 2 **(Cancellation law)**
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* $\mathbf{AB}$ *and* $\mathbf{BA}$.

Proof. **(a)** Premultiply $\mathbf{AB} = \mathbf{AC}$ on both sides by $\mathbf{A}^{-1}$, which exists by Theorem 1.

(b) Premultiply $\mathbf{AB} = \mathbf{0}$ on both sides by $\mathbf{A}^{-1}$.

(c_1) Rank $\mathbf{A} < n$ by Theorem 1. Hence $\mathbf{Ax} = \mathbf{0}$ has nontrivial solutions, by Theorem 2 in Sec. 7.6. Multiplication gives $\mathbf{BAx} = \mathbf{0}$. Hence those solutions also satisfy $\mathbf{BAx} = \mathbf{0}$. Hence rank $(\mathbf{BA}) < n$ by Theorem 2 in Sec. 7.6, and $\mathbf{BA}$ is singular by Theorem 1.

(c_2) $\mathbf{A}^{\mathsf{T}}$ is singular by Theorem 1 in Sec. 7.5. Hence $\mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}$ is singular by part (c_1). But $\mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}} = (\mathbf{AB})^{\mathsf{T}}$; see Sec. 7.3. Hence $\mathbf{AB}$ is singular by Theorem 1 in Sec. 7.5. ∎

Problem Set 7.7

Find the inverse and check the result, or state that the inverse does not exist, giving a reason.

1. $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ **2.** $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ **3.** $\begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}$

Find the inverse and check the result, or state that the inverse does not exist, giving a reason.

4. $\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$ **5.** $\begin{bmatrix} 6 & -2 & \frac{1}{2} \\ 1 & 5 & 2 \\ -8 & 24 & 7 \end{bmatrix}$ **6.** $\begin{bmatrix} 0 & 0 & 5 \\ 0 & -1 & 0 \\ 3 & 0 & 0 \end{bmatrix}$

7. $\begin{bmatrix} 0.5 & 0 & -0.5 \\ -0.1 & 0.2 & 0.3 \\ 0.5 & 0 & -1.5 \end{bmatrix}$ **8.** $\begin{bmatrix} 7 & 9 & 11 \\ 8 & -8 & 5 \\ 4 & 60 & 29 \end{bmatrix}$ **9.** $\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 5 & 2 \end{bmatrix}$

10. $\begin{bmatrix} 0 & 2 & 0 \\ 4 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ **11.** $\begin{bmatrix} 1 & 4 & 8 \\ 0 & 5 & 2 \\ 0 & 0 & 10 \end{bmatrix}$ **12.** $\begin{bmatrix} 10 & 0 & 0 \\ 0 & 9 & 17 \\ 0 & 4 & 8 \end{bmatrix}$

13. $\begin{bmatrix} 3 & -1 & 1 \\ -15 & 6 & -5 \\ 5 & -2 & 2 \end{bmatrix}$ **14.** $\begin{bmatrix} 19 & 2 & -9 \\ -4 & -1 & 2 \\ -2 & 0 & 1 \end{bmatrix}$ **15.** $\begin{bmatrix} 371 & -76 & -40 \\ 36 & -7 & -4 \\ -176 & 36 & 19 \end{bmatrix}$

16. Prove (6).

17. Verify (4) and (5) by showing that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

18. Show that $(\mathbf{A}^{-1})^{\mathsf{T}} = (\mathbf{A}^{\mathsf{T}})^{-1}$.

19. Show that the inverse of a nonsingular symmetric matrix is symmetric.

20. Show that $(\mathbf{A}^2)^{-1} = (\mathbf{A}^{-1})^2$. Find $(\mathbf{A}^2)^{-1}$ for the matrix in Prob. 14.

7.8 Determinants

Determinants were first defined for solving linear systems and, although ***impractical in computations***,[10] they have important engineering applications in eigenvalue problems (Sec. 7.10), differential equations (Chaps. 2, 4), vector algebra (vector products, scalar triple products, Sec. 8.3), and so on. They can be introduced in several equivalent ways, and our definition is particularly practical in connection with those systems.

An *nth-order determinant* is an expression associated with an $n \times n$ (hence square!) matrix $\mathbf{A} = [a_{jk}]$, as we now explain, beginning with $n = 2$.

Second-Order Determinants

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

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

[10] In numerical work, use a method from Secs. 7.4, 19.1–19.3; ***do not use*** **Cramer's rule** (see Sec. 7.9).

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

$$\begin{vmatrix} 4 & 3 \\ 2 & 5 \end{vmatrix} = 4 \cdot 5 - 3 \cdot 2 = 14.$$

This definition is suggested by systems

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

whose solution can be written $x_1 = D_1/D$, $x_2 = D_2/D$ with D as in (1) and

$$D_1 = \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} = b_1a_{22} - a_{12}b_2,$$

$$D_2 = \begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix} = a_{11}b_2 - b_1a_{21},$$

provided $D \neq 0$; this is called **Cramer's rule.**[11] It follows by the usual elimination. Indeed, to eliminate x_2, multiply (2a) by a_{22} and (2b) by $-a_{12}$ and add, finding

$$(a_{11}a_{22} - a_{12}a_{21})x_1 = b_1a_{22} - a_{12}b_2, \qquad \text{thus} \qquad Dx_1 = D_1.$$

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

$$(a_{11}a_{22} - a_{12}a_{21})x_2 = a_{11}b_2 - b_1a_{21}, \qquad \text{thus} \qquad Dx_2 = D_2.$$

Now divide by D (if $D \neq 0$) to get x_1 and x_2.

EXAMPLE 1 **Use of second-order determinants**

If

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

then

$$D = \begin{vmatrix} 4 & 3 \\ 2 & 5 \end{vmatrix} = 14, \qquad D_1 = \begin{vmatrix} 12 & 3 \\ -8 & 5 \end{vmatrix} = 84, \qquad D_2 = \begin{vmatrix} 4 & 12 \\ 2 & -8 \end{vmatrix} = -56,$$

so that $x_1 = 84/14 = 6$ and $x_2 = -56/14 = -4$. ∎

[11]GABRIEL CRAMER (1704—1752). Swiss mathematician.

If the system (2) is homogeneous ($b_1 = b_2 = 0$) and $D \neq 0$, it has only the trivial solution $x_1 = x_2 = 0$, and if $D = 0$, it also has nontrivial solutions.

Third-Order Determinants

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

$$(3) \qquad D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

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 by deleting from D the row and column of that entry (thus for a_{11} delete the first row and first column, etc.).

If we write out the minors, we get

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

For linear systems of three equations in three unknowns

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

Cramer's rule is

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

with the "*determinant of the system*" D given by (3) and

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

This could be derived by elimination similarly as above, but, instead, we shall obtain Cramer's rule for general n in the next section.

Determinant of Any Order n

A **determinant of order** n is a scalar associated with an $n \times n$ matrix $\mathbf{A} = [a_{jk}]$, which is written

$$(7) \qquad D = \det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

and is defined for $n = 1$ by

$$(8) \qquad D = a_{11}$$

and for $n \geqq 2$ by

$$\textbf{(9a)} \qquad \boxed{D = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn}} \qquad (j = 1, 2, \cdots, \text{ or } n)$$

or

$$\textbf{(9b)} \qquad \boxed{D = a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk}} \qquad (k = 1, 2, \cdots, \text{ or } n)$$

where

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

and M_{jk} is a determinant of order $n - 1$, namely, the determinant of the submatrix of **A** obtained from **A** by deleting the row and column of the entry a_{jk} (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; we finally arrive at second-order determinants, in which those submatrices consist of single entries whose determinant is defined by (8).

From the definition it follows that *we may* **expand** *D by any row or column*, that is, choose in (9) the entries in any row or column, similarly when expanding the C_{jk}'s in (9), and so on.

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

Terms used in connection with determinants are taken from matrices: in D we have n^2 **entries** or *elements* a_{jk}, also n **rows** and n **columns**, a **main diagonal** or *principal diagonal* on which $a_{11}, a_{22}, \cdots, a_{nn}$ stand. Two names 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 (9) may also be written in terms of minors

$$\textbf{(10a)} \qquad D = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (j = 1, 2, \cdots, \text{ or } n)$$

$$\textbf{(10b)} \qquad D = \sum_{j=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (k = 1, 2, \cdots, \text{ or } n).$$

EXAMPLE 2 **Determinant of second order**

For

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

formula (9) gives four possibilities of expanding it, namely, by

the first row: $D = a_{11}a_{22} + a_{12}(-a_{21})$,

the second row: $D = a_{21}(-a_{12}) + a_{22}a_{11}$,

the first column: $D = a_{11}a_{22} + a_{21}(-a_{12})$,

the second column: $D = a_{12}(-a_{21}) + a_{22}a_{11}$.

They all give the same value $D = a_{11}a_{22} - a_{12}a_{21}$, which agrees with (1). ∎

EXAMPLE 3 **Minors and cofactors of a third-order determinant**

In the third-order determinant

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

the minors are

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

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

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

and the cofactors are

$$C_{11} = +M_{11}, \quad C_{12} = -M_{12}, \quad C_{13} = +M_{13},$$

$$C_{21} = -M_{21}, \quad C_{22} = +M_{22}, \quad C_{23} = -M_{23},$$

$$C_{31} = +M_{31}, \quad C_{32} = -M_{32}, \quad C_{33} = +M_{33}.$$

Hence the signs form a **checkerboard pattern:**

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

∎

EXAMPLE 4 **A third-order determinant**

Let

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

The expansion by the first row is

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

The expansion by the third column is

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

etc. ∎

EXAMPLE 5 **Determinant of a triangular matrix**

The determinant of any triangular matrix equals the product of all the entries on the main diagonal. To see this, expand by rows if the matrix is lower triangular, and by columns if it is upper triangular. For instance,

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

∎

General Properties of Determinants

From our definition we may now readily obtain the most important properties of determinants, as follows.

Since the same value is obtained whether we expand a determinant by any row or any column, we have

Theorem 1 **(Transposition)**

The value of a determinant is not altered if its rows are written as columns, in the same order.

EXAMPLE 6 **Transposition**

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

∎

Theorem 2 (Multiplication by a constant)

If all the entries in one row (or one column) of a determinant are multiplied by the same factor k, the value of the new determinant is k times the value of the given determinant.

Proof. Expand the determinant by that row (or column) whose entries are multiplied by k. ∎

Caution! $\det k\mathbf{A} = k^n \det \mathbf{A}$ (not $k \det \mathbf{A}$). Explain why.

EXAMPLE 7 Application of Theorem 2

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

From Theorem 2, with $k = 0$, or directly by expanding, we obtain

Theorem 3 *If all the entries in a row (or a column) of a determinant are zero, the value of the determinant is zero.*

Theorem 4 *If each entry in a row (or a column) of a determinant is expressed as a binomial, the determinant can be written as the sum of two determinants.*

Proof. Expand the determinant by the row (or column) whose entries are binomials. ∎

EXAMPLE 8 Illustration of Theorem 4

$$\begin{vmatrix} a_1 + d_1 & b_1 & c_1 \\ a_2 + d_2 & b_2 & c_2 \\ a_3 + d_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}$$ ∎

Theorem 5 (Interchange of rows or columns)

If any two rows (or two columns) of a determinant are interchanged, the value of the determinant is multiplied by −1.

Proof. The proof is by induction. We see that the theorem holds for determinants of order $n = 2$. Assuming that it holds for determinants of order $n - 1$, we will show that it holds for determinants of order n.

Let D be of order n and E obtained from D by interchanging two rows. Expand D and E by a row that is not one of those interchanged, call it the jth row. Then, by (10a),

$$D = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk}, \qquad E = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} N_{jk} \tag{11}$$

where N_{jk} is obtained from the minor M_{jk} of a_{jk} in D by interchanging two rows. Since these minors are of order $n - 1$, the induction hypothesis applies and gives $N_{jk} = -M_{jk}$. Hence $E = -D$ by (11). This proves the statement for *rows*. To get it for *columns*, apply Theorem 1. ∎

EXAMPLE 9 **Interchange of two rows**

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

Theorem 6 **(Proportional rows or columns)**
If corresponding entries in two rows (or two columns) of a determinant are proportional, the value of the determinant is zero.

Proof. Let the entries in the ith and jth rows of D be proportional, say, $a_{ik} = ca_{jk}$, $k = 1, \cdots, n$. If $c = 0$, then $D = 0$. Now let $c \neq 0$. By Theorem 2,

$$D = cB$$

where the ith and jth rows of B are identical. Interchange these rows. Then B goes over into $-B$, by Theorem 5. On the other hand, since the rows are identical, the new determinant is still B. Thus $B = -B$, $B = 0$, and $D = 0$. ∎

EXAMPLE 10 **Proportional rows**

$$\begin{vmatrix} 3 & 6 & -4 \\ 1 & -1 & 3 \\ -6 & -12 & 8 \end{vmatrix} = 0$$ ∎

Theorem 7 **(Addition of a row or column)**
The value of a determinant is left unchanged if the entries in a row (or column) are altered by adding to them any constant multiple of the corresponding entries in any other row (or column, respectively).

Proof. Apply Theorem 4 to the determinant that results from the given addition. This yields a sum of two determinants; one is the original determinant and the other contains two proportional rows. According to Theorem 6, the second determinant is zero, and the proof is complete. ∎

Theorem 7 shows that we can evaluate a determinant by first creating zeros as in the Gauss elimination (Sec. 7.4), a method that can be programmed easily.[12] We explain it in terms of an example.

[12] In specific cases, selecting rows or columns by inspection may save work (e.g., in pocket computations), an art which old-fashioned texts emphasize to this day.

EXAMPLE 11 **Evaluation of a determinant by reduction to "triangular form"**

Explanations of computations, such as "Row 2 − 2 Row 1," always refer to the preceding determinant; they are placed behind the row where the result goes.

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

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

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

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

$$= 2 \times 5 \times 2.4 \times 47.25 = 1134.$$

In work with pencil and paper, one writes down lower order determinants when they appear, instead of carrying along zeros,

$$D = \begin{vmatrix} 2 & 0 & -4 & 6 \\ 4 & 5 & 1 & 0 \\ 0 & 2 & 6 & -1 \\ -3 & 8 & 9 & 1 \end{vmatrix} = \cdots = 2\begin{vmatrix} 5 & 9 & -12 \\ 2 & 6 & -1 \\ 8 & 3 & 10 \end{vmatrix} = \cdots = 10\begin{vmatrix} 2.4 & 3.8 \\ -11.4 & 29.2 \end{vmatrix} = 1134. \blacksquare$$

For determinants of products of matrices, there is a very useful formula that has various applications. A proof of this formula will be given in the next section.

Theorem 8 **(Determinant of a product of matrices)**

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

$$\det(\mathbf{AB}) = \det(\mathbf{BA}) = \det \mathbf{A} \det \mathbf{B}. \tag{12}$$

EXAMPLE 12 **Illustration of Theorem 8**

$$\begin{vmatrix} 2 & 4 & 3 \\ 6 & 10 & 14 \\ 4 & 7 & 9 \end{vmatrix} \begin{vmatrix} 4 & 0 & 5 \\ -2 & 1 & -1 \\ 3 & 0 & 4 \end{vmatrix} = \begin{vmatrix} 9 & 4 & 18 \\ 46 & 10 & 76 \\ 29 & 7 & 49 \end{vmatrix}$$ ∎

If the entries of a square matrix are scalars (numbers), so is the determinant. If they are functions, so is the determinant, and in this case, one occasionally needs the following theorem, which can be obtained by the product rule.

Theorem 9 **(Derivative of a determinant)**
The derivative D' of a determinant D of order n whose entries are differentiable functions can be written

$$D' = D_{(1)} + D_{(2)} + \cdots + D_{(n)} \tag{13}$$

where $D_{(j)}$ is obtained from D by differentiating the entries in the jth row.

