CHAPTER

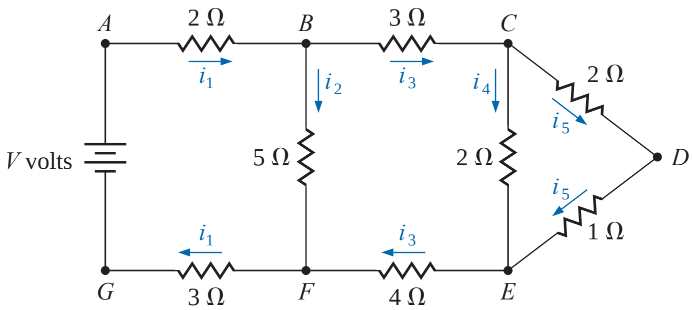
6

Direct Methods for Solving Linear Systems

Introduction

Kirchhoff's laws of electrical circuits state that both the net flow of current through each junction and the net voltage drop around each closed loop of a circuit are zero. Suppose that a potential of V volts is applied between the points A and G in the circuit and that i_1, i_2, i_3, i_4, and i_5 represent current flow as shown in the diagram. Using G as a reference point, Kirchhoff's laws imply that the currents satisfy the following system of linear equations:

$$\begin{aligned} 5i_1 + 5i_2 &= V, \\ i_3 - i_4 - i_5 &= 0, \\ 2i_4 - 3i_5 &= 0, \\ i_1 - i_2 - i_3 &= 0, \\ 5i_2 - 7i_3 - 2i_4 &= 0. \end{aligned}$$



The solution of systems of this type will be considered in this chapter. This application is discussed in Exercise 29 of Section 6.6.

Linear systems of equations are associated with many problems in engineering and science, as well as with applications of mathematics to the social sciences and the quantitative study of business and economic problems.

In this chapter we consider *direct methods* for solving a linear system of n equations in n variables. Such a system has the form

$$\begin{aligned} E_1 &: \quad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ E_2 &: \quad a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ &\qquad\qquad \vdots \\ E_n &: \quad a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n. \end{aligned} \tag{6.1}$$

In this system we are given the constants a_{ij}, for each $i, j = 1, 2, \ldots, n$, and b_i, for each $i = 1, 2, \ldots, n$, and we need to determine the unknowns $x_1, \ldots, x_n$.

Direct techniques are methods that theoretically give the exact solution to the system in a finite number of steps. In practice, of course, the solution obtained will be contaminated by the round-off error that is involved with the arithmetic being used. Analyzing the effect of this round-off error and determining ways to keep it under control will be a major component of this chapter.

A course in linear algebra is not assumed to be prerequisite for this chapter, so we will include a number of the basic notions of the subject. These results will also be used in Chapter 7, where we consider methods of approximating the solution to linear systems using iterative methods.

6.1 Linear Systems of Equations

We use three operations to simplify the linear system given in (6.1):

1. Equation E_i can be multiplied by any nonzero constant λ with the resulting equation used in place of E_i. This operation is denoted $(\lambda E_i) \to (E_i)$.

2. Equation E_j can be multiplied by any constant λ and added to equation E_i with the resulting equation used in place of E_i. This operation is denoted $(E_i + \lambda E_j) \to (E_i)$.

3. Equations E_i and E_j can be transposed in order. This operation is denoted $(E_i) \leftrightarrow (E_j)$.

By a sequence of these operations, a linear system will be systematically transformed into to a new linear system that is more easily solved and has the same solutions. The sequence of operations is illustrated in the following.

Illustration The four equations

$$\begin{aligned}
E_1 &: \quad x_1 + x_2 \phantom{{}- x_3} + 3x_4 = 4,\\
E_2 &: \quad 2x_1 + x_2 - x_3 + x_4 = 1,\\
E_3 &: \quad 3x_1 - x_2 - x_3 + 2x_4 = -3,\\
E_4 &: \quad -x_1 + 2x_2 + 3x_3 - x_4 = 4,
\end{aligned} \tag{6.2}$$

will be solved for x_1, x_2, x_3, and x_4. We first use equation E_1 to eliminate the unknown x_1 from equations E_2, E_3, and E_4 by performing $(E_2 - 2E_1) \to (E_2)$, $(E_3 - 3E_1) \to (E_3)$, and $(E_4 + E_1) \to (E_4)$. For example, in the second equation

$$(E_2 - 2E_1) \to (E_2)$$

produces

$$(2x_1 + x_2 - x_3 + x_4) - 2(x_1 + x_2 + 3x_4) = 1 - 2(4).$$

which simplifies to the result shown as E_2 in

$$\begin{aligned}
E_1 &: \quad x_1 + x_2 \phantom{{}- x_3} + 3x_4 = 4,\\
E_2 &: \quad -x_2 - x_3 - 5x_4 = -7,\\
E_3 &: \quad -4x_2 - x_3 - 7x_4 = -15,\\
E_4 &: \quad 3x_2 + 3x_3 + 2x_4 = 8.
\end{aligned}$$

For simplicity, the new equations are again labeled E_1, E_2, E_3, and E_4.

In the new system, E_2 is used to eliminate the unknown x_2 from E_3 and E_4 by performing $(E_3 - 4E_2) \rightarrow (E_3)$ and $(E_4 + 3E_2) \rightarrow (E_4)$. This results in

$$
\begin{aligned}
E_1 &: \quad x_1 + x_2 \qquad\quad + \ \ 3x_4 = \ \ \ 4,\\
E_2 &: \qquad\ - x_2 - \ x_3 - \ \ 5x_4 = -7,\\
E_3 &: \qquad\qquad\quad\ \ 3x_3 + 13x_4 = \ \ 13,\\
E_4 &: \qquad\qquad\qquad\quad - 13x_4 = -13.
\end{aligned}
\tag{6.3}
$$

The system of equations (6.3) is now in **triangular** (or **reduced) form** and can be solved for the unknowns by a **backward-substitution process**. Since E_4 implies $x_4 = 1$, we can solve E_3 for x_3 to give

$$x_3 = \frac{1}{3}(13 - 13x_4) = \frac{1}{3}(13 - 13) = 0.$$

Continuing, E_2 gives

$$x_2 = -(-7 + 5x_4 + x_3) = -(-7 + 5 + 0) = 2,$$

and E_1 gives

$$x_1 = 4 - 3x_4 - x_2 = 4 - 3 - 2 = -1.$$

The solution to system (6.3), and consequently to system (6.2), is therefore, $x_1 = -1$, $x_2 = 2$, $x_3 = 0$, and $x_4 = 1$. □

Matrices and Vectors

When performing the calculations in the Illustration, we would not need to write out the full equations at each step or to carry the variables x_1, x_2, x_3, and x_4 through the calculations, if they always remained in the same column. The only variation from system to system occurs in the coefficients of the unknowns and in the values on the right side of the equations. For this reason, a linear system is often replaced by a *matrix*, which contains all the information about the system that is necessary to determine its solution, but in a compact form, and one that is easily represented in a computer.

Definition 6.1 An $\boldsymbol{n \times m}$ (**n** by **m**) **matrix** is a rectangular array of elements with n rows and m columns in which not only is the value of an element important, but also its position in the array. ■

The notation for an $n \times m$ matrix will be a capital letter such as A for the matrix and lowercase letters with double subscripts, such as a_{ij}, to refer to the entry at the intersection of the ith row and jth column; that is,

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}.$$

Example 1 Determine the size and respective entries of the matrix

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

Solution The matrix has two rows and three columns so it is of size 2×3. It entries are described by $a_{11} = 2$, $a_{12} = -1$, $a_{13} = 7$, $a_{21} = 3$, $a_{22} = 1$, and $a_{23} = 0$. ■

The $1 \times n$ matrix

$$A = [a_{11} \;\; a_{12} \;\; \cdots \;\; a_{1n}]$$

is called an ***n*-dimensional row vector**, and an $n \times 1$ matrix

$$A = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}$$

is called an ***n*-dimensional column vector**. Usually the unnecessary subscripts are omitted for vectors, and a boldface lowercase letter is used for notation. Thus

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

denotes a column vector, and

$$\mathbf{y} = [y_1 \;\; y_2 \;\; \ldots \;\; y_n]$$

a row vector. In addition, row vectors often have commas inserted between the entries to make the separation clearer. So you might see $\mathbf{y}$ written as $\mathbf{y} = [y_1, \; y_2, \; \ldots, \; y_n]$.

An $n \times (n + 1)$ matrix can be used to represent the linear system

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

by first constructing

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$[A, \mathbf{b}] = \left[\begin{array}{cccc:c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right],$$

Augmented refers to the fact that the right-hand side of the system has been included in the matrix.

where the vertical dotted line is used to separate the coefficients of the unknowns from the values on the right-hand side of the equations. The array $[A, \mathbf{b}]$ is called an **augmented matrix**.

Repeating the operations involved in Example 1 with the matrix notation results in first considering the augmented matrix:

$$\left[\begin{array}{rrrr:r} 1 & 1 & 0 & 3 & 4 \\ 2 & 1 & -1 & 1 & 1 \\ 3 & -1 & -1 & 2 & -3 \\ -1 & 2 & 3 & -1 & 4 \end{array}\right].$$

Performing the operations as described in that example produces the augmented matrices

$$\left[\begin{array}{rrrr:r} 1 & 1 & 0 & 3 & 4 \\ 0 & -1 & -1 & -5 & -7 \\ 0 & -4 & -1 & -7 & -15 \\ 0 & 3 & 3 & 2 & 8 \end{array}\right] \quad \text{and} \quad \left[\begin{array}{rrrr:r} 1 & 1 & 0 & 3 & 4 \\ 0 & -1 & -1 & -5 & -7 \\ 0 & 0 & 3 & 13 & 13 \\ 0 & 0 & 0 & -13 & -13 \end{array}\right].$$

A technique similar to Gaussian elimination first appeared during the Han dynasty in China in the text *Nine Chapters on the Mathematical Art*, which was written about 200 B.C.E. Joseph Louis Lagrange (1736–1813) described a technique similar to this procedure in 1778 for the case when the value of each equation is 0. Gauss gave a more general description in *Theoria Motus corporum coelestium sectionibus solem ambientium*, which described the least squares technique he used in 1801 to determine the orbit of the minor planet Ceres.

The final matrix can now be transformed into its corresponding linear system, and solutions for x_1, x_2, x_3, and x_4, can be obtained. The procedure is called **Gaussian elimination with backward substitution**.

The general Gaussian elimination procedure applied to the linear system

$$\begin{aligned} E_1: &\quad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ E_2: &\quad a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ &\quad \vdots \qquad\qquad\qquad\qquad\quad \vdots \\ E_n: &\quad a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n, \end{aligned} \tag{6.4}$$

is handled in a similar manner. First form the augmented matrix $\tilde{A}$:

$$\tilde{A} = [A, \mathbf{b}] = \left[\begin{array}{cccc:c} a_{11} & a_{12} & \cdots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{2,n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & a_{n,n+1} \end{array}\right], \tag{6.5}$$

where A denotes the matrix formed by the coefficients. The entries in the $(n+1)$st column are the values of $\mathbf{b}$; that is, $a_{i,n+1} = b_i$ for each $i = 1, 2, \cdots, n$.

Provided $a_{11} \neq 0$, we perform the operations corresponding to

$$(E_j - (a_{j1}/a_{11})E_1) \rightarrow (E_j) \quad \text{for each } j = 2, 3, \ldots, n$$

to eliminate the coefficient of x_1 in each of these rows. Although the entries in rows $2, 3, \ldots, n$ are expected to change, for ease of notation we again denote the entry in the ith row and the jth column by a_{ij}. With this in mind, we follow a sequential procedure for $i = 2, 3, \ldots, n-1$ and perform the operation

$$(E_j - (a_{ji}/a_{ii})E_i) \rightarrow (E_j) \quad \text{for each } j = i+1, i+2, \ldots, n,$$

provided $a_{ii} \neq 0$. This eliminates (changes the coefficient to zero) x_i in each row below the ith for all values of $i = 1, 2, \ldots, n-1$. The resulting matrix has the form:

$$\tilde{\tilde{A}} = \left[\begin{array}{cccc:c} a_{11} & a_{12} & \cdots & a_{1n} & a_{1,n+1} \\ 0 & a_{22} & \cdots & a_{2n} & a_{2,n+1} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{nn} & a_{n,n+1} \end{array}\right],$$

where, except in the first row, the values of a_{ij} are not expected to agree with those in the original matrix $\tilde{A}$. The matrix $\tilde{\tilde{A}}$ represents a linear system with the same solution set as the original system.

The new linear system is triangular,

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= a_{1,n+1},\\
a_{22}x_2 + \cdots + a_{2n}x_n &= a_{2,n+1},\\
\ddots \qquad \vdots \qquad &\quad \vdots\\
a_{nn}x_n &= a_{n,n+1},
\end{aligned}$$

so *backward substitution* can be performed. Solving the nth equation for x_n gives

$$x_n = \frac{a_{n,n+1}}{a_{nn}}.$$

Solving the $(n-1)$st equation for x_{n-1} and using the known value for x_n yields

$$x_{n-1} = \frac{a_{n-1,n+1} - a_{n-1,n}x_n}{a_{n-1,n-1}}.$$

Continuing this process, we obtain

$$x_i = \frac{a_{i,n+1} - a_{i,n}x_n - a_{i,n-1}x_{n-1} - \cdots - a_{i,i+1}x_{i+1}}{a_{ii}} = \frac{a_{i,n+1} - \sum_{j=i+1}^{n} a_{ij}x_j}{a_{ii}},$$

for each $i = n-1, n-2, \cdots, 2, 1$.

Gaussian elimination procedure is described more precisely, although more intricately, by forming a sequence of augmented matrices $\tilde{A}^{(1)}, \tilde{A}^{(2)}, \ldots, \tilde{A}^{(n)}$, where $\tilde{A}^{(1)}$ is the matrix $\tilde{A}$ given in (6.5) and $\tilde{A}^{(k)}$, for each $k = 2, 3, \ldots, n$, has entries $a_{ij}^{(k)}$, where:

$$a_{ij}^{(k)} = \begin{cases}
a_{ij}^{(k-1)}, & \text{when } i = 1, 2, \ldots, k-1 \text{ and } j = 1, 2, \ldots, n+1,\\
0, & \text{when } i = k, k+1, \ldots, n \text{ and } j = 1, 2, \cdots, k-1,\\
a_{ij}^{(k-1)} - \dfrac{a_{i,k-1}^{(k-1)}}{a_{k-1,k-1}^{(k-1)}} a_{k-1,j}^{(k-1)}, & \text{when } i = k, k+1, \ldots, n \text{ and } j = k, k+1, \ldots, n+1.
\end{cases}$$

Thus

$$\tilde{A}^{(k)} = \left[\begin{array}{cccccccc:c}
a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1,k-1}^{(1)} & a_{1k}^{(1)} & \cdots & a_{1n}^{(1)} & a_{1,n+1}^{(1)}\\
0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2,k-1}^{(2)} & a_{2k}^{(2)} & \cdots & a_{2n}^{(2)} & a_{2,n+1}^{(2)}\\
\vdots & \ddots & \ddots & & \vdots & \vdots & & \vdots & \vdots\\
\vdots & & & \ddots & a_{k-1,k-1}^{(k-1)} & a_{k-1,k}^{(k-1)} & \cdots & a_{k-1,n}^{(k-1)} & a_{k-1,n+1}^{(k-1)}\\
\vdots & & & & 0 & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} & a_{k,n+1}^{(k)}\\
\vdots & & & & \vdots & \vdots & & \vdots & \vdots\\
0 & \cdots & \cdots & \cdots & 0 & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} & a_{n,n+1}^{(k)}
\end{array}\right] \tag{6.6}$$

represents the equivalent linear system for which the variable x_{k-1} has just been eliminated from equations $E_k, E_{k+1}, \ldots, E_n$.

The procedure will fail if one of the elements $a_{11}^{(1)}, a_{22}^{(2)}, a_{33}^{(3)}, \ldots, a_{n-1,n-1}^{(n-1)}, a_{nn}^{(n)}$ is zero because the step

$$\left(E_i - \frac{a_{i,k}^{(k)}}{a_{kk}^{(k)}}(E_k)\right) \to E_i$$

either cannot be performed (this occurs if one of $a_{11}^{(1)}, \ldots, a_{n-1,n-1}^{(n-1)}$ is zero), or the backward substitution cannot be accomplished (in the case $a_{nn}^{(n)} = 0$). The system may still have a solution, but the technique for finding the solution must be altered. An illustration is given in the following example.

Example 2 Represent the linear system

$$\begin{aligned}
E_1: &\quad x_1 - x_2 + 2x_3 - x_4 = -8,\\
E_2: &\quad 2x_1 - 2x_2 + 3x_3 - 3x_4 = -20,\\
E_3: &\quad x_1 + x_2 + x_3 = -2,\\
E_4: &\quad x_1 - x_2 + 4x_3 + 3x_4 = 4,
\end{aligned}$$

as an augmented matrix and use Gaussian Elimination to find its solution.

Solution The augmented matrix is

$$\tilde{A} = \tilde{A}^{(1)} = \left[\begin{array}{rrrr:r} 1 & -1 & 2 & -1 & -8 \\ 2 & -2 & 3 & -3 & -20 \\ 1 & 1 & 1 & 0 & -2 \\ 1 & -1 & 4 & 3 & 4 \end{array}\right].$$

Performing the operations

$$(E_2 - 2E_1) \to (E_2),\ (E_3 - E_1) \to (E_3), \quad \text{and} \quad (E_4 - E_1) \to (E_4),$$

gives

$$\tilde{A}^{(2)} = \left[\begin{array}{rrrr:r} 1 & -1 & 2 & -1 & -8 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 2 & 4 & 12 \end{array}\right].$$

The pivot element for a specific column is the entry that is used to place zeros in the other entries in that column.

The diagonal entry $a_{22}^{(2)}$, called the **pivot element**, is 0, so the procedure cannot continue in its present form. But operations $(E_i) \leftrightarrow (E_j)$ are permitted, so a search is made of the elements $a_{32}^{(2)}$ and $a_{42}^{(2)}$ for the first nonzero element. Since $a_{32}^{(2)} \neq 0$, the operation $(E_2) \leftrightarrow (E_3)$ is performed to obtain a new matrix,

$$\tilde{A}^{(2)'} = \left[\begin{array}{rrrr:r} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 2 & 4 & 12 \end{array}\right].$$

Since x_2 is already eliminated from E_3 and E_4, $\tilde{A}^{(3)}$ will be $\tilde{A}^{(2)'}$, and the computations continue with the operation $(E_4 + 2E_3) \to (E_4)$, giving

$$\tilde{A}^{(4)} = \left[\begin{array}{rrrr:r} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 0 & 2 & 4 \end{array}\right].$$

Finally, the matrix is converted back into a linear system that has a solution equivalent to the solution of the original system and the backward substitution is applied:

$$
\begin{aligned}
x_4 &= \frac{4}{2} = 2,\\
x_3 &= \frac{[-4-(-1)x_4]}{-1} = 2,\\
x_2 &= \frac{[6-x_4-(-1)x_3]}{2} = 3,\\
x_1 &= \frac{[-8-(-1)x_4-2x_3-(-1)x_2]}{1} = -7.
\end{aligned}
$$

■

Example 2 illustrates what is done if $a_{kk}^{(k)} = 0$ for some $k = 1, 2, \ldots, n-1$. The kth column of $\tilde{A}^{(k-1)}$ from the kth row to the nth row is searched for the first nonzero entry. If $a_{pk}^{(k)} \neq 0$ for some p,with $k+1 \leq p \leq n$, then the operation $(E_k) \leftrightarrow (E_p)$ is performed to obtain $\tilde{A}^{(k-1)'}$. The procedure can then be continued to form $\tilde{A}^{(k)}$, and so on. If $a_{pk}^{(k)} = 0$ for each p, it can be shown (see Theorem 6.17 on page 398) that the linear system does not have a unique solution and the procedure stops. Finally, if $a_{nn}^{(n)} = 0$, the linear system does not have a unique solution, and again the procedure stops.

Algorithm 6.1 summarizes Gaussian elimination with backward substitution. The algorithm incorporates pivoting when one of the pivots $a_{kk}^{(k)}$ is 0 by interchanging the kth row with the pth row, where p is the smallest integer greater than k for which $a_{pk}^{(k)} \neq 0$.

ALGORITHM 6.1

Gaussian Elimination with Backward Substitution

To solve the $n \times n$ linear system

$$
\begin{array}{lllllll}
E_1: & a_{11}x_1 & + a_{12}x_2 & + \cdots + & a_{1n}x_n & = & a_{1,n+1}\\
E_2: & a_{21}x_1 & + a_{22}x_2 & + \cdots + & a_{2n}x_n & = & a_{2,n+1}\\
\vdots & \vdots & \vdots & & \vdots & & \vdots\\
E_n: & a_{n1}x_1 & + a_{n2}x_2 & + \cdots + & a_{nn}x_n & = & a_{n,n+1}
\end{array}
$$

INPUT number of unknowns and equations n; augmented matrix $A = [a_{ij}]$, where $1 \leq i \leq n$ and $1 \leq j \leq n+1$.

OUTPUT solution $x_1, x_2, \ldots, x_n$ or message that the linear system has no unique solution.

Step 1 For $i = 1, \ldots, n-1$ do Steps 2–4. (*Elimination process.*)

  Step 2 Let p be the smallest integer with $i \leq p \leq n$ and $a_{pi} \neq 0$.
  If no integer p can be found
   then OUTPUT ('no unique solution exists');
    STOP.

  Step 3 If $p \neq i$ then perform $(E_p) \leftrightarrow (E_i)$.

  Step 4 For $j = i+1, \ldots, n$ do Steps 5 and 6.

   Step 5 Set $m_{ji} = a_{ji}/a_{ii}$.

   Step 6 Perform $(E_j - m_{ji}E_i) \rightarrow (E_j)$;

Step 7 If $a_{nn} = 0$ then OUTPUT ('no unique solution exists');
STOP.

Step 8 Set $x_n = a_{n,n+1}/a_{nn}$. (*Start backward substitution.*)

Step 9 For $i = n-1, \ldots, 1$ set $x_i = \left[a_{i,n+1} - \sum_{j=i+1}^{n} a_{ij}x_j\right] \Big/ a_{ii}$.

Step 10 OUTPUT $(x_1, \ldots, x_n)$; (*Procedure completed successfully.*)
STOP. ■

To define matrices and perform Gaussian elimination using Maple, first access the *LinearAlgebra* library using the command

$with(LinearAlgebra)$

To define the matrix $\tilde{A}^{(1)}$ of Example 2, which we will call AA, use the command

$AA := Matrix([[1, -1, 2, -1, -8], [2, -2, 3, -3, -20], [1, 1, 1, 0, -2], [1, -1, 4, 3, 4]])$

This lists the entries, by row, of the augmented matrix $AA \equiv \tilde{A}^{(1)}$.

The function $RowOperation(AA, [i, j], m)$ performs the operation $(E_j + mE_i) \to (E_j)$, and the same command without the last parameter, that is, $RowOperation(AA, [i, j])$ performs the operation $(E_i) \leftrightarrow (E_j)$. So the sequence of operations

$AA1 := RowOperation(AA, [2, 1], -2)$
$AA2 := RowOperation(AA1, [3, 1], -1)$
$AA3 := RowOperation(AA2, [4, 1], -1)$
$AA4 := RowOperation(AA3, [2, 3])$
$AA5 := RowOperation(AA4, [4, 3], 2)$

gives the reduction to $AA5 \equiv \tilde{A}^{(4)}$.

Gaussian Elimination is a standard routine in the *LinearAlgebra* package of Maple, and the single command

$AA5 := GaussianElimination(AA)$

returns this same reduced matrix. In either case, the final operation

$x := BackwardSubstitute(AA5)$

gives the solution $\mathbf{x}$ which has $x_1 = -7$, $x_2 = 3$, $x_3 = 2$, and $x_4 = 2$.