EXAMPLE 13 **Derivative of a third-order determinant**

$$\frac{d}{dx}\begin{vmatrix} f & g & h \\ p & q & r \\ u & v & w \end{vmatrix} = \begin{vmatrix} f' & g' & h' \\ p & q & r \\ u & v & w \end{vmatrix} + \begin{vmatrix} f & g & h \\ p' & q' & r' \\ u & v & w \end{vmatrix} + \begin{vmatrix} f & g & h \\ p & q & r \\ u' & v' & w' \end{vmatrix}$$ ∎

Problem Set 7.8

Evaluate

1. $\begin{vmatrix} 17 & 9 \\ -4 & 13 \end{vmatrix}$ **2.** $\begin{vmatrix} \cos n\theta & \sin n\theta \\ -\sin n\theta & \cos n\theta \end{vmatrix}$ **3.** $\begin{vmatrix} 4.3 & 0.7 \\ 0.8 & -9.2 \end{vmatrix}$

4. $\begin{vmatrix} 1.0 & 0.2 & 1.6 \\ 3.0 & 0.6 & 1.2 \\ 2.0 & 0.8 & 0.4 \end{vmatrix}$ **5.** $\begin{vmatrix} 5 & 1 & 8 \\ 15 & 3 & 6 \\ 10 & 4 & 2 \end{vmatrix}$ **6.** $\begin{vmatrix} 4 & 6 & 5 \\ 0 & 1 & -7 \\ 0 & 0 & 6 \end{vmatrix}$

7. $\begin{vmatrix} 16 & 22 & 4 \\ 4 & -3 & 2 \\ 12 & 25 & 2 \end{vmatrix}$ **8.** $\begin{vmatrix} a & b & c \\ c & a & b \\ b & c & a \end{vmatrix}$ **9.** $\begin{vmatrix} 1.1 & 8.7 & 3.6 \\ 0 & 9.1 & -1.7 \\ 0 & 0 & 4.5 \end{vmatrix}$

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

Evaluate

13. $\begin{vmatrix} 4 & 3 & 9 & 9 \\ -8 & 3 & 5 & -4 \\ -8 & 0 & -2 & -8 \\ -16 & 6 & 14 & -5 \end{vmatrix}$ 14. $\begin{vmatrix} 4 & 3 & 0 & 0 \\ -8 & 1 & 2 & 0 \\ 0 & -7 & 3 & -6 \\ 0 & 0 & 5 & -5 \end{vmatrix}$ 15. $\begin{vmatrix} 12 & 6 & 1 & 11 \\ 4 & 4 & 1 & 4 \\ 7 & 4 & 3 & 7 \\ 8 & 2 & 3 & 9 \end{vmatrix}$

16. Show that $\det(k\mathbf{A}) = k^n \det \mathbf{A}$ (*not* $k \det \mathbf{A}$), where $\mathbf{A}$ is any $n \times n$ matrix.
17. Write the product of the determinants in Probs. 5 and 6 as a determinant.
18. Do the same task as in Prob. 17, taking the determinants in reverse order.
19. Verify that the answer to Prob. 11 equals the product of the determinants of the 2×2 submatrices containing no zero entries. Explain why.
20. Show that the straight line through two points P_1: (x_1, y_1) and P_2: (x_2, y_2) in the xy-plane is given by formula (a) (below), and derive from (a) the familiar formula (b).

$$\text{(a)} \quad \begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0 \qquad \text{(b)} \quad \frac{x - x_1}{x_1 - x_2} = \frac{y - y_1}{y_1 - y_2}$$

7.9 Rank in Terms of Determinants. Cramer's Rule

In this section we first show that the rank of a matrix $\mathbf{A}$ (defined as the maximum number of linearly independent row or column vectors of $\mathbf{A}$; see Sec. 7.5) can also be characterized in terms of determinants. This remarkable property is often used for *defining* rank. We formulate this as follows, assuming rank $\mathbf{A} > 0$ (since rank $\mathbf{A} = 0$ if and only if $\mathbf{A} = \mathbf{0}$; see Sec. 7.5).

Theorem 1 **(Rank in terms of determinants)**

An $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ has rank $r \geqq 1$ if and only if $\mathbf{A}$ has an $r \times r$ submatrix with nonzero determinant, whereas the determinant of every square submatrix with $r + 1$ or more rows that $\mathbf{A}$ has (or does not have!) is zero.

In particular, if $\mathbf{A}$ is a square matrix, $\mathbf{A}$ is nonsingular, so that the inverse $\mathbf{A}^{-1}$ of $\mathbf{A}$ exists, if and only if

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

Proof. The key lies in the fact that elementary row operations (Sec. 7.4), which do not alter the rank (by Theorem 3 in Sec. 7.5), also do not alter the property of a determinant of being zero or not zero, since the determinant is multiplied by

(*i*) -1 if we interchange two rows (Theorem 5, Sec. 7.8),
(*ii*) $c \neq 0$ if we multiply a row by $c \neq 0$ (Theorem 2, Sec. 7.8),
(*iii*) 1 if we add a multiple of a row to another row (Theorem 7, Sec. 7.8).

Let $\hat{\mathbf{A}}$ denote the echelon form of $\mathbf{A}$ (see Sec. 7.4). $\hat{\mathbf{A}}$ has r nonzero row vectors (which are the first r row vectors) if and only if rank $\mathbf{A} = r$. Let $\hat{\mathbf{R}}$ be the $r \times r$ submatrix of $\hat{\mathbf{A}}$ consisting of the r^2 entries that are simultaneously in the first r rows and the first r columns of $\hat{\mathbf{A}}$. Since $\hat{\mathbf{R}}$ is triangular and has all diagonal entries different from zero, det $\hat{\mathbf{R}} \neq 0$. Since $\hat{\mathbf{R}}$ is obtained from the corresponding $r \times r$ submatrix $\mathbf{R}$ of $\mathbf{A}$ by elementary operations, we have det $\mathbf{R} \neq 0$. Similarly, det $\mathbf{S} = 0$ for a square submatrix $\mathbf{S}$ of $r + 1$ or more rows possibly contained in $\mathbf{A}$, since the corresponding submatrix $\hat{\mathbf{S}}$ of $\hat{\mathbf{A}}$ must contain a row of zeros, so that det $\hat{\mathbf{S}} = 0$ by Theorem 3 in Sec. 7.8. This proves the assertion of the theorem for an $m \times n$ matrix.

If $\mathbf{A}$ is square, say, an $n \times n$ matrix, the statement just proved implies that rank $\mathbf{A} = n$ if and only if $\mathbf{A}$ has an $n \times n$ submatrix with a nonzero determinant; but this submatrix is $\mathbf{A}$ itself, so that det $\mathbf{A} \neq 0$. ∎

Using this theorem, we shall now derive Cramer's rule, which gives solutions of linear systems as quotients of determinants. ***Cramer's rule is not practical in computations*** (for which the methods in Secs. 7.4 and 19.1–19.3 are suitable), but is of ***theoretical interest*** in differential equations (Sec. 3.5) and other theories that have engineering applications.

Theorem 2 Cramer's Theorem (Solution of linear systems by determinants)

(a) *If the determinant* $D = \det \mathbf{A}$ *of a linear system of* n *equations*

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

in the same number of unknowns $x_1, \cdots, x_n$ *is not zero, the system has precisely one solution. This solution is given by the formulas*

$$
x_1 = \frac{D_1}{D}, \quad x_2 = \frac{D_2}{D}, \quad \cdots, \quad x_n = \frac{D_n}{D} \qquad \text{(Cramer's rule)} \tag{2}
$$

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* (1) *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. From Theorem 1 and the Fundamental Theorem in Sec. 7.6 it follows that (1) has a unique solution, because if

$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \neq 0, \tag{3}$$

then rank $\mathbf{A} = n$. We prove (2). Expanding D by the kth column, we obtain

$$D = a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk}, \tag{4}$$

where C_{ik} is the cofactor of the 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, $\tilde{D}$. Clearly, its expansion by the kth column will be of the form (4), 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 $a_{1l}, \cdots, a_{nl}$ in the lth column of D (where $l \neq k$), then the expansion of the resulting determinant $\tilde{D}$ becomes

$$a_{1l}C_{1k} + a_{2l}C_{2k} + \cdots + a_{nl}C_{nk} = 0 \qquad (l \neq k) \tag{5}$$

because $\tilde{D}$ has two identical columns and is zero (by Theorem 6 in Sec. 7.8). If we multiply the first equation in (1) by C_{1k}, the second by $C_{2k}, \cdots$, the last by C_{nk} and add the resulting equations, we first obtain

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

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

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

From this we see that x_k is multiplied by

$$a_{1k}C_{1k} + \cdots + a_{nk}C_{nk},$$

which equals D by (4), and x_l is multiplied by

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

which is zero by (5) when $l \neq k$. Hence the left side equals x_kD, and we thus have

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

The right side is D_k (as defined in the theorem) expanded by its kth column. Division by D ($\neq 0$) gives (2).

If (1) is homogeneous and $D \neq 0$, then each D_k has a column of zeros, so that $D_k = 0$ by Theorem 3 in Sec. 7.8, and (2) gives the trivial solution.

Finally, if (1) is homogeneous and $D = 0$, then rank $\mathbf{A} < n$ by Theorem 1, so that nontrivial solutions exist by Theorem 2 in Sec. 7.6. ∎

An example is included in Sec. 7.8 (Example 1).

As an important consequence of Cramer's theorem, we may now express the entries of the inverse of a matrix as follows.

Theorem 3 (Inverse of a matrix)

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

$$\text{(6)} \qquad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}[A_{jk}]^{\mathsf{T}} = \frac{1}{\det \mathbf{A}}\begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdot & \cdot & \cdots & \cdot \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix},$$

where A_{jk} is the cofactor of a_{jk} in det $\mathbf{A}$ (see Sec. 7.8). *Note well that in $\mathbf{A}^{-1}$, the cofactor A_{jk} occupies the same place as a_{kj} (not a_{jk}) does in* $\mathbf{A}$.

Proof. We denote the right side of (6) by $\mathbf{B}$ and show that $\mathbf{BA} = \mathbf{I}$. We write

$$\text{(7)} \qquad \mathbf{BA} = \mathbf{G} = [g_{kl}].$$

Here, by the definition of matrix multiplication, and by the form of the entries of $\mathbf{B}$ as given in (6),

$$\text{(8)} \qquad g_{kl} = \sum_{s=1}^{n} \frac{A_{sk}}{\det \mathbf{A}} a_{sl} = \frac{1}{\det \mathbf{A}}(a_{1l}A_{1k} + \cdots + a_{nl}A_{nk}).$$

Now (3) and (4) (with C_{jk} written in our present notation A_{jk}) show that the sum $(\cdots)$ on the right is $D = \det \mathbf{A}$ when $l = k$ and zero when $l \neq k$. Hence

$$g_{kk} = \frac{1}{\det \mathbf{A}} \det \mathbf{A} = 1, \qquad g_{kl} = 0 \quad (l \neq k),$$

so that $\mathbf{G} = [g_{kl}] = \mathbf{BA} = \mathbf{I}$ in (7). Similarly $\mathbf{AB} = \mathbf{I}$. Hence $\mathbf{B} = \mathbf{A}^{-1}$. ∎

The explicit formula (6) is often useful in theoretical studies, as opposed to *methods of actually computing inverses* (see Sec. 7.7).

EXAMPLE 1 **Illustration of Theorem 3**

Using (6), find the inverse of

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

Solution. We get $\det \mathbf{A} = -1(-7) - 1 \cdot 13 + 2 \cdot 8 = 10$, and in (6),

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

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

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

so that by (6), in agreement with Example 1 in Sec. 7.7,

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

Using Theorem 1, we may now also prove the theorem on determinants of matrix products (Theorem 8 in Sec. 7.8), which we first restate.

Theorem 4 **(Determinant of a product of matrices)**

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

$$\det(\mathbf{AB}) = \det(\mathbf{BA}) = \det \mathbf{A} \det \mathbf{B}. \tag{9}$$

Proof. If **A** is singular, so is **AB** by Theorem 2(c) in Sec. 7.7. Hence we have $\det \mathbf{A} = 0$, $\det(\mathbf{AB}) = 0$ by Theorem 1, and (9) is $0 = 0$, which holds.

Let **A** be nonsingular. Then we can reduce **A** to a diagonal matrix $\hat{\mathbf{A}} = [\hat{a}_{jk}]$ by Gauss–Jordan steps (Sec. 7.7). Under these operations, det **A** retains its value, by Theorem 7 in Sec. 7.8, except perhaps for a sign reversal if we have to interchange two rows to get a nonzero pivot (see Theorem 5 in Sec. 7.8). But the same operations reduce **AB** to $\hat{\mathbf{A}}\mathbf{B}$ with the same effect on det (**AB**). Hence it remains to prove (9) 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 $\hat{\mathbf{A}}$, since $\hat{\mathbf{A}}$ is diagonal. The remaining determinant is det **B**, and (9) is proved. ∎

This completes our discussion of linear systems (Secs. 7.4–7.9). (For numerical methods, see Secs. 19.1–19.4, which are independent of other sections on numerical methods.) Beginning with Sec. 7.10, we turn to *eigenvalue problems*, whose importance in engineering and physics can hardly be overestimated.

Problem Set 7.9

Using Theorem 1, find the rank of the following matrices.

1. $\begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix}$ **2.** $\begin{bmatrix} 30 & -70 & 50 \\ -36 & 84 & -60 \end{bmatrix}$ **3.** $\begin{bmatrix} 3 & 6 & 12 & 10 \\ 2 & 4 & 8 & 7 \end{bmatrix}$

4. $\begin{bmatrix} 0.4 & 2.0 \\ 3.2 & 1.6 \\ 0 & 1.1 \end{bmatrix}$ **5.** $\begin{bmatrix} 0 & 5 & 2 \\ 3 & 0 & -1 \\ 7 & 9 & 0 \end{bmatrix}$ **6.** $\begin{bmatrix} 21 & -3 & 17 & 13 \\ 46 & 11 & 52 & 14 \\ 33 & 48 & 71 & -23 \end{bmatrix}$

Using Theorem 3, find the inverse. Check your answer.

7. $\begin{bmatrix} 9 & 5 \\ 25 & 14 \end{bmatrix}$ **8.** $\begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix}$ **9.** $\begin{bmatrix} \cos 3\theta & \sin 3\theta \\ -\sin 3\theta & \cos 3\theta \end{bmatrix}$

10. $\begin{bmatrix} -3 & 1 & -1 \\ 15 & -6 & 5 \\ -5 & 2 & -2 \end{bmatrix}$ **11.** $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & -4 & 1 \end{bmatrix}$ **12.** $\begin{bmatrix} 2 & 5 & 4 \\ 0 & 1 & 8 \\ 0 & 0 & 10 \end{bmatrix}$

13. $\begin{bmatrix} 0 & -0.4 & 0.2 \\ 0.1 & 0.1 & -0.1 \\ -0.2 & 0.4 & 0 \end{bmatrix}$ **14.** $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **15.** $\begin{bmatrix} 19 & 2 & -9 \\ -4 & -1 & 2 \\ -2 & 0 & 1 \end{bmatrix}$

Solve by Cramer's rule and by the Gauss elimination:

16. $\begin{aligned} 4x - y &= 3 \\ -2x + 5y &= 21 \end{aligned}$ **17.** $\begin{aligned} -x + 3y - 2z &= 7 \\ 3x + 3z &= -3 \\ 2x + y + 2z &= -1 \end{aligned}$ **18.** $\begin{aligned} 2x + 5y + 3z &= 1 \\ -x + 2y + z &= 2 \\ x + y + z &= 0 \end{aligned}$

19. Using $\mathbf{A}^{-1}$ as given by (6), show that $\mathbf{AA}^{-1} = \mathbf{I}$.
20. Obtain (4) and (5) in Sec. 7.7 from the present Theorem 3.
21. Show that the product of two $n \times n$ matrices is singular if and only if at least one of the two matrices is singular.

Geometrical applications. Using Cramer's theorem, part (b), show:

22. The **plane through three points** (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3) in space is given by the formula below.
23. The **circle through three points** (x_1, y_1), (x_2, y_2), (x_3, y_3) in the plane is given by the formula below.

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0 \qquad \begin{vmatrix} x^2 + y^2 & x & y & 1 \\ x_1^2 + y_1^2 & x_1 & y_1 & 1 \\ x_2^2 + y_2^2 & x_2 & y_2 & 1 \\ x_3^2 + y_3^2 & x_3 & y_3 & 1 \end{vmatrix} = 0$$

Problem 22 **Problem 23**

24. Find the plane through (1, 1, 1), (5, 0, 5), (3, 2, 6).
25. Find the circle through (2, 6), (6, 4), (7, 1).

7.10 Eigenvalues, Eigenvectors

From the standpoint of engineering applications, eigenvalue problems are among the most important problems in connection with matrices, and the student should follow our present discussion with particular attention. We first define the basic concepts and explain them in terms of typical examples. Then we shall turn to practical applications.

Let $\mathbf{A} = [a_{jk}]$ be a given $n \times n$ matrix and consider the vector equation

$$\mathbf{Ax} = \lambda\mathbf{x} \tag{1}$$

where λ is a number.

It is clear that the zero vector $\mathbf{x} = \mathbf{0}$ is a solution of (1) for any value of λ. A value of λ for which (1) has a solution $\mathbf{x} \neq \mathbf{0}$ is called an **eigenvalue**[13] or **characteristic value** (or *latent root*) of the matrix **A**. The corresponding solutions $\mathbf{x} \neq \mathbf{0}$ of (1) are called **eigenvectors** or **characteristic vectors** of **A** corresponding to that eigenvalue λ. The set of the eigenvalues is called the **spectrum** of **A**. The largest of the absolute values of the eigenvalues of **A** is called the **spectral radius** of **A**.

The set of all eigenvectors corresponding to an eigenvalue of **A**, together with **0**, forms a vector space (Sec. 7.5), called the **eigenspace** of **A** corresponding to this eigenvalue.

The problem of determining the eigenvalues and eigenvectors of a matrix is called an *eigenvalue problem*.[14] Problems of this type occur in connection with physical and technical applications, as we shall see.

[13]German: *Eigenwert*; "*eigen*" means "proper"; "*wert*" means "value."

[14]More precisely: an *algebraic* eigenvalue problem, because there are other eigenvalue problems involving a differential equation (see Secs. 5.8 and 11.3) or an integral equation.

How to Find Eigenvalues and Eigenvectors

The following example illustrates all steps.

EXAMPLE 1 **Determination of eigenvalues and eigenvectors**

Find the eigenvalues and eigenvectors of the matrix

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

Solution. **(a)** ***Eigenvalues.*** These must be determined *first*. Equation (1) is

$$\mathbf{Ax} = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix};$$

written out in components,

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

Transferring the terms on the right to the left, we get

$$\text{(2*)} \qquad \begin{aligned} (-5 - \lambda)x_1 + \quad 2x_2 \quad &= 0 \\ 2x_1 \quad + (-2 - \lambda)x_2 &= 0. \end{aligned}$$

This can be written in matrix notation

$$\text{(3*)} \qquad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}.$$

[Indeed, (1) is $\mathbf{Ax} - \lambda\mathbf{x} = \mathbf{0}$ or $\mathbf{Ax} - \lambda\mathbf{Ix} = 0$, which gives (3).] We see that this is a *homogeneous* linear system. By Cramer's Theorem in Sec. 7.9 it has a nontrivial solution $\mathbf{x} \neq \mathbf{0}$ (an eigenvector of **A** we are looking for) if and only if its coefficient determinant is zero,

$$\text{(4*)} \qquad \begin{aligned} D(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) &= \begin{vmatrix} -5-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} \\ &= (-5-\lambda)(-2-\lambda) - 4 = \lambda^2 + 7\lambda + 6 = 0. \end{aligned}$$

We call $D(\lambda)$ the **characteristic determinant** or, if expanded, the **characteristic polynomial**, and $D(\lambda) = 0$ the **characteristic equation** of **A**. The solutions of this quadratic equation are $\lambda_1 = -1$ and $\lambda_2 = -6$. These are the eigenvalues of **A**.

(b_1) Eigenvector of **A** *corresponding to* $\boldsymbol{\lambda_1}$. 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_1 arbitrary, $x_2 = 2x_1$. If we choose $x_1 = 1$, then $x_2 = 2$, and an eigenvector of **A** corresponding to $\lambda_1 = -1$ is

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

We can easily check this:

$$\mathbf{Ax}_1 = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix} = (-1)\mathbf{x}_1 = \lambda_1\mathbf{x}_1.$$

***(b_2) Eigenvector of* A *corresponding to* λ_2.** For $\lambda = \lambda_2 = -6$, equation (2*) becomes

$$x_1 + 2x_2 = 0$$

$$2x_1 + 4x_2 = 0.$$

A solution is $x_2 = -x_1/2$. If we choose $x_1 = 2$, we get $x_2 = -1$, and an eigenvector of **A** corresponding to $\lambda_2 = -6$ is

$$\mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

Check 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

$$\text{(2)} \quad \begin{aligned} (a_{11} - \lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots&\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n &= 0. \end{aligned}$$

In matrix notation,

$$\text{(3)} \quad \boxed{(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}.}$$

By Cramer's Theorem in Sec. 7.9, this homogeneous linear system of equations has a nontrivial solution if and only if the corresponding determinant of the coefficients is zero:

$$\text{(4)} \quad D(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0.$$

$D(\lambda)$ is called the **characteristic determinant,** and (4) is called the **characteristic equation** corresponding to the matrix $\mathbf{A}$. By developing $D(\lambda)$ we obtain a polynomial of nth degree in λ. This is called the **characteristic polynomial** corresponding to $\mathbf{A}$.

This proves the following important theorem.

Theorem 1 **(Eigenvalues)**

The eigenvalues of a square matrix $\mathbf{A}$ *are the roots of the corresponding characteristic equation* (4).

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. 18.2) or another numerical approximation method in Secs. 19.7–19.10. Sometimes it may also help to observe that the product and sum of the eigenvalues are the constant term and $(-1)^{n-1}$ times the coefficient of the second highest term, respectively, of the characteristic polynomial (why?). Once an eigenvalue λ_1 has been found, one may divide the characteristic polynomial by $\lambda - \lambda_1$.

The eigen*values* 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.

Theorem 2 **(Eigenvectors)**

If $\mathbf{x}$ *is an eigenvector of a matrix* $\mathbf{A}$ *corresponding to an eigenvalue* λ, *so is* $k\mathbf{x}$ *with any* $k \neq 0$.

Proof. $\mathbf{Ax} = \lambda\mathbf{x}$ implies $k(\mathbf{Ax}) = \mathbf{A}(k\mathbf{x}) = \lambda(k\mathbf{x})$. ∎

Examples 2 and 3 will illustrate that an $n \times n$ matrix may have n linearly independent eigenvectors,[15] 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 the matrix

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

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 $\lambda_1 = 5$, $\lambda_2 = \lambda_3 = -3$. To find eigenvectors, we apply the Gauss elimination (Sec. 7.4) to the system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$, first with $\lambda = 5$ and then with $\lambda = -3$. We find that the vector

[15] A property that will play a role in Sec. 7.14.

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

is an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue 5, and the vectors

$$\mathbf{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$$

are two linearly independent eigenvectors of $\mathbf{A}$ corresponding to the eigenvalue -3. This agrees with the fact that, for $\lambda = -3$, the matrix $\mathbf{A} - \lambda\mathbf{I}$ has rank 1 and so, by Theorem 2 in Sec. 7.6, a basis of solutions of the corresponding system (2) with $\lambda = -3$,

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

consists of two linearly independent vectors. ∎

If an eigenvalue λ of a matrix $\mathbf{A}$ is a root of order M_λ of the characteristic polynomial of $\mathbf{A}$, then M_λ is called the **algebraic multiplicity** of λ, as opposed to the **geometric multiplicity** m_λ of λ, which is defined to be the number of linearly independent eigenvectors corresponding to λ, thus, the dimension of the corresponding eigenspace. Since the characteristic polynomial has degree n, the sum of all algebraic multiplicities equals 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. We convince ourselves that $m_\lambda < M_\lambda$ is possible:

EXAMPLE 3 **Algebraic and geometric multiplicity**

The characteristic equation of the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{is} \quad \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & 1 \\ 0 & -\lambda \end{vmatrix} = \lambda^2 = 0.$$

Hence $\lambda = 0$ is an eigenvalue of algebraic multiplicity 2. But its geometric multiplicity is only 1, since eigenvectors result from $-0x_1 + x_2 = 0$, hence $x_2 = 0$, 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

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{is} \quad \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0$$

and 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

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

The reader may show that, more generally, these vectors are eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \qquad (a, b \text{ real})$$

and that $\mathbf{A}$ has the eigenvalues $a + ib$ and $a - ib$. ∎

Having gained a first impression of matrix eigenvalue problems, in the next section we illustrate their importance by some typical applications.

Problem Set 7.10

Find the eigenvalues and eigenvectors of the following matrices.

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

5. $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ **6.** $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ **7.** $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ **8.** $\begin{bmatrix} 14 & -10 \\ 5 & -1 \end{bmatrix}$

9. $\begin{bmatrix} -5 & 2 \\ -9 & 6 \end{bmatrix}$ **10.** $\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$ **11.** $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ **12.** $\begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}$

13. $\begin{bmatrix} 4 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 6 \end{bmatrix}$ **14.** $\begin{bmatrix} 3 & 0 & 0 \\ 5 & 4 & 0 \\ 3 & 6 & 1 \end{bmatrix}$ **15.** $\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

16. $\begin{bmatrix} 2 & -2 & 3 \\ -2 & -1 & 6 \\ 1 & 2 & 0 \end{bmatrix}$ **17.** $\begin{bmatrix} 6 & 10 & 6 \\ 0 & 8 & 12 \\ 0 & 0 & 2 \end{bmatrix}$ **18.** $\begin{bmatrix} 32 & -24 & -8 \\ 16 & -11 & -4 \\ 72 & -57 & -18 \end{bmatrix}$

19. $\begin{bmatrix} 8 & 0 & 3 \\ 2 & 2 & 1 \\ 2 & 0 & 3 \end{bmatrix}$ **20.** $\begin{bmatrix} 5 & 0 & -15 \\ -3 & -4 & 9 \\ 5 & 0 & -15 \end{bmatrix}$ **21.** $\begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{bmatrix}$

Some general properties of the spectrum. Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of a given matrix $\mathbf{A} = [a_{jk}]$. In each case prove the proposition and illustrate it with an example.

22. **(Trace)** The so-called *trace* of $\mathbf{A}$, given by trace $\mathbf{A} = a_{11} + a_{22} + \cdots + a_{nn}$, is equal to $\lambda_1 + \cdots + \lambda_n$. The constant term of $D(\lambda)$ equals det $\mathbf{A}$.
23. If $\mathbf{A}$ is real, the eigenvalues are real or complex conjugates in pairs.
24. **(Inverse)** The inverse $\mathbf{A}^{-1}$ exists if and only if $\lambda_j \neq 0$ $(j = 1, \cdots, n)$.
25. The inverse $\mathbf{A}^{-1}$ has the eigenvalues $1/\lambda_1, \cdots, 1/\lambda_n$.
26. **(Triangular matrix)** If $\mathbf{A}$ is triangular, the entries on the main diagonal are the eigenvalues of $\mathbf{A}$.
27. **("Spectral shift")** The matrix $\mathbf{A} - k\mathbf{I}$ has the eigenvalues $\lambda_1 - k, \cdots, \lambda_n - k$.
28. The matrix $k\mathbf{A}$ has the eigenvalues $k\lambda_1, \cdots, k\lambda_n$.
29. The matrix $\mathbf{A}^m$ (m a nonnegative integer) has the eigenvalues $\lambda_1^m, \cdots, \lambda_n^m$.

30. (Spectral mapping theorem) The matrix

$$k_m \mathbf{A}^m + k_{m-1}\mathbf{A}^{m-1} + \cdots + k_1 \mathbf{A} + k_0 \mathbf{I},$$

which is called a **polynomial matrix,** has the eigenvalues

$$k_m \lambda_j^{\ m} + k_{m-1}\lambda_j^{\ m-1} + \cdots + k_1\lambda_j + k_0 \quad (j = 1, \cdots, n).$$

(This proposition is called the *spectral mapping theorem for polynomial matrices.*) The eigenvectors of that matrix are the same as those of **A**.

7.11 Some Applications of Eigenvalue Problems

In this section we discuss a few typical examples from the range of applications of matrix eigenvalue problems, which is incredibly large. Chapter 4 shows matrix eigenvalue problems related to differential equations governing mechanical systems and electrical networks. To keep our present discussion independent, for students not familiar with Chap. 4 we include a typical application of that kind as our last example.

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. 136) is stretched so that a point P: (x_1, x_2) goes over into the point Q: (y_1, y_2) given by

$$(1) \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{A}\mathbf{x} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \qquad \text{in components,} \qquad \begin{aligned} y_1 &= 5x_1 + 3x_2 \\ y_2 &= 3x_1 + 5x_2. \end{aligned}$$

Find the "*principal directions*," that is, directions of the position vector **x** of P for which the direction of the position vector **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 **x** such that $\mathbf{y} = \lambda\mathbf{x}$. Since $\mathbf{y} = \mathbf{A}\mathbf{x}$, this gives $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, an equation of the form (1), an eigenvalue problem. In components, $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ is

$$(2) \qquad \begin{aligned} 5x_1 + 3x_2 &= \lambda x_1 \\ 3x_1 + 5x_2 &= \lambda x_2 \end{aligned} \qquad \text{or} \qquad \begin{aligned} (5-\lambda)x_1 + \qquad 3x_2 &= 0 \\ 3x_1 + (5-\lambda)x_2 &= 0. \end{aligned}$$

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

$$\begin{aligned} -3x_1 + 3x_2 &= 0, \\ 3x_1 - 3x_2 &= 0. \end{aligned} \qquad \begin{aligned} &\text{Solution } x_2 = x_1, \quad x_1 \text{ arbitrary,} \\ &\text{for instance, } x_1 = x_2 = 1. \end{aligned}$$

For $\lambda_2 = 2$, our system (2) becomes

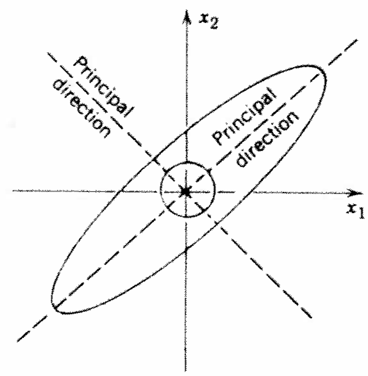


Fig. 136. Undeformed and deformed membrane in Example 1

$$\begin{aligned} 3x_1 + 3x_2 &= 0, \\ 3x_1 + 3x_2 &= 0. \end{aligned} \quad \Bigg| \quad \begin{aligned} &\text{Solution } x_2 = -x_1, \quad x_1 \text{ arbitrary,} \\ &\text{for instance, } x_1 = 1, x_2 = -1. \end{aligned}$$

We thus obtain as eigenvectors of **A**, for instance,

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ corresponding to } \lambda_1; \qquad \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ corresponding to } \lambda_2;$$

(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. 136.

Accordingly, if we choose the principal directions as directions of a new Cartesian u_1u_2-coordinate system, say, with the positive u_1-semiaxis in the first quadrant and the positive u_2-semiaxis in the second quadrant of the x_1x_2-system, and if we set

$$u_1 = r \cos \phi, \qquad u_2 = r \sin \phi,$$

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. 136)

$$\frac{z_1^2}{8^2} + \frac{z_2^2}{2^2} = 1$$

with principal semiaxes 8 and 2 in the principal directions. ∎

EXAMPLE 2 **Eigenvalue problems arising from Markov processes**

As another application, let us show that Markov processes also lead to eigenvalue problems. To see this, let us determine the limit state of the land-use succession in Example 8, Sec. 7.3.