Illustration The purpose of this illustration is to show what can happen if Algorithm 6.1 fails. The computations will be done simultaneously on two linear systems:

$$\begin{aligned} x_1 + x_2 + x_3 &= 4, \\ 2x_1 + 2x_2 + x_3 &= 6, \\ x_1 + x_2 + 2x_3 &= 6, \end{aligned} \qquad \text{and} \qquad \begin{aligned} x_1 + x_2 + x_3 &= 4, \\ 2x_1 + 2x_2 + x_3 &= 4, \\ x_1 + x_2 + 2x_3 &= 6. \end{aligned}$$

These systems produce the augmented matrices

$$\tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 2 & 2 & 1 & 6 \\ 1 & 1 & 2 & 6 \end{array}\right] \quad \text{and} \quad \tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 2 & 2 & 1 & 4 \\ 1 & 1 & 2 & 6 \end{array}\right].$$

Since $a_{11} = 1$, we perform $(E_2 - 2E_1) \to (E_2)$ and $(E_3 - E_1) \to (E_3)$ to produce

$$\tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad \text{and} \quad \tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 0 & 0 & -1 & -4 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

At this point, $a_{22} = a_{32} = 0$. The algorithm requires that the procedure be halted, and no solution to either system is obtained. Writing the equations for each system gives

$$\begin{aligned} x_1 + x_2 + \quad x_3 &= \quad 4, \\ -x_3 &= -2, \\ x_3 &= \quad 2, \end{aligned} \qquad \text{and} \qquad \begin{aligned} x_1 + x_2 + \quad x_3 &= \quad 4, \\ -x_3 &= -4, \\ x_3 &= \quad 2. \end{aligned}$$

The first linear system has an infinite number of solutions, which can be described by $x_3 = 2$, $x_2 = 2 - x_1$, and x_1 arbitrary.

The second system leads to the contradiction $x_3 = 2$ and $x_3 = 4$, so no solution exists. In each case, however, there is no *unique* solution, as we conclude from Algorithm 6.1. □

Although Algorithm 6.1 can be viewed as the construction of the augmented matrices $\tilde{A}^{(1)}, \ldots, \tilde{A}^{(n)}$, the computations can be performed in a computer using only one $n \times (n+1)$ array for storage. At each step we simply replace the previous value of a_{ij} by the new one. In addition, we can store the multipliers m_{ji} in the locations of a_{ji} because a_{ji} has the value 0 for each $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$. Thus A can be overwritten by the multipliers in the entries that are below the main diagonal (that is, the entries of the form a_{ji}, with $j > i$) and by the newly computed entries of $\tilde{A}^{(n)}$ on and above the main diagonal (the entries of the form a_{ij}, with $j \le i$). These values can be used to solve other linear systems involving the original matrix A, as we will see in Section 6.5.

Operation Counts

Both the amount of time required to complete the calculations and the subsequent round-off error depend on the number of floating-point arithmetic operations needed to solve a routine problem. In general, the amount of time required to perform a multiplication or division on a computer is approximately the same and is considerably greater than that required to perform an addition or subtraction. The actual differences in execution time, however, depend on the particular computing system. To demonstrate the counting operations for a given method, we will count the operations required to solve a typical linear system of n equations in n unknowns using Algorithm 6.1. We will keep the count of the additions/subtractions separate from the count of the multiplications/divisions because of the time differential.

No arithmetic operations are performed until Steps 5 and 6 in the algorithm. Step 5 requires that $(n-i)$ divisions be performed. The replacement of the equation E_j by $(E_j - m_{ji}E_i)$ in Step 6 requires that m_{ji} be multiplied by each term in E_i, resulting in a total of $(n-i)(n-i+1)$ multiplications. After this is completed, each term of the resulting equation is subtracted from the corresponding term in E_j. This requires $(n-i)(n-i+1)$ subtractions. For each $i = 1, 2, \ldots, n-1$, the operations required in Steps 5 and 6 are as follows.

Multiplications/divisions

$$(n-i) + (n-i)(n-i+1) = (n-i)(n-i+2).$$

Additions/subtractions

$$(n-i)(n-i+1).$$

The total number of operations required by Steps 5 and 6 is obtained by summing the operation counts for each i. Recalling from calculus that

$$\sum_{j=1}^{m} 1 = m, \quad \sum_{j=1}^{m} j = \frac{m(m+1)}{2}, \quad \text{and} \quad \sum_{j=1}^{m} j^2 = \frac{m(m+1)(2m+1)}{6},$$

we have the following operation counts.

Multiplications/divisions

$$\begin{aligned}\sum_{i=1}^{n-1}(n-i)(n-i+2) &= \sum_{i=1}^{n-1}(n^2 - 2ni + i^2 + 2n - 2i)\\ &= \sum_{i=1}^{n-1}(n-i)^2 + 2\sum_{i=1}^{n-1}(n-i) = \sum_{i=1}^{n-1} i^2 + 2\sum_{i=1}^{n-1} i\\ &= \frac{(n-1)n(2n-1)}{6} + 2\frac{(n-1)n}{2} = \frac{2n^3+3n^2-5n}{6}.\end{aligned}$$

Additions/subtractions

$$\begin{aligned}\sum_{i=1}^{n-1}(n-i)(n-i+1) &= \sum_{i=1}^{n-1}(n^2 - 2ni + i^2 + n - i)\\ &= \sum_{i=1}^{n-1}(n-i)^2 + \sum_{i=1}^{n-1}(n-i) = \sum_{i=1}^{n-1} i^2 + \sum_{i=1}^{n-1} i\\ &= \frac{(n-1)n(2n-1)}{6} + \frac{(n-1)n}{2} = \frac{n^3-n}{3}.\end{aligned}$$

The only other steps in Algorithm 6.1 that involve arithmetic operations are those required for backward substitution, Steps 8 and 9. Step 8 requires one division. Step 9 requires $(n-i)$ multiplications and $(n-i-1)$ additions for each summation term and then one subtraction and one division. The total number of operations in Steps 8 and 9 is as follows.

Multiplications/divisions

$$1 + \sum_{i=1}^{n-1}((n-i)+1) = 1 + \left(\sum_{i=1}^{n-1}(n-i)\right) + n - 1$$

$$1 = n + \sum_{i=1}^{n-1}(n-i) = n + \sum_{i=1}^{n-1} i = \frac{n^2+n}{2}.$$

Additions/subtractions

$$\sum_{i=1}^{n-1}((n-i-1)+1) = \sum_{i=1}^{n-1}(n-i) = \sum_{i=1}^{n-1} i = \frac{n^2-n}{2}$$

The total number of arithmetic operations in Algorithm 6.1 is, therefore:

Multiplications/divisions

$$\frac{2n^3 + 3n^2 - 5n}{6} + \frac{n^2 + n}{2} = \frac{n^3}{3} + n^2 - \frac{n}{3}.$$

Additions/subtractions

$$\frac{n^3 - n}{3} + \frac{n^2 - n}{2} = \frac{n^3}{3} + \frac{n^2}{2} - \frac{5n}{6}.$$

For large n, the total number of multiplications and divisions is approximately $n^3/3$, as is the total number of additions and subtractions. Thus the amount of computation and the time required increases with n in proportion to n^3, as shown in Table 6.1.

Table 6.1

n	Multiplications/Divisions	Additions/Subtractions
3	17	11
10	430	375
50	44,150	42,875
100	343,300	338,250

EXERCISE SET 6.1

1. For each of the following linear systems, obtain a solution by graphical methods, if possible. Explain the results from a geometrical standpoint.

   a. $x_1 + 2x_2 = 3,$ $x_1 - x_2 = 0.$

   b. $x_1 + 2x_2 = 3,$ $2x_1 + 4x_2 = 6.$

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

   d. $2x_1 + x_2 = -1,$ $4x_1 + 2x_2 = -2,$ $x_1 - 3x_2 = 5.$

2. For each of the following linear systems, obtain a solution by graphical methods, if possible. Explain the results from a geometrical standpoint.

   a. $x_1 + 2x_2 = 0,$ $x_1 - x_2 = 0.$

   b. $x_1 + 2x_2 = 3,$ $-2x_1 - 4x_2 = 6.$

   c. $2x_1 + x_2 = -1,$ $x_1 + x_2 = 2,$ $x_1 - 3x_2 = 5.$

   d. $2x_1 + x_2 + x_3 = 1,$ $2x_1 + 4x_2 - x_3 = -1.$

3. Use Gaussian elimination with backward substitution and two-digit rounding arithmetic to solve the following linear systems. Do not reorder the equations. (The exact solution to each system is $x_1 = 1, x_2 = -1, x_3 = 3$.)

   a.
   $$\begin{aligned} 4x_1 - x_2 + x_3 &= 8, \\ 2x_1 + 5x_2 + 2x_3 &= 3, \\ x_1 + 2x_2 + 4x_3 &= 11. \end{aligned}$$

   b.
   $$\begin{aligned} 4x_1 + x_2 + 2x_3 &= 9, \\ 2x_1 + 4x_2 - x_3 &= -5, \\ x_1 + x_2 - 3x_3 &= -9. \end{aligned}$$

4. Use Gaussian elimination with backward substitution and two-digit rounding arithmetic to solve the following linear systems. Do not reorder the equations. (The exact solution to each system is $x_1 = -1, x_2 = 1, x_3 = 3$.)

   a.
   $$\begin{aligned} -x_1 + 4x_2 + x_3 &= 8, \\ \tfrac{5}{3}x_1 + \tfrac{2}{3}x_2 + \tfrac{2}{3}x_3 &= 1, \\ 2x_1 + x_2 + 4x_3 &= 11. \end{aligned}$$

   b.
   $$\begin{aligned} 4x_1 + 2x_2 - x_3 &= -5, \\ \tfrac{1}{9}x_1 + \tfrac{1}{9}x_2 - \tfrac{1}{3}x_3 &= -1, \\ x_1 + 4x_2 + 2x_3 &= 9. \end{aligned}$$

5. Use the Gaussian Elimination Algorithm to solve the following linear systems, if possible, and determine whether row interchanges are necessary:

a.
$$\begin{aligned} x_1 - x_2 + 3x_3 &= 2, \\ 3x_1 - 3x_2 + x_3 &= -1, \\ x_1 + x_2 &= 3. \end{aligned}$$

b.
$$\begin{aligned} 2x_1 - 1.5x_2 + 3x_3 &= 1, \\ -x_1 + 2x_3 &= 3, \\ 4x_1 - 4.5x_2 + 5x_3 &= 1. \end{aligned}$$

c.
$$\begin{aligned} 2x_1 &= 3, \\ x_1 + 1.5x_2 &= 4.5, \\ -3x_2 + 0.5x_3 &= -6.6, \\ 2x_1 - 2x_2 + x_3 + x_4 &= 0.8. \end{aligned}$$

d.
$$\begin{aligned} x_1 + x_2 + x_4 &= 2, \\ 2x_1 + x_2 - x_3 + x_4 &= 1, \\ 4x_1 - x_2 - 2x_3 + 2x_4 &= 0, \\ 3x_1 - x_2 - x_3 + 2x_4 &= -3. \end{aligned}$$

6. Use the Gaussian Elimination Algorithm to solve the following linear systems, if possible, and determine whether row interchanges are necessary:

a.
$$\begin{aligned} x_2 - 2x_3 &= 4, \\ x_1 - x_2 + x_3 &= 6, \\ x_1 - x_3 &= 2. \end{aligned}$$

b.
$$\begin{aligned} x_1 - \tfrac{1}{2}x_2 + x_3 &= 4, \\ 2x_1 - x_2 - x_3 + x_4 &= 5, \\ x_1 + x_2 + \tfrac{1}{2}x_3 &= 2, \\ x_1 - \tfrac{1}{2}x_2 + x_3 + x_4 &= 5. \end{aligned}$$

c.
$$\begin{aligned} 2x_1 - x_2 + x_3 - x_4 &= 6, \\ x_2 - x_3 + x_4 &= 5, \\ x_4 &= 5, \\ x_3 - x_4 &= 3. \end{aligned}$$

d.
$$\begin{aligned} x_1 + x_2 + x_4 &= 2, \\ 2x_1 + x_2 - x_3 + x_4 &= 1, \\ -x_1 + 2x_2 + 3x_3 - x_4 &= 4, \\ 3x_1 - x_2 - x_3 + 2x_4 &= -3. \end{aligned}$$

7. Use Algorithm 6.1 and Maple with *Digits*:= 10 to solve the following linear systems.

a.
$$\begin{aligned} \tfrac{1}{4}x_1 + \tfrac{1}{5}x_2 + \tfrac{1}{6}x_3 &= 9, \\ \tfrac{1}{3}x_1 + \tfrac{1}{4}x_2 + \tfrac{1}{5}x_3 &= 8, \\ \tfrac{1}{2}x_1 + x_2 + 2x_3 &= 8. \end{aligned}$$

b.
$$\begin{aligned} 3.333x_1 + 15920x_2 - 10.333x_3 &= 15913, \\ 2.222x_1 + 16.71x_2 + 9.612x_3 &= 28.544, \\ 1.5611x_1 + 5.1791x_2 + 1.6852x_3 &= 8.4254. \end{aligned}$$

c.
$$\begin{aligned} x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{3}x_3 + \tfrac{1}{4}x_4 &= \tfrac{1}{6}, \\ \tfrac{1}{2}x_1 + \tfrac{1}{3}x_2 + \tfrac{1}{4}x_3 + \tfrac{1}{5}x_4 &= \tfrac{1}{7}, \\ \tfrac{1}{3}x_1 + \tfrac{1}{4}x_2 + \tfrac{1}{5}x_3 + \tfrac{1}{6}x_4 &= \tfrac{1}{8}, \\ \tfrac{1}{4}x_1 + \tfrac{1}{5}x_2 + \tfrac{1}{6}x_3 + \tfrac{1}{7}x_4 &= \tfrac{1}{9}. \end{aligned}$$

d.
$$\begin{aligned} 2x_1 + x_2 - x_3 + x_4 - 3x_5 &= 7, \\ x_1 + 2x_3 - x_4 + x_5 &= 2, \\ -2x_2 - x_3 + x_4 - x_5 &= -5, \\ 3x_1 + x_2 - 4x_3 + 5x_5 &= 6, \\ x_1 - x_2 - x_3 - x_4 + x_5 &= 3. \end{aligned}$$

8. Use Algorithm 6.1 and Maple with *Digits*:= 10 to solve the following linear systems.

a.
$$\begin{aligned} \tfrac{1}{2}x_1 + \tfrac{1}{4}x_2 - \tfrac{1}{8}x_3 &= 0, \\ \tfrac{1}{3}x_1 - \tfrac{1}{6}x_2 + \tfrac{1}{9}x_3 &= 1, \\ \tfrac{1}{7}x_1 + \tfrac{1}{7}x_2 + \tfrac{1}{10}x_3 &= 2. \end{aligned}$$

b.
$$\begin{aligned} 2.71x_1 + x_2 + 1032x_3 &= 12, \\ 4.12x_1 - x_2 + 500x_3 &= 11.49, \\ 3.33x_1 + 2x_2 - 200x_3 &= 41. \end{aligned}$$

c.
$$\begin{aligned} \pi x_1 + \sqrt{2}x_2 - x_3 + x_4 &= 0, \\ ex_1 - x_2 + x_3 + 2x_4 &= 1, \\ x_1 + x_2 - \sqrt{3}x_3 + x_4 &= 2, \\ -x_1 - x_2 + x_3 - \sqrt{5}x_4 &= 3. \end{aligned}$$

d.
$$\begin{aligned} x_1 + x_2 - x_3 + x_4 - x_5 &= 2, \\ 2x_1 + 2x_2 + x_3 - x_4 + x_5 &= 4, \\ 3x_1 + x_2 - 3x_3 - 2x_4 + 3x_5 &= 8, \\ 4x_1 + x_2 - x_3 + 4x_4 - 5x_5 &= 16, \\ 16x_1 - x_2 + x_3 - x_4 - x_5 &= 32. \end{aligned}$$

9. Given the linear system

$$\begin{aligned} 2x_1 - 6\alpha x_2 &= 3, \\ 3\alpha x_1 - x_2 &= \tfrac{3}{2}. \end{aligned}$$

a. Find value(s) of α for which the system has no solutions.

b. Find value(s) of α for which the system has an infinite number of solutions.

c. Assuming a unique solution exists for a given α, find the solution.

10. Given the linear system

$$\begin{aligned} x_1 - x_2 + \alpha x_3 &= -2, \\ -x_1 + 2x_2 - \alpha x_3 &= 3, \\ \alpha x_1 + x_2 + x_3 &= 2. \end{aligned}$$

a. Find value(s) of α for which the system has no solutions.

b. Find value(s) of α for which the system has an infinite number of solutions.

c. Assuming a unique solution exists for a given α, find the solution.

11. Show that the operations

a. $(\lambda E_i) \to (E_i)$ **b.** $(E_i + \lambda E_j) \to (E_i)$ **c.** $(E_i) \leftrightarrow (E_j)$

do not change the solution set of a linear system.

12. Gauss-Jordan Method: This method is described as follows. Use the ith equation to eliminate not only x_i from the equations $E_{i+1}, E_{i+2}, \dots, E_n$, as was done in the Gaussian elimination method, but also from $E_1, E_2, \dots, E_{i-1}$. Upon reducing $[A, \mathbf{b}]$ to:

$$\left[\begin{array}{cccc:c} a_{11}^{(1)} & 0 & \cdots & 0 & a_{1,n+1}^{(1)} \\ 0 & a_{22}^{(2)} & \ddots & \vdots & a_{2,n+1}^{(2)} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & a_{nn}^{(n)}) & a_{n,n+1}^{(n)} \end{array}\right],$$

the solution is obtained by setting

$$x_i = \frac{a_{i,n+1}^{(i)}}{a_{ii}^{(i)}},$$

for each $i = 1, 2, \dots, n$. This procedure circumvents the backward substitution in the Gaussian elimination. Construct an algorithm for the Gauss-Jordan procedure patterned after that of Algorithm 6.1.

13. Use the Gauss-Jordan method and two-digit rounding arithmetic to solve the systems in Exercise 3.

14. Repeat Exercise 7 using the Gauss-Jordan method.

15. **a.** Show that the Gauss-Jordan method requires

$$\frac{n^3}{2} + n^2 - \frac{n}{2} \quad \text{multiplications/divisions}$$

and

$$\frac{n^3}{2} - \frac{n}{2} \quad \text{additions/subtractions.}$$

b. Make a table comparing the required operations for the Gauss-Jordan and Gaussian elimination methods for $n = 3, 10, 50, 100$. Which method requires less computation?

16. Consider the following Gaussian-elimination-Gauss-Jordan hybrid method for solving the system (6.4). First, apply the Gaussian-elimination technique to reduce the system to triangular form. Then use the nth equation to eliminate the coefficients of x_n in each of the first $n - 1$ rows. After this is completed use the $(n-1)$st equation to eliminate the coefficients of x_{n-1} in the first $n - 2$ rows, etc. The system will eventually appear as the reduced system in Exercise 12.

a. Show that this method requires

$$\frac{n^3}{3} + \frac{3}{2}n^2 - \frac{5}{6}n \quad \text{multiplications/divisions}$$

and

$$\frac{n^3}{3} + \frac{n^2}{2} - \frac{5}{6}n \quad \text{additions/subtractions.}$$

b. Make a table comparing the required operations for the Gaussian elimination, Gauss-Jordan, and hybrid methods, for $n = 3, 10, 50, 100$.

17. Use the hybrid method described in Exercise 16 and two-digit rounding arithmetic to solve the systems in Exercise 3.

18. Repeat Exercise 7 using the method described in Exercise 16.

19. Suppose that in a biological system there are n species of animals and m sources of food. Let x_j represent the population of the jth species, for each $j = 1, \cdots, n$; b_i represent the available daily supply of the ith food; and a_{ij} represent the amount of the ith food consumed on the average by a member of the jth species. The linear system

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

represents an equilibrium where there is a daily supply of food to precisely meet the average daily consumption of each species.

a. Let

$$A = [a_{ij}] = \left[\begin{array}{ccc:c} 1 & 2 & 0 & 3 \\ 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right],$$

$\mathbf{x} = (x_j) = [1000, 500, 350, 400]$, and $\mathbf{b} = (b_i) = [3500, 2700, 900]$. Is there sufficient food to satisfy the average daily consumption?

b. What is the maximum number of animals of each species that could be individually added to the system with the supply of food still meeting the consumption?

c. If species 1 became extinct, how much of an individual increase of each of the remaining species could be supported?

d. If species 2 became extinct, how much of an individual increase of each of the remaining species could be supported?

20. A Fredholm integral equation of the second kind is an equation of the form

$$u(x) = f(x) + \int_a^b K(x, t)u(t)\, dt,$$

where a and b and the functions f and K are given. To approximate the function u on the interval $[a, b]$, a partition $x_0 = a < x_1 < \cdots < x_{m-1} < x_m = b$ is selected and the equations

$$u(x_i) = f(x_i) + \int_a^b K(x_i, t)u(t)\, dt, \quad \text{for each } i = 0, \cdots, m,$$

are solved for $u(x_0), u(x_1), \cdots, u(x_m)$. The integrals are approximated using quadrature formulas based on the nodes $x_0, \cdots, x_m$. In our problem, $a = 0$, $b = 1$, $f(x) = x^2$, and $K(x, t) = e^{|x-t|}$.

a. Show that the linear system

$$u(0) = f(0) + \frac{1}{2}[K(0, 0)u(0) + K(0, 1)u(1)],$$

$$u(1) = f(1) + \frac{1}{2}[K(1, 0)u(0) + K(1, 1)u(1)]$$

must be solved when the Trapezoidal rule is used.

b. Set up and solve the linear system that results when the Composite Trapezoidal rule is used with $n = 4$.

c. Repeat part (b) using the Composite Simpson's rule.

6.2 Pivoting Strategies

In deriving Algorithm 6.1, we found that a row interchange was needed when one of the pivot elements $a_{kk}^{(k)}$ is 0. This row interchange has the form $(E_k) \leftrightarrow (E_p)$, where p is the smallest integer greater than k with $a_{pk}^{(k)} \neq 0$. To reduce round-off error, it is often necessary to perform row interchanges even when the pivot elements are not zero.

If $a_{kk}^{(k)}$ is small in magnitude compared to $a_{jk}^{(k)}$, then the magnitude of the multiplier

$$m_{jk} = \frac{a_{jk}^{(k)}}{a_{kk}^{(k)}}$$

will be much larger than 1. Round-off error introduced in the computation of one of the terms $a_{kl}^{(k)}$ is multiplied by m_{jk} when computing $a_{jl}^{(k+1)}$, which compounds the original error. Also, when performing the backward substitution for

$$x_k = \frac{a_{k,n+1}^{(k)} - \sum_{j=k+1}^{n} a_{kj}^{(k)}}{a_{kk}^{(k)}},$$

with a small value of $a_{kk}^{(k)}$, any error in the numerator can be dramatically increased because of the division by $a_{kk}^{(k)}$. In our next example, we will see that even for small systems, round-off error can dominate the calculations.

Example 1 Apply Gaussian elimination to the system

$$\begin{aligned} E_1: &\quad 0.003000x_1 + 59.14x_2 = 59.17 \\ E_2: &\quad 5.291x_1 - 6.130x_2 = 46.78, \end{aligned}$$

using four-digit arithmetic with rounding, and compare the results to the exact solution $x_1 = 10.00$ and $x_2 = 1.000$.