Solution. We recall that Example 8 in Sec. 7.3 concerns a *Markov process* and that such a transition process is governed by a **stochastic matrix** $\mathbf{A} = [a_{jk}]$, that is, a square matrix with nonnegative entries a_{jk} (giving transition probabilities) and all row sums equal to 1. Furthermore, state $\mathbf{y}$ (a column vector) is obtained from state $\mathbf{x}$ according to $\mathbf{y}^T = \mathbf{x}^T\mathbf{A}$, equivalently, $\mathbf{y} = \mathbf{A}^T\mathbf{x}$. A limit is reached if states remain unchanged, if $\mathbf{x}^T = \mathbf{x}^T\mathbf{A}$ or

$$\mathbf{A}^T\mathbf{x} = \mathbf{x}. \tag{4}$$

This means that $\mathbf{A}^T$ should have the eigenvalue 1. But $\mathbf{A}^T$ has the same eigenvalues as $\mathbf{A}$ (by Theorem 1 in Sec. 7.8); and $\mathbf{A}$ has the eigenvalue 1, with eigenvector $\mathbf{v}^T = [1 \cdots 1]$, because the row sums of $\mathbf{A}$ equal 1. In our example,

$$\mathbf{A}\mathbf{v} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0 & 0.1 & 0.9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

as claimed. Hence (4) has a nontrivial solution $\mathbf{x} \neq \mathbf{0}$, which is an eigenvector of $\mathbf{A}^T$ corresponding to $\lambda = 1$. Now (4) is $(\mathbf{A}^T - \mathbf{I})\mathbf{x} = \mathbf{0}$; written out,

$$\begin{aligned} -0.2x_1 + 0.1x_2 \qquad\quad &= 0 \\ 0.1x_1 - 0.3x_2 + 0.1x_3 &= 0 \\ 0.1x_1 + 0.2x_2 - 0.1x_3 &= 0. \end{aligned}$$

A solution is $x^T = [12.5 \quad 25 \quad 62.5]$. *Answer.* Assuming that the probabilities remain the same as time progresses, we see that the states tend to 12.5% residentially, 25% commercially, and 62.5% industrially used area. ∎

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 6 years. Divide the population into three age classes of 2 years each. Let the "*Leslie matrix*" be

$$\text{(5)} \qquad \mathbf{L} = [l_{jk}] = \begin{bmatrix} 0 & 2.3 & 0.4 \\ 0.6 & 0 & 0 \\ 0 & 0.3 & 0 \end{bmatrix}$$

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 2, 4, 6 years if each class initially consists of 500 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, $\mathbf{x}_{(0)}^T = [500 \quad 500 \quad 500]$. After 2 years,

$$\mathbf{x}_{(2)} = \mathbf{L}\mathbf{x}_{(0)} = \begin{bmatrix} 0 & 2.3 & 0.4 \\ 0.6 & 0 & 0 \\ 0 & 0.3 & 0 \end{bmatrix} \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix} = \begin{bmatrix} 1350 \\ 300 \\ 150 \end{bmatrix}.$$

Similarly, after 4 years we have $\mathbf{x}_{(4)}^T = (\mathbf{L}\mathbf{x}_{(2)})^T = [750 \quad 810 \quad 90]$ and after 6 years we have $\mathbf{x}_{(6)}^T = (\mathbf{L}\mathbf{x}_{(4)})^T = [1899 \quad 450 \quad 243]$.

(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

$$\det(\mathbf{L} - \lambda\mathbf{I}) = -\lambda^3 - 0.6(-2.3\lambda - 0.3 \cdot 0.4) = -\lambda^3 + 1.38\lambda + 0.072 = 0.$$

A positive root is found to be (for instance, by Newton's method, Sec. 18.2) $\lambda = 1.2$. A corresponding eigenvector can be determined from $0.6x_1 - 1.2x_2 = 0$, $0.3x_2 - 1.2x_3 = 0$. Thus, $\mathbf{x}^T = [1 \quad 0.5 \quad 0.125]$. To get an initial population of 1500, as before, we multiply $\mathbf{x}^T$ by 923. *Answer.* 923 females in class 1, 462 in class 2, 115 in class 3. Growth rate 1.2. ∎

EXAMPLE 4 Vibrating system of two masses on two springs (Fig. 57 in Sec. 4.1)

Mass–spring systems involving several masses and springs can be treated as eigenvalue problems. For instance, the mechanical system in Fig. 57 (Sec. 4.1) is governed by the differential equations

$$(6)\qquad \begin{aligned} y_1'' &= -5y_1 + 2y_2 \\ y_2'' &= 2y_1 - 2y_2 \end{aligned}$$

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

$$(7)\qquad \mathbf{y}'' = \begin{bmatrix} y_1'' \\ y_2'' \end{bmatrix} = \mathbf{Ay} = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

We try a vector solution of the form

$$(8)\qquad \mathbf{y} = \mathbf{x}e^{\omega t}.$$

This is suggested by a mechanical system of a single mass on a spring (Sec. 2.5), whose motion is given by exponential functions (and sines and cosines). Substitution into (7) gives

$$\omega^2 \mathbf{x}e^{\omega t} = \mathbf{Ax}e^{\omega t}.$$

Dividing by $e^{\omega t}$ and writing $\omega^2 = \lambda$, we see that our mechanical system leads to the eigenvalue problem

$$(9)\qquad \mathbf{Ax} = \lambda \mathbf{x} \qquad \text{where } \lambda = \omega^2.$$

From Example 1 in Sec. 7.10 we see that $\mathbf{A}$ has the eigenvalues $\lambda_1 = -1$; consequently, $\omega = \sqrt{-1} = \pm i$, and $\lambda_2 = -6$, thus $\omega = \sqrt{-6} = \pm i\sqrt{6}$, and corresponding eigenvectors

$$(10)\qquad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

From (8) we thus obtain the four complex solutions [see (7), Sec. 2.3]

$$\begin{aligned} \mathbf{x}_1 e^{\pm it} &= \mathbf{x}_1(\cos t \pm i \sin t), \\ \mathbf{x}_2 e^{\pm i\sqrt{6}t} &= \mathbf{x}_2(\cos \sqrt{6}\,t \pm i \sin \sqrt{6}\,t), \end{aligned}$$

and by addition and subtraction (see Sec. 2.3) 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

$$\begin{aligned} y_1 &= a_1 \cos t + b_1 \sin t + 2a_2 \cos \sqrt{6}\,t + 2b_2 \sin \sqrt{6}\,t \\ y_2 &= 2a_1 \cos t + 2b_1 \sin t - a_2 \cos \sqrt{6}\,t - b_2 \sin \sqrt{6}\,t. \end{aligned}$$

These functions describe harmonic oscillations of the two masses. ∎

Problem Set 7.11

Find the principal directions and corresponding factors of extension or contraction of the elastic deformation $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{A}$ equals

1. $\begin{bmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{bmatrix}$ **2.** $\begin{bmatrix} 3/2 & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 \end{bmatrix}$ **3.** $\begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$

4. $\begin{bmatrix} 2.00 & 1.75 \\ 2.00 & 2.25 \end{bmatrix}$ **5.** $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ **6.** $\begin{bmatrix} 3.50 & 1.00 \\ 0.75 & 2.50 \end{bmatrix}$

Find limit states of the Markov processes governed by the following stochastic matrices.

7. $\begin{bmatrix} 0.3 & 0.7 \\ 0.5 & 0.5 \end{bmatrix}$ **8.** $\begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.25 & 0.50 & 0.25 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}$ **9.** $\begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.1 & 0.1 & 0.8 \\ 0 & 1.0 & 0 \end{bmatrix}$

Find the growth rate in the Leslie model with Leslie matrix

10. $\begin{bmatrix} 0 & 5.2 & 2.125 \\ 0.4 & 0 & 0 \\ 0 & 0.3 & 0 \end{bmatrix}$ **11.** $\begin{bmatrix} 0 & 8 & 0 \\ 0.5 & 0 & 0 \\ 0 & 0.2 & 0 \end{bmatrix}$ **12.** $\begin{bmatrix} 0 & 4.5 & 2.5 \\ 0.2 & 0 & 0 \\ 0 & 0.2 & 0 \end{bmatrix}$

13. **(Leontief[16] 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**

$$\mathbf{A} = [a_{jk}] = \begin{bmatrix} 0.2 & 0.5 & 0 \\ 0.6 & 0 & 0.4 \\ 0.2 & 0.5 & 0.6 \end{bmatrix}$$

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 $\mathbf{p} = [p_1 \;\; p_2 \;\; p_3]^\mathsf{T}$, 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 is consumed by the industries themselves (as in Prob. 13), then instead of $\mathbf{Ax} = \mathbf{x}$ we have $\mathbf{x} - \mathbf{Ax} = \mathbf{y}$, where $\mathbf{x} = [x_1 \;\; x_2 \;\; x_3]^\mathsf{T}$ 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 $\mathbf{y} = [0.1 \;\; 0.3 \;\; 0.1]^\mathsf{T}$ 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] WASSILY LEONTIEF (born 1906). American economist. For his work he was awarded the Nobel Prize in 1973.

16. **(Perron–Frobenius theorem)** Show that a Leslie matrix **L** with positive l_{12}, l_{13}, l_{21}, l_{32} has a positive eigenvalue. *Hint.* Use Probs. 22, 23 in Sec. 7.10. (This is a special case of the famous *Perron–Frobenius theorem* in Sec. 19.7, which is difficult to prove in its general form.)

For **differential equations** and related matrix eigenvalue problems, see Chap. 5.

7.12 Symmetric, Skew-Symmetric, and Orthogonal Matrices

We consider three classes of real square matrices that occur quite frequently in applications. These are defined as follows.

Definitions of symmetric, skew-symmetric, and orthogonal matrices

A ***real*** square matrix $\mathbf{A} = [a_{jk}]$ is called

symmetric if transposition leaves it unchanged,

$$\boxed{\mathbf{A}^\mathsf{T} = \mathbf{A},} \qquad \text{thus} \qquad a_{kj} = a_{jk}, \tag{1}$$

skew-symmetric if transposition gives the negative of **A**,

$$\boxed{\mathbf{A}^\mathsf{T} = -\mathbf{A},} \qquad \text{thus} \qquad a_{kj} = -a_{jk}, \tag{2}$$

orthogonal if transposition gives the inverse of **A**,

$$\boxed{\mathbf{A}^\mathsf{T} = \mathbf{A}^{-1}.} \tag{3}$$

EXAMPLE 1 **Symmetric, skew-symmetric, and orthogonal matrices**

The matrices

$$\begin{bmatrix} -3 & 1 & 5 \\ 1 & 0 & -2 \\ 5 & -2 & 4 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 9 & -12 \\ -9 & 0 & 20 \\ 12 & -20 & 0 \end{bmatrix}, \qquad \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{bmatrix}$$

are symmetric, skew-symmetric, and orthogonal, respectively, as the student 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)**

$$\mathbf{A} = \begin{bmatrix} 3 & -4 & -1 \\ 6 & 0 & -1 \\ -3 & 13 & -4 \end{bmatrix} = \mathbf{R} + \mathbf{S} = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & 6 \\ -2 & 6 & -4 \end{bmatrix} + \begin{bmatrix} 0 & -5 & 1 \\ 5 & 0 & -7 \\ -1 & 7 & 0 \end{bmatrix}$$ ∎

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.* (Proofs see in the next section.)

EXAMPLE 3 **Eigenvalues of symmetric and skew-symmetric matrices**

The matrices in (1) and (7) of Sec. 7.11 are symmetric and have real eigenvalues. The skew-symmetric matrix in Example 1 has the eigenvalues 0, $-25i$, and $25i$. (Verify this.) The matrix

$$\begin{bmatrix} 3 & 4 \\ 1 & 3 \end{bmatrix}$$

has the real eigenvalues 1 and 5 and is not symmetric. Does this contradict Theorem 1? ∎

Orthogonal Transformations and Matrices

Orthogonal transformations are transformations

(5) $$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad \text{with } \mathbf{A} \text{ an orthogonal matrix.}$$

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

(6) $$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is an orthogonal transformation, and one can show 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 following property of orthogonal transformations is the main reason for the importance of orthogonal matrices.

Theorem 2 **(Invariance of inner product)**

An orthogonal transformation preserves the value of the inner product of vectors (see Sec. 7.3)

(7) $$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^{\mathsf{T}}\mathbf{b},$$

hence also the **length** *or* **norm** *of a vector in R^n given by*

(8) $$\|\mathbf{a}\| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{\mathbf{a}^{\mathsf{T}}\mathbf{a}}.$$

Proof. Let $\mathbf{u} = \mathbf{Aa}$ and $\mathbf{v} = \mathbf{Ab}$, where $\mathbf{A}$ is orthogonal. We must show that $\mathbf{u}\cdot\mathbf{v} = \mathbf{a}\cdot\mathbf{b}$. Now (5) in Sec. 7.3 gives $\mathbf{u}^T = (\mathbf{Aa})^T = \mathbf{a}^T\mathbf{A}^T$. Also, $\mathbf{A}^T\mathbf{A} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ by (3). Hence

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = (\mathbf{Aa})^T\mathbf{Ab} = \mathbf{a}^T\mathbf{A}^T\mathbf{Ab} = \mathbf{a}^T\mathbf{Ib} = \mathbf{a}^T\mathbf{b} = \mathbf{a}\cdot\mathbf{b}. \tag{9}$$ ∎

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,*

$$\mathbf{a}_j\cdot\mathbf{a}_k = \mathbf{a}_j^T\mathbf{a}_k = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k. \end{cases} \tag{10}$$

Proof. **(a)** Let $\mathbf{A}$ be orthogonal. Then $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}^T\mathbf{A} = \mathbf{I}$, in terms of column vectors,

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}^T\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} [\mathbf{a}_1 \cdots \mathbf{a}_n] = \begin{bmatrix} \mathbf{a}_1^T\mathbf{a}_1 & \mathbf{a}_1^T\mathbf{a}_2 & \cdots & \mathbf{a}_1^T\mathbf{a}_n \\ \mathbf{a}_2^T\mathbf{a}_1 & \mathbf{a}_2^T\mathbf{a}_2 & \cdots & \mathbf{a}_2^T\mathbf{a}_n \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{a}_n^T\mathbf{a}_1 & \mathbf{a}_n^T\mathbf{a}_2 & \cdots & \mathbf{a}_n^T\mathbf{a}_n \end{bmatrix} = \mathbf{I}, \tag{11}$$

where 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 Prob. 22), and the column vectors of $\mathbf{A}^{-1}$ $(= \mathbf{A}^T)$ 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 the big matrix in (11) are 0 and the diagonal entries are 1. Hence $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, as (11) shows. Similarly, $\mathbf{AA}^T = \mathbf{I}$. This implies $\mathbf{A}^T = \mathbf{A}^{-1}$ because also $\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I}$ 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. This follows from $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$ and $\det \mathbf{A}^T = \det \mathbf{A}$ (Theorems 1 and 8 in Sec. 7.8). Indeed, if $\mathbf{A}$ is orthogonal, then

$$1 = \det \mathbf{I} = \det (\mathbf{AA}^{-1}) = \det (\mathbf{AA}^T) = \det \mathbf{A} \det \mathbf{A}^T = (\det \mathbf{A})^2.$$ ∎

EXAMPLE 4 Illustration of Theorems 3 and 4

The last matrix in Example 1 and the matrix in (6) illustrate Theorems 3 and 4, their determinants being -1 and $+1$, as the student 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 the next section. ∎

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 $(5 - i\sqrt{11})/6$. Verify all of this. ∎

Problem Set 7.12

Write the following matrices as the sum of a symmetric and a skew-symmetric matrix.

1. $\begin{bmatrix} 4 & -3 \\ 7 & -1 \end{bmatrix}$ **2.** $\begin{bmatrix} -2 & -2 & 3 \\ 4 & 0 & -5 \\ -1 & -3 & 5 \end{bmatrix}$ **3.** $\begin{bmatrix} 0 & -1 & 4 \\ 9 & 1 & -7 \\ -10 & 11 & -1 \end{bmatrix}$

4. Show that the main diagonal entries of a skew-symmetric matrix are all zero.

Are the following matrices symmetric? Skew-symmetric? Orthogonal? Find their eigenvalues (thereby illustrating Theorems 1 and 5).

5. $\begin{bmatrix} 6 & 8 \\ 8 & -6 \end{bmatrix}$ **6.** $\begin{bmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{bmatrix}$ **7.** $\begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix}$

8. $\begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$ **9.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$ **10.** $\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

11. $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ **12.** $\begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.25 & 0.50 & 0.25 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}$ **13.** $\begin{bmatrix} 0 & 18 & -24 \\ -18 & 0 & 40 \\ 24 & -40 & 0 \end{bmatrix}$

14. **(Symmetric matrix)** Prove that eigenvectors of a symmetric matrix corresponding to different eigenvalues are orthogonal. Give an example.

15. Find a real matrix that has real eigenvalues but is not symmetric. Does this contradict Theorem 1?

16. Show that (6) is an orthogonal transformation. Verify that Theorem 3 holds. Find the inverse transformation.

17. Let $\mathbf{v}^T = [4 \;\; 2]$, $\mathbf{x}^T = [-2 \;\; 1]$, $\mathbf{w} = \mathbf{Av}$, $\mathbf{y} = \mathbf{Ax}$ with $\mathbf{A}$ given in (6). Find $|\mathbf{v}|$, $|\mathbf{x}|$, $|\mathbf{w}|$, $|\mathbf{y}|$. Which theorem do the results illustrate?

18. Find $\mathbf{A}$ such that $\mathbf{y} = \mathbf{Ax}$ is a counterclockwise rotation through 30° in the plane.

19. Interpret the transformation $\mathbf{y} = \mathbf{Ax}$ geometrically, where $\mathbf{A}$ is the matrix in Prob. 10 and the components of $\mathbf{x}$ and $\mathbf{y}$ are Cartesian coordinates.

20. Find a 2×2 matrix that is both orthogonal and skew-symmetric. Find its eigenvalues.

21. Does there exist an orthogonal skew-symmetric 3×3 matrix? An orthogonal symmetric 3×3 matrix? (Give a reason.)

22. Show that the inverse of an orthogonal matrix is orthogonal.

23. Show that the product of two orthogonal $n \times n$ matrices is orthogonal.

24. Is the sum of two orthogonal matrices orthogonal?

25. Show that the inverse of a nonsingular skew-symmetric matrix is skew-symmetric.

7.13 Hermitian, Skew-Hermitian, and Unitary Matrices

We shall now introduce three classes of complex square matrices that generalize the three classes of real matrices just considered and have important applications, for instance, in quantum mechanics.

In this connection we use the standard notation

$$\overline{\mathbf{A}} = [\overline{a}_{jk}]$$

for the matrix obtained from $\mathbf{A} = [a_{jk}]$ by replacing each entry by its complex conjugate, and we also use the notation

$$\boxed{\overline{\mathbf{A}}^{\mathsf{T}} = [\overline{a}_{kj}]}$$

for the conjugate transpose. For example, if

$$\mathbf{A} = \begin{bmatrix} 3+4i & -5i \\ -7 & 6-2i \end{bmatrix}, \qquad \text{then} \qquad \overline{\mathbf{A}}^{\mathsf{T}} = \begin{bmatrix} 3-4i & -7 \\ 5i & 6+2i \end{bmatrix}.$$

Definitions of Hermitian,[17] Skew-Hermitian, and unitary matrices

A square matrix $\mathbf{A} = [a_{jk}]$ is called

Hermitian if $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}$, that is, $\overline{a}_{kj} = a_{jk}$

skew-Hermitian if $\overline{\mathbf{A}}^{\mathsf{T}} = -\mathbf{A}$, that is, $\overline{a}_{kj} = -a_{jk}$

unitary if $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{-1}$.

From these definitions we see the following. If $\mathbf{A}$ is Hermitian, the entries on the main diagonal must satisfy $\overline{a}_{jj} = a_{jj}$, that is, they are real. Similarly, if $\mathbf{A}$ is skew-Hermitian, then $\overline{a}_{jj} = -a_{jj}$ or, if we set $a_{jj} = \alpha + i\beta$, this becomes $\alpha - i\beta = -(\alpha + i\beta)$, so that $\alpha = 0$ and a_{jj} is pure imaginary or 0.

[17] See footnote 22 in Problem Set 5.9.

EXAMPLE 1 **Hermitian, skew-Hermitian, and unitary matrices**

The matrices

$$\mathbf{A} = \begin{bmatrix} 4 & 1-3i \\ 1+3i & 7 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3i & 2+i \\ -2+i & -i \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \frac{1}{2}i & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2}i \end{bmatrix}$$

are Hermitian, skew-Hermitian, and unitary, respectively, as the reader may verify. ∎

If a Hermitian matrix is real, then $\bar{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} = \mathbf{A}$. Hence a real Hermitian matrix is a symmetric matrix (Sec. 7.12).

Similarly, if a skew-Hermitian matrix is real, then $\bar{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} = -\mathbf{A}$. Hence a real skew-Hermitian matrix is a skew-symmetric matrix.

Finally, if a unitary matrix is real, then $\bar{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} = \mathbf{A}^{-1}$. 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 and in part accounts for the importance of the matrices under consideration that their spectra (their sets of eigenvalues; see Sec. 7.10) can be characterized in a general way as follows (see Fig. 137).

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.

Proof. Let λ be an eigenvalue of $\mathbf{A}$ and $\mathbf{x}$ a corresponding eigenvector. Then

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \tag{1}$$

(a) Let $\mathbf{A}$ be Hermitian. Multiplying (1) by $\bar{\mathbf{x}}^{\mathsf{T}}$ from the left, we obtain

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \bar{\mathbf{x}}^{\mathsf{T}}\lambda\mathbf{x} = \lambda\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}.$$

Now $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x} = \bar{x}_1 x_1 + \cdots + \bar{x}_n x_n = |x_1|^2 + \cdots + |x_n|^2$ is real, and is not 0 since $\mathbf{x} \neq \mathbf{0}$. Hence we may divide to get

$$\lambda = \frac{\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x}}{\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}}. \tag{2}$$

We see that λ is real if the numerator is real. We prove that the numerator is real by showing that it is equal to its complex conjugate, using $\bar{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}$ or $\bar{\mathbf{A}} = \mathbf{A}^{\mathsf{T}}$ and (5) in Sec. 7.3. Indeed, beginning with the application of a transposition, which has no effect on a number (the numerator), we get

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = (\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x})^{\mathsf{T}} = \mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\bar{\mathbf{x}} = \mathbf{x}^{\mathsf{T}}\bar{\mathbf{A}}\bar{\mathbf{x}} = \overline{(\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x})}. \tag{3}$$

From this and (2), whose denominator is real, we see that λ is real.

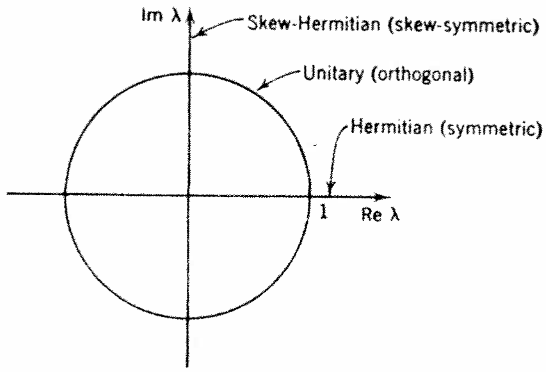


Fig. 137. Location of the eigenvalues of Hermitian, skew-Hermitian, and unitary matrices in the complex λ-plane

(b) If **A** is skew-Hermitian, then $\bar{\mathbf{A}}^{\mathsf{T}} = -\mathbf{A}$, thus $\bar{\mathbf{A}} = -\mathbf{A}^{\mathsf{T}}$, so that we get a minus sign in (3),

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = -(\overline{\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x}}). \tag{4}$$

So this is a complex number $c = a + ib$ that equals minus its conjugate $\bar{c} = a - ib$, that is, $a + ib = -(a - ib)$. Hence $a = 0$, so that c is pure imaginary or zero, and division by the real $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}$ in (2) gives a pure imaginary λ or $\lambda = 0$.

(c) Let **A** be unitary. We take (1) and its conjugate transpose,

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \qquad \text{and} \qquad (\overline{\mathbf{A}\mathbf{x}})^{\mathsf{T}} = (\overline{\lambda\mathbf{x}})^{\mathsf{T}} = \bar{\lambda}\bar{\mathbf{x}}^{\mathsf{T}}$$

and multiply the two left sides and the two right sides,

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

But **A** is unitary, $\bar{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{-1}$, so that on the left we obtain

$$(\overline{\mathbf{A}\mathbf{x}})^{\mathsf{T}}\mathbf{A}\mathbf{x} = \bar{\mathbf{x}}^{\mathsf{T}}\bar{\mathbf{A}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{I}\mathbf{x} = \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}.$$

Together, $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x} = |\lambda|^2\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}$. Now divide by $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{x}$ ($\neq 0$) to get $|\lambda|^2 = 1$, hence $|\lambda| = 1$.

This proves our present theorem as well as Theorems 1 and 5 in the previous section. ∎

EXAMPLE 2 **Illustration of Theorem 1**

For the matrices in Example 1 we find by direct calculation

	Matrix	Characteristic Equation	Eigenvalues
A	Hermitian	$\lambda^2 - 11\lambda + 18 = 0$	$9,\ 2$
B	Skew-Hermitian	$\lambda^2 - 2i\lambda + 8 = 0$	$4i,\ -2i$
C	Unitary	$\lambda^2 - i\lambda - 1 = 0$	$\frac{1}{2}\sqrt{3} + \frac{1}{2}i,\ -\frac{1}{2}\sqrt{3} + \frac{1}{2}i$

and $|\pm\frac{1}{2}\sqrt{3} + \frac{1}{2}i|^2 = \frac{3}{4} + \frac{1}{4} = 1.$ ∎

Forms

We mention that the numerator $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x}$ in (2) is called a **form** in the components $x_1, \cdots, x_n$ of $\mathbf{x}$, and $\mathbf{A}$ is called its *coefficient matrix*. When $n = 2$, we get

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = [\bar{x}_1 \quad \bar{x}_2]\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [\bar{x}_1 \quad \bar{x}_2]\begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$$

$$= \begin{cases} a_{11}\bar{x}_1x_1 + a_{12}\bar{x}_1x_2 \\ + a_{21}\bar{x}_2x_1 + a_{22}\bar{x}_2x_2. \end{cases}$$

Similarly for general n,

$$\begin{aligned} \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}\bar{x}_jx_k = \quad & a_{11}\bar{x}_1x_1 + \cdots + a_{1n}\bar{x}_1x_n \\ & + a_{21}\bar{x}_2x_1 + \cdots + a_{2n}\bar{x}_2x_n \\ & + \cdots\cdots\cdots\cdots\cdots \\ & + a_{n1}\bar{x}_nx_1 + \cdots + a_{nn}\bar{x}_nx_n. \end{aligned} \tag{5}$$

If $\mathbf{x}$ and $\mathbf{A}$ are real, then (5) becomes

$$\begin{aligned} \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_jx_k = \quad & a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\ & + \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2 \end{aligned} \tag{6}$$

and is called a **quadratic form.** Without restriction we may then assume the coefficient matrix to be ***symmetric***, because we can take off-diagonal terms together in pairs and then write the result as a sum of two equal terms, as the next example (Example 3) illustrates. 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. (Their "transformation to principal axes" will be discussed in the next section.)

EXAMPLE 3 **Quadratic form. Symmetric coefficient matrix C**

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 $c_{jk} = \frac{1}{2}(a_{jk} + a_{jk})$, thus $c_{11} = 3$, $c_{12} = c_{21} = 5$, $c_{22} = 2$, we get the same result

$$\mathbf{x}^{\mathsf{T}}\mathbf{C}\mathbf{x} = [x_1 \quad x_2]\begin{bmatrix} 3 & 5 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 5x_1x_2 + 5x_2x_1 + 2x_2^2 = 3x_1^2 + 10x_1x_2 + 2x_2^2. \quad \blacksquare$$

If the matrix $\mathbf{A}$ in (5) is Hermitian or skew-Hermitian, the form (5) is called a **Hermitian form** or **skew-Hermitian form,** respectively. These forms have the following property, which accounts for their importance in physics.

Theorem 1* **(Hermitian and skew-Hermitian forms)**
For every choice of the vector $\mathbf{x}$ *the value of a Hermitian form is real, and the value of a skew-Hermitian form is pure imaginary or* 0.

Proof. In proving (3) and (4), we made no use of the fact that $\mathbf{x}$ was an eigenvector, and the proofs remain valid for any vectors (and Hermitian or skew-Hermitian matrices). From this, our present theorem follows. ∎

EXAMPLE 4 **Hermitian form**
If

$$\mathbf{A} = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1+i \\ 2i \end{bmatrix},$$

then

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x} = [1-i \quad -2i]\begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix}\begin{bmatrix} 1+i \\ 2i \end{bmatrix}$$

$$= [1-i \quad -2i]\begin{bmatrix} 3(1+i)+(2-i)2i \\ (2+i)(1+i)+4\cdot 2i \end{bmatrix} = 34.$$ ∎

Properties of Unitary Matrices. Complex Vector Space C^n

We now extend our discussion of orthogonal matrices in Sec. 7.12 to unitary matrices. Instead of the real vector space R^n of all real vectors with n components and real numbers as scalars, 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

$$\mathbf{a}\cdot\mathbf{b} = \bar{\mathbf{a}}^{\mathsf{T}}\mathbf{b} \tag{7}$$

and the **length** or **norm** of a vector by

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}\cdot\mathbf{a}} = \sqrt{\bar{\mathbf{a}}^{\mathsf{T}}\mathbf{a}} = \sqrt{\bar{a}_1 a_1 + \cdots + \bar{a}_n a_n} = \sqrt{|a_1|^2 + \cdots + |a_n|^2}. \tag{8}$$

Note that for *real* vectors this reduces to the inner product as defined in Sec. 7.3.

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* (7), *hence also the norm* (8).

Proof. The proof is the same as that of Theorem 2 in Sec. 7.12, which the theorem generalizes; in the analog of (9), Sec. 7.12, we now have bars,

$$\mathbf{u}\cdot\mathbf{v} = \bar{\mathbf{u}}^{\mathsf{T}}\mathbf{v} = (\overline{\mathbf{A}\mathbf{a}})^{\mathsf{T}}\mathbf{A}\mathbf{b} = \bar{\mathbf{a}}^{\mathsf{T}}\bar{\mathbf{A}}^{\mathsf{T}}\mathbf{A}\mathbf{b} = \bar{\mathbf{a}}^{\mathsf{T}}\mathbf{I}\mathbf{b} = \bar{\mathbf{a}}^{\mathsf{T}}\mathbf{b} = \mathbf{a}\cdot\mathbf{b}.$$ ∎

The complex analog of an *orthonormal system* of real vectors (see Sec. 7.12) is a **unitary system,** defined by

$$\mathbf{a}_j \cdot \mathbf{a}_k = \bar{\mathbf{a}}_j^{\mathsf{T}} \mathbf{a}_k = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k, \end{cases} \tag{9}$$

and the extension of Theorem 3, Sec. 7.12, to complex is as follows.

Theorem 3 **(Unitary systems of column and row vectors)**
A 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. 7.12, except for the bars required by the definitions $\overline{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{-1}$ and (7) and (9). ∎

Theorem 4 **(Determinant of a unitary matrix)**
The determinant of a unitary matrix has absolute value 1.

Proof. Similarly as in Sec. 7.12 we get

$$1 = \det \mathbf{A}\mathbf{A}^{-1} = \det(\mathbf{A}\overline{\mathbf{A}}^{\mathsf{T}}) = \det \mathbf{A} \det \overline{\mathbf{A}}^{\mathsf{T}} = \det \mathbf{A} \det \overline{\mathbf{A}}$$
$$= \det \mathbf{A}\, \overline{\det \mathbf{A}} = |\det \mathbf{A}|^2.$$

Hence $|\det \mathbf{A}| = 1$ (where det **A** may now be complex). ∎

EXAMPLE 5 **Unitary matrix illustrating Theorems 2–4**
For the vectors $\mathbf{a}^{\mathsf{T}} = [1 \quad i]$ and $\mathbf{b}^{\mathsf{T}} = [3i \quad 2 + i]$ we get $\bar{\mathbf{a}}^{\mathsf{T}}\mathbf{b} = 3i - i(2 + i) = 1 + i$, and with

$$\mathbf{A} = \begin{bmatrix} 0.6i & 0.8 \\ 0.8 & 0.6i \end{bmatrix} \quad \text{also} \quad \mathbf{A}\mathbf{a} = \begin{bmatrix} 1.4i \\ 0.2 \end{bmatrix} \quad \text{and} \quad \mathbf{A}\mathbf{b} = \begin{bmatrix} -0.2 + 0.8i \\ -0.6 + 3.6i \end{bmatrix},$$

as one can readily verify. This gives $(\overline{\mathbf{A}\mathbf{a}})^{\mathsf{T}}\mathbf{A}\mathbf{b} = 1 + i$, illustrating Theorem 2. The matrix is unitary. Its columns form a unitary system, and so do the rows, as we see. Also, $\det \mathbf{A} = -1$. ∎

Problem Set 7.13

1. Verify the eigenvalues in Example 2.

In Examples 1 and 2, find eigenvectors of

2. The matrix **A**
3. The matrix **B**
4. The matrix **C**

Indicate whether the following matrices are Hermitian, skew-Hermitian, or unitary and find their eigenvalues (thereby verifying Theorem 1) and eigenvectors.