Solution The first pivot element, $a_{11}^{(1)} = 0.003000$, is small, and its associated multiplier,

$$m_{21} = \frac{5.291}{0.003000} = 1763.\overline{6},$$

rounds to the large number 1764. Performing $(E_2 - m_{21}E_1) \rightarrow (E_2)$ and the appropriate rounding gives the system

$$\begin{aligned} 0.003000x_1 + 59.14x_2 &\approx 59.17 \\ -104300x_2 &\approx -104400, \end{aligned}$$

instead of the exact system, which is

$$\begin{aligned} 0.003000x_1 + 59.14x_2 &= 59.17 \\ -104309.37\overline{6}x_2 &= -104309.37\overline{6}. \end{aligned}$$

The disparity in the magnitudes of $m_{21}a_{13}$ and a_{23} has introduced round-off error, but the round-off error has not yet been propagated. Backward substitution yields

$$x_2 \approx 1.001,$$

which is a close approximation to the actual value, $x_2 = 1.000$. However, because of the small pivot $a_{11} = 0.003000$,

$$x_1 \approx \frac{59.17 - (59.14)(1.001)}{0.003000} = -10.00$$

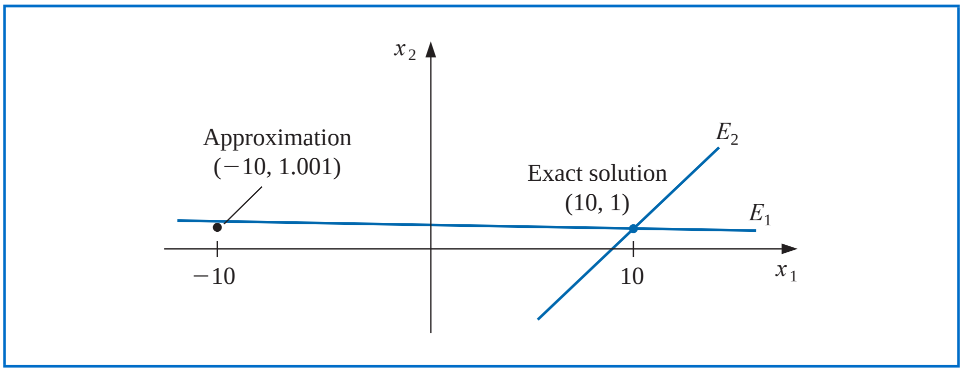
contains the small error of 0.001 multiplied by

$$\frac{59.14}{0.003000} \approx 20000.$$

This ruins the approximation to the actual value $x_1 = 10.00$.

This is clearly a contrived example and the graph in Figure 6.1. shows why the error can so easily occur. For larger systems it is much more difficult to predict in advance when devastating round-off error might occur. ■

Figure 6.1



Partial Pivoting

Example 1 shows how difficulties can arise when the pivot element $a_{kk}^{(k)}$ is small relative to the entries $a_{ij}^{(k)}$, for $k \le i \le n$ and $k \le j \le n$. To avoid this problem, pivoting is performed by selecting an element $a_{pq}^{(k)}$ with a larger magnitude as the pivot, and interchanging the kth and pth rows. This can be followed by the interchange of the kth and qth columns, if necessary.

The simplest strategy is to select an element in the same column that is below the diagonal and has the largest absolute value; specifically, we determine the smallest $p \ge k$ such that

$$|a_{pk}^{(k)}| = \max_{k \le i \le n} |a_{ik}^{(k)}|$$

and perform $(E_k) \leftrightarrow (E_p)$. In this case no interchange of columns is used.

Example 2 Apply Gaussian elimination to the system

$$\begin{aligned} E_1: &\quad 0.003000x_1 + 59.14x_2 = 59.17 \\ E_2: &\quad 5.291x_1 - 6.130x_2 = 46.78, \end{aligned}$$

using partial pivoting and four-digit arithmetic with rounding, and compare the results to the exact solution $x_1 = 10.00$ and $x_2 = 1.000$.

Solution The partial-pivoting procedure first requires finding

$$\max\left\{|a_{11}^{(1)}|, |a_{21}^{(1)}|\right\} = \max\{|0.003000|, |5.291|\} = |5.291| = |a_{21}^{(1)}|.$$

This requires that the operation $(E_2) \leftrightarrow (E_1)$ be performed to produce the equivalent system

$$\begin{aligned} E_1: &\quad 5.291x_1 - 6.130x_2 = 46.78, \\ E_2: &\quad 0.003000x_1 + 59.14x_2 = 59.17. \end{aligned}$$

The multiplier for this system is

$$m_{21} = \frac{a_{21}^{(1)}}{a_{11}^{(1)}} = 0.0005670,$$

and the operation $(E_2 - m_{21}E_1) \rightarrow (E_2)$ reduces the system to

$$\begin{aligned} 5.291x_1 - 6.130x_2 &\approx 46.78, \\ 59.14x_2 &\approx 59.14. \end{aligned}$$

The four-digit answers resulting from the backward substitution are the correct values $x_1 = 10.00$ and $x_2 = 1.000$. ■

The technique just described is called **partial pivoting** (or *maximal column pivoting*) and is detailed in Algorithm 6.2. The actual row interchanging is simulated in the algorithm by interchanging the values of $NROW$ in Step 5.

ALGORITHM 6.2

Gaussian Elimination with Partial Pivoting

To solve the $n \times n$ linear system

$$\begin{aligned} E_1 :&\quad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = a_{1,n+1} \\ E_2 :&\quad a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = a_{2,n+1} \\ &\qquad\qquad \vdots \qquad\qquad\qquad\quad \vdots \\ E_n :&\quad a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = a_{n,n+1} \end{aligned}$$

INPUT number of unknowns and equations n; augmented matrix $A = [a_{ij}]$ where $1 \leq i \leq n$ and $1 \leq j \leq n+1$.

OUTPUT solution $x_1, \ldots, x_n$ or message that the linear system has no unique solution.

Step 1 For $i = 1, \ldots, n$ set $NROW(i) = i$. (*Initialize row pointer.*)

Step 2 For $i = 1, \ldots, n-1$ do Steps 3–6. (*Elimination process.*)

Step 3 Let p be the smallest integer with $i \leq p \leq n$ and $|a(NROW(p), i)| = \max_{i \leq j \leq n} |a(NROW(j), i)|$.
(*Notation:* $a(NROW(i), j) \equiv a_{NROW_i, j}$.)

Step 4 If $a(NROW(p), i) = 0$ then OUTPUT ('no unique solution exists');
STOP.

Step 5 If $NROW(i) \neq NROW(p)$ then set $NCOPY = NROW(i)$;
$NROW(i) = NROW(p)$;
$NROW(p) = NCOPY$.
(*Simulated row interchange.*)

Step 6 For $j = i+1, \ldots, n$ do Steps 7 and 8.

Step 7 Set $m(NROW(j), i) = a(NROW(j), i)/a(NROW(i), i)$.

Step 8 Perform $(E_{NROW(j)} - m(NROW(j), i) \cdot E_{NROW(i)}) \rightarrow (E_{NROW(j)})$.

Step 9 If $a(NROW(n), n) = 0$ then OUTPUT ('no unique solution exists');
STOP.

Step 10 Set $x_n = a(NROW(n), n+1)/a(NROW(n), n)$.
(*Start backward substitution.*)

Step 11 For $i = n-1, \ldots, 1$

$$\text{set } x_i = \frac{a(NROW(i), n+1) - \sum_{j=i+1}^{n} a(NROW(i), j) \cdot x_j}{a(NROW(i), i)}.$$

Step 12 OUTPUT $(x_1, \ldots, x_n)$; (*Procedure completed successfully.*)
STOP. ■

Each multiplier m_{ji} in the partial pivoting algorithm has magnitude less than or equal to 1. Although this strategy is sufficient for many linear systems, situations do arise when it is inadequate.

Illustration The linear system

$$\begin{aligned} E_1: &\quad 30.00x_1 + 591400x_2 = 591700, \\ E_2: &\quad 5.291x_1 - 6.130x_2 = 46.78, \end{aligned}$$

is the same as that in Examples 1 and 2 except that all the entries in the first equation have been multiplied by 10^4. The partial pivoting procedure described in Algorithm 6.2 with four-digit arithmetic leads to the same results as obtained in Example 1. The maximal value in the first column is 30.00, and the multiplier

$$m_{21} = \frac{5.291}{30.00} = 0.1764$$

leads to the system

$$\begin{aligned} 30.00x_1 + 591400x_2 &\approx 591700, \\ -104300x_2 &\approx -104400, \end{aligned}$$

which has the same inaccurate solutions as in Example 1: $x_2 \approx 1.001$ and $x_1 \approx -10.00$. □

Scaled Partial Pivoting

Scaled partial pivoting (or *scaled-column pivoting*) is needed for the system in the Illustration. It places the element in the pivot position that is largest relative to the entries in its row. The first step in this procedure is to define a scale factor s_i for each row as

$$s_i = \max_{1 \le j \le n} |a_{ij}|.$$

If we have $s_i = 0$ for some i, then the system has no unique solution since all entries in the ith row are 0. Assuming that this is not the case, the appropriate row interchange to place zeros in the first column is determined by choosing the least integer p with

$$\frac{|a_{p1}|}{s_p} = \max_{1 \le k \le n} \frac{|a_{k1}|}{s_k}$$

and performing $(E_1) \leftrightarrow (E_p)$. The effect of scaling is to ensure that the largest element in each row has a *relative* magnitude of 1 before the comparison for row interchange is performed.

In a similar manner, before eliminating the variable x_i using the operations

$$E_k - m_{ki}E_i, \quad \text{for } k = i+1, \ldots, n,$$

we select the smallest integer $p \geq i$ with

$$\frac{|a_{pi}|}{s_p} = \max_{i \leq k \leq n} \frac{|a_{ki}|}{s_k}$$

and perform the row interchange $(E_i) \leftrightarrow (E_p)$ if $i \neq p$. The scale factors $s_1, \ldots, s_n$ are computed only once, at the start of the procedure. They are row dependent, so they must also be interchanged when row interchanges are performed.

Illustration Applying scaled partial pivoting to the previous Illustration gives

$$s_1 = \max\{|30.00|, |591400|\} = 591400$$

and

$$s_2 = \max\{|5.291|, |-6.130|\} = 6.130.$$

Consequently

$$\frac{|a_{11}|}{s_1} = \frac{30.00}{591400} = 0.5073 \times 10^{-4}, \qquad \frac{|a_{21}|}{s_2} = \frac{5.291}{6.130} = 0.8631,$$

and the interchange $(E_1) \leftrightarrow (E_2)$ is made.

Applying Gaussian elimination to the new system

$$5.291x_1 - 6.130x_2 = 46.78$$

$$30.00x_1 + 591400x_2 = 591700$$

produces the correct results: $x_1 = 10.00$ and $x_2 = 1.000$. □

Algorithm 6.3 implements scaled partial pivoting.

ALGORITHM 6.3

Gaussian Elimination with Scaled Partial Pivoting

The only steps in this algorithm that differ from those of Algorithm 6.2 are:

Step 1 For $i = 1, \ldots, n$ set $s_i = \max_{1 \leq j \leq n} |a_{ij}|$;
if $s_i = 0$ then OUTPUT ('no unique solution exists');
STOP.
set $NROW(i) = i$.

Step 2 For $i = 1, \ldots, n-1$ do Steps 3–6. (*Elimination process.*)

Step 3 Let p be the smallest integer with $i \leq p \leq n$ and

$$\frac{|a(NROW(p), i)|}{s(NROW(p))} = \max_{i \leq j \leq n} \frac{|a(NROW(j), i)|}{s(NROW(j))}.$$

■

The next example demonstrates using Maple and the *LinearAlgebra* library to perform scaled partial pivoting with finite-digit rounding arithmetic.

Example 3 Solve the linear system using three-digit rounding arithmetic in Maple with the *LinearAlgebra* library.

$$\begin{aligned}
2.11x_1 - \ \ 4.21x_2 + 0.921x_3 &= \ \ 2.01,\\
4.01x_1 + \ \ 10.2x_2 - \ \ 1.12x_3 &= -3.09,\\
1.09x_1 + 0.987x_2 + 0.832x_3 &= \ \ 4.21.
\end{aligned}$$

Solution To obtain three-digit rounding arithmetic, enter

$Digits := 3$

We have $s_1 = 4.21$, $s_2 = 10.2$, and $s_3 = 1.09$. So

$$\frac{|a_{11}|}{s_1} = \frac{2.11}{4.21} = 0.501, \quad \frac{|a_{21}|}{s_1} = \frac{4.01}{10.2} = 0.393, \quad \text{and} \quad \frac{|a_{31}|}{s_3} = \frac{1.09}{1.09} = 1.$$

Next we load the *LinearAlgebra* library.

with(*LinearAlgebra*)

The augmented matrix *AA* is defined by

$AA := Matrix([[2.11, -4.21, 0.921, 2.01], [4.01, 10.2, -1.12, -3.09],$
$[1.09, 0.987, 0.832, 4.21]])$

which gives

$$\begin{bmatrix} 2.11 & -4.21 & .921 & 2.01 \\ 4.01 & 10.2 & -1.12 & -3.09 \\ 1.09 & .987 & .832 & 4.21 \end{bmatrix}.$$

Since $|a_{31}|/s_3$ is largest, we perform $(E_1) \leftrightarrow (E_3)$ using

$AA1 := RowOperation(AA, [1, 3])$

to obtain

$$\begin{bmatrix} 1.09 & .987 & .832 & 4.21 \\ 4.01 & 10.2 & -1.12 & -3.09 \\ 2.11 & -4.21 & .921 & 2.01 \end{bmatrix}.$$

Compute the multipliers

$$m21 := \frac{AA1[2, 1]}{AA1[1, 1]}; m31 := \frac{AA1[3, 1]}{AA1[1, 1]}$$

giving

$$3.68$$

$$1.94$$

Perform the first two eliminations using

$AA2 := RowOperation(AA1, [2, 1], -m21) : AA3 := RowOperation(AA2, [3, 1], -m31)$

to produce

$$\begin{bmatrix} 1.09 & .987 & .832 & 4.21 \\ 0 & 6.57 & -4.18 & -18.6 \\ 0 & -6.12 & -.689 & -6.16 \end{bmatrix}.$$

Since

$$\frac{|a_{22}|}{s_2} = \frac{6.57}{10.2} = 0.644 \quad \text{and} \quad \frac{|a_{32}|}{s_3} = \frac{6.12}{4.21} = 1.45,$$

we perform

$AA4 := RowOperation(AA3, [2, 3])$

giving

$$\begin{bmatrix} 1.09 & .987 & .832 & 4.21 \\ 0 & -6.12 & -.689 & -6.16 \\ 0 & 6.57 & -4.18 & -18.6 \end{bmatrix}.$$

The multiplier m_{32} is computed by

$m32 := \dfrac{AA4[3, 2]}{AA4[2, 2]}$

$$-1.07$$

and the elimination step

$AA5 := RowOperation(AA4, [3, 2], -m32)$

results in the matrix

$$\begin{bmatrix} 1.09 & .987 & .832 & 4.21 \\ 0 & -6.12 & -.689 & -6.16 \\ 0 & .02 & -4.92 & -25.2 \end{bmatrix}.$$

We cannot use *BackwardSubstitute* on this matrix because of the entry .02 in the last row of the second column, that is, which Maple knows as the $(3, 2)$ position. This entry is nonzero due to rounding, but we can remedy this minor problem setting it to 0 with the command

$AA5[3, 2] := 0$

You can verify this is correct with the command $evalm(AA5)$

Finally, backward substitution gives the solution $\mathbf{x}$, which to 3 decimal digits is $x_1 = -0.436$, $x_2 = 0.430$, and $x_3 = 5.12$. ■

The first additional computations required for scaled partial pivoting result from the determination of the scale factors; there are $(n - 1)$ comparisons for each of the n rows, for a total of

$$n(n - 1) \text{ comparisons.}$$

To determine the correct first interchange, n divisions are performed, followed by $n - 1$ comparisons. So the first interchange determination adds

$$n \text{ divisions and } (n - 1) \text{ comparisons.}$$

The scaling factors are computed only once, so the second step requires

$$(n - 1) \text{ divisions and } (n - 2) \text{ comparisons.}$$

We proceed in a similar manner until there are zeros below the main diagonal in all but the nth row. The final step requires that we perform

$$2 \text{ divisions and } 1 \text{ comparison.}$$

As a consequence, scaled partial pivoting adds a total of

$$n(n - 1) + \sum_{k=1}^{n-1} k = n(n - 1) + \frac{(n - 1)n}{2} = \frac{3}{2}n(n - 1) \quad \text{comparisons} \tag{6.7}$$

and

$$\sum_{k=2}^{n} k = \left(\sum_{k=1}^{n} k\right) - 1 = \frac{n(n+1)}{2} - 1 = \frac{1}{2}(n-1)(n+2) \quad \text{divisions}$$

to the Gaussian elimination procedure. The time required to perform a comparison is about the same as an addition/subtraction. Since the total time to perform the basic Gaussian elimination procedure is $O(n^3/3)$ multiplications/divisions and $O(n^3/3)$ additions/subtractions, scaled partial pivoting does not add significantly to the computational time required to solve a system for large values of n.

To emphasize the importance of choosing the scale factors only once, consider the amount of additional computation that would be required if the procedure were modified so that new scale factors were determined each time a row interchange decision was to be made. In this case, the term $n(n-1)$ in Eq. (6.7) would be replaced by

$$\sum_{k=2}^{n} k(k-1) = \frac{1}{3}n(n^2-1).$$

As a consequence, this pivoting technique would add $O(n^3/3)$ comparisons, in addition to the $[n(n+1)/2] - 1$ divisions.

Complete Pivoting

Pivoting can incorporate the interchange of both rows and columns. **Complete** (or *maximal*) **pivoting** at the kth step searches all the entries a_{ij}, for $i = k, k+1, \ldots, n$ and $j = k, k+1, \ldots, n$, to find the entry with the largest magnitude. Both row and column interchanges are performed to bring this entry to the pivot position. The first step of total pivoting requires that $n^2 - 1$ comparisons be performed, the second step requires $(n-1)^2 - 1$ comparisons, and so on. The total additional time required to incorporate complete pivoting into Gaussian elimination is

$$\sum_{k=2}^{n} (k^2 - 1) = \frac{n(n-1)(2n+5)}{6}$$

comparisons. Complete pivoting is, consequently, the strategy recommended only for systems where accuracy is essential and the amount of execution time needed for this method can be justified.

EXERCISE SET 6.2

1. Find the row interchanges that are required to solve the following linear systems using Algorithm 6.1.

a.
$$\begin{aligned} x_1 - 5x_2 + x_3 &= 7, \\ 10x_1 + 20x_3 &= 6, \\ 5x_1 - x_3 &= 4. \end{aligned}$$

b.
$$\begin{aligned} x_1 + x_2 - x_3 &= 1, \\ x_1 + x_2 + 4x_3 &= 2, \\ 2x_1 - x_2 + 2x_3 &= 3. \end{aligned}$$

c.
$$\begin{aligned} 2x_1 - 3x_2 + 2x_3 &= 5, \\ -4x_1 + 2x_2 - 6x_3 &= 14, \\ 2x_1 + 2x_2 + 4x_3 &= 8. \end{aligned}$$

d.
$$\begin{aligned} x_2 + x_3 &= 6, \\ x_1 - 2x_2 - x_3 &= 4, \\ x_1 - x_2 + x_3 &= 5. \end{aligned}$$

2. Find the row interchanges that are required to solve the following linear systems using Algorithm 6.1.

a. $$\begin{aligned} 13x_1 + 17x_2 + x_3 &= 5, \\ x_2 + 19x_3 &= 1, \\ 12x_2 - x_3 &= 0. \end{aligned}$$

b. $$\begin{aligned} x_1 + x_2 - x_3 &= 0, \\ 12x_2 - x_3 &= 4, \\ 2x_1 + x_2 + x_3 &= 5. \end{aligned}$$

c. $$\begin{aligned} 5x_1 + x_2 - 6x_3 &= 7, \\ 2x_1 + x_2 - x_3 &= 8, \\ 6x_1 + 12x_2 + x_3 &= 9. \end{aligned}$$

d. $$\begin{aligned} x_1 - x_2 + x_3 &= 5, \\ 7x_1 + 5x_2 - x_3 &= 8, \\ 2x_1 + x_2 + x_3 &= 7. \end{aligned}$$

3. Repeat Exercise 1 using Algorithm 6.2.
4. Repeat Exercise 2 using Algorithm 6.2.
5. Repeat Exercise 1 using Algorithm 6.3.
6. Repeat Exercise 2 using Algorithm 6.3.
7. Repeat Exercise 1 using complete pivoting.
8. Repeat Exercise 2 using complete pivoting.
9. Use Gaussian elimination and three-digit chopping arithmetic to solve the following linear systems, and compare the approximations to the actual solution.

a. $$\begin{aligned} 0.03x_1 + 58.9x_2 &= 59.2, \\ 5.31x_1 - 6.10x_2 &= 47.0. \end{aligned}$$
Actual solution $[10, 1]$.

b. $$\begin{aligned} 3.03x_1 - 12.1x_2 + 14x_3 &= -119, \\ -3.03x_1 + 12.1x_2 - 7x_3 &= 120, \\ 6.11x_1 - 14.2x_2 + 21x_3 &= -139. \end{aligned}$$
Actual solution $[0, 10, \frac{1}{7}]$.

c. $$\begin{aligned} 1.19x_1 + 2.11x_2 - 100x_3 + x_4 &= 1.12, \\ 14.2x_1 - 0.122x_2 + 12.2x_3 - x_4 &= 3.44, \\ 100x_2 - 99.9x_3 + x_4 &= 2.15, \\ 15.3x_1 + 0.110x_2 - 13.1x_3 - x_4 &= 4.16. \end{aligned}$$
Actual solution $[0.176, 0.0126, -0.0206, -1.18]$.

d. $$\begin{aligned} \pi x_1 - e x_2 + \sqrt{2}x_3 - \sqrt{3}x_4 &= \sqrt{11}, \\ \pi^2 x_1 + e x_2 - e^2 x_3 + \tfrac{3}{7}x_4 &= 0, \\ \sqrt{5}x_1 - \sqrt{6}x_2 + x_3 - \sqrt{2}x_4 &= \pi, \\ \pi^3 x_1 + e^2 x_2 - \sqrt{7}x_3 + \tfrac{1}{9}x_4 &= \sqrt{2}. \end{aligned}$$
Actual solution $[0.788, -3.12, 0.167, 4.55]$.

10. Use Gaussian elimination and three-digit chopping arithmetic to solve the following linear systems, and compare the approximations to the actual solution.

a. $$\begin{aligned} 58.9x_1 + 0.03x_2 &= 59.2, \\ -6.10x_1 + 5.31x_2 &= 47.0. \end{aligned}$$
Actual solution $[1, 10]$.

b. $$\begin{aligned} 3.3330x_1 + 15920x_2 + 10.333x_3 &= 7953, \\ 2.2220x_1 + 16.710x_2 + 9.6120x_3 &= 0.965, \\ -1.5611x_1 + 5.1792x_2 - 1.6855x_3 &= 2.714. \end{aligned}$$
Actual solution $[1, 0.5, -1]$.

c. $$\begin{aligned} 2.12x_1 - 2.12x_2 + 51.3x_3 + 100x_4 &= \pi, \\ 0.333x_1 - 0.333x_2 - 12.2x_3 + 19.7x_4 &= \sqrt{2}, \\ 6.19x_1 + 8.20x_2 - 1.00x_3 - 2.01x_4 &= 0, \\ -5.73x_1 + 6.12x_2 + x_3 - x_4 &= -1. \end{aligned}$$
Actual solution $[0.0998, -0.0683, -0.0363, 0.0465]$.

d. $$\begin{aligned} \pi x_1 + \sqrt{2}x_2 - x_3 + x_4 &= 0, \\ e x_1 - x_2 + x_3 + 2x_4 &= 1, \\ x_1 + x_2 - \sqrt{3}x_3 + x_4 &= 2, \\ -x_1 - x_2 + x_3 - \sqrt{5}x_4 &= 3. \end{aligned}$$
Actual solution $[1.35, -4.68, -4.03, -1.66]$.

11. Repeat Exercise 9 using three-digit rounding arithmetic.
12. Repeat Exercise 10 using three-digit rounding arithmetic.
13. Repeat Exercise 9 using Gaussian elimination with partial pivoting.
14. Repeat Exercise 10 using Gaussian elimination with partial pivoting.
15. Repeat Exercise 9 using Gaussian elimination with partial pivoting and three-digit rounding arithmetic.
16. Repeat Exercise 10 using Gaussian elimination with partial pivoting and three-digit rounding arithmetic.
17. Repeat Exercise 9 using Gaussian elimination with scaled partial pivoting.
18. Repeat Exercise 10 using Gaussian elimination with scaled partial pivoting.
19. Repeat Exercise 9 using Gaussian elimination with scaled partial pivoting and three-digit rounding arithmetic.
20. Repeat Exercise 10 using Gaussian elimination with scaled partial pivoting and three-digit rounding arithmetic.
21. Repeat Exercise 9 using Algorithm 6.1 in Maple with *Digits*:= 10.
22. Repeat Exercise 10 using Algorithm 6.1 in Maple with *Digits*:= 10.
23. Repeat Exercise 9 using Algorithm 6.2 in Maple with *Digits*:= 10.
24. Repeat Exercise 10 using Algorithm 6.2 in Maple with *Digits*:= 10.
25. Repeat Exercise 9 using Algorithm 6.3 in Maple with *Digits*:= 10.
26. Repeat Exercise 10 using Algorithm 6.3 in Maple with *Digits*:= 10.
27. Repeat Exercise 9 using Gaussian elimination with complete pivoting.
28. Repeat Exercise 10 using Gaussian elimination with complete pivoting.
29. Repeat Exercise 9 using Gaussian elimination with complete pivoting and three-digit rounding arithmetic.
30. Repeat Exercise 10 using Gaussian elimination with complete pivoting and three-digit rounding arithmetic.
31. Suppose that

$$2x_1 + x_2 + 3x_3 = 1,$$
$$4x_1 + 6x_2 + 8x_3 = 5,$$
$$6x_1 + \alpha x_2 + 10x_3 = 5,$$

with $|\alpha| < 10$. For which of the following values of α will there be no row interchange required when solving this system using scaled partial pivoting?

a. $\alpha = 6$ **b.** $\alpha = 9$ **c.** $\alpha = -3$

32. Construct an algorithm for the complete pivoting procedure discussed in the text.
33. Use the complete pivoting algorithm to repeat Exercise 9 Maple with *Digits*:= 10.
34. Use the complete pivoting algorithm to repeat Exercise 10 Maple with *Digits*:= 10.

6.3 Linear Algebra and Matrix Inversion

Matrices were introduced in Section 6.1 as a convenient method for expressing and manipulating linear systems. In this section we consider some algebra associated with matrices and show how it can be used to solve problems involving linear systems.

Definition 6.2 Two matrices A and B are **equal** if they have the same number of rows and columns, say $n \times m$, and if $a_{ij} = b_{ij}$, for each $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. ■

This definition means, for example, that

$$\begin{bmatrix} 2 & -1 & 7 \\ 3 & 1 & 0 \end{bmatrix} \neq \begin{bmatrix} 2 & 3 \\ -1 & 1 \\ 7 & 0 \end{bmatrix},$$

because they differ in dimension.

Matrix Arithmetic

Two important operations performed on matrices are the sum of two matrices and the multiplication of a matrix by a real number.

Definition 6.3 If A and B are both $n \times m$ matrices, then the **sum** of A and B, denoted $A + B$, is the $n \times m$ matrix whose entries are $a_{ij} + b_{ij}$, for each $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. ■

Definition 6.4 If A is an $n \times m$ matrix and λ is a real number, then the **scalar multiplication** of λ and A, denoted λA, is the $n \times m$ matrix whose entries are λa_{ij}, for each $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. ■

Example 1 Determine $A + B$ and λA when

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

Solution We have

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

and

$$\lambda A = \begin{bmatrix} -2(2) & -2(-1) & -2(7) \\ -2(3) & -2(1) & -2(0) \end{bmatrix} = \begin{bmatrix} -4 & 2 & -14 \\ -6 & -2 & 0 \end{bmatrix}.$$ ■

We have the following general properties for matrix addition and scalar multiplication. These properties are sufficient to classify the set of all $n \times m$ matrices with real entries as a **vector space** over the field of real numbers.

- We let O denote a matrix all of whose entries are 0 and $-A$ denote the matrix whose entries are $-a_{ij}$.

Theorem 6.5 Let A, B, and C be $n \times m$ matrices and λ and μ be real numbers. The following properties of addition and scalar multiplication hold:

(i) $A + B = B + A$, **(ii)** $(A + B) + C = A + (B + C)$,

(iii) $A + O = O + A = A$, **(iv)** $A + (-A) = -A + A = 0$,

(v) $\lambda(A + B) = \lambda A + \lambda B$, **(vi)** $(\lambda + \mu)A = \lambda A + \mu A$,

(vii) $\lambda(\mu A) = (\lambda\mu)A$, **(viii)** $1A = A$.

All these properties follow from similar results concerning the real numbers. ■

Matrix-Vector Products

The product of matrices can also be defined in certain instances. We will first consider the product of an $n \times m$ matrix and a $m \times 1$ column vector.

Definition 6.6 Let A be an $n \times m$ matrix and $\mathbf{b}$ an m-dimensional column vector. The **matrix-vector product** of A and $\mathbf{b}$, denoted $A\mathbf{b}$, is an n-dimensional column vector given by

$$A\mathbf{b} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} a_{1i}b_i \\ \sum_{i=1}^{m} a_{2i}b_i \\ \vdots \\ \sum_{i=1}^{m} a_{ni}b_i \end{bmatrix}.$$ ■

For this product to be defined the number of columns of the matrix A must match the number of rows of the vector $\mathbf{b}$, and the result is another column vector with the number of rows matching the number of rows in the matrix.

Example 2 Determine the product $A\mathbf{b}$ if $A = \begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 6 & 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

Solution Because A has dimension 3×2 and $\mathbf{b}$ has dimension 2×1, the product is defined and is a vector with three rows. These are

$$3(3) + 2(-1) = 7, \quad (-1)(3) + 1(-1) = -4, \quad \text{and} \quad 6(3) + 4(-1) = 14.$$

That is,

$$A\mathbf{b} = \begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 7 \\ -4 \\ 14 \end{bmatrix}$$ ■

The introduction of the matrix-vector product permits us to view the linear system

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

as the matrix equation

$$A\mathbf{x} = \mathbf{b},$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

because all the entries in the product $A\mathbf{x}$ must match the corresponding entries in the vector $\mathbf{b}$. In essence, then, an $n \times m$ matrix is a function with domain the set of m-dimensional column vectors and range a subset of the n-dimensional column vectors.

Matrix-Matrix Products

We can use this matrix-vector multiplication to define general matrix-matrix multiplication.

Definition 6.7 Let A be an $n \times m$ matrix and B an $m \times p$ matrix. The **matrix product** of A and B, denoted AB, is an $n \times p$ matrix C whose entries c_{ij} are

$$c_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj},$$

for each $i = 1, 2, \cdots n$, and $j = 1, 2, \cdots, p$. ■

The computation of c_{ij} can be viewed as the multiplication of the entries of the ith row of A with corresponding entries in the jth column of B, followed by a summation; that is,

$$[a_{i1}, a_{i2}, \cdots, a_{im}] \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{bmatrix} = c_{ij},$$

where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj} = \sum_{k=1}^{m} a_{ik}b_{kj}.$$

This explains why the number of columns of A must equal the number of rows of B for the product AB to be defined.

The following example should serve to clarify the matrix multiplication process.

Example 3 Determine all possible products of the matrices

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

$$C = \begin{bmatrix} 2 & 1 & 0 & 1 \\ -1 & 3 & 2 & 1 \\ 1 & 1 & 2 & 0 \end{bmatrix}, \quad \text{and} \quad D = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix}.$$

Solution The size of the matrices are

$$A : 3 \times 2, \quad B : 2 \times 3, \quad C : 3 \times 4, \quad \text{and} \quad D : 2 \times 2.$$

The products that can be defined, and their dimensions, are:

$$AB : 3 \times 3, \quad BA : 2 \times 2, \quad AD : 3 \times 2, \quad BC : 2 \times 4, \quad DB : 2 \times 3, \quad \text{and} \quad DD : 2 \times 2.$$

These products are

$$AB = \begin{bmatrix} 12 & 5 & 1 \\ 1 & 0 & 3 \\ 14 & 5 & 7 \end{bmatrix}, \quad BA = \begin{bmatrix} 4 & 1 \\ 10 & 15 \end{bmatrix}, \quad AD = \begin{bmatrix} 7 & -5 \\ 1 & 0 \\ 9 & -5 \end{bmatrix},$$

$$BC = \begin{bmatrix} 2 & 4 & 0 & 3 \\ 7 & 8 & 6 & 4 \end{bmatrix}, \quad DB = \begin{bmatrix} -1 & 0 & -3 \\ 1 & 1 & -4 \end{bmatrix}, \quad \text{and} \quad DD = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

■

Notice that although the matrix products AB and BA are both defined, their results are very different; they do not even have the same dimension. In mathematical language, we say that the matrix product operation is *not commutative*, that is, products in reverse order can differ. This is the case even when both products are defined and are of the same dimension. Almost any example will show this, for example,

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad \text{whereas} \quad \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$

Certain important operations involving matrix product do hold, however, as indicated in the following result.

Theorem 6.8 Let A be an $n \times m$ matrix, B be an $m \times k$ matrix, C be a $k \times p$ matrix, D be an $m \times k$ matrix, and λ be a real number. The following properties hold:

(a) $A(BC) = (AB)C$; **(b)** $A(B + D) = AB + AD$; **(c)** $\lambda(AB) = (\lambda A)B = A(\lambda B)$. ■

Proof The verification of the property in part **(a)** is presented to show the method involved. The other parts can be shown in a similar manner.

To show that $A(BC) = (AB)C$, compute the sj-entry of each side of the equation. BC is an $m \times p$ matrix with sj-entry

$$(BC)_{sj} = \sum_{l=1}^{k} b_{sl} c_{lj}.$$

Thus, $A(BC)$ is an $n \times p$ matrix with entries

$$[A(BC)]_{ij} = \sum_{s=1}^{m} a_{is}(BC)_{sj} = \sum_{s=1}^{m} a_{is}\left(\sum_{l=1}^{k} b_{sl} c_{lj}\right) = \sum_{s=1}^{m}\sum_{l=1}^{k} a_{is} b_{sl} c_{lj}.$$

Similarly, AB is an $n \times k$ matrix with entries

$$(AB)_{il} = \sum_{s=1}^{m} a_{is} b_{sl},$$

so $(AB)C$ is an $n \times p$ matrix with entries

$$[(AB)C]_{ij} = \sum_{l=1}^{k} (AB)_{il} c_{lj} = \sum_{l=1}^{k}\left(\sum_{s=1}^{m} a_{is} b_{sl}\right) c_{lj} = \sum_{l=1}^{k}\sum_{s=1}^{m} a_{is} b_{sl} c_{lj}.$$

Interchanging the order of summation on the right side gives

$$[(AB)C]_{ij} = \sum_{s=1}^{m}\sum_{l=1}^{k} a_{is} b_{sl} c_{lj} = [A(BC)]_{ij},$$

for each $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$. So $A(BC) = (AB)C$. ■ ■ ■

Square Matrices

Matrices that have the same number of rows as columns are important in applications.

Definition 6.9

The term diagonal applied to a matrix refers to the entries in the diagonal that runs from the top left entry to the bottom right entry.

(i) A **square** matrix has the same number of rows as columns.

(ii) A **diagonal** matrix $D = [d_{ij}]$ is a square matrix with $d_{ij} = 0$ whenever $i \neq j$.

(iii) The **identity matrix of order** n, $I_n = [\delta_{ij}]$, is a diagonal matrix whose diagonal entries are all 1s. When the size of I_n is clear, this matrix is generally written simply as I. ■

For example, the identity matrix of order three is

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

Definition 6.10 An **upper-triangular** $n \times n$ matrix $U = [u_{ij}]$ has, for each $j = 1, 2, \cdots, n$, the entries

$$u_{ij} = 0, \quad \text{for each } i = j+1, j+2, \cdots, n;$$

A triangular matrix is one that has all zero entries except either on and above (upper) or on and below (lower) the main diagonal.

and a **lower-triangular** matrix $L = [l_{ij}]$ has, for each $j = 1, 2, \cdots, n$, the entries

$$l_{ij} = 0, \quad \text{for each } i = 1, 2, \cdots, j-1.$$ ■

A diagonal matrix, then, is both both upper triangular and lower triangular because its only nonzero entries must lie on the main diagonal.

Illustration Consider the identity matrix of order three,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If A is any 3×3 matrix, then

$$AI_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = A.$$ □

The identity matrix I_n commutes with any $n \times n$ matrix A; that is, the order of multiplication does not matter,

$$I_n A = A = AI_n.$$

Keep in mind that this property is not true in general, even for square matrices.

Inverse Matrices

Related to the linear systems is the **inverse of a matrix**.

Definition 6.11 An $n \times n$ matrix A is said to be **nonsingular** (or *invertible*) if an $n \times n$ matrix A^{-1} exists with $AA^{-1} = A^{-1}A = I$. The matrix A^{-1} is called the **inverse** of A. A matrix without an inverse is called **singular** (or *noninvertible*). ■

The word singular means something that deviates from the ordinary. Hence a singular matrix does *not* have an inverse.

The following properties regarding matrix inverses follow from Definition 6.11. The proofs of these results are considered in Exercise 5.

Theorem 6.12 For any nonsingular $n \times n$ matrix A:

(i) A^{-1} is unique.

(ii) A^{-1} is nonsingular and $(A^{-1})^{-1} = A$.

(iii) If B is also a nonsingular $n \times n$ matrix, then $(AB)^{-1} = B^{-1}A^{-1}$. ■

Example 4 Let

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Show that $B = A^{-1}$, and that the solution to the linear system described by

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 2, \\ 2x_1 + x_2 \phantom{{}+2x_3} &= 3, \\ -x_1 + x_2 + 2x_3 &= 4. \end{aligned}$$

is given by the entries in $B\mathbf{b}$, where $\mathbf{b}$ is the column vector with entries 2, 3, and 4.

Solution First note that

$$AB = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3.$$

In a similar manner, $BA = I_3$, so A and B are both nonsingular with $B = A^{-1}$ and $A = B^{-1}$.

Now convert the given linear system to the matrix equation

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

and multiply both sides by B, the inverse of A. Because we have both

$$B(A\mathbf{x}) = (BA)\mathbf{x} = I_3\mathbf{x} = \mathbf{x} \quad \text{and} \quad B(A\mathbf{x}) = \mathbf{b},$$

we have

$$BA\mathbf{x} = \left(\begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{3}{9} & \frac{3}{9} & \frac{3}{9} \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \right) \mathbf{x} = \mathbf{x}$$

and

$$BA\mathbf{x} = B(\mathbf{b}) = \begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{7}{9} \\ \frac{13}{9} \\ \frac{5}{3} \end{bmatrix}$$

This implies that $\mathbf{x} = B\mathbf{b}$ and gives the solution $x_1 = 7/9$, $x_2 = 13/9$, and $x_3 = 5/3$. ■

Although it is easy to solve a linear system of the form $A\mathbf{x} = \mathbf{b}$ if A^{-1} is known, it is not computationally efficient to determine A^{-1} in order to solve the system. (See

Exercise 8.) Even so, it is useful from a conceptual standpoint to describe a method for determining the inverse of a matrix.

To find a method of computing A^{-1} assuming A is nonsingular, let us look again at matrix multiplication. Let B_j be the jth column of the $n \times n$ matrix B,

$$B_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}.$$

If $AB = C$, then the jth column of C is given by the product

$$\begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{nj} \end{bmatrix} = C_j = AB_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}b_{kj} \\ \sum_{k=1}^{n} a_{2k}b_{kj} \\ \vdots \\ \sum_{k=1}^{n} a_{nk}b_{kj} \end{bmatrix}.$$

Suppose that A^{-1} exists and that $A^{-1} = B = (b_{ij})$. Then $AB = I$ and

$$AB_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \text{where the value 1 appears in the } j\text{th row.}$$

To find B we need to solve n linear systems in which the jth column of the inverse is the solution of the linear system with right-hand side the jth column of I. The next illustration demonstrates this method.

Illustration To determine the inverse of the matrix

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

let us first consider the product AB, where B is an arbitrary 3×3 matrix.

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \\ &= \begin{bmatrix} b_{11} + 2b_{21} - b_{31} & b_{12} + 2b_{22} - b_{32} & b_{13} + 2b_{23} - b_{33} \\ 2b_{11} + b_{21} & 2b_{12} + b_{22} & 2b_{13} + b_{23} \\ -b_{11} + b_{21} + 2b_{31} & -b_{12} + b_{22} + 2b_{32} & -b_{13} + b_{23} + 2b_{33} \end{bmatrix}. \end{aligned}$$

If $B = A^{-1}$, then $AB = I$, so

$$\begin{aligned} b_{11} + 2b_{21} - b_{31} &= 1, & b_{12} + 2b_{22} - b_{32} &= 0, & & b_{13} + 2b_{23} - b_{33} = 0, \\ 2b_{11} + b_{21} &= 0, & 2b_{12} + b_{22} &= 1, & \text{and} \quad & 2b_{13} + b_{23} = 0, \\ -b_{11} + b_{21} + 2b_{31} &= 0, & -b_{12} + b_{22} + 2b_{32} &= 0, & & -b_{13} + b_{23} + 2b_{33} = 1. \end{aligned}$$

Notice that the coefficients in each of the systems of equations are the same, the only change in the systems occurs on the right side of the equations. As a consequence, Gaussian elimination can be performed on a larger augmented matrix formed by combining the matrices for each of the systems:

$$\left[\begin{array}{rrrrr:r} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 2 & 0 & 0 & 1 \end{array}\right].$$

First, performing $(E_2-2E_1) \to (E_2)$ and $(E_3+E_1) \to (E_3)$, followed by $(E_3+E_2) \to (E_3)$ produces

$$\left[\begin{array}{rrrrr:r} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 3 & 1 & 1 & 0 & 1 \end{array}\right] \quad \text{and} \quad \left[\begin{array}{rrrrr:r} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array}\right].$$

Backward substitution is performed on each of the three augmented matrices,

$$\left[\begin{array}{rrr:r} 1 & 2 & -1 & 1 \\ 0 & -3 & 2 & -2 \\ 0 & 0 & 3 & -1 \end{array}\right], \left[\begin{array}{rrr:r} 1 & 2 & -1 & 0 \\ 0 & -3 & 2 & 1 \\ 0 & 0 & 3 & 1 \end{array}\right], \left[\begin{array}{rrr:r} 1 & 2 & -1 & 0 \\ 0 & -3 & 2 & 0 \\ 0 & 0 & 3 & 1 \end{array}\right],$$

to eventually give

$$\begin{array}{llll} b_{11} = -\frac{2}{9}, & b_{12} = \frac{5}{9}, & & b_{13} = -\frac{1}{9}, \\ b_{21} = \frac{4}{9}, & b_{22} = -\frac{1}{9}, & \text{and} & b_{23} = \frac{2}{9}, \\ b_{31} = -\frac{1}{3}, & b_{32} = \frac{1}{3}, & & b_{32} = \frac{1}{3}. \end{array}$$

As shown in Example 4, these are the entries of A^{-1}:

$$B = A^{-1} = \begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$ □

As we saw in the illustration, in order to compute A^{-1} it is convenient to set up a larger augmented matrix,

$$[\, A \;\vdots\; I \,].$$

Upon performing the elimination in accordance with Algorithm 6.1, we obtain an augmented matrix of the form

$$[\, U \;\vdots\; Y \,],$$

where U is an upper-triangular matrix and Y is the matrix obtained by performing the same operations on the identity I that were performed to take A into U.

Gaussian elimination with backward substitution requires

$$\frac{4}{3}n^3 - \frac{1}{3}n \text{ multiplications/divisions} \quad \text{and} \quad \frac{4}{3}n^3 - \frac{3}{2}n^2 + \frac{n}{6} \text{ additions/subtractions.}$$

to solve the n linear systems (see Exercise 8(a)). Special care can be taken in the implementation to note the operations that need *not* be performed, as, for example, a multiplication when one of the multipliers is known to be unity or a subtraction when the subtrahend is known to be 0. The number of multiplications/divisions required can then be reduced to n^3 and the number of additions/subtractions reduced to $n^3 - 2n^2 + n$ (see Exercise 8(d)).

Transpose of a Matrix

Another important matrix associated with a given matrix A is its *transpose*, denoted A^t.

Definition 6.13 The **transpose** of an $n \times m$ matrix $A = [a_{ij}]$ is the $m \times n$ matrix $A^t = [a_{ji}]$, where for each i, the ith column of A^t is the same as the ith row of A. A square matrix A is called **symmetric** if $A = A^t$. ■

Illustration The matrices

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

have transposes

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

The matrix C is symmetric because $C^t = C$. The matrices A and B are not symmetric. □

The proof of the next result follows directly from the definition of the transpose.

Theorem 6.14 The following operations involving the transpose of a matrix hold whenever the operation is possible:

(i) $(A^t)^t = A$,

(ii) $(A + B)^t = A^t + B^t$,

(iii) $(AB)^t = B^t A^t$,

(iv) if A^{-1} exists, then $(A^{-1})^t = (A^t)^{-1}$. ■

Matrix arithmetic is performed in Maple using the *LinearAlgebra* package whenever the operations are defined. For example, the addition of two $n \times m$ matrices A and B is done in Maple with the command $A + B$, and scalar multiplication by a number c is defined by $c\,A$.

If A is $n \times m$ and B is $m \times p$, then the $n \times p$ matrix AB is produced with the command $A.B$. Matrix transposition is achieved with *Transpose*(A) and matrix inversion, with *MatrixInverse*(A).

EXERCISE SET 6.3

1. Perform the following matrix-vector multiplications:

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

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

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

d. $[-4 \quad 0 \quad 0]\begin{bmatrix} 1 & -2 & 4 \\ -2 & 3 & 1 \\ 4 & 1 & 0 \end{bmatrix}$

2. Perform the following matrix-vector multiplications:

a. $\begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ **b.** $\begin{bmatrix} 3 & 2 \\ 6 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

c. $\begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & 2 \\ 0 & 2 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$ **d.** $[2 \quad -2 \quad 1]\begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & -2 \end{bmatrix}$

3. Perform the following matrix-matrix multiplications:

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

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

4. Perform the following matrix-matrix multiplications:

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

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

5. Determine which of the following matrices are nonsingular, and compute the inverse of these matrices:

a. $\begin{bmatrix} 4 & 2 & 6 \\ 3 & 0 & 7 \\ -2 & -1 & -3 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & -1 \\ 3 & 1 & 1 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 2 & -4 & -2 \\ 2 & 1 & 1 & 5 \\ -1 & 0 & -2 & -4 \end{bmatrix}$ **d.** $\begin{bmatrix} 4 & 0 & 0 & 0 \\ 6 & 7 & 0 & 0 \\ 9 & 11 & 1 & 0 \\ 5 & 4 & 1 & 1 \end{bmatrix}$

6. Determine which of the following matrices are nonsingular, and compute the inverse of these matrices:

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

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

7. Given the two 4×4 linear systems having the same coefficient matrix:

$$\begin{aligned} x_1 - x_2 + 2x_3 - x_4 &= 6, \\ x_1 \qquad - x_3 + x_4 &= 4, \\ 2x_1 + x_2 + 3x_3 - 4x_4 &= -2, \\ -x_2 + x_3 - x_4 &= 5; \end{aligned} \qquad \begin{aligned} x_1 - x_2 + 2x_3 - x_4 &= 1, \\ x_1 - x_3 + x_4 &= 1, \\ 2x_1 + x_2 + 3x_3 - 4x_4 &= 2, \\ -x_2 + x_3 - x_4 &= -1. \end{aligned}$$

a. Solve the linear systems by applying Gaussian elimination to the augmented matrix

$$\left[\begin{array}{rrrr:rr} 1 & -1 & 2 & -1 & 6 & 1 \\ 1 & 0 & -1 & 1 & 4 & 1 \\ 2 & 1 & 3 & -4 & -2 & 2 \\ 0 & -1 & 1 & -1 & 5 & -1 \end{array}\right].$$

b. Solve the linear systems by finding and multiplying by the inverse of

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

c. Which method requires more operations?