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

8. $\begin{bmatrix} 1/\sqrt{3} & i\sqrt{2/3} \\ -i\sqrt{2/3} & -1/\sqrt{3} \end{bmatrix}$ 9. $\begin{bmatrix} 1/\sqrt{2} & i/\sqrt{2} \\ -i/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$ 10. $\begin{bmatrix} i & 0 & 0 \\ 0 & 0 & i \\ 0 & i & 0 \end{bmatrix}$

11. Show that the product of two $n \times n$ unitary matrices is unitary.
12. Show that the inverse of a unitary matrix is unitary. Verify this for the matrix in Prob. 10.
13. Verify Theorems 3 and 4 for the matrix in Prob. 9.
14. Show that any square matrix may be written as the sum of a Hermitian matrix and a skew-Hermitian matrix.
15. **(Normal matrix)** By definition, a *normal matrix* is a square matrix that commutes with its conjugate transpose,

$$\mathbf{A}\bar{\mathbf{A}}^{\mathsf{T}} = \bar{\mathbf{A}}^{\mathsf{T}}\mathbf{A}.$$

Show that Hermitian, skew-Hermitian, and unitary matrices are normal.

Quadratic forms. Find a symmetric matrix **C** such that $Q = \mathbf{x}^{\mathsf{T}}\mathbf{C}\mathbf{x}$, where Q equals

16. $x_1^2 - 4x_1x_2 + 7x_2^2$
17. $(x_1 - 3x_2)^2$
18. $(x_1 + x_2 + x_3)^2$
19. $-3x_1^2 + 4x_1x_2 - x_2^2 + 2x_1x_3 - 5x_3^2$
20. $(x_1 - x_2 + 2x_3 - 2x_4)^2$
21. $(x_1 + x_2)^2 + (x_3 + x_4)^2$
22. **(Definiteness)** A real quadratic form $Q = \mathbf{x}^{\mathsf{T}}\mathbf{C}\mathbf{x}$ and its symmetric matrix $\mathbf{C} = [c_{jk}]$ are said to be **positive definite** if $Q > 0$ for all $[x_1 \cdots x_n] \neq [0 \cdots 0]$. A necessary and sufficient condition for positive definiteness is that all the determinants

$$C_1 = c_{11}, \quad C_2 = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix}, \quad C_3 = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix}, \cdots, \quad C_n = \det \mathbf{C}$$

are positive (see Ref. [B2], vol. 1, p. 306). Show that the form in Prob. 16 is positive definite, whereas that in Example 3 is not positive definite.

Hermitian and skew-Hermitian forms. Is **A** Hermitian or skew-Hermitian? Find $\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\mathbf{x}$.

23. $\mathbf{A} = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ i \end{bmatrix}$
24. $\mathbf{A} = \begin{bmatrix} 2 & 1+i \\ 1-i & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
25. $\mathbf{A} = \begin{bmatrix} i & 1 \\ -1 & 2i \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ i \end{bmatrix}$
26. $\mathbf{A} = \begin{bmatrix} a & b+ic \\ b-ic & k \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
27. $\mathbf{A} = \begin{bmatrix} 0 & i & 0 \\ -i & 1 & -2i \\ 0 & 2i & 2 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} i \\ 1 \\ -i \end{bmatrix}$
28. $\mathbf{A} = \begin{bmatrix} i & 1+i & 2 \\ -1+i & -3i & 3+i \\ -2 & -3+i & 0 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} i \\ 1 \\ -i \end{bmatrix}$

Is **A** Hermitian or skew-Hermitian? Find $\bar{\mathbf{x}}^T\mathbf{A}\mathbf{x}$.

29. $\mathbf{A} = \begin{bmatrix} 3 & -i & 0 \\ i & 0 & 2i \\ 0 & -2i & 4 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

30. $\mathbf{A} = \begin{bmatrix} 2i & 0 & 4 \\ 0 & i & 5-i \\ -4 & -5-i & 4i \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 0 \\ 2i \\ -3 \end{bmatrix}$

7.14 Properties of Eigenvectors. Diagonalization

In our discussion of eigenvalue problems so far we have emphasized properties of eigen*values*. We now turn to eigen*vectors* and their properties. Eigenvectors of an $n \times n$ matrix **A** may (or may not!) form a basis for R^n or C^n (see Sec. 7.13), and if they do, we can use them for "*diagonalizing*" **A**, that is, for transforming it into diagonal form with the eigenvalues on the main diagonal. These are the key issues in this section.

We begin with a concept of central interest in eigenvalue problems:

Similarity of Matrices

An $n \times n$ matrix $\hat{\mathbf{A}}$ is called **similar** to an $n \times n$ matrix **A** if

$$\hat{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} \tag{1}$$

for some (nonsingular!) $n \times n$ matrix **T**. This transformation, which gives $\hat{\mathbf{A}}$ from **A**, is called a **similarity transformation.**

Similarity transformations are important since they preserve eigenvalues:

Theorem 1 **(Eigenvalues and eigenvectors of similar matrices)**

If $\hat{\mathbf{A}}$ is similar to **A**, *then $\hat{\mathbf{A}}$ has the same eigenvalues as* **A**.

Furthermore, if **x** *is an eigenvector of* **A**, *then* $\mathbf{y} = \mathbf{T}^{-1}\mathbf{x}$ *is an eigenvector of $\hat{\mathbf{A}}$ corresponding to the same eigenvalue.*

Proof. From $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ (λ an eigenvalue, $\mathbf{x} \neq \mathbf{0}$) we get $\mathbf{T}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{T}^{-1}\mathbf{x}$. Now $\mathbf{I} = \mathbf{T}\mathbf{T}^{-1}$, so that

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{x} = \mathbf{T}^{-1}\mathbf{A}\mathbf{I}\mathbf{x} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{T}^{-1}\mathbf{x} = \hat{\mathbf{A}}(\mathbf{T}^{-1}\mathbf{x}) = \lambda\mathbf{T}^{-1}\mathbf{x}.$$

Hence λ is an eigenvalue of $\hat{\mathbf{A}}$ and $\mathbf{T}^{-1}\mathbf{x}$ a corresponding eigenvector, because $\mathbf{T}^{-1}\mathbf{x} = \mathbf{0}$ would give $\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{T}\mathbf{T}^{-1}\mathbf{x} = \mathbf{T}\mathbf{0} = \mathbf{0}$, contradicting $\mathbf{x} \neq \mathbf{0}$. ∎

Properties of Eigenvectors

The next theorem is of interest in itself and of help in connection with bases of eigenvectors.

Theorem 2 (Linear independence of eigenvectors)
Let $\lambda_1, \lambda_2, \cdots, \lambda_k$ be ***distinct*** *eigenvalues of an $n \times n$ matrix. Then corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$ form a linearly independent set.*

Proof. Suppose that the conclusion is false. Let r be the largest integer such that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is a linearly independent set. Then $r < k$ and the set $\{\mathbf{x}_1, \cdots, \mathbf{x}_{r+1}\}$ 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. 7.5). Multiplying both sides by $\mathbf{A}$ and using $\mathbf{A}\mathbf{x}_j = \lambda_j\mathbf{x}_j$, we obtain

$$c_1\lambda_1\mathbf{x}_1 + \cdots + c_{r+1}\lambda_{r+1}\mathbf{x}_{r+1} = \mathbf{0}. \tag{3}$$

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. ∎

This theorem immediately implies the following.

Theorem 3 (Basis of eigenvectors)
If an $n \times n$ matrix $\mathbf{A}$ has n ***distinct*** *eigenvalues, then $\mathbf{A}$ has a basis of eigenvectors for C^n (or R^n).*

EXAMPLE 1 Basis of eigenvectors

The matrix

$$\mathbf{A} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} \quad \text{has a basis of eigenvectors} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

corresponding to the eigenvalues $\lambda_1 = 8$, $\lambda_2 = 2$. (See Example 1 in Sec. 7.11.) ∎

EXAMPLE 2 Basis when not all eigenvalues are distinct. Nonexistence of basis

Even if not all n eigenvalues are different, a matrix $\mathbf{A}$ may still provide a basis of eigenvectors for C^n or R^n. This is illustrated by Example 2 in Sec. 7.10, where $n = 3$. On the other hand, $\mathbf{A}$ may not have enough linearly independent eigenvectors to make up a basis. For instance, the matrix in Example 3, Sec. 7.10,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{has only one eigenvector} \quad \begin{bmatrix} k \\ 0 \end{bmatrix},$$

where k is arbitrary, not zero. Hence $\mathbf{A}$ does not provide a basis of eigenvectors for R^2. ∎

Actually, bases of eigenvectors exist under much more general conditions than those given in Theorem 3, and for the matrices in the previous section we can even choose a unitary system of eigenvectors, as follows.

Theorem 4 **(Basis of eigenvectors)**
A Hermitian, skew-Hermitian, or unitary matrix has a basis of eigenvectors for C^n that is a unitary system (see Sec. 7.13). *A symmetric matrix has an orthonormal basis of eigenvectors for R^n.* (Proof see Ref. [B2], vol. 1, pp. 270–272.)

EXAMPLE 3 **Orthonormal basis of eigenvectors**
The matrix in Example 1 is symmetric, and an orthonormal basis of eigenvectors is $[1/\sqrt{2} \quad 1/\sqrt{2}]^T$, $[1/\sqrt{2} \quad -1/\sqrt{2}]^T$. ∎

A basis of eigenvectors of a matrix $\mathbf{A}$ is of great advantage if we are interested in a transformation $\mathbf{y} = \mathbf{Ax}$ because then we can represent any $\mathbf{x}$ uniquely as

$$\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n$$

in terms of such a basis $\mathbf{x}_1, \cdots, \mathbf{x}_n$, and if these eigenvectors of $\mathbf{A}$ correspond to (not necessarily distinct) eigenvalues $\lambda_1, \cdots, \lambda_n$ of $\mathbf{A}$, then we get

$$\begin{aligned} \mathbf{y} = \mathbf{Ax} &= \mathbf{A}(c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n) \\ &= c_1\mathbf{Ax}_1 + \cdots + c_n\mathbf{Ax}_n \qquad (4) \\ &= c_1\lambda_1\mathbf{x}_1 + \cdots + c_n\lambda_n\mathbf{x}_n. \end{aligned}$$

This shows the advantage: we have decomposed the complicated action of $\mathbf{A}$ on arbitrary vectors $\mathbf{x}$ into a sum of simple actions (multiplication by scalars) on the eigenvectors of $\mathbf{A}$.

Diagonalization

Bases of eigenvectors also play a central role in the diagonalization of an $n \times n$ matrix $\mathbf{A}$, as the following theorem explains.

Theorem 5 **(Diagonalization of a matrix)**
If an $n \times n$ matrix $\mathbf{A}$ has a basis of eigenvectors, then

$$\boxed{\mathbf{D} = \mathbf{X}^{-1}\mathbf{AX}} \qquad (5)$$

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,

$$\mathbf{D}^m = \mathbf{X}^{-1}\mathbf{A}^m\mathbf{X}. \qquad (5^*)$$

Proof. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ form a basis of eigenvectors of $\mathbf{A}$ for C^n (or R^n) corresponding to the eigenvalues $\lambda_1, \cdots, \lambda_n$, respectively, of $\mathbf{A}$. Then $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]$ has rank n, by Theorem 1 in Sec. 7.5. Hence $\mathbf{X}^{-1}$ exists by Theorem 1 in Sec. 7.7. Now (9) in Sec. 7.3 and $\mathbf{A}\mathbf{x}_j = \lambda_j \mathbf{x}_j$ give

$$\mathbf{AX} = \mathbf{A}[\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n] = [\mathbf{A}\mathbf{x}_1 \quad \cdots \quad \mathbf{A}\mathbf{x}_n] = [\lambda_1\mathbf{x}_1 \quad \cdots \quad \lambda_n\mathbf{x}_n].$$

Together, $\mathbf{AX} = \mathbf{XD}$. This we multiply on both sides by $\mathbf{X}^{-1}$ from the left to get $\mathbf{X}^{-1}\mathbf{AX} = \mathbf{X}^{-1}\mathbf{XD} = \mathbf{D}$, which is (5). Also, (5*) follows by noting that

$$\mathbf{D}^2 = \mathbf{DD} = \mathbf{X}^{-1}\mathbf{AXX}^{-1}\mathbf{AX} = \mathbf{X}^{-1}\mathbf{AAX} = \mathbf{X}^{-1}\mathbf{A}^2\mathbf{X}, \qquad \text{etc.} \quad ∎$$

EXAMPLE 4 **Diagonalization**

Calculation as in the examples in Sec. 7.10, etc. shows that the matrix

$$\mathbf{A} = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix} \text{ has eigenvectors } \begin{bmatrix} 4 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \qquad \text{Hence } \mathbf{X} = \begin{bmatrix} 4 & 1 \\ 1 & -1 \end{bmatrix}$$

and, using (4) in Sec. 7.7, we obtain

$$\mathbf{X}^{-1}\mathbf{AX} = \frac{1}{-5}\begin{bmatrix} -1 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 4 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.2 \\ 0.2 & -0.8 \end{bmatrix}\begin{bmatrix} 24 & 1 \\ 6 & -1 \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}.$$

The student may show that an interchange of the columns of $\mathbf{X}$ results in an interchange of the eigenvalues 6 and 1 in the diagonal matrix. ∎

EXAMPLE 5 **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. 7.7, Example 1). The results are

$$\begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}, \quad \mathbf{X}^{-1} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix}.$$

Calculating $\mathbf{AX}$ and multiplying by $\mathbf{X}^{-1}$ from the left, we thus obtain

$$\mathbf{D} = \mathbf{X}^{-1}\mathbf{AX} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix}\begin{bmatrix} -3 & -4 & 0 \\ 9 & 4 & 0 \\ -3 & -12 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad ∎$$

Transformation of Forms to Principal Axes

This is an important practical task related to the diagonalization of matrices. We explain the idea for quadratic forms (see Sec. 7.13)

$$Q = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}. \tag{6}$$

Without restriction we can assume that $\mathbf{A}$ is real ***symmetric*** (see Sec. 7.13). Then $\mathbf{A}$ has an orthonormal basis of n eigenvectors, by Theorem 4. Hence the matrix $\mathbf{X}$ with these vectors as column vectors 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 (6) gives

$$Q = \mathbf{x}^{\mathsf{T}}\mathbf{X}\mathbf{D}\mathbf{X}^{\mathsf{T}}\mathbf{x}.$$

If we set $\mathbf{X}^{\mathsf{T}}\mathbf{x} = \mathbf{y}$, then, since $\mathbf{X}^{\mathsf{T}} = \mathbf{X}^{-1}$, we get

$$\mathbf{x} = \mathbf{X}\mathbf{y} \tag{7}$$

and Q becomes simply

$$Q = \mathbf{y}^{\mathsf{T}}\mathbf{D}\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2. \tag{8}$$

This proves

Theorem 6 (Principal axes theorem)
The substitution (7) *transforms a quadratic form*

$$Q = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_j x_k$$

to the principal axes form (8), *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 (8) 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 (7). 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. 7.11, Example 1, except for the notations. See also Fig. 136 in that example. ∎

Problem Set 7.14

Similarity transformations

Find $\hat{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$. Find the eigenvalues of $\hat{\mathbf{A}}$ and $\mathbf{A}$ and verify that they are the same. Find corresponding eigenvectors $\mathbf{y}$ of $\hat{\mathbf{A}}$, compute $\mathbf{x} = \mathbf{T}\mathbf{y}$, and verify that they are eigenvectors of $\mathbf{A}$.

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

2. $\mathbf{A} = \begin{bmatrix} 6 & 8 \\ 8 & -6 \end{bmatrix}$, $\mathbf{T} = \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}$

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

4. $\mathbf{A} = \begin{bmatrix} -2 & 0 \\ -4 & 2 \end{bmatrix}$, $\mathbf{T} = \begin{bmatrix} 10 & 3 \\ 0 & 1 \end{bmatrix}$

5. $\mathbf{A} = \begin{bmatrix} 5 & 10 \\ 4 & -1 \end{bmatrix}$, $\mathbf{T} = \begin{bmatrix} 5 & 1 \\ 2 & -1 \end{bmatrix}$

6. $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix}$, $\mathbf{T} = \begin{bmatrix} 7 & -5 \\ 10 & -7 \end{bmatrix}$

7. $\mathbf{A} = \begin{bmatrix} 8 & 0 & 3 \\ 2 & 2 & 1 \\ 2 & 0 & 3 \end{bmatrix}$, $\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

8. $\mathbf{A} = \begin{bmatrix} 5 & 0 & -15 \\ -3 & -4 & 9 \\ 5 & 0 & -15 \end{bmatrix}$, $\mathbf{T} = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 1 & 0 \\ 3 & 0 & 5 \end{bmatrix}$

Traces of similar matrices

The sum of the entries on the main diagonal of an $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ is called the **trace** of $\mathbf{A}$; thus trace $\mathbf{A} = a_{11} + a_{22} + \cdots + a_{nn}$.