8. Consider the four 3×3 linear systems having the same coefficient matrix:

$$\begin{aligned} 2x_1 - 3x_2 + x_3 &= 2, \\ x_1 + x_2 - x_3 &= -1, \\ -x_1 + x_2 - 3x_3 &= 0; \end{aligned} \qquad \begin{aligned} 2x_1 - 3x_2 + x_3 &= 6, \\ x_1 + x_2 - x_3 &= 4, \\ -x_1 + x_2 - 3x_3 &= 5; \end{aligned}$$

$$\begin{aligned} 2x_1 - 3x_2 + x_3 &= 0, \\ x_1 + x_2 - x_3 &= 1, \\ -x_1 + x_2 - 3x_3 &= -3; \end{aligned} \qquad \begin{aligned} 2x_1 - 3x_2 + x_3 &= -1, \\ x_1 + x_2 - x_3 &= 0, \\ -x_1 + x_2 - 3x_3 &= 0. \end{aligned}$$

a. Solve the linear systems by applying Gaussian elimination to the augmented matrix

$$\left[\begin{array}{ccc|cccc} 2 & -3 & 1 & 2 & 6 & 0 & -1 \\ 1 & 1 & -1 & -1 & 4 & 1 & 0 \\ -1 & 1 & -3 & 0 & 5 & -3 & 0 \end{array}\right].$$

b. Solve the linear systems by finding and multiplying by the inverse of

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

c. Which method requires more operations?

9. The following statements are needed to prove Theorem 6.12.

a. Show that if A^{-1} exists, it is unique.

b. Show that if A is nonsingular, then $(A^{-1})^{-1} = A$.

c. Show that if A and B are nonsingular $n{\times}n$ matrices, then $(AB)^{-1} = B^{-1}A^{-1}$.

10. Prove the following statements or provide counterexamples to show they are not true.

a. The product of two symmetric matrices is symmetric.

b. The inverse of a nonsingular symmetric matrix is a nonsingular symmetric matrix.

c. If A and B are $n \times n$ matrices, then $(AB)^t = A^tB^t$.

11. a. Show that the product of two $n \times n$ lower triangular matrices is lower triangular.

b. Show that the product of two $n \times n$ upper triangular matrices is upper triangular.

c. Show that the inverse of a nonsingular $n \times n$ lower triangular matrix is lower triangular.

12. Suppose m linear systems

$$A\mathbf{x}^{(p)} = \mathbf{b}^{(p)}, \quad p = 1, 2, \ldots, m,$$

are to be solved, each with the $n \times n$ coefficient matrix A.

a. Show that Gaussian elimination with backward substitution applied to the aug- mented matrix

$$[A : \quad \mathbf{b}^{(1)}\mathbf{b}^{(2)} \cdots \mathbf{b}^{(m)}]$$

requires

$$\frac{1}{3}n^3 + mn^2 - \frac{1}{3}n \quad \text{multiplications/ divisions}$$

and

$$\frac{1}{3}n^3 + mn^2 - \frac{1}{2}n^2 - mn + \frac{1}{6}n \quad \text{additions/subtractions.}$$

b. Show that the Gauss-Jordan method (see Exercise 12, Section 6.1) applied to the augmented matrix

$$[A: \quad \mathbf{b}^{(1)}\mathbf{b}^{(2)} \cdots \mathbf{b}^{(m)}]$$

requires

$$\frac{1}{2}n^3 + mn^2 - \frac{1}{2}n \quad \text{multiplications/divisions}$$

and

$$\frac{1}{2}n^3 + (m-1)n^2 + \left(\frac{1}{2} - m\right)n \quad \text{additions/subtractions.}$$

c. For the special case

$$\mathbf{b}^{(p)} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \\ \\ \\ \leftarrow p\text{th row,} \\ \\ \\ \end{array}$$

for each $p = 1, \ldots, m$, with $m = n$, the solution $x^{(p)}$ is the pth column of A^{-1}. Show that Gaussian elimination with backward substitution requires

$$\frac{4}{3}n^3 - \frac{1}{3}n \quad \text{multiplications/divisions}$$

and

$$\frac{4}{3}n^3 - \frac{3}{2}n^2 + \frac{1}{6}n \quad \text{additions/subtractions}$$

for this application, and that the Gauss-Jordan method requires

$$\frac{3}{2}n^3 - \frac{1}{2}n \quad \text{multiplications/divisions}$$

and

$$\frac{3}{2}n^3 - 2n^2 + \frac{1}{2}n \quad \text{additions/subtractions.}$$

d. Construct an algorithm using Gaussian elimination to find A^{-1}, but do not per- form multiplications when one of the multipliers is known to be 1, and do not per- form additions/subtractions when one of the elements involved is known to be 0. Show that the required computations are reduced to n^3 multiplications/divisions and $n^3 - 2n^2 + n$ additions/subtractions.

e. Show that solving the linear system $Ax = \mathbf{b}$, when A^{-1} is known, still requires n^2 multiplications/divisions and $n^2 - n$ additions/subtractions.

f. Show that solving m linear systems $Ax^{(p)} = \mathbf{b}^{(p)}$, for $p = 1, 2, \ldots, m$, by the method $x^{(p)} = A^{-1}\mathbf{b}(p)$ requires mn^2 multiplications and $m(n^2 - n)$ additions, if A^{-1} is known.

g. Let A be an $n \times n$ matrix. Compare the number of operations required to solve n linear systems involving A by Gaussian elimination with backward substitution and by first inverting A and then multiplying $Ax = \mathbf{b}$ by A^{-1}, for $n = 3, 10, 50, 100$. Is it ever advantageous to compute A^{-1} for the purpose of solving linear systems?

13. Use the algorithm developed in Exercise 8(d) to find the inverses of the nonsingular matrices in Exercise 1.

14. It is often useful to partition matrices into a collection of submatrices. For example, the matrices

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

can be partitioned into

$$\left[\begin{array}{cc:c} 1 & 2 & -1 \\ 3 & -4 & -3 \\ \hdashline 6 & 5 & 0 \end{array}\right] = \left[\begin{array}{c:c} A_{11} & A_{12} \\ \hdashline A_{21} & A_{22} \end{array}\right] \quad \text{and} \quad \left[\begin{array}{ccc:c} 2 & -1 & 7 & 0 \\ 3 & 0 & 4 & 5 \\ \hdashline -2 & 1 & -3 & 1 \end{array}\right] = \left[\begin{array}{c:c} B_{11} & B_{12} \\ \hdashline B_{21} & B_{22} \end{array}\right]$$

a. Show that the product of A and B in this case is

$$AB = \left[\begin{array}{c:c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hdashline A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array}\right]$$

b. If B were instead partitioned into

$$B = \left[\begin{array}{ccc:c} 2 & -1 & 7 & 0 \\ \hdashline 3 & 0 & 4 & 5 \\ -2 & 1 & -3 & 1 \end{array}\right] = \left[\begin{array}{c:c} B_{11} & B_{12} \\ \hdashline B_{21} & B_{22} \end{array}\right],$$

would the result in part (a) hold?

c. Make a conjecture concerning the conditions necessary for the result in part (a) to hold in the general case.

15. In a paper entitled "Population Waves," Bernadelli [Ber] (see also [Se]) hypothesizes a type of simplified beetle that has a natural life span of 3 years. The female of this species has a survival rate of $\frac{1}{2}$ in the first year of life, has a survival rate of $\frac{1}{3}$ from the second to third years, and gives birth to an average of six new females before expiring at the end of the third year. A matrix can be used to show the contribution an individual female beetle makes, in a probabilistic sense, to the female population of the species by letting a_{ij} in the matrix $A = [a_{ij}]$ denote the contribution that a single female beetle of age j will make to the next year's female population of age i; that is,

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

a. The contribution that a female beetle makes to the population 2 years hence is determined from the entries of A^2, of 3 years hence from A^3, and so on. Construct A^2 and A^3, and try to make a general statement about the contribution of a female beetle to the population in n years' time for any positive integral value of n.

b. Use your conclusions from part (a) to describe what will occur in future years to a population of these beetles that initially consists of 6000 female beetles in each of the three age groups.

c. Construct A^{-1}, and describe its significance regarding the population of this species.

16. The study of food chains is an important topic in the determination of the spread and accumulation of environmental pollutants in living matter. Suppose that a food chain has three links. The first link consists of vegetation of types $v_1, v_2, \ldots, v_n$, which provide all the food requirements for herbivores of species $h_1, h_2, \ldots, h_m$ in the second link. The third link consists of carnivorous animals $c_1, c_2, \ldots, c_k$, which depend entirely on the herbivores in the second link for their food supply. The coordinate a_{ij} of the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

represents the total number of plants of type v_i eaten by the herbivores in the species h_j, whereas b_{ij} in

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{bmatrix}$$

describes the number of herbivores in species h_i that are devoured by the animals of type c_j.

a. Show that the number of plants of type v_i that eventually end up in the animals of species c_j is given by the entry in the ith row and jth column of the matrix AB.

b. What physical significance is associated with the matrices $A-1, B-1$, and $(AB)-1 = B-1A-1$?

17. In Section 3.6 we found that the parametric form $(x(t), y(t))$ of the cubic Hermite polynomials through $(x(0), y(0)) = (x_0, y_0)$ and $(x(1), y(1)) = (x_1, y_1)$ with guide points $(x_0+\alpha_0, y_0+\beta_0)$ and $(x_1-\alpha_1, y_1-\beta_1)$, respectively, are given by

$$x(t) = (2(x_0 - x_1) + (\alpha_0 + \alpha_1))t^3 + (3(x_1 - x_0) - \alpha_1 - 2\alpha_0)t^2 + \alpha_0 t + x_0,$$

and

$$y(t) = (2(y_0 - y_1) + (\beta_0 + \beta_1))t^3 + (3(y_1 - y_0) - \beta_1 - 2\beta_0)t^2 + \beta_0 t + y_0.$$

The Bézier cubic polynomials have the form

$$\hat{x}(t) = (2(x_0 - x_1) + 3(\alpha_0 + \alpha_1))t^3 + (3(x_1 - x_0) - 3(\alpha_1 + 2\alpha_0))t^2, +3\alpha_0 t + x_0$$

and

$$\hat{y}(t) = (2(y_0 - y_1) + 3(\beta_0 + \beta_1))t^3 + (3(y_1 - y_0) - 3(\beta_1 + 2\beta_0))t^2 + 3\beta_0 t + y_0.$$

a. Show that the matrix

$$A = \begin{bmatrix} 7 & 4 & 4 & 0 \\ -6 & -3 & -6 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

transforms the Hermite polynomial coefficients into the Bézier polynomial coefficients.

b. Determine a matrix B that transforms the Bézier polynomial coefficients into the Hermite polynomial coefficients.

18. Consider the 2×2 linear system $(A + iB)(x + iy) = c + i\mathbf{d}$ with complex entries in component form:

$$(a_{11} + ib_{11})(x_1 + iy_1) + (a_{12} + ib_{12})(x_2 + iy_2) = c_1 + id_1,$$
$$(a_{11} + ib_{21})(x_1 + iy_1) + (a_{22} + ib_{22})(x_2 + iy_2) = c_2 + id_2.$$

a. Use the properties of complex numbers to convert this system to the equivalent 4×4 real linear system

$$A\mathbf{x} - B\mathbf{y} = \mathbf{c},$$
$$B\mathbf{x} + A\mathbf{y} = \mathbf{d}.$$

b. Solve the linear system

$$(1 - 2i)(x_1 + iy_1) + (3 + 2i)(x_2 + iy_2) = 5 + 2i,$$
$$(2 + i)(x_1 + iy_1) + (4 + 3i)(x_2 + iy_2) = 4 - i.$$

6.4 The Determinant of a Matrix

The *determinant* of a matrix provides existence and uniqueness results for linear systems having the same number of equations and unknowns. We will denote the determinant of a square matrix A by $\det A$, but it is also common to use the notation $|A|$.

Definition 6.15 Suppose that A is a square matrix.

(i) If $A = [a]$ is a 1×1 matrix, then $\det A = a$.

(ii) If A is an $n \times n$ matrix, with $n > 1$ the **minor** M_{ij} is the determinant of the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of the matrix A.

(iii) The **cofactor** A_{ij} associated with M_{ij} is defined by $A_{ij} = (-1)^{i+j} M_{ij}$.

(iv) The **determinant** of the $n \times n$ matrix A, when $n > 1$, is given either by

$$\det A = \sum_{j=1}^{n} a_{ij} A_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}, \quad \text{for any } i = 1, 2, \cdots, n,$$

or by

$$\det A = \sum_{i=1}^{n} a_{ij} A_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}, \quad \text{for any } j = 1, 2, \cdots, n.$$ ■

The notion of a determinant appeared independently in 1683 both in Japan and Europe, although neither Takakazu Seki Kowa (1642–1708) nor Gottfried Leibniz (1646–1716) appear to have used the term determinant.

It can be shown (see Exercise 9) that to calculate the determinant of a general $n \times n$ matrix by this definition requires $O(n!)$ multiplications/divisions and additions/subtractions. Even for relatively small values of n, the number of calculations becomes unwieldy.

Although it appears that there are $2n$ different definitions of $\det A$, depending on which row or column is chosen, all definitions give the same numerical result. The flexibility in the definition is used in the following example. It is most convenient to compute $\det A$ across the row or down the column with the most zeros.

Example 1 Find the determinant of the matrix

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

using the row or column with the most zero entries.

Solution To compute $\det A$, it is easiest to use the fourth column:

$$\det A = a_{14}A_{14} + a_{24}A_{24} + a_{34}A_{34} + a_{44}A_{44} = 5A_{34} = -5M_{34}.$$

Eliminating the third row and the fourth column gives

$$\begin{aligned} \det A &= -5 \det \begin{bmatrix} 2 & -1 & 3 \\ 4 & -2 & 7 \\ 6 & -6 & 8 \end{bmatrix} \\ &= -5 \left\{ 2 \det \begin{bmatrix} -2 & 7 \\ -6 & 8 \end{bmatrix} - (-1) \det \begin{bmatrix} 4 & 7 \\ 6 & 8 \end{bmatrix} + 3 \det \begin{bmatrix} 4 & -2 \\ 6 & -6 \end{bmatrix} \right\} = -30. \end{aligned}$$

■

The determinant of an $n \times n$ matrix of can be computed in Maple with the *LinearAlgebra* package using the command *Determinant*(A).

The following properties are useful in relating linear systems and Gaussian elimination to determinants. These are proved in any standard linear algebra text.

Theorem 6.16 Suppose A is an $n \times n$ matrix:

(i) If any row or column of A has only zero entries, then $\det A = 0$.

(ii) If A has two rows or two columns the same, then $\det A = 0$.

(iii) If $\tilde{A}$ is obtained from A by the operation $(E_i) \leftrightarrow (E_j)$, with $i \neq j$, then $\det \tilde{A} = -\det\ A$.

(iv) If $\tilde{A}$ is obtained from A by the operation $(\lambda E_i) \to (E_i)$, then $\det \tilde{A} = \lambda \det A$.

(v) If $\tilde{A}$ is obtained from A by the operation $(E_i + \lambda E_j) \to (E_i)$ with $i \neq j$, then $\det \tilde{A} = \det A$.

(vi) If B is also an $n \times n$ matrix, then $\det AB = \det A \det B$.

(vii) $\det A^t = \det A$.

(viii) When A^{-1} exists, $\det A^{-1} = (\det A)^{-1}$.

(ix) If A is an upper triangular, lower triangular, or diagonal matrix, then $\det A = \prod_{i=1}^{n} a_{ii}$. ■

As part (ix) of Theorem 6.16 indicates, the determinant of a triangular matrix is simply the product of its diagonal elements. By employing the row operations given in parts (iii), f(iv), and (v) we can reduce a given square matrix to triangular form to find its determinant.

Example 2 Compute the determinant of the matrix

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

using parts (iii), (iv), and (v) of Theorem 6.16, doing the computations in Maple with the *LinearAlgebra* package.

Solution Matrix A is defined in Maple by

$A := Matrix([[2, 1, -1, 1], [1, 1, 0, 3], [-1, 2, 3, -1], [3, -1, -1, 2]])$

The sequence of operations in Table 6.2 produces the matrix

$$A8 = \begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 1 & 5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix}.$$

By part (ix), $\det A8 = -39$, so $\det A = 39$. ■

Table 6.2

Operation	Maple	Effect
$\frac{1}{2}E_1 \to E_1$	$A1 := RowOperation(A, 1, \frac{1}{2})$	$\det A1 = \frac{1}{2} \det A$
$E_2 - E_1 \to E_2$	$A2 := RowOperation(A1, [2, 1], -1)$	$\det A2 = \det A1 = \frac{1}{2} \det A$
$E_3 + E_1 \to E_3$	$A3 := RowOperation(A2, [3, 1], 1)$	$\det A3 = \det A2 = \frac{1}{2} \det A$
$E_4 - 3E_1 \to E_4$	$A4 := RowOperation(A3, [4, 1], -3)$	$\det A4 = \det A3 = \frac{1}{2} \det A$
$2E_2 \to E_2$	$A5 := RowOperation(A4, 2, 2)$	$\det A5 = 2 \det A4 = \det A$
$E_3 - \frac{5}{2}E_2 \to E_3$	$A6 := RowOperation(A5, [3, 2], -\frac{5}{2})$	$\det A6 = \det A5 = \det A$
$E_4 + \frac{5}{2}E_2 \to E_4$	$A7 := RowOperation(A6, [4, 2], \frac{5}{2})$	$\det A7 = \det A6 = \det A$
$E_3 \leftrightarrow E_4$	$A8 := RowOperation(A7, [3, 4])$	$\det A8 = -\det A7 = -\det A$

The key result relating nonsingularity, Gaussian elimination, linear systems, and determinants is that the following statements are equivalent.

Theorem 6.17 The following statements are equivalent for any $n \times n$ matrix A:

(i) The equation $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$.

(ii) The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for any n-dimensional column vector $\mathbf{b}$.

(iii) The matrix A is nonsingular; that is, A^{-1} exists.

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

(v) Gaussian elimination with row interchanges can be performed on the system $A\mathbf{x} = \mathbf{b}$ for any n-dimensional column vector $\mathbf{b}$. ■

The following Corollary to Theorem 6.17 illustrates how the determinant can be used to show important properties about square matrices.

Corollary 6.18 Suppose that A and B are both $n \times n$ matrices with either $AB = I$ or $BA = I$. Then $B = A^{-1}$ (and $A = B^{-1}$). ■

Proof Suppose that $AB = I$. Then by part (vi) of Theorem 6.16,

$$1 = \det(I) = \det(AB) = \det(A) \cdot \det(B), \quad \text{so} \quad \det(A) \neq 0 \text{ and } \det(B) \neq 0.$$

The equivalence of parts (iii) and (iv) of Theorem 6.17 imply that both A^{-1} and B^{-1} exist. Hence

$$A^{-1} = A^{-1} \cdot I = A^{-1} \cdot (AB) = \left(A^{-1}A\right) \cdot B = I \cdot B = B.$$

The roles of A and B are similar, so this also establishes that $BA = I$. Hence $B = A^{-1}$. ■ ■ ■

EXERCISE SET 6.4

1. Use Definition 6.15 to compute the determinants of the following matrices:

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

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

2. Use Definition 6.15 to compute the determinants of the following matrices:

a. $\begin{bmatrix} 4 & 2 & 6 \\ -1 & 0 & 4 \\ 2 & 1 & 7 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 & 2 & 1 \\ 3 & 4 & -1 \\ 3 & 0 & 5 \end{bmatrix}$

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

3. Repeat Exercise 1 using the method of Example 2.
4. Repeat Exercise 2 using the method of Example 2.
5. Find all values of α that make the following matrix singular.

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

6. Find all values of α that make the following matrix singular.

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

7. Find all values of α so that the following linear system has no solutions.

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 5, \\ 4x_1 + 2x_2 + 2x_3 &= 6, \\ -2x_1 + \alpha x_2 + 3x_3 &= 4. \end{aligned}$$

8. Find all values of α so that the following linear system has an infinite number of solutions.

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 5, \\ 4x_1 + 2x_2 + 2x_3 &= 6, \\ -2x_1 + \alpha x_2 + 3x_3 &= 1. \end{aligned}$$

9. Use mathematical induction to show that when $n > 1$, the evaluation of the determinant of an $n \times n$ matrix using the definition requires

$$n! \sum_{k=1}^{n-1} \frac{1}{k!} \text{ multiplications/divisions} \quad \text{and} \quad n! - 1 \text{ additions/subtractions.}$$

10. Let A be a 3×3 matrix. Show that if $\tilde{A}$ is the matrix obtained from A using any of the operations

$$(E_1) \leftrightarrow (E_2), \quad (E_1) \leftrightarrow (E_3), \quad \text{or} \quad (E_2) \leftrightarrow (E_3),$$

then $\det \tilde{A} = -\det A$.

11. Prove that AB is nonsingular if and only if both A and B are nonsingular.

12. The solution by **Cramer's rule** to the linear system

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

has

$$x_1 = \frac{1}{D}\det\begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} \equiv \frac{D_1}{D}, \quad x_2 = \frac{1}{D}\det\begin{bmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{bmatrix} \equiv \frac{D_2}{D},$$

and

$$x_3 = \frac{1}{D}\det\begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{bmatrix} \equiv \frac{D_3}{D}, \quad \text{where} \quad D = \det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

a. Find the solution to the linear system

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 4, \\ x_1 - 2x_2 + x_3 &= 6, \\ x_1 - 12x_2 + 5x_3 &= 10, \end{aligned}$$

by Cramer's rule.

b. Show that the linear system

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 4, \\ x_1 - 2x_2 + x_3 &= 6, \\ -x_1 - 12x_2 + 5x_3 &= 9 \end{aligned}$$

does not have a solution. Compute D_1, D_2, and D_3.

c. Show that the linear system

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 4, \\ x_1 - 2x_2 + x_3 &= 6, \\ -x_1 - 12x_2 + 5x_3 &= 10 \end{aligned}$$

has an infinite number of solutions. Compute D_1, D_2, and D_3.

d. Prove that if a 3×3 linear system with $D = 0$ has solutions, then $D_1 = D_2 = D_3 = 0$.

e. Determine the number of multiplications/divisions and additions/subtractions required for Cramer's rule on a 3×3 system.

13. **a.** Generalize Cramer's rule to an $n \times n$ linear system.

b. Use the result in Exercise 9 to determine the number of multiplications/divisions and additions/subtractions required for Cramer's rule on an $n \times n$ system.

6.5 Matrix Factorization

Gaussian elimination is the principal tool in the direct solution of linear systems of equations, so it should be no surprise that it appears in other guises. In this section we will see that the steps used to solve a system of the form $A\mathbf{x} = \mathbf{b}$ can be used to factor a matrix. The factorization is particularly useful when it has the form $A = LU$, where L is lower triangular and U is upper triangular. Although not all matrices have this type of representation, many do that occur frequently in the application of numerical techniques.

In Section 6.1 we found that Gaussian elimination applied to an arbitrary linear system $A\mathbf{x} = \mathbf{b}$ requires $O(n^3/3)$ arithmetic operations to determine $\mathbf{x}$. However, to solve a linear system that involves an upper-triangular system requires only backward substitution, which takes $O(n^2)$ operations. The number of operations required to solve a lower-triangular systems is similar.

Suppose that A has been factored into the triangular form $A = LU$, where L is lower triangular and U is upper triangular. Then we can solve for $\mathbf{x}$ more easily by using a two-step process.

- First we let $\mathbf{y} = U\mathbf{x}$ and solve the lower triangular system $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$. Since L is triangular, determining $\mathbf{y}$ from this equation requires only $O(n^2)$ operations.

- Once $\mathbf{y}$ is known, the upper triangular system $U\mathbf{x} = \mathbf{y}$ requires only an additional $O(n^2)$ operations to determine the solution $\mathbf{x}$.

Solving a linear system $A\mathbf{x} = \mathbf{b}$ in factored form means that the number of operations needed to solve the system $A\mathbf{x} = \mathbf{b}$ is reduced from $O(n^3/3)$ to $O(2n^2)$.

Example 1 Compare the approximate number of operations required to determine the solution to a linear system using a technique requiring $O(n^3/3)$ operations and one requiring $O(2n^2)$ when $n = 20$, $n = 100$, and $n = 1000$.

Solution Table 6.3 gives the results of these calculations. ■

Table 6.3

n	$n^3/3$	$2n^2$	% Reduction
10	$3.\overline{3} \times 10^2$	2×10^2	40
100	$3.\overline{3} \times 10^5$	2×10^4	94
1000	$3.\overline{3} \times 10^8$	2×10^6	99.4

As the example illustrates, the reduction factor increases dramatically with the size of the matrix. Not surprisingly, the reductions from the factorization come at a cost; determining the specific matrices L and U requires $O(n^3/3)$ operations. But once the factorization is determined, systems involving the matrix A can be solved in this simplified manner for any number of vectors $\mathbf{b}$.

To see which matrices have an LU factorization and to find how it is determined, first suppose that Gaussian elimination can be performed on the system $A\mathbf{x} = \mathbf{b}$ without row interchanges. With the notation in Section 6.1, this is equivalent to having nonzero pivot elements $a_{ii}^{(i)}$, for each $i = 1, 2, \ldots, n$.

The first step in the Gaussian elimination process consists of performing, for each $j = 2, 3, \ldots, n$, the operations

$$(E_j - m_{j,1}E_1) \to (E_j), \quad \text{where} \quad m_{j,1} = \frac{a_{j1}^{(1)}}{a_{11}^{(1)}}. \tag{6.8}$$

These operations transform the system into one in which all the entries in the first column below the diagonal are zero.

The system of operations in (6.8) can be viewed in another way. It is simultaneously accomplished by multiplying the original matrix A on the left by the matrix

$$M^{(1)} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -m_{21} & 1 & \ddots & & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -m_{n1} & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Matrix factorization is another of the important techniques that Gauss seems to be the first to have discovered. It is included in his two-volume treatise on celestial mechanics *Theoria motus corporum coelestium in sectionibus conicis Solem ambientium*, which was published in 1809.

This is called the **first Gaussian transformation matrix**. We denote the product of this matrix with $A^{(1)} \equiv A$ by $A^{(2)}$ and with $\mathbf{b}$ by $\mathbf{b}^{(2)}$, so

$$A^{(2)}\mathbf{x} = M^{(1)}A\mathbf{x} = M^{(1)}\mathbf{b} = \mathbf{b}^{(2)}.$$

In a similar manner we construct $M^{(2)}$, the identity matrix with the entries below the diagonal in the second column replaced by the negatives of the multipliers

$$m_{j,2} = \frac{a_{j2}^{(2)}}{a_{22}^{(2)}}.$$

The product of this matrix with $A^{(2)}$ has zeros below the diagonal in the first two columns, and we let

$$A^{(3)}\mathbf{x} = M^{(2)}A^{(2)}\mathbf{x} = M^{(2)}M^{(1)}A\mathbf{x} = M^{(2)}M^{(1)}\mathbf{b} = \mathbf{b}^{(3)}.$$

In general, with $A^{(k)}\mathbf{x} = \mathbf{b}^{(k)}$ already formed, multiply by the ***k*th Gaussian transformation matrix**

$$M^{(k)} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & & \vdots \\ \vdots & & 0 & \ddots & \ddots & & & \vdots \\ \vdots & & \vdots & -m_{k+1,k} & \ddots & \ddots & & \vdots \\ \vdots & & \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -m_{n,k} & 0 & \cdots & 0 & 1 \end{bmatrix},$$

to obtain

$$A^{(k+1)}\mathbf{x} = M^{(k)}A^{(k)}\mathbf{x} = M^{(k)} \cdots M^{(1)}A\mathbf{x} = M^{(k)}\mathbf{b}^{(k)} = \mathbf{b}^{(k+1)} = M^{(k)} \cdots M^{(1)}\mathbf{b}. \quad (6.9)$$

The process ends with the formation of $A^{(n)}\mathbf{x} = \mathbf{b}^{(n)}$, where $A^{(n)}$ is the upper triangular matrix

$$A^{(n)} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n}^{(n-1)} \\ 0 & \cdots & 0 & a_{nn}^{(n)} \end{bmatrix},$$

given by

$$A^{(n)} = M^{(n-1)}M^{(n-2)}\cdots M^{(1)}A.$$

This process forms the $U = A^{(n)}$ portion of the matrix factorization $A = LU$. To determine the complementary lower triangular matrix L, first recall the multiplication of $A^{(k)}\mathbf{x} = \mathbf{b}^{(k)}$ by the Gaussian transformation of $M^{(k)}$ used to obtain (6.9):

$$A^{(k+1)}\mathbf{x} = M^{(k)}A^{(k)}\mathbf{x} = M^{(k)}\mathbf{b}^{(k)} = \mathbf{b}^{(k+1)},$$

where $M^{(k)}$ generates the row operations

$$(E_j - m_{j,k}E_k) \to (E_j), \quad \text{for } j = k+1, \ldots, n.$$

To reverse the effects of this transformation and return to $A^{(k)}$ requires that the operations $(E_j + m_{j,k}E_k) \to (E_j)$ be performed for each $j = k+1, \ldots, n$. This is equivalent to multiplying by the inverse of the matrix $M^{(k)}$, the matrix

$$L^{(k)} = [M^{(k)}]^{-1} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 0 & & & & \vdots \\ \vdots & & \vdots & m_{k+1,k} & & & \vdots \\ \vdots & & \vdots & \vdots & 0 & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & m_{n,k} & 0 & \cdots\, 0 & 1 \end{bmatrix}.$$

The lower-triangular matrix L in the factorization of A, then, is the product of the matrices $L^{(k)}$:

$$L = L^{(1)}L^{(2)}\cdots L^{(n-1)} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ m_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ m_{n1} & \cdots & m_{n,n-1} & 1 \end{bmatrix},$$

since the product of L with the upper-triangular matrix $U = M^{(n-1)}\cdots M^{(2)}M^{(1)}A$ gives

$$\begin{aligned} LU &= L^{(1)}L^{(2)}\cdots L^{(n-3)}L^{(n-2)}L^{(n-1)} \cdot M^{(n-1)}M^{(n-2)}M^{(n-3)}\cdots M^{(2)}M^{(1)}A \\ &= [M^{(1)}]^{-1}[M^{(2)}]^{-1}\cdots[M^{(n-2)}]^{-1}[M^{(n-1)}]^{-1} \cdot M^{(n-1)}M^{(n-2)}\cdots M^{(2)}M^{(1)}A = A. \end{aligned}$$

Theorem 6.19 follows from these observations.

Theorem 6.19 If Gaussian elimination can be performed on the linear system $A\mathbf{x} = \mathbf{b}$ without row interchanges, then the matrix A can be factored into the product of a lower-triangular matrix L and an upper-triangular matrix U, that is, $A = LU$, where $m_{ji} = a_{ji}^{(i)}/a_{ii}^{(i)}$,

$$U = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n}^{(n-1)} \\ 0 & \cdots & 0 & a_{nn}^{(n)} \end{bmatrix}, \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ m_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ m_{n1} & \cdots & m_{n,n-1} & 1 \end{bmatrix}.$$

■

Example 2 **(a)** Determine the LU factorization for matrix A in the linear system $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ -3 \\ 4 \end{bmatrix}.$$

(b) Then use the factorization to solve the system

$$\begin{aligned} x_1 + x_2 \phantom{{}-x_3} + 3x_4 &= 8, \\ 2x_1 + x_2 - x_3 + x_4 &= 7, \\ 3x_1 - x_2 - x_3 + 2x_4 &= 14, \\ -x_1 + 2x_2 + 3x_3 - x_4 &= -7. \end{aligned}$$

Solution **(a)** The original system was considered in Section 6.1, where we saw that the sequence of operations $(E_2 - 2E_1) \rightarrow (E_2)$, $(E_3 - 3E_1) \rightarrow (E_3)$, $(E_4 - (-1)E_1) \rightarrow (E_4)$, $(E_3 - 4E_2) \rightarrow (E_3)$, $(E_4 - (-3)E_2) \rightarrow (E_4)$ converts the system to the triangular system

$$\begin{aligned} x_1 + x_2 \phantom{{}-x_3} + 3x_4 &= 4, \\ -x_2 - x_3 - 5x_4 &= -7, \\ 3x_3 + 13x_4 &= 13, \\ -13x_4 &= -13. \end{aligned}$$

The multipliers m_{ij} and the upper triangular matrix produce the factorization

$$A = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} = LU.$$

(b) To solve

$$A\mathbf{x} = LU\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ 14 \\ -7 \end{bmatrix},$$

we first introduce the substitution $\mathbf{y} = U\mathbf{x}$. Then $\mathbf{b} = L(U\mathbf{x}) = L\mathbf{y}$. That is,

$$L\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ 14 \\ -7 \end{bmatrix}.$$

This system is solved for $\mathbf{y}$ by a simple forward-substitution process:

$$\begin{aligned} y_1 &= 8; \\ 2y_1 + y_2 &= 7, \quad \text{so } y_2 = 7 - 2y_1 = -9; \\ 3y_1 + 4y_2 + y_3 &= 14, \quad \text{so } y_3 = 14 - 3y_1 - 4y_2 = 26; \\ -y_1 - 3y_2 + y_4 &= -7, \quad \text{so } y_4 = -7 + y_1 + 3y_2 = -26. \end{aligned}$$

We then solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, the solution of the original system; that is,

$$\begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ -9 \\ 26 \\ -26 \end{bmatrix}.$$

Using backward substitution we obtain $x_4 = 2, x_3 = 0, x_2 = -1, x_1 = 3$. ■

The *NumericalAnalysis* subpackage of Maple can be used to perform the matrix factorization in Example 2. First load the package

$with(Student[NumericalAnalysis])$

and the matrix A

$A := Matrix([[1, 1, 0, 3], [2, 1, -1, 1], [3, -1, -1, 2], [-1, 2, 3, -1]])$

The factorization is performed with the command

$Lower, Upper := MatrixDecomposition(A, method = LU, output = ['L', 'U'])$

giving

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix}$$

To use the factorization to solve the system $A\mathbf{x} = \mathbf{b}$, define $\mathbf{b}$ by

$b := Vector([8, 7, 14, -7])$

Then perform the forward substitution to determine $\mathbf{y}$ with $U\mathbf{x} = y$, followed by backward substitution to determine $\mathbf{x}$ with $U\mathbf{x} = \mathbf{y}$.

$y := ForwardSubstitution(Lower, b)$: $x := BackSubstitution(Upper, y)$

The solution agrees with that in Example 2.

The factorization used in Example 2 is called *Doolittle's method* and requires that 1s be on the diagonal of L, which results in the factorization described in Theorem 6.19. In Section 6.6, we consider *Crout's method*, a factorization which requires that 1s be on the diagonal elements of U, and *Cholesky's method*, which requires that $l_{ii} = u_{ii}$, for each i.

A general procedure for factoring matrices into a product of triangular matrices is contained in Algorithm 6.4. Although new matrices L and U are constructed, the generated values can replace the corresponding entries of A that are no longer needed.

Algorithm 6.4 permits either the diagonal of L or the diagonal of U to be specified.

ALGORITHM 6.4

LU Factorization

To factor the $n \times n$ matrix $A = [a_{ij}]$ into the product of the lower-triangular matrix $L = [l_{ij}]$ and the upper-triangular matrix $U = [u_{ij}]$; that is, $A = LU$, where the main diagonal of either L or U consists of all ones:

INPUT dimension n; the entries a_{ij}, $1 \le i, j \le n$ of A; the diagonal $l_{11} = \cdots = l_{nn} = 1$ of L or the diagonal $u_{11} = \cdots = u_{nn} = 1$ of U.

OUTPUT the entries l_{ij}, $1 \le j \le i$, $1 \le i \le n$ of L and the entries, u_{ij}, $i \le j \le n$, $1 \le i \le n$ of U.

Step 1 Select l_{11} and u_{11} satisfying $l_{11}u_{11} = a_{11}$.
If $l_{11}u_{11} = 0$ then OUTPUT ('Factorization impossible');
STOP.

Step 2 For $j = 2, \ldots, n$ set $u_{1j} = a_{1j}/l_{11}$; (*First row of* U.)
$l_{j1} = a_{j1}/u_{11}$. (*First column of* L.)

Step 3 For $i = 2, \ldots, n-1$ do Steps 4 and 5.

Step 4 Select l_{ii} and u_{ii} satisfying $l_{ii}u_{ii} = a_{ii} - \sum_{k=1}^{i-1} l_{ik}u_{ki}$.
If $l_{ii}u_{ii} = 0$ then OUTPUT ('Factorization impossible');
STOP.

Step 5 For $j = i+1, \ldots, n$

set $u_{ij} = \frac{1}{l_{ii}}\left[a_{ij} - \sum_{k=1}^{i-1} l_{ik}u_{kj}\right]$; (*ith row of* U.)

$l_{ji} = \frac{1}{u_{ii}}\left[a_{ji} - \sum_{k=1}^{i-1} l_{jk}u_{ki}\right]$. (*ith column of* L.)

Step 6 Select l_{nn} and u_{nn} satisfying $l_{nn}u_{nn} = a_{nn} - \sum_{k=1}^{n-1} l_{nk}u_{kn}$.
(*Note: If* $l_{nn}u_{nn} = 0$, *then* $A = LU$ *but* A *is singular.*)

Step 7 OUTPUT (l_{ij} for $j = 1, \ldots, i$ and $i = 1, \ldots, n$);
OUTPUT (u_{ij} for $j = i, \ldots, n$ and $i = 1, \ldots, n$);
STOP. ■

Once the matrix factorization is complete, the solution to a linear system of the form $A\mathbf{x} = LU\mathbf{x} = \mathbf{b}$ is found by first letting $\mathbf{y} = U\mathbf{x}$ and solving $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$. Since L is lower triangular, we have

$$y_1 = \frac{b_1}{l_{11}},$$

and, for each $i = 2, 3, \ldots, n$,

$$y_i = \frac{1}{l_{ii}}\left[b_i - \sum_{j=1}^{i-1} l_{ij}y_j\right].$$

After $\mathbf{y}$ is found by this forward-substitution process, the upper-triangular system $U\mathbf{x} = \mathbf{y}$ is solved for $\mathbf{x}$ by backward substitution using the equations

$$x_n = \frac{y_n}{u_{nn}} \quad \text{and} \quad x_i = \frac{1}{u_{ii}}\left[y_i - \sum_{j=i+1}^{n} u_{ij}x_j\right].$$

Permutation Matrices

In the previous discussion we assumed that $A\mathbf{x} = \mathbf{b}$ can be solved using Gaussian elimination without row interchanges. From a practical standpoint, this factorization is useful only when row interchanges are not required to control the round-off error resulting from the use of finite-digit arithmetic. Fortunately, many systems we encounter when using approximation methods are of this type, but we will now consider the modifications that must be made when row interchanges are required. We begin the discussion with the introduction of a class of matrices that are used to rearrange, or permute, rows of a given matrix.

An $n \times n$ **permutation matrix** $P = [p_{ij}]$ is a matrix obtained by rearranging the rows of I_n, the identity matrix. This gives a matrix with precisely one nonzero entry in each row and in each column, and each nonzero entry is a 1.

Illustration The matrix

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

is a 3×3 permutation matrix. For any 3×3 matrix A, multiplying on the left by P has the effect of interchanging the second and third rows of A:

$$PA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}.$$

Similarly, multiplying A on the right by P interchanges the second and third columns of A. □

Two useful properties of permutation matrices relate to Gaussian elimination, the first of which is illustrated in the previous example. Suppose $k_1, \cdots, k_n$ is a permutation of the integers $1, \cdots, n$ and the permutation matrix $P = (p_{ij})$ is defined by

$$p_{ij} = \begin{cases} 1, & \text{if } j = k_i, \\ 0, & \text{otherwise.} \end{cases}$$

Then

The matrix multiplication AP permutes the columns of A.

- PA permutes the rows of A; that is,

$$PA = \begin{bmatrix} a_{k_1 1} & a_{k_1 2} & \cdots & a_{k_1 n} \\ a_{k_2 1} & a_{k_2 2} & \cdots & a_{k_2 n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k_n 1} & a_{k_n 2} & \cdots & a_{k_n n} \end{bmatrix}.$$

- P^{-1} exists and $P^{-1} = P^t$.

At the end of Section 6.4 we saw that for any nonsingular matrix A, the linear system $A\mathbf{x} = \mathbf{b}$ can be solved by Gaussian elimination, with the possibility of row interchanges. If we knew the row interchanges that were required to solve the system by Gaussian elimination, we could arrange the original equations in an order that would ensure that no row interchanges are needed. Hence there *is* a rearrangement of the equations in the system that permits Gaussian elimination to proceed *without* row interchanges. This implies that for any nonsingular matrix A, a permutation matrix P exists for which the system

$$PA\mathbf{x} = P\mathbf{b}$$

can be solved without row interchanges. As a consequence, this matrix PA can be factored into

$$PA = LU,$$

where L is lower triangular and U is upper triangular. Because $P^{-1} = P^t$, this produces the factorization

$$A = P^{-1}LU = (P^tL)U.$$

The matrix U is still upper triangular, but P^tL is not lower triangular unless $P = I$.

Example 3 Determine a factorization in the form $A = (P^tL)U$ for the matrix

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

Solution The matrix A cannot have an LU factorization because $a_{11} = 0$. However, using the row interchange $(E_1) \leftrightarrow (E_2)$, followed by $(E_3 + E_1) \rightarrow (E_3)$ and $(E_4 - E_1) \rightarrow (E_4)$, produces

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

Then the row interchange $(E_2) \leftrightarrow (E_4)$, followed by $(E_4 + E_3) \rightarrow (E_4)$, gives the matrix

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

The permutation matrix associated with the row interchanges $(E_1) \leftrightarrow (E_2)$ and $(E_2) \leftrightarrow (E_4)$ is

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

and

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

Gaussian elimination is performed on PA using the same operations as on A, except without the row interchanges. That is, $(E_2 - E_1) \rightarrow (E_2)$, $(E_3 + E_1) \rightarrow (E_3)$, followed by $(E_4 + E_3) \rightarrow (E_4)$. The nonzero multipliers for PA are consequently,

$$m_{21} = 1, \quad m_{31} = -1, \quad \text{and} \quad m_{43} = -1,$$

and the LU factorization of PA is

$$PA = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix} = LU.$$

Multiplying by $P^{-1} = P^t$ produces the factorization

$$A = P^{-1}(LU) = P^t(LU) = (P^tL)U = \begin{bmatrix} 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}. \quad \blacksquare$$

A matrix factorization of the form $A = PLU$ for a matrix A can be obtained using the *LinearAlgebra* package of Maple with the command

$LUDecomposition(A)$

The function call

$(P, L, U) := LUDecomposition(A)$

gives the factorization, and stores the permutation matrix as P, the lower triangular matrix as L, and the upper triangular matrix as U.

EXERCISE SET 6.5

1. Solve the following linear systems:

a. $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & -1 \\ 0 & -2 & 1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$

b. $\begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}$

2. Solve the following linear systems:

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

b. $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}$

3. Consider the following matrices. Find the permutation matrix P so that PA can be factored into the product LU, where L is lower triangular with 1s on its diagonal and U is upper triangular for these matrices.