9. Show that trace $\mathbf{AB} = \sum_{i=1}^{n} \sum_{l=1}^{n} a_{il}b_{li} =$ trace $\mathbf{BA}$, where $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ are $n \times n$ matrices.
10. Using Prob. 9, show that similar matrices have equal traces.

11. **(Sum of the eigenvalues)** Show that trace $\mathbf{A}$ equals the sum of the eigenvalues of $\mathbf{A}$, each counted as often as its algebraic multiplicity indicates.
12. Using Prob. 11, show that trace $\hat{\mathbf{A}}$ = trace $\mathbf{A}$ when $\hat{\mathbf{A}}$ is similar to $\mathbf{A}$.

Find a basis of eigenvectors that form a unitary system (or a real orthogonal system).

13. $\begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$
14. $\begin{bmatrix} 0 & 2i \\ 2i & 0 \end{bmatrix}$
15. $\begin{bmatrix} 0.8i & 0.6i \\ 0.6i & -0.8i \end{bmatrix}$
16. $\begin{bmatrix} 1 & 1+i \\ 1-i & 1 \end{bmatrix}$
17. $\begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix}$
18. $\begin{bmatrix} \frac{1}{2}i & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2}i \end{bmatrix}$

Diagonalization

Find a basis of eigenvectors and diagonalize:

19. $\begin{bmatrix} 0 & 16 \\ 4 & 0 \end{bmatrix}$
20. $\begin{bmatrix} 3 & 3 \\ 0 & 0 \end{bmatrix}$
21. $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
22. $\begin{bmatrix} 3 & 4 \\ 1 & 3 \end{bmatrix}$
23. $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 2 \end{bmatrix}$
24. $\begin{bmatrix} 5 & 0 & -6 \\ 2 & 1 & -4 \\ 3 & 0 & -4 \end{bmatrix}$

Transformation of quadratic forms to principal axes. Conic sections

Find out what type of conic section (or pair of straight lines) is represented by the given quadratic form. Transform it to principal axes. Express $\mathbf{x}^T = [x_1 \ \ x_2]$ in terms of the new coordinate vector $\mathbf{y}^T = [y_1 \ \ y_2]$, as in Example 6.

25. $x_1^2 + 24x_1x_2 - 6x_2^2 = 5$
26. $2x_1^2 + 2\sqrt{3}x_1x_2 + 4x_2^2 = 5$
27. $3x_1^2 + 4\sqrt{3}x_1x_2 + 7x_2^2 = 9$
28. $-3x_1^2 + 8x_1x_2 + 3x_2^2 = 0$
29. $x_1^2 + 6x_1x_2 + 9x_2^2 = 10$
30. $6x_1^2 + 16x_1x_2 - 6x_2^2 = 10$

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

From Sec. 7.5 we recall that the **real vector space** R^n is the set of all real vectors with n components (thus, each such vector is an ordered n-tuple of real numbers), with the two algebraic operations of vector addition and multiplication by scalars (real numbers). Similarly, taking ordered n-tuples of *complex* numbers as vectors and *complex* numbers as scalars, we obtain the **complex vector space** C^n (see Sec. 7.13).

There are other sets of practical interest (sets of matrices, functions, transformations, etc.) for which an addition and a scalar multiplication can be defined in a natural way. The desire to treat such sets as "vector spaces" suggests we create from the "**concrete model**" R^n the "**abstract concept**" of a "real vector space" V by taking the most basic properties of R^n as axioms by which V is defined, properties without which one would not be able to create a useful and applicable theory of those more general situations. Selecting good axioms is not easy, but needs experience, sometimes gained only over a long period of time. In the present case, the following system of axioms turned out to be useful; note that each axiom expresses a simple property of R^n, or of R^3, as a matter of fact.

Definition of a real vector space

A nonempty set V of elements **a, b,** · · · is called a *real vector space* (or *real linear space*), and these elements are called **vectors,**[18] 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 **a** and **b** of V a unique vector of V, called the *sum* of **a** and **b** and denoted by **a** + **b**, such that the following axioms are satisfied.

I.1 *Commutativity*. For any two vectors **a** and **b** of V,

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

I.2 *Associativity*. For any three vectors **u, v, w** of V,

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) \qquad \text{(written } \mathbf{u} + \mathbf{v} + \mathbf{w}\text{)}.$$

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

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

I.4 For every **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 **a** in V and every scalar c a unique vector of V, called the *product* of c and **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 **a** and **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 **a** in V,

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

[18] Regardless of what they actually are; this convention causes no confusion because in any specific case the nature of those elements is clear from the context.

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,

$$1\mathbf{a} = \mathbf{a}.$$

A **complex vector space** is obtained if, instead of real numbers, we take complex numbers as scalars.

Basic Concepts Related to Vector Space

These concepts are defined as in Sec. 7.5.

A **linear combination** of vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ 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

$$(1) \qquad c_1\mathbf{a}_{(1)} + \cdots + c_m\mathbf{a}_{(m)} = \mathbf{0}$$

implies that $c_1 = 0, \cdots, c_m = 0$; otherwise they 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, called a **basis** for V, whereas any set of more than n vectors in V is linearly dependent. Then every vector in V can be uniquely written as a linear combination of the basis vectors.

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 $\mathbf{A} = [a_{jk}] = a_{11}\mathbf{B}_{11} + a_{12}\mathbf{B}_{12} + a_{21}\mathbf{B}_{21} + a_{22}\mathbf{B}_{22}$ in a unique fashion. 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 skew-symmetric 3×3 matrices? Can you find a basis?

EXAMPLE 2 **Polynomials**

The set of all constant, linear and quadratic polynomials in x together forms a vector space under the usual addition and multiplication by a real number, since these two operations give polynomials of degree not exceeding 2, and the axioms in our definition follow by direct calculation. This space has dimension 3. A basis is $\{1, \ x, \ x^2\}$.

EXAMPLE 3 **Second-order homogeneous linear differential equations**

The solutions of such an equation on a fixed interval $a < x < b$ form a vector space under the usual addition and multiplication by a number since these two operations give again such a solution, by Fundamental Theorem 1 in Sec. 2.1, and I.1 to II.4 follow by direct calculation. Do the solutions of a nonhomogeneous linear differential equation form a vector space?

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 as defined before. An example is the space of all continuous functions on some interval $[a, b]$ of the x-axis, as we mention without proof.

Inner Product Spaces

From Sec. 7.3 we know that for vectors $\mathbf{a}$ and $\mathbf{b}$ in R^n we can define an inner product $\mathbf{a}\cdot\mathbf{b} = \mathbf{a}^T\mathbf{b}$. This definition can be extended to general real vector spaces by taking basic properties of $\mathbf{a}\cdot\mathbf{b}$ as axioms for an "abstract" inner product, denoted by $(\mathbf{a}, \mathbf{b})$.

Definition of a real inner product space

A real vector space V is called a *real inner product space* (or *real pre-Hilbert*[19] *space*) if it has the following property. With every pair of vectors $\mathbf{a}$ and $\mathbf{b}$ in V there is associated a real number, which is denoted by $(\mathbf{a}, \mathbf{b})$ and is called the **inner product** of $\mathbf{a}$ and $\mathbf{b}$, such that the following axioms are satisfied.

I. For all scalars q_1 and q_2 and all vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 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 $\mathbf{a}$ and $\mathbf{b}$ in V,

$$(\mathbf{a}, \mathbf{b}) = (\mathbf{b}, \mathbf{a}) \qquad \textit{(Symmetry)}.$$

III. For every $\mathbf{a}$ in V,

$$\left.\begin{array}{r} (\mathbf{a}, \mathbf{a}) \geqq 0, \quad \text{and} \\ (\mathbf{a}, \mathbf{a}) = 0 \quad \text{if and only if} \quad \mathbf{a} = \mathbf{0} \end{array}\right\} \textit{(Positive-definiteness)}.$$

∎

Vectors whose inner product is zero are called **orthogonal.**

The *length* or **norm** of a vector in V is now defined by

$$\|\mathbf{a}\| = \sqrt{(\mathbf{a}, \mathbf{a})} \qquad (\geqq 0). \tag{2}$$

This generalizes (8) in Sec. 7.12.

A vector of norm 1 is called a **unit vector.**

[19] 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 Ref. [9], p. 73, listed in Appendix 1.

From these axioms and from (2) one can derive the basic inequality

(3) $$|(\mathbf{a}, \mathbf{b})| \leqq \|\mathbf{a}\| \, \|\mathbf{b}\| \qquad (\textit{Schwarz}^{20} \textit{ inequality}),$$

from this

(4) $$\|\mathbf{a} + \mathbf{b}\| \leqq \|\mathbf{a}\| + \|\mathbf{b}\| \qquad (\textit{Triangle inequality}),$$

and by a simple direct calculation

(5) $$\|\mathbf{a} + \mathbf{b}\|^2 + \|\mathbf{a} - \mathbf{b}\|^2 = 2(\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2) \quad (\textit{Parallelogram equality}).$$

EXAMPLE 4 **n-dimensional Euclidean space**

R^n with the inner product of Sec. 7.3,

(6) $$(\mathbf{a}, \mathbf{b}) = \mathbf{a}^{\mathsf{T}}\mathbf{b} = a_1 b_1 + \cdots + a_n b_n,$$

is called *n-dimensional Euclidean space* and denoted by E^n or again simply by R^n. Axioms I–III hold, as direct calculation shows. Equation (2) gives the "**Euclidean norm**" (8) of Sec. 7.12,

(7) $$\|\mathbf{a}\| = \sqrt{(\mathbf{a}, \mathbf{a})} = \sqrt{\mathbf{a}^{\mathsf{T}}\mathbf{a}} = \sqrt{a_1^{\,2} + \cdots + a_n^{\,2}}.$$ ∎

EXAMPLE 5 **An inner product for functions**

The set of all real-valued continuous functions $f(x)$, $g(x)$, $\cdots$ on a given interval $\alpha \leqq x \leqq \beta$ forms a real vector space under the usual addition of functions and multiplication by scalars (real numbers). On this space we can define an inner product by the integral

(8) $$(f, g) = \int_\alpha^\beta f(x) g(x) \, dx.$$

Axioms I–III can be verified by direct calculation. Equation (2) gives the norm

(9) $$\|f\| = \sqrt{(f, f)} = \sqrt{\int_\alpha^\beta f(x)^2 \, dx}.$$ ∎

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 Ref. [9] listed in Appendix 1) and cannot be discussed here. Instead we now take up a related topic where matrices play a central role.

[20] HERMANN AMANDUS SCHWARZ (1843—1921). German mathematician, successor of Weierstrass at Berlin, known by his work in complex analysis (conformal mapping), differential geometry, and the calculus of variations (minimal surfaces).

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 $\mathbf{x}$ and is denoted by $F(\mathbf{x})$ [or $F\mathbf{x}$, without parentheses].

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,

$$\mathbf{y} = \mathbf{Ax}. \tag{11}$$

Since $\mathbf{A}(\mathbf{u} + \mathbf{x}) = \mathbf{Au} + \mathbf{Ax}$ and $\mathbf{A}(c\mathbf{x}) = c\mathbf{Ax}$, 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 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**"

$$\mathbf{e}_{(1)} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_{(2)} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad \mathbf{e}_{(n)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{12}$$

where $\mathbf{e}_{(j)}$ has its jth component equal to 1 and all others 0. 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,

$$\mathbf{y} = F(\mathbf{x}) = \mathbf{Ax}.$$

Indeed, from the image $\mathbf{y}^{(1)} = F(\mathbf{e}_{(1)})$ of $\mathbf{e}_{(1)}$ we get the condition

$$\mathbf{y}^{(1)} = \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_m^{(1)} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

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.

The standard basis (12) for R^3 is usually written $\mathbf{e}_{(1)} = \mathbf{i}$, $\mathbf{e}_{(2)} = \mathbf{j}$, $\mathbf{e}_{(3)} = \mathbf{k}$; thus

$$\textbf{(13)} \qquad \mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

EXAMPLE 6 **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 7 **Linear transformations**

Our discussion preceding Example 6 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. Since

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ are mapped onto } \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ and } \begin{bmatrix} -5 \\ 4 \end{bmatrix},$$

respectively, we obtain, according to our discussion,

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

We check this, finding

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & -5 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 - 5x_2 \\ 3x_1 + 4x_2 \end{bmatrix}.$$

The reader may obtain **A** also at once by writing the given transformation in the form

$$y_1 = 2x_1 - 5x_2$$
$$y_2 = 3x_1 + 4x_2.$$

∎

If **A** in (11) is square, $n \times n$, then (11) maps R^n into R^n. If this **A** is nonsingular, so that $\mathbf{A}^{-1}$ exists (see Sec. 7.7), 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 **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.

This is the end of Chap. 7 on linear algebra, which was concerned with algebraic operations on matrices and vectors and applications to systems of linear equations and eigenvalue problems. The next chapter is devoted to the application of differential calculus to vector functions in 3-space, a field of basic importance in engineering and physics.

Problem Set 7.15

Vector spaces, bases

Is the given set (taken with the usual addition and scalar multiplication) a vector space or not? (Give reason.) If your answer is yes, determine the dimension and find a basis. (See Sec. 7.5 for similar problems.)

1. All polynomials in x, of degree not exceeding 4.
2. All symmetric real 3×3 matrices.
3. All skew-symmetric real 2×2 matrices.
4. All vectors in R^3 satisfying $v_1 - 2v_2 + v_3 = 0$.
5. All real 4×4 matrices with positive entries.
6. All real 2×3 matrices with the first row any multiple of $[1 \quad 0 \quad 2]$.
7. All functions $f(x) = (ax + b)e^{-x}$ with arbitrary constant a and b.
8. All ordered quintuples of nonnegative real numbers.

Linear independence, bases. (See Sec. 7.5 for further problems.)

9. If a subset S_0 of a set S is linearly dependent, show that S is itself linearly dependent.
10. Show that a subset of a linearly independent set is itself linearly independent.
11. Find three different bases for R^2.
12. **(Uniqueness)** Show that the representation $\mathbf{v} = c_1\mathbf{a}_{(1)} + \cdots + c_n\mathbf{a}_{(n)}$ of any given vector **v** in an n-dimensional vector space V in terms of a basis $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(n)}$ for V is unique.

Inner product, orthogonality

Find the Euclidean norm of the vectors:

13. $[1 \quad 3 \quad -1]^{\mathsf{T}}$ **14.** $[2 \quad 4 \quad 1 \quad 3]^{\mathsf{T}}$ **15.** $[3 \quad 0 \quad 0 \quad 4]^{\mathsf{T}}$

16. $[4 \quad 2 \quad 0]^{\mathsf{T}}$ **17.** $[5 \quad 1 \quad 0 \quad 6]^{\mathsf{T}}$ **18.** $[\frac{1}{2} \quad 3 \quad \frac{1}{4} \quad 2]^{\mathsf{T}}$

Using the Euclidean norm, verify:

19. The Schwarz inequality for the vectors in Probs. 13 and 16.

20. The Schwarz inequality for the vectors in Probs. 14 and 18.

21. The triangle inequality for the vectors in Probs. 15 and 17.

22. The parallelogram equality for the vectors $[2 \quad 1]^{\mathsf{T}}$ and $[1 \quad 3]^{\mathsf{T}}$. Graph the parallelogram with these sides and explain the geometric meaning of this equality.

23. Find all vectors $\mathbf{v}$ orthogonal to $\mathbf{a} = [1 \quad 2 \quad 0]^{\mathsf{T}}$. Do they form a vector space?