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

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

4. Consider the following matrices. Find the permutation matrix P so that PA can be factored into the product LU, where L is lower triangular with 1s on its diagonal and U is upper triangular for these matrices.

a. $A = \begin{bmatrix} 0 & 2 & -1 \\ 1 & -1 & 2 \\ 1 & -1 & 4 \end{bmatrix}$ **b.** $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & 7 \\ -1 & 2 & 5 \end{bmatrix}$

c. $A = \begin{bmatrix} 1 & 1 & -1 & 2 \\ -1 & -1 & 1 & 5 \\ 2 & 2 & 3 & 7 \\ 2 & 3 & 4 & 5 \end{bmatrix}$ **d.** $A = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 2 & 2 & 4 & 5 \\ 1 & -1 & 1 & 7 \\ 2 & 3 & 4 & 6 \end{bmatrix}$

5. Factor the following matrices into the LU decomposition using the LU Factorization Algorithm with $l_{ii} = 1$ for all i.

a. $\begin{bmatrix} 2 & -1 & 1 \\ 3 & 3 & 9 \\ 3 & 3 & 5 \end{bmatrix}$ **b.** $\begin{bmatrix} 1.012 & -2.132 & 3.104 \\ -2.132 & 4.096 & -7.013 \\ 3.104 & -7.013 & 0.014 \end{bmatrix}$

c. $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 1.5 & 0 & 0 \\ 0 & -3 & 0.5 & 0 \\ 2 & -2 & 1 & 1 \end{bmatrix}$ **d.** $\begin{bmatrix} 2.1756 & 4.0231 & -2.1732 & 5.1967 \\ -4.0231 & 6.0000 & 0 & 1.1973 \\ -1.0000 & -5.2107 & 1.1111 & 0 \\ 6.0235 & 7.0000 & 0 & -4.1561 \end{bmatrix}$

6. Factor the following matrices into the LU decomposition using the LU Factorization Algorithm with $l_{ii} = 1$ for all i.

a. $\begin{bmatrix} 1 & -1 & 0 \\ 2 & 2 & 3 \\ -1 & 3 & 2 \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{3} & \frac{1}{2} & -\frac{1}{4} \\ \frac{1}{5} & \frac{2}{3} & \frac{3}{8} \\ \frac{2}{5} & -\frac{2}{3} & \frac{5}{8} \end{bmatrix}$

c. $\begin{bmatrix} 2 & 1 & 0 & 0 \\ -1 & 3 & 3 & 0 \\ 2 & -2 & 1 & 4 \\ -2 & 2 & 2 & 5 \end{bmatrix}$ **d.** $\begin{bmatrix} 2.121 & -3.460 & 0 & 5.217 \\ 0 & 5.193 & -2.197 & 4.206 \\ 5.132 & 1.414 & 3.141 & 0 \\ -3.111 & -1.732 & 2.718 & 5.212 \end{bmatrix}$

7. Modify the LU Factorization Algorithm so that it can be used to solve a linear system, and then solve the following linear systems.

a.
$$\begin{aligned} 2x_1 - x_2 + x_3 &= -1, \\ 3x_1 + 3x_2 + 9x_3 &= 0, \\ 3x_1 + 3x_2 + 5x_3 &= 4. \end{aligned}$$

b.
$$\begin{aligned} 1.012x_1 - 2.132x_2 + 3.104x_3 &= 1.984, \\ -2.132x_1 + 4.096x_2 - 7.013x_3 &= -5.049, \\ 3.104x_1 - 7.013x_2 + 0.014x_3 &= -3.895. \end{aligned}$$

c.
$$\begin{aligned} 2x_1 \qquad\qquad\qquad\qquad &= 3, \\ x_1 + 1.5x_2 \qquad\qquad &= 4.5, \\ -\ 3x_2 + 0.5x_3 \qquad &= -6.6, \\ 2x_1 - 2x_2 + x_3 + x_4 &= 0.8. \end{aligned}$$

d.
$$\begin{aligned} 2.1756x_1 + 4.0231x_2 - 2.1732x_3 + 5.1967x_4 &= 17.102, \\ -4.0231x_1 + 6.0000x_2 \qquad\qquad + 1.1973x_4 &= -6.1593, \\ -1.0000x_1 - 5.2107x_2 + 1.1111x_3 \qquad\qquad &= 3.0004, \\ 6.0235x_1 + 7.0000x_2 \qquad\qquad - 4.1561x_4 &= 0.0000. \end{aligned}$$

8. Modify the *LU* Factorization Algorithm so that it can be used to solve a linear system, and then solve the following linear systems.

a.
$$\begin{aligned} x_1 - x_2 &= 2,\\ 2x_1 + 2x_2 + 3x_3 &= -1,\\ -x_1 + 3x_2 + 2x_3 &= 4. \end{aligned}$$

b.
$$\begin{aligned} \tfrac{1}{3}x_1 + \tfrac{1}{2}x_2 - \tfrac{1}{4}x_3 &= 1,\\ \tfrac{1}{5}x_1 + \tfrac{2}{3}x_2 + \tfrac{3}{8}x_3 &= 2,\\ \tfrac{2}{5}x_1 - \tfrac{2}{3}x_2 + \tfrac{5}{8}x_3 &= -3. \end{aligned}$$

b.
$$\begin{aligned} 2x_1 + x_2 &= 0,\\ -x_1 + 3x_2 + 3x_3 &= 5,\\ 2x_1 - 2x_2 + x_3 + 4x_4 &= -2,\\ -2x_1 + 2x_2 + 2x_3 + 5x_4 &= 6. \end{aligned}$$

d.
$$\begin{aligned} 2.121x_1 - 3.460x_2 + 5.217x_4 &= 1.909,\\ 5.193x_2 - 2.197x_3 + 4.206x_4 &= 0,\\ 5.132x_1 + 1.414x_2 + 3.141x_3 &= -2.101,\\ -3.111x_1 - 1.732x_2 + 2.718x_3 + 5.212x_4 &= 6.824. \end{aligned}$$

9. Obtain factorizations of the form $A = P^tLU$ for the following matrices.

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

c. $A = \begin{bmatrix} 1 & -2 & 3 & 0 \\ 3 & -6 & 9 & 3 \\ 2 & 1 & 4 & 1 \\ 1 & -2 & 2 & -2 \end{bmatrix}$ **d.** $A = \begin{bmatrix} 1 & -2 & 3 & 0 \\ 1 & -2 & 3 & 1 \\ 1 & -2 & 2 & -2 \\ 2 & 1 & 3 & -1 \end{bmatrix}$

10. Suppose $A = P^tLU$, where P is a permutation matrix, L is a lower-triangular matrix with ones on the diagonal, and U is an upper-triangular matrix.

a. Count the number of operations needed to compute P^tLU for a given matrix A.

b. Show that if P contains k row interchanges, then
$$\det P = \det P^t = (-1)^k.$$

c. Use $\det A = \det P^t \det L \det U = (-1)^k \det U$ to count the number of operations for determining $\det A$ by factoring.

d. Compute $\det A$ and count the number of operations when
$$A = \begin{bmatrix} 0 & 2 & 1 & 4 & -1 & 3 \\ 1 & 2 & -1 & 3 & 4 & 0 \\ 0 & 1 & 1 & -1 & 2 & -1 \\ 2 & 3 & -4 & 2 & 0 & 5 \\ 1 & 1 & 1 & 3 & 0 & 2 \\ -1 & -1 & 2 & -1 & 2 & 0 \end{bmatrix}.$$

11. a. Show that the *LU* Factorization Algorithm requires
$$\tfrac{1}{3}n^3 - \tfrac{1}{3}n \text{ multiplications/divisions} \quad \text{and} \quad \tfrac{1}{3}n^3 - \tfrac{1}{2}n^2 + \tfrac{1}{6}n \text{ additions/subtractions.}$$

b. Show that solving $L\mathbf{y} = \mathbf{b}$, where L is a lower-triangular matrix with $l_{ii} = 1$ for all i, requires
$$\tfrac{1}{2}n^2 - \tfrac{1}{2}n \text{ multiplications/divisions} \quad \text{and} \quad \tfrac{1}{2}n^2 - \tfrac{1}{2}n \text{ additions/subtractions.}$$

c. Show that solving $A\mathbf{x} = \mathbf{b}$ by first factoring A into $A = LU$ and then solving $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$ requires the same number of operations as the Gaussian Elimination Algorithm 6.1.

d. Count the number of operations required to solve m linear systems $A\mathbf{x}^{(k)} = \mathbf{b}^{(k)}$ for $k = 1, \ldots, m$ by first factoring A and then using the method of part (c) m times.

6.6 Special Types of Matrices

We now turn attention to two classes of matrices for which Gaussian elimination can be performed effectively without row interchanges.

Diagonally Dominant Matrices

The first class is described in the following definition.

Definition 6.20 The $n \times n$ matrix A is said to be **diagonally dominant** when

$$|a_{ii}| \geq \sum_{\substack{j=1,\\ j\neq i}}^{n} |a_{ij}| \quad \text{holds for each } i = 1, 2, \cdots, n. \tag{6.10}$$

Each main diagonal entry in a strictly diagonally dominant matrix has a magnitude that is strictly greater that the sum of the magnitudes of all the other entries in that row.

A diagonally dominant matrix is said to be **strictly diagonally dominant** when the inequality in (6.10) is strict for each n, that is, when

$$|a_{ii}| > \sum_{\substack{j=1,\\ j\neq i}}^{n} |a_{ij}| \quad \text{holds for each } i = 1, 2, \cdots, n.$$ ■

Illustration Consider the matrices

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

The nonsymmetric matrix A is strictly diagonally dominant because

$$|7| > |2| + |0|, \quad |5| > |3| + |-1|, \quad \text{and} \quad |-6| > |0| + |5|.$$

The symmetric matrix B is not strictly diagonally dominant because, for example, in the first row the absolute value of the diagonal element is $|6| < |4| + |-3| = 7$. It is interesting to note that A^t is not strictly diagonally dominant, because the middle row of A^t is [2 5 5], nor, of course, is B^t because $B^t = B$. □

The following theorem was used in Section 3.5 to ensure that there are unique solutions to the linear systems needed to determine cubic spline interpolants.

Theorem 6.21 A strictly diagonally dominant matrix A is nonsingular. Moreover, in this case, Gaussian elimination can be performed on any linear system of the form $A\mathbf{x} = \mathbf{b}$ to obtain its unique solution without row or column interchanges, and the computations will be stable with respect to the growth of round-off errors. ■

Proof We first use proof by contradiction to show that A is nonsingular. Consider the linear system described by $A\mathbf{x} = \mathbf{0}$, and suppose that a nonzero solution $\mathbf{x} = (x_i)$ to this system exists. Let k be an index for which

$$0 < |x_k| = \max_{1\leq j\leq n} |x_j|.$$

Because $\sum_{j=1}^{n} a_{ij}x_j = 0$ for each $i = 1, 2, \ldots, n$, we have, when $i = k$,

$$a_{kk}x_k = -\sum_{\substack{j=1,\\ j\neq k}}^{n} a_{kj}x_j.$$

From the triangle inequality we have

$$|a_{kk}||x_k| \le \sum_{\substack{j=1,\\ j\ne k}}^{n} |a_{kj}||x_j|, \quad \text{so} \quad |a_{kk}| \le \sum_{\substack{j=1,\\ j\ne k}}^{n} |a_{kj}|\frac{|x_j|}{|x_k|} \le \sum_{\substack{j=1,\\ j\ne k}}^{n} |a_{kj}|.$$

This inequality contradicts the strict diagonal dominance of A. Consequently, the only solution to $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. This is shown in Theorem 6.17 on page 398 to be equivalent to the nonsingularity of A.

To prove that Gaussian elimination can be performed without row interchanges, we show that each of the matrices $A^{(2)}, A^{(3)}, \ldots, A^{(n)}$ generated by the Gaussian elimination process (and described in Section 6.5) is strictly diagonally dominant. This will ensure that at each stage of the Gaussian elimination process the pivot element is nonzero.

Since A is strictly diagonally dominant, $a_{11} \ne 0$ and $A^{(2)}$ can be formed. Thus for each $i = 2, 3, \ldots, n$,

$$a_{ij}^{(2)} = a_{ij}^{(1)} - \frac{a_{1j}^{(1)}a_{i1}^{(1)}}{a_{11}^{(1)}}, \quad \text{for} \quad 2 \le j \le n.$$

First, $a_{i1}^{(2)} = 0$. The triangle inequality implies that

$$\sum_{\substack{j=2\\ j\ne i}}^{n} |a_{ij}^{(2)}| = \sum_{\substack{j=2\\ j\ne i}}^{n} \left| a_{ij}^{(1)} - \frac{a_{1j}^{(1)}a_{i1}^{(1)}}{a_{11}^{(1)}} \right| \le \sum_{\substack{j=2\\ j\ne i}}^{n} |a_{ij}^{(1)}| + \sum_{\substack{j=2\\ j\ne i}}^{n} \left| \frac{a_{1j}^{(1)}a_{i1}^{(1)}}{a_{11}^{(1)}} \right|.$$

But since A is strictly diagonally dominant,

$$\sum_{\substack{j=2\\ j\ne i}}^{n} |a_{ij}^{(1)}| < |a_{ii}^{(1)}| - |a_{i1}^{(1)}| \quad \text{and} \quad \sum_{\substack{j=2\\ j\ne i}}^{n} |a_{1j}^{(1)}| < |a_{11}^{(1)}| - |a_{1i}^{(1)}|,$$

so

$$\sum_{\substack{j=2\\ j\ne i}}^{n} |a_{ij}^{(2)}| < |a_{ii}^{(1)}| - |a_{i1}^{(1)}| + \frac{|a_{i1}^{(1)}|}{|a_{11}^{(1)}|}(|a_{11}^{(1)}| - |a_{1i}^{(1)}|) = |a_{ii}^{(1)}| - \frac{|a_{i1}^{(1)}||a_{1i}^{(1)}|}{|a_{11}^{(1)}|}.$$

The triangle inequality also implies that

$$|a_{ii}^{(1)}| - \frac{|a_{i1}^{(1)}||a_{1i}^{(1)}|}{|a_{11}^{(1)}|} \le \left| a_{ii}^{(1)} - \frac{|a_{i1}^{(1)}||a_{1i}^{(1)}|}{|a_{11}^{(1)}|} \right| = |a_{ii}^{(2)}|.$$

which gives

$$\sum_{\substack{j=2\\ j\ne i}}^{n} |a_{ij}^{(2)}| < |a_{ii}^{(2)}|.$$

This establishes the strict diagonal dominance for rows $2, \ldots, n$. But the first row of $A^{(2)}$ and A are the same, so $A^{(2)}$ is strictly diagonally dominant.

This process is continued inductively until the upper-triangular and strictly diagonally dominant $A^{(n)}$ is obtained. This implies that all the diagonal elements are nonzero, so Gaussian elimination can be performed without row interchanges.

The demonstration of stability for this procedure can be found in [We]. ■ ■ ■

Positive Definite Matrices

The next special class of matrices is called *positive definite*.

Definition 6.22 A matrix A is **positive definite** if it is symmetric and if $\mathbf{x}^t A\mathbf{x} > 0$ for every n-dimensional vector $\mathbf{x} \neq \mathbf{0}$. ■

The name positive definite refers to the fact that the number $\mathbf{x}^t A\mathbf{x}$ must be positive whenever $\mathbf{x} \neq \mathbf{0}$.

Not all authors require symmetry of a positive definite matrix. For example, Golub and Van Loan [GV], a standard reference in matrix methods, requires only that $\mathbf{x}^t A\mathbf{x} > 0$ for each $\mathbf{x} \neq 0$. Matrices we call positive definite are called symmetric positive definite in [GV]. Keep this discrepancy in mind if you are using material from other sources.

To be precise, Definition 6.22 should specify that the 1×1 matrix generated by the operation $\mathbf{x}^t A\mathbf{x}$ has a positive value for its only entry since the operation is performed as follows:

$$\mathbf{x}^t A\mathbf{x} = [x_1, x_2, \cdots, x_n] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= [x_1, x_2, \cdots, x_n] \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{bmatrix} = \left[\sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j \right].$$

Example 1 Show that the matrix

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

is positive definite

Solution Suppose $\mathbf{x}$ is any three-dimensional column vector. Then

$$\mathbf{x}^t A\mathbf{x} = [x_1, x_2, x_3] \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= [x_1, x_2, x_3] \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - x_3 \\ -x_2 + 2x_3 \end{bmatrix}$$

$$= 2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2.$$

Rearranging the terms gives

$$\begin{aligned} \mathbf{x}^t A\mathbf{x} &= x_1^2 + (x_1^2 - 2x_1x_2 + x_2^2) + (x_2^2 - 2x_2x_3 + x_3^2) + x_3^2 \\ &= x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2, \end{aligned}$$

which implies that

$$x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 > 0$$

unless $x_1 = x_2 = x_3 = 0$. ■

It should be clear from the example that using the definition to determine if a matrix is positive definite can be difficult. Fortunately, there are more easily verified criteria, which are presented in Chapter 9, for identifying members of this important class. The next result provides some necessary conditions that can be used to eliminate certain matrices from consideration.

Theorem 6.23 If A is an $n \times n$ positive definite matrix, then

(i) A has an inverse; **(ii)** $a_{ii} > 0$, for each $i = 1, 2, \ldots, n$;

(iii) $\max_{1\le k,j\le n} |a_{kj}| \le \max_{1\le i\le n} |a_{ii}|$; **(iv)** $(a_{ij})^2 < a_{ii}a_{jj}$, for each $i \ne j$. ■

Proof

(i) If $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x}^t A\mathbf{x} = 0$. Since A is positive definite, this implies $\mathbf{x} = \mathbf{0}$. Consequently, $A\mathbf{x} = \mathbf{0}$ has only the zero solution. By Theorem 6.17 on page 398, this is equivalent to A being nonsingular.

(ii) For a given i, let $\mathbf{x} = (x_j)$ be defined by $x_i = 1$ and $x_j = 0$, if $j \ne i$. Since $\mathbf{x} \ne \mathbf{0}$,

$$0 < \mathbf{x}^t A\mathbf{x} = a_{ii}.$$

(iii) For $k \ne j$, define $\mathbf{x} = (x_i)$ by

$$x_i = \begin{cases} 0, & \text{if } i \ne j \text{ and } i \ne k, \\ 1, & \text{if } i = j, \\ -1, & \text{if } i = k. \end{cases}$$

Since $\mathbf{x} \ne \mathbf{0}$,

$$0 < \mathbf{x}^t A\mathbf{x} = a_{jj} + a_{kk} - a_{jk} - a_{kj}.$$

But $A^t = A$, so $a_{jk} = a_{kj}$, which implies that

$$2a_{kj} < a_{jj} + a_{kk}. \tag{6.11}$$

Now define $\mathbf{z} = (z_i)$ by

$$z_i = \begin{cases} 0, & \text{if } i \ne j \text{ and } i \ne k, \\ 1, & \text{if } i = j \text{ or } i = k. \end{cases}$$

Then $\mathbf{z}^t A\mathbf{z} > 0$, so

$$-2a_{kj} < a_{kk} + a_{jj}. \tag{6.12}$$

Equations (6.11) and (6.12) imply that for each $k \ne j$,

$$|a_{kj}| < \frac{a_{kk} + a_{jj}}{2} \le \max_{1\le i\le n} |a_{ii}|, \quad \text{so} \quad \max_{1\le k,j\le n} |a_{kj}| \le \max_{1\le i\le n} |a_{ii}|.$$

(iv) For $i \ne j$, define $\mathbf{x} = (x_k)$ by

$$x_k = \begin{cases} 0, & \text{if } k \ne j \text{ and } k \ne i, \\ \alpha, & \text{if } k = i, \\ 1, & \text{if } k = j, \end{cases}$$

where α represents an arbitrary real number. Because $\mathbf{x} \ne \mathbf{0}$,

$$0 < \mathbf{x}^t A\mathbf{x} = a_{ii}\alpha^2 + 2a_{ij}\alpha + a_{jj}.$$

As a quadratic polynomial in α with no real roots, the discriminant of $P(\alpha) = a_{ii}\alpha^2 + 2a_{ij}\alpha + a_{jj}$ must be negative. Thus

$$4a_{ij}^2 - 4a_{ii}a_{jj} < 0 \quad \text{and} \quad a_{ij}^2 < a_{ii}a_{jj}.$$ ■ ■ ■

Although Theorem 6.23 provides some important conditions that must be true of positive definite matrices, it does not ensure that a matrix satisfying these conditions is positive definite.

The following notion will be used to provide a necessary and sufficient condition.

Definition 6.24 A **leading principal submatrix** of a matrix A is a matrix of the form

$$A_k = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix},$$

for some $1 \le k \le n$. ■

A proof of the following result can be found in [Stew2], p. 250.

Theorem 6.25 A symmetric matrix A is positive definite if and only if each of its leading principal submatrices has a positive determinant. ■

Example 2 In Example 1 we used the definition to show that the symmetric matrix

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

is positive definite. Confirm this using Theorem 6.25.

Solution Note that

$$\det A_1 = \det[2] = 2 > 0,$$

$$\det A_2 = \det \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 4 - 1 = 3 > 0,$$

and

$$\det A_3 = \det \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = 2 \det \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} - (-1) \det \begin{bmatrix} -1 & -1 \\ 0 & 2 \end{bmatrix}$$

$$= 2(4 - 1) + (-2 + 0) = 4 > 0.$$

in agreement with Theorem 6.25. ■

The next result extends part **(i)** of Theorem 6.23 and parallels the strictly diagonally dominant results presented in Theorem 6.21 on page 412. We will not give a proof of this theorem because it requires introducing terminology and results that are not needed for any other purpose. The development and proof can be found in [We], pp. 120 ff.

Theorem 6.26 The symmetric matrix A is positive definite if and only if Gaussian elimination without row interchanges can be performed on the linear system $A\mathbf{x} = \mathbf{b}$ with all pivot elements positive. Moreover, in this case, the computations are stable with respect to the growth of round-off errors. ■

Some interesting facts that are uncovered in constructing the proof of Theorem 6.26 are presented in the following corollaries.

Corollary 6.27 The matrix A is positive definite if and only if A can be factored in the form LDL^t, where L is lower triangular with 1s on its diagonal and D is a diagonal matrix with positive diagonal entries. ■

Corollary 6.28 The matrix A is positive definite if and only if A can be factored in the form LL^t, where L is lower triangular with nonzero diagonal entries. ■

The matrix L in this Corollary is not the same as the matrix L in Corollary 6.27. A relationship between them is presented in Exercise 32.

Algorithm 6.5 is based on the *LU* Factorization Algorithm 6.4 and obtains the LDL^t factorization described in Corollary 6.27.

ALGORITHM
6.5

LDL^t Factorization

To factor the positive definite $n \times n$ matrix A into the form LDL^t, where L is a lower triangular matrix with 1s along the diagonal and D is a diagonal matrix with positive entries on the diagonal:

INPUT the dimension n; entries a_{ij}, for $1 \le i, j \le n$ of A.

OUTPUT the entries l_{ij}, for $1 \le j < i$ and $1 \le i \le n$ of L, and d_i, for $1 \le i \le n$ of D.

Step 1 For $i = 1, \ldots, n$ do Steps 2–4.

Step 2 For $j = 1, \ldots, i-1$, set $v_j = l_{ij} d_j$.

Step 3 Set $d_i = a_{ii} - \sum_{j=1}^{i-1} l_{ij} v_j$.

Step 4 For $j = i+1, \ldots, n$ set $l_{ji} = (a_{ji} - \sum_{k=1}^{i-1} l_{jk} v_k)/d_i$.

Step 5 OUTPUT (l_{ij} for $j = 1, \ldots, i-1$ and $i = 1, \ldots, n$);
OUTPUT (d_i for $i = 1, \ldots, n$);
STOP. ■

The *NumericalAnalysis* subpackage factors a positive definite matrix A as LDL^t with the command

$L, DD, Lt := MatrixDecomposition(A, method = LDLt)$

Corollary 6.27 has a counterpart when A is symmetric but not necessarily positive definite. This result is widely applied because symmetric matrices are common and easily recognized.

Corollary 6.29 Let A be a symmetric $n \times n$ matrix for which Gaussian elimination can be applied without row interchanges. Then A can be factored into LDL^t, where L is lower triangular with 1s on its diagonal and D is the diagonal matrix with $a_{11}^{(1)}, \ldots, a_{nn}^{(n)}$ on its diagonal. ■

Example 3 Determine the LDL^t factorization of the positive definite matrix

$$A = \begin{bmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix}.$$

Solution The LDL^t factorization has 1s on the diagonal of the lower triangular matrix L so we need to have

$$A = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} 1 & l_{21} & l_{31} \\ 0 & 1 & l_{32} \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} d_1 & d_1 l_{21} & d_1 l_{31} \\ d_1 l_{21} & d_2 + d_1 l_{21}^2 & d_2 l_{32} + d_1 l_{21} l_{31} \\ d_1 l_{31} & d_1 l_{21} l_{31} + d_2 l_{32} & d_1 l_{31}^2 + d_2 l_{32}^2 + d_3 \end{bmatrix}$$

Thus

$$a_{11} : 4 = d_1 \Longrightarrow d_1 = 4, \qquad a_{21} : -1 = d_1 l_{21} \Longrightarrow l_{21} = -0.25$$

$$a_{31} : 1 = d_1 l_{31} \Longrightarrow l_{31} = 0.25, \qquad a_{22} : 4.25 = d_2 + d_1 l_{21}^2 \Longrightarrow d_2 = 4$$

$$a_{32} : 2.75 = d_1 l_{21} l_{31} + d_2 l_{32} \Longrightarrow l_{32} = 0.75, \quad a_{33} : 3.5 = d_1 l_{31}^2 + d_2 l_{32}^2 + d_3 \Longrightarrow d_3 = 1,$$

and we have

$$A = LDL^t = \begin{bmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0.25 & 0.75 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -0.25 & 0.25 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{bmatrix}.$$ ■

Algorithm 6.5 is easily modified to factor the symmetric matrices described in Corollary 6.29. It simply requires adding a check to ensure that the diagonal elements are nonzero. The Cholesky Algorithm 6.6 produces the LL^t factorization described in Corollary 6.28.

ALGORITHM **6.6**

Cholesky

To factor the positive definite $n \times n$ matrix A into LL^t, where L is lower triangular:

INPUT the dimension n; entries a_{ij}, for $1 \le i, j \le n$ of A.

OUTPUT the entries l_{ij}, for $1 \le j \le i$ and $1 \le i \le n$ of L. (*The entries of* $U = L^t$ *are* $u_{ij} = l_{ji}$, *for* $i \le j \le n$ *and* $1 \le i \le n$.)

Step 1 Set $l_{11} = \sqrt{a_{11}}$.

Step 2 For $j = 2, \ldots, n$, set $l_{j1} = a_{j1}/l_{11}$.