24. Using (6), find all unit vectors $\mathbf{v} = [v_1 \quad v_2]$ orthogonal to $[3 \quad -4]$.

Linear transformations

Find the inverse transformation:

25. $y_1 = 3x_1 - x_2$, $y_2 = -5x_1 + 2x_2$

26. $y_1 = 4x_1 + 3x_2$, $y_2 = 3x_1 + 2x_2$

27. $y_1 = 2x_1 + 4x_2 + x_3$, $y_2 = x_1 + 2x_2 + x_3$, $y_3 = 3x_1 + 4x_2 + 2x_3$

28. $y_1 = x_1 + 3x_3$, $y_2 = 2x_2 + x_3$, $y_3 = 3x_1 + x_2 + 10x_3$

29. $y_1 = 0.2x_1 - 0.1x_2$, $y_2 = -0.2x_2 + 0.1x_3$, $y_3 = 0.1x_1 + 0.1x_3$

30. $y_1 = 3x_1 - x_2 + x_3$, $y_2 = -15x_1 + 6x_2 - 5x_3$, $y_3 = 5x_1 - 2x_2 + 2x_3$

Review Questions and Problems for Chapter 7

1. Let $\mathbf{A}$ be a 20×20 matrix and $\mathbf{B}$ a 20×10 matrix. Indicate whether or not the following expressions are defined: $\mathbf{A} + \mathbf{B}$, $\mathbf{AB}$, $\mathbf{A}^{\mathsf{T}}\mathbf{B}$, $\mathbf{AB}^{\mathsf{T}}$, $\mathbf{B}^{\mathsf{T}}\mathbf{A}$, $\mathbf{A}^2$, $\mathbf{AA}^{\mathsf{T}}$, $\mathbf{B}^2$, $\mathbf{BB}^{\mathsf{T}}$, $\mathbf{B}^{\mathsf{T}}\mathbf{BA}$. (Give reasons.)
2. What properties of matrix multiplication are "unusual" (i.e., differ from those for the multiplication of numbers)?
3. How can the rank of a matrix be given in terms of row vectors? Column vectors? Determinants?
4. What do you know about the existence and number of solutions of a nonhomogeneous system of linear equations? A homogeneous system?
5. What is the Gauss elimination good for? What is its basic idea? Why is this elimination generally better than Cramer's rule? What does pivoting mean?
6. What is the inverse of a matrix? When does it exist? How would you practically determine it?

7. Give simple examples of linear systems of equation without solutions. With a unique solution. With more than one solution.
8. Write down the formulas for $(\mathbf{AB})^T$ and $(\mathbf{AB})^{-1}$ from memory. Work an example.
9. Can the row vectors of an 8×6 matrix be linearly independent?
10. What are symmetric, skew-symmetric, and orthogonal matrices? Hermitian, skew-Hermitian, and unitary matrices?
11. Show that the symmetric 4×4 matrices form a vector space. What is its dimension? Find a basis.
12. What is the nullity of a matrix **A**? The row space of **A**? The column space of **A**?
13. State the definitions of an eigenvalue and an eigenvector of a matrix **A**. What is the spectrum of **A**? The characteristic equation?
14. What do you know about the eigenvalues of the matrices in Prob. 10?
15. Do there exist square matrices without eigenvalues? Can a real matrix have complex eigenvalues? Does a real 3×3 matrix always have a real eigenvalue?

Let $\mathbf{a} = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 6 \\ -4 \\ 3 \end{bmatrix}$, $\mathbf{C} = \begin{bmatrix} 0 & 0 & 5 \\ 0 & 4 & -2 \\ 1 & 3 & 2 \end{bmatrix}$, $\mathbf{D} = \begin{bmatrix} 6 & -2 & 2 \\ 8 & 3 & 2 \\ 1 & 5 & -9 \end{bmatrix}$. Find

16. $8\mathbf{C} - 5\mathbf{D}$
17. $\mathbf{a} + 4\mathbf{b}$
18. $\mathbf{Ca}$, $\mathbf{a}^T\mathbf{C}$
19. $\mathbf{Cb}$
20. $\mathbf{CD}$, $\mathbf{D}^T\mathbf{C}^T$
21. $\mathbf{C}^{-1}$
22. rank $\mathbf{D}$
23. rank $(\mathbf{b}^T\mathbf{D})$
24. $\mathbf{C} + \mathbf{C}^T$
25. $\mathbf{CC}^T$
26. $\mathbf{DCab}^T$
27. $\mathbf{a} + \mathbf{Db}$
28. $\mathbf{C}^4$
29. $(\mathbf{C} - \mathbf{D})\,\mathbf{a}$
30. $\mathbf{a}^T\mathbf{b}$, $\mathbf{ab}^T$
31. $\mathbf{aa}^T$, $\mathbf{a}^T\mathbf{a}$

Solve the following systems of equations or indicate that no solutions exist.

32. $\begin{aligned} 5x + 3y - 3z &= -1 \\ 3x + 2y - 2z &= -1 \\ 2x - y + 2z &= 8 \end{aligned}$

33. $\begin{aligned} 2y - z &= -1 \\ x + 3z &= 11 \\ 2x - 4y + 2z &= 6 \end{aligned}$

34. $\begin{aligned} 7x - 4y - 2z &= -6 \\ 16x + 2y + z &= 3 \end{aligned}$

35. $\begin{aligned} 6x + 4y &= 4 \\ 8x - 6z &= 7 \\ -8y - 2z &= -1 \end{aligned}$

36. $\begin{aligned} 4y + 7z &= -13 \\ 5x - 3y + 4z &= -23 \\ -x + 2y - 8z &= 29 \end{aligned}$

37. $\begin{aligned} 4x + 2y - 6z &= -6 \\ 5x + 3y - 8z &= -9 \end{aligned}$

38. $\begin{aligned} 3x - 2y &= -2 \\ 5x + 4y &= 26 \\ x - 3y &= 8 \end{aligned}$

39. $\begin{aligned} 3x + 4y + 6z &= 1 \\ -2x + 8y - 4z &= 2 \\ 4x - 8y + 8z &= -2 \end{aligned}$

40. $\begin{aligned} 4x - y + z &= 0 \\ x + 2y - z &= 0 \\ 3x + y + 5z &= 0 \end{aligned}$

Determine the rank of the following matrices.

41. $\begin{bmatrix} 9 & 1 \\ 0 & 4 \\ 2 & 6 \end{bmatrix}$

42. $\begin{bmatrix} 3 & 1 & 4 \\ 0 & 5 & 8 \\ -3 & 4 & 4 \end{bmatrix}$

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

Find the inverse or indicate that it does not exist. Check your result.

44. $\begin{bmatrix} 0.3 & 0.1 \\ -0.4 & 0.2 \end{bmatrix}$

45. $\begin{bmatrix} 0 & 5 \\ 1 & 8 \end{bmatrix}$

46. $\begin{bmatrix} 0.6 & 0.3 \\ -1.6 & -0.8 \end{bmatrix}$

Find the inverse or indicate that it does not exist. Check your result.

47. $\begin{bmatrix} 2 & 4 & 1 \\ 1 & 2 & 1 \\ 3 & 4 & 2 \end{bmatrix}$ **48.** $\begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}$ **49.** $\begin{bmatrix} 0.2 & 0.1 & 0.2 \\ 0 & 0.5 & 0.4 \\ 0 & 0 & 0.1 \end{bmatrix}$

Find the eigenvalues and eigenvectors:

50. $\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$ **51.** $\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$ **52.** $\begin{bmatrix} 1.4 & 0.5 \\ -1.0 & -0.1 \end{bmatrix}$

53. $\begin{bmatrix} 15 & 0 & -15 \\ -3 & 6 & 9 \\ 5 & 0 & -5 \end{bmatrix}$ **54.** $\begin{bmatrix} 3 & 1 & 4 \\ 0 & 2 & 6 \\ 0 & 0 & 5 \end{bmatrix}$ **55.** $\begin{bmatrix} 1 & 2 & 4 \\ -2 & -4 & 2 \\ 2 & 4 & 3 \end{bmatrix}$

56. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **57.** $\begin{bmatrix} 0 & 2 & 0 \\ 3 & -2 & 3 \\ 0 & 3 & 0 \end{bmatrix}$ **58.** $\begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 2 & 0 & 0 \end{bmatrix}$

Find a basis of eigenvectors and diagonalize:

59. $\begin{bmatrix} 0 & 4 \\ 9 & 0 \end{bmatrix}$ **60.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 0 & 0 \end{bmatrix}$ **61.** $\begin{bmatrix} -2 & 0 & 0 \\ -2 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

What kind of conic section is represented by the given quadratic form? Transform it to principal axes. Express $[x_1 \quad x_2]^T$ in terms of the new coordinates.

62. $10x_1^2 - 9x_1x_2 + \frac{25}{4}x_2^2 = 13$ **63.** $7x_1^2 + 48x_1x_2 - 7x_2^2 = 25$

64. $4x_1x_2 + 3x_2^2 = 1$ **65.** $801x_1^2 - 600x_1x_2 + 1396x_2^2 = 169$

Find $\bar{\mathbf{x}}^T\mathbf{A}\mathbf{x}$ where

66. $\mathbf{A} = \begin{bmatrix} 5i & 2+i \\ -2+i & i \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 2i \\ 3 \end{bmatrix}$

67. $\mathbf{A} = \begin{bmatrix} 2i & 4 \\ -4 & 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

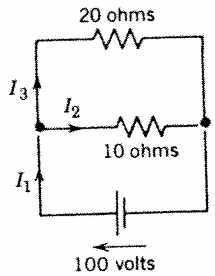
68. $\mathbf{A} = \begin{bmatrix} 1 & i & 0 \\ -i & 0 & 3 \\ 0 & 3 & 2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ **69.** $\mathbf{A} = \begin{bmatrix} 1 & -i & 2i \\ i & 1 & 0 \\ -2i & 0 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ i \end{bmatrix}$

70. **(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

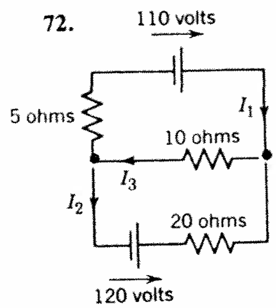
$$\mathbf{S}_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{S}_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \mathbf{S}_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Networks. Find the currents in the following networks.

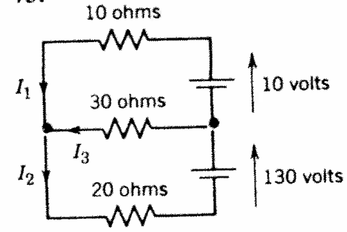
71.



72.



73.



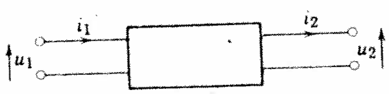


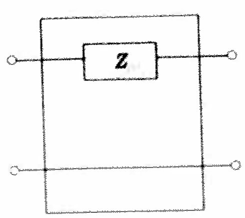
Fig. 138. Four-terminal network

Four-terminal networks. Assume that the input current i_1 and voltage u_1 of the four-terminal network in Fig. 138 are related to the output current i_2 and voltage u_2 according to

$$\mathbf{v}_1 = \mathbf{T}\mathbf{v}_2, \quad \text{where} \quad \mathbf{v}_1 = \begin{bmatrix} u_1 \\ i_1 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} u_2 \\ i_2 \end{bmatrix}$$

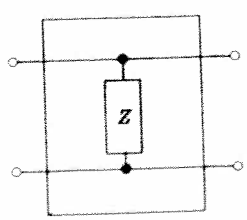
and where **T** is called the transmission matrix of the network. Verify the form of **T**:

74.



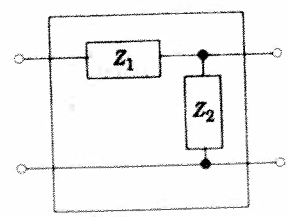
$$\mathbf{T} = \begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$$

75.



$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ 1/Z & 1 \end{bmatrix}$$

76.



$$\mathbf{T} = \begin{bmatrix} 1 + Z_1/Z_2 & Z_1 \\ 1/Z_2 & 1 \end{bmatrix}$$

77. Show that for the networks in cascade in Fig. 139 we have $\mathbf{v}_1 = \mathbf{T}\mathbf{v}_2$ with $\mathbf{T} = \mathbf{T}_1\mathbf{T}_2$ and $\mathbf{v}_1$ and $\mathbf{v}_2$ as before.
78. Use Prob. 77 to get the matrix in Prob. 76 from those in Probs. 74 and 75.

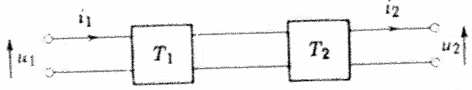


Fig. 139. Four-terminal networks in cascade

Summary of Chapter 7 Linear Algebra: Matrices, Vectors, Determinants

An $m \times n$ **matrix** $\mathbf{A} = [a_{jk}]$ is a rectangular array of numbers ("entries" or "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** (see 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 (see Sec. 7.2).

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

$$c_{jk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk} \qquad \text{(row } j \text{ of } \mathbf{A} \text{ times column } k \text{ of } \mathbf{B}). \tag{1}$$

This multiplication is motivated by the composition of **linear transformations** (Secs. 7.3, 7.15). 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. 7.3, 7.7):

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

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

$$[1 \quad 2] \begin{bmatrix} 3 \\ 4 \end{bmatrix} = [11], \qquad \begin{bmatrix} 3 \\ 4 \end{bmatrix} [1 \quad 2] = \begin{bmatrix} 3 & 6 \\ 4 & 8 \end{bmatrix}.$$

A main application of matrices concerns **linear systems of equations**

$$\mathbf{Ax} = \mathbf{b} \qquad \text{(Sec. 7.4)} \tag{2}$$

(m equations in n unknowns $x_1, \cdots, x_n$; $\mathbf{b}$ given). The most important method of solution is the **Gauss elimination** (Sec. 7.4), which reduces the system to "triangular" form by *elementary row operations,* which leave the set of solutions unchanged. (Numerical aspects and variants, such as *Doolittle's method* are discussed in Secs. 19.1 and 19.2.)

Cramer's rule (Sec. 7.9) represents the unknowns in a system (2) of n equations in n unknowns as quotients of determinants; for numerical work it is impractical. **Determinants** (Sec. 7.8) 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*. See Secs. 7.7, 7.9.

The **rank** r of a matrix $\mathbf{A}$ is the maximum number of linearly independent rows or columns of $\mathbf{A}$, equivalently, the number of rows of the largest square submatrix of $\mathbf{A}$ with nonzero determinant (Secs. 7.5, 7.9). The system (2) has solutions if and only if rank $\mathbf{A}$ = rank $[\mathbf{A} \quad \mathbf{b}]$, $[\mathbf{A} \quad \mathbf{b}]$ the *augmented matrix* (Fundamental Theorem, Sec. 7.6). The *homogeneous system*

$$\text{(3)} \qquad \mathbf{A}\mathbf{x} = \mathbf{0}$$

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 $\det \mathbf{A} = 0$ (Secs. 7.6, 7.9).

Hence the system of n equations in n unknowns

$$\text{(4)} \qquad \mathbf{A}\mathbf{x} = \lambda\mathbf{x} \qquad \text{or} \qquad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

has solutions $\mathbf{x} \neq \mathbf{0}$ if and only if λ is a root of the *characteristic equation*

$$\text{(5)} \qquad \det(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0 \qquad \text{(Sec. 7.10)}.$$

Such a (real or complex) number λ is called an **eigenvalue** of $\mathbf{A}$ and that solution $\mathbf{x} \neq \mathbf{0}$ an **eigenvector** of $\mathbf{A}$ corresponding to this λ. Equation (4) is called an (algebraic) **eigenvalue problem** (Sec. 7.10). Eigenvalue problems are of great importance in physics and engineering (Sec. 7.11), and they also have applications in economics and statistics. They are basic in solving systems of differential equations (Chap. 4).

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. 7.1). For a product, $(\mathbf{A}\mathbf{B})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$ (Sec. 7.3). The *complex conjugate* of $\mathbf{A}$ is $\overline{\mathbf{A}} = [\overline{a}_{jk}]$. Six classes of square matrices of practical importance arise from this: a real matrix $\mathbf{A}$ is called *real* **symmetric** if $\mathbf{A}^\mathsf{T} = \mathbf{A}$, *real* **skew-symmetric** if $\mathbf{A}^\mathsf{T} = -\mathbf{A}$, **orthogonal** if $\mathbf{A}^\mathsf{T} = \mathbf{A}^{-1}$ (Secs. 7.2, 7.12); a complex matrix is called **Hermitian** if $\overline{\mathbf{A}}^\mathsf{T} = \mathbf{A}$, **skew-Hermitian** if $\overline{\mathbf{A}}^\mathsf{T} = -\mathbf{A}$, and **unitary** if $\overline{\mathbf{A}}^\mathsf{T} = \mathbf{A}^{-1}$ (Sec. 7.13). The eigenvalues of Hermitian (and real-symmetric) matrices are real; those of skew-Hermitian (and real skew-symmetric) are pure imaginary or 0; those of unitary (and orthogonal) matrices have absolute value 1 (Sec. 7.13).

The diagonalization of matrices and the transformation of quadratic forms to principal axes are discussed in Sec. 7.14.

General vector spaces and inner product spaces are discussed in Sec. 7.15. For R^n and C^n, see also Secs. 7.5 and 7.13.