Step 3 For $i = 2, \ldots, n-1$ do Steps 4 and 5.

Step 4 Set $l_{ii} = \left(a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2\right)^{1/2}$.

Step 5 For $j = i+1, \ldots, n$

set $l_{ji} = \left(a_{ji} - \sum_{k=1}^{i-1} l_{jk} l_{ik}\right) / l_{ii}$.

Step 6 Set $l_{nn} = \left(a_{nn} - \sum_{k=1}^{n-1} l_{nk}^2\right)^{1/2}$.

Step 7 OUTPUT (l_{ij} for $j = 1, \ldots, i$ and $i = 1, \ldots, n$);
STOP. ■

Andre-Louis Cholesky (1875-1918) was a French military officer involved in geodesy and surveying in the early 1900s. He developed this factorization method to compute solutions to least squares problems.

The Cholesky factorization of A is computed in the *LinearAlgebra* library of Maple using the statement

$L := LUDecomposition(A, method = \ 'Cholesky')$

and gives the lower triangular matrix L as its output.

Example 4 Determine the Cholesky LL^t factorization of the positive definite matrix

$$A = \begin{bmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix}.$$

Solution The LL^t factorization does not necessarily has 1s on the diagonal of the lower triangular matrix L so we need to have

$$A = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

$$= \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Thus

$$a_{11}: \quad 4 = l_{11}^2 \implies l_{11} = 2, \qquad a_{21}: \quad -1 = l_{11}l_{21} \implies l_{21} = -0.5$$

$$a_{31}: \quad 1 = l_{11}l_{31} \implies l_{31} = 0.5, \qquad a_{22}: \quad 4.25 = l_{21}^2 + l_{22}^2 \implies l_{22} = 2$$

$$a_{32}: \quad 2.75 = l_{21}l_{31} + l_{22}l_{32} \implies l_{32} = 1.5, \qquad a_{33}: \quad 3.5 = l_{31}^2 + l_{32}^2 + l_{33}^2 \implies l_{33} = 1,$$

and we have

$$A = LL^t = \begin{bmatrix} 2 & 0 & 0 \\ -0.5 & 2 & 0 \\ 0.5 & 1.5 & 1 \end{bmatrix} \begin{bmatrix} 2 & -0.5 & 0.5 \\ 0 & 2 & 1.5 \\ 0 & 0 & 1 \end{bmatrix}.$$ ■

The LDL^t factorization described in Algorithm 6.5 requires

$$\frac{1}{6}n^3 + n^2 - \frac{7}{6}n \text{ multiplications/divisions} \quad \text{and} \quad \frac{1}{6}n^3 - \frac{1}{6}n \text{ additions/subtractions.}$$

The LL^t Cholesky factorization of a positive definite matrix requires only

$$\frac{1}{6}n^3 + \frac{1}{2}n^2 - \frac{2}{3}n \text{ multiplications/divisions} \quad \text{and} \quad \frac{1}{6}n^3 - \frac{1}{6}n \text{ additions/subtractions.}$$

This computational advantage of Cholesky's factorization is misleading, because it requires extracting n square roots. However, the number of operations required for computing the n square roots is a linear factor of n and will decrease in significance as n increases.

Algorithm 6.5 provides a stable method for factoring a positive definite matrix into the form $A = LDL^t$, but it must be modified to solve the linear system $A\mathbf{x} = \mathbf{b}$. To do this, we delete the STOP statement from Step 5 in the algorithm and add the following steps to solve the lower triangular system $L\mathbf{y} = \mathbf{b}$:

Step 6 Set $y_1 = b_1$.

Step 7 For $i = 2, \ldots, n$ set $y_i = b_i - \sum_{j=1}^{i-1} l_{ij} y_j$.

The linear system $D\mathbf{z} = \mathbf{y}$ can then be solved by

Step 8 For $i = 1, \ldots, n$ set $z_i = y_i / d_i$.

Finally, the upper-triangular system $L^t\mathbf{x} = \mathbf{z}$ is solved with the steps given by

Step 9 Set $x_n = z_n$.

Step 10 For $i = n-1, \ldots, 1$ set $x_i = z_i - \sum_{j=i+1}^{n} l_{ji} x_j$.

Step 11 OUTPUT (x_i for $i = 1, \ldots, n$);
STOP.

Table 6.4 shows the additional operations required to solve the linear system.

Table 6.4

Step	Multiplications/Divisions	Additions/Subtractions
6	0	0
7	$n(n-1)/2$	$n(n-1)/2$
8	n	0
9	0	0
10	$n(n-1)/2$	$n(n-1)/2$
Total	n^2	$n^2 - n$

If the Cholesky factorization given in Algorithm 6.6 is preferred, the additional steps for solving the system $A\mathbf{x} = \mathbf{b}$ are as follows. First delete the STOP statement from Step 7. Then add

Step 8 Set $y_1 = b_1 / l_{11}$.

Step 9 For $i = 2, \ldots, n$ set $y_i = \left(b_i - \sum_{j=1}^{i-1} l_{ij} y_j\right) \Big/ l_{ii}$.

Step 10 Set $x_n = y_n / l_{nn}$.

Step 11 For $i = n-1, \ldots, 1$ set $x_i = \left(y_i - \sum_{j=i+1}^{n} l_{ji} x_j\right) \Big/ l_{ii}$.

Step 12 OUTPUT (x_i for $i = 1, \ldots, n$);
STOP.

Steps 8–12 require $n^2 + n$ multiplications/divisions and $n^2 - n$ additions/ subtractions.

Band Matrices

The last class of matrices considered are *band matrices*. In many applications, the band matrices are also strictly diagonally dominant or positive definite.

Definition 6.30 An $n \times n$ matrix is called a **band matrix** if integers p and q, with $1 < p, q < n$, exist with the property that $a_{ij} = 0$ whenever $p \le j - i$ or $q \le i - j$. The **band width** of a band matrix is defined as $w = p + q - 1$. ■

The name for a band matrix comes from the fact that all the nonzero entries lie in a band which is centered on the main diagonal.

The number p describes the number of diagonals above, and including, the main diagonal on which nonzero entries may lie. The number q describes the number of diagonals below, and including, the main diagonal on which nonzero entries may lie. For example, the matrix

$$A = \begin{bmatrix} 7 & 2 & 0 \\ 3 & 5 & -1 \\ 0 & -5 & -6 \end{bmatrix}$$

is a band matrix with $p = q = 2$ and bandwidth $2 + 2 - 1 = 3$.

The definition of band matrix forces those matrices to concentrate all their nonzero entries about the diagonal. Two special cases of band matrices that occur frequently have $p = q = 2$ and $p = q = 4$.

Tridiagonal Matrices

Matrices of bandwidth 3 occurring when $p = q = 2$ are called **tridiagonal** because they have the form

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & \cdots & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \ddots & & \vdots \\ 0 & a_{32} & a_{33} & a_{34} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & 0 & & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

Tridiagonal matrices are also considered in Chapter 11 in connection with the study of piecewise linear approximations to boundary-value problems. The case of $p = q = 4$ will be used for the solution of boundary-value problems when the approximating functions assume the form of cubic splines.

The factorization algorithms can be simplified considerably in the case of band matrices because a large number of zeros appear in these matrices in regular patterns. It is particularly interesting to observe the form the Crout or Doolittle method assumes in this case.

To illustrate the situation, suppose a tridiagonal matrix A can be factored into the triangular matrices L and U. Then A has at most $(3n - 2)$ nonzero entries. Then there are only $(3n - 2)$ conditions to be applied to determine the entries of L and U, provided, of course, that the zero entries of A are also obtained.

Suppose that the matrices L and U also have tridiagonal form, that is,

$$L = \begin{bmatrix} l_{11} & 0 & \cdots & \cdots & 0 \\ l_{21} & l_{22} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & l_{n,n-1} & l_{nn} \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 1 & u_{12} & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & u_{n-1,n} \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}.$$

There are $(2n-1)$ undetermined entries of L and $(n-1)$ undetermined entries of U, which totals $(3n-2)$, the number of possible nonzero entries of A. The 0 entries of A are obtained automatically.

The multiplication involved with $A = LU$ gives, in addition to the 0 entries,

$$a_{11} = l_{11};$$

$$a_{i,i-1} = l_{i,i-1}, \quad \text{for each } i = 2, 3, \ldots, n; \tag{6.13}$$

$$a_{ii} = l_{i,i-1}u_{i-1,i} + l_{ii}, \quad \text{for each } i = 2, 3, \ldots, n; \tag{6.14}$$

and

$$a_{i,i+1} = l_{ii}u_{i,i+1}, \quad \text{for each } i = 1, 2, \ldots, n-1. \tag{6.15}$$

A solution to this system is found by first using Eq. (6.13) to obtain all the nonzero off-diagonal terms in L and then using Eqs. (6.14) and (6.15) to alternately obtain the remainder of the entries in U and L. Once an entry L or U is computed, the corresponding entry in A is not needed. So the entries in A can be overwritten by the entries in L and U with the result that no new storage is required.

Algorithm 6.7 solves an $n \times n$ system of linear equations whose coefficient matrix is tridiagonal. This algorithm requires only $(5n-4)$ multiplications/divisions and $(3n-3)$ additions/subtractions. Consequently, it has considerable computational advantage over the methods that do not consider the tridiagonality of the matrix.

ALGORITHM
6.7

Crout Factorization for Tridiagonal Linear Systems

To solve the $n \times n$ linear system

$$\begin{aligned}
E_1 &: \quad a_{11}x_1 + a_{12}x_2 &&= a_{1,n+1},\\
E_2 &: \quad a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &&= a_{2,n+1},\\
&\vdots \qquad\qquad \vdots && \vdots\\
E_{n-1} &: \quad a_{n-1,n-2}x_{n-2} + a_{n-1,n-1}x_{n-1} + a_{n-1,n}x_n &&= a_{n-1,n+1},\\
E_n &: \quad a_{n,n-1}x_{n-1} + a_{nn}x_n &&= a_{n,n+1},
\end{aligned}$$

which is assumed to have a unique solution:

INPUT the dimension n; the entries of A.

OUTPUT the solution $x_1, \ldots, x_n$.

(*Steps 1–3 set up and solve* $L\mathbf{z} = \mathbf{b}$.)

Step 1 Set $l_{11} = a_{11}$;
$u_{12} = a_{12}/l_{11}$;
$z_1 = a_{1,n+1}/l_{11}$.

Step 2 For $i = 2, \ldots, n-1$ set $l_{i,i-1} = a_{i,i-1}$; (*ith row of* L.)
$l_{ii} = a_{ii} - l_{i,i-1}u_{i-1,i}$;
$u_{i,i+1} = a_{i,i+1}/l_{ii}$; ($(i+1)$*th column of* U.)
$z_i = (a_{i,n+1} - l_{i,i-1}z_{i-1})/l_{ii}$.

Step 3 Set $l_{n,n-1} = a_{n,n-1}$; (*nth row of* L.)
$l_{nn} = a_{nn} - l_{n,n-1}u_{n-1,n}$.
$z_n = (a_{n,n+1} - l_{n,n-1}z_{n-1})/l_{nn}$.

(*Steps 4 and 5 solve* $U\mathbf{x} = \mathbf{z}$.)

Step 4 Set $x_n = z_n$.

Step 5 For $i = n-1, \ldots, 1$ set $x_i = z_i - u_{i,i+1}x_{i+1}$.

Step 6 OUTPUT $(x_1, \ldots, x_n)$;
STOP. ■

Example 5 Determine the Crout factorization of the symmetric tridiagonal matrix

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

and use this factorization to solve the linear system

$$\begin{aligned} 2x_1 - x_2 \phantom{{}+2x_3 - x_4} &= 1, \\ -x_1 + 2x_2 - x_3 \phantom{{}-x_4} &= 0, \\ -x_2 + 2x_3 - x_4 &= 0, \\ -x_3 + 2x_4 &= 1. \end{aligned}$$

Solution The LU factorization of A has the form

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ 0 & l_{32} & l_{33} & 0 \\ 0 & 0 & l_{43} & l_{44} \end{bmatrix} \begin{bmatrix} 1 & u_{12} & 0 & 0 \\ 0 & 1 & u_{23} & 0 \\ 0 & 0 & 1 & u_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} l_{11} & l_{11}u_{12} & 0 & 0 \\ l_{21} & l_{22} + l_{21}u_{12} & l_{22}u_{23} & 0 \\ 0 & l_{32} & l_{33} + l_{32}u_{23} & l_{33}u_{34} \\ 0 & 0 & l_{43} & l_{44} + l_{43}u_{34} \end{bmatrix}.$$

Thus

$$\begin{aligned}
&a_{11}: \quad 2 = l_{11} \implies l_{11} = 2, && a_{12}: \quad -1 = l_{11}u_{12} \implies u_{12} = -\tfrac{1}{2}, \\
&a_{21}: \quad -1 = l_{21} \implies l_{21} = -1, && a_{22}: \quad 2 = l_{22} + l_{21}u_{12} \implies l_{22} = -\tfrac{3}{2}, \\
&a_{23}: \quad -1 = l_{22}u_{23} \implies u_{23} = -\tfrac{2}{3}, && a_{32}: \quad -1 = l_{32} \implies l_{32} = -1, \\
&a_{33}: \quad 2 = l_{33} + l_{32}u_{23} \implies l_{33} = \tfrac{4}{3}, && a_{34}: \quad -1 = l_{33}u_{34} \implies u_{34} = -\tfrac{3}{4}, \\
&a_{43}: \quad -1 = l_{43} \implies l_{43} = -1, && a_{44}: \quad 2 = l_{44} + l_{43}u_{34} \implies l_{44} = \tfrac{5}{4}.
\end{aligned}$$

This gives the Crout factorization

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

Solving the system

$$L\mathbf{z} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & \frac{3}{2} & 0 & 0 \\ 0 & -1 & \frac{4}{3} & 0 \\ 0 & 0 & -1 & \frac{5}{4} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{gives} \quad \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{4} \\ 1 \end{bmatrix},$$

and then solving

$$U\mathbf{x} = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 & -\frac{3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{4} \\ 1 \end{bmatrix} \quad \text{gives} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \quad ■$$

The Crout Factorization Algorithm can be applied whenever $l_{ii} \neq 0$ for each $i = 1, 2, \ldots, n$. Two conditions, either of which ensure that this is true, are that the coefficient matrix of the system is positive definite or that it is strictly diagonally dominant. An additional condition that ensures this algorithm can be applied is given in the next theorem, whose proof is considered in Exercise 28.

Theorem 6.31 Suppose that $A = [a_{ij}]$ is tridiagonal with $a_{i,i-1}a_{i,i+1} \neq 0$, for each $i = 2, 3, \ldots, n-1$. If $|a_{11}| > |a_{12}|$, $|a_{ii}| \geq |a_{i,i-1}| + |a_{i,i+1}|$, for each $i = 2, 3, \ldots, n-1$, and $|a_{nn}| > |a_{n,n-1}|$, then A is nonsingular and the values of l_{ii} described in the Crout Factorization Algorithm are nonzero for each $i = 1, 2, \ldots, n$. ■

The *LinearAlgebra* package of Maple supports a number of commands that test properties for matrices. The return in each case is *true* if the property holds for the matrix and is *false* if it does not hold. For example,

IsDefinite$(A, query = {'positive\_definite'})$

would return *true* for the positive matrix

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

but would return *false* for the matrix

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

Consistent with our definition, symmetry is required for a *true* result.

The *NumericalAnalysis* subpackage also has query commands for matrices. Some of these are

IsMatrixShape$(A, {'diagonal'})$
IsMatrixShape$(A, {'symmetric'})$
IsMatrixShape$(A, {'positivedefinite'})$
IsMatrixShape$(A, {'diagonallydominant'})$
IsMatrixShape$(A, {'strictlydiagonallydominant'})$
IsMatrixShape$(A, {'triangular'}_{'upper'})$
IsMatrixShape$(A, {'triangular'}_{'lower'})$

EXERCISE SET 6.6

1. Determine which of the following matrices are (i) symmetric, (ii) singular, (iii) strictly diagonally dominant, (iv) positive definite.

a. $\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$

b. $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 0 \\ 1 & 0 & 4 \end{bmatrix}$

c. $\begin{bmatrix} 4 & 2 & 6 \\ 3 & 0 & 7 \\ -2 & -1 & -3 \end{bmatrix}$

d. $\begin{bmatrix} 4 & 0 & 0 & 0 \\ 6 & 7 & 0 & 0 \\ 9 & 11 & 1 & 0 \\ 5 & 4 & 1 & 1 \end{bmatrix}$

2. Determine which of the following matrices are (i) symmetric, (ii) singular, (iii) strictly diagonally dominant, (iv) positive definite.

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

b. $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 2 \\ 1 & 2 & 4 \end{bmatrix}$

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

d. $\begin{bmatrix} 2 & 3 & 1 & 2 \\ -2 & 4 & -1 & 5 \\ 3 & 7 & 1.5 & 1 \\ 6 & -9 & 3 & 7 \end{bmatrix}$

3. Use the LDL^t Factorization Algorithm to find a factorizaton of the form $A = LDL^t$ for the following matrices:

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

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

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

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

4. Use the LDL^t Factorization Algorithm to find a factorization of the form $A = LDL^t$ for the following matrices:

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

b. $A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 5 \end{bmatrix}$

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

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

5. Use the Cholesky Algorithm to find a factorization of the form $A = LL^t$ for the matrices in Exercise 3.

6. Use the Cholesky Algorithm to find a factorization of the form $A = LL^t$ for the matrices in Exercise 4.

7. Modify the LDL^t Factorization Algorithm as suggested in the text so that it can be used to solve linear systems. Use the modified algorithm to solve the following linear systems.

a.
$$\begin{aligned} 2x_1 - x_2 \qquad &= 3, \\ -x_1 + 2x_2 - x_3 &= -3, \\ - x_2 + 2x_3 &= 1. \end{aligned}$$

b.
$$\begin{aligned} 4x_1 + x_2 + x_3 + x_4 &= 0.65, \\ x_1 + 3x_2 - x_3 + x_4 &= 0.05, \\ x_1 - x_2 + 2x_3 \qquad &= 0, \\ x_1 + x_2 \qquad + 2x_4 &= 0.5. \end{aligned}$$

c. $$\begin{aligned} 4x_1 + x_2 - x_3 &= 7,\\ x_1 + 3x_2 - x_3 &= 8,\\ -x_1 - x_2 + 5x_3 + 2x_4 &= -4,\\ 2x_3 + 4x_4 &= 6. \end{aligned}$$

d. $$\begin{aligned} 6x_1 + 2x_2 + x_3 - x_4 &= 0,\\ 2x_1 + 4x_2 + x_3 &= 7,\\ x_1 + x_2 + 4x_3 - x_4 &= -1,\\ -x_1 - x_3 + 3x_4 &= -2. \end{aligned}$$

8. Use the modified algorithm from Exercise 7 to solve the following linear systems.

a. $$\begin{aligned} 4x_1 - x_2 + x_3 &= -1,\\ -x_1 + 3x_2 &= 4,\\ x_1 + 2x_3 &= 5. \end{aligned}$$

b. $$\begin{aligned} 4x_1 + 2x_2 + 2x_3 &= 0,\\ 2x_1 + 6x_2 + 2x_3 &= 1,\\ 2x_1 + 2x_2 + 5x_3 &= 0. \end{aligned}$$

c. $$\begin{aligned} 4x_1 + 2x_3 + x_4 &= -2,\\ 3x_2 - x_3 + x_4 &= 0,\\ 2x_1 - x_2 + 6x_3 + 3x_4 &= 7,\\ x_1 + x_2 + 3x_3 + 8x_4 &= -2. \end{aligned}$$

d. $$\begin{aligned} 4x_1 + x_2 + x_3 + x_4 &= 2,\\ x_1 + 3x_2 - x_4 &= 2,\\ x_1 + 2x_3 + x_4 &= 1,\\ x_1 - x_2 + x_3 + 4x_4 &= 1. \end{aligned}$$

9. Modify the Cholesky Algorithm as suggested in the text so that it can be used to solve linear systems, and use the modified algorithm to solve the linear systems in Exercise 7.

10. Use the modified algorithm developed in Exercise 9 to solve the linear systems in Exercise 8.

11. Use Crout factorization for tridiagonal systems to solve the following linear systems.

a. $$\begin{aligned} x_1 - x_2 &= 0,\\ -2x_1 + 4x_2 - 2x_3 &= -1,\\ -x_2 + 2x_3 &= 1.5. \end{aligned}$$

b. $$\begin{aligned} 3x_1 + x_2 &= -1,\\ 2x_1 + 4x_2 + x_3 &= 7,\\ 2x_2 + 5x_3 &= 9. \end{aligned}$$

c. $$\begin{aligned} 2x_1 - x_2 &= 3,\\ -x_1 + 2x_2 - x_3 &= -3,\\ -x_2 + 2x_3 &= 1. \end{aligned}$$

d. $$\begin{aligned} 0.5x_1 + 0.25x_2 &= 0.35,\\ 0.35x_1 + 0.8x_2 + 0.4x_3 &= 0.77,\\ 0.25x_2 + x_3 + 0.5x_4 &= -0.5,\\ x_3 - 2x_4 &= -2.25. \end{aligned}$$

12. Use Crout factorization for tridiagonal systems to solve the following linear systems.

a. $$\begin{aligned} 2x_1 + x_2 &= 3,\\ x_1 + 2x_2 + x_3 &= -2,\\ 2x_2 + 3x_3 &= 0. \end{aligned}$$

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

c. $$\begin{aligned} 2x_1 - x_2 &= 3,\\ x_1 + 2x_2 - x_3 &= 4,\\ x_2 - 2x_3 + x_4 &= 0,\\ x_3 + 2x_4 &= 6. \end{aligned}$$

d. $$\begin{aligned} 2x_1 - x_2 &= 1,\\ x_1 + 2x_2 - x_3 &= 2,\\ 2x_2 + 4x_3 - x_4 &= -1,\\ 2x_4 - x_5 &= -2,\\ x_4 + 2x_5 &= -1. \end{aligned}$$

13. Let A be the 10×10 tridiagonal matrix given by $a_{ii} = 2, a_{i,i+1} = a_{i,i-1} = -1$, for each $i = 2, \cdots, 9$, and $a_{11} = a_{10,10} = 2, a_{12} = a_{10,9} = -1$. Let $\mathbf{b}$ be the ten-dimensional column vector given by $b_1 = b_{10} = 1$ and $b_i = 0$, for each $i = 2, 3, \cdots, 9$. Solve $A\mathbf{x} = \mathbf{b}$ using the Crout factorization for tridiagonal systems.

14. Modify the LDL^t factorization to factor a symmetric matrix A. [*Note:* The factorization may not always be possible.] Apply the new algorithm to the following matrices:

a. $A = \begin{bmatrix} 3 & -3 & 6\\ -3 & 2 & -7\\ 6 & -7 & 13 \end{bmatrix}$

b. $A = \begin{bmatrix} 3 & -6 & 9\\ -6 & 14 & -20\\ 9 & -20 & 29 \end{bmatrix}$

c. $A = \begin{bmatrix} -1 & 2 & 0 & 1\\ 2 & -3 & 2 & -1\\ 0 & 2 & 5 & 6\\ 1 & -1 & 6 & 12 \end{bmatrix}$

d. $A = \begin{bmatrix} 2 & -2 & 4 & -4\\ -2 & 3 & -4 & 5\\ 4 & -4 & 10 & -10\\ -4 & 5 & -10 & 14 \end{bmatrix}$

15. Which of the symmetric matrices in Exercise 14 are positive definite?

16. Find all α so that $A = \begin{bmatrix} \alpha & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 4 \end{bmatrix}$ is positive definite.

17. Find all α so that $A = \begin{bmatrix} 2 & \alpha & -1 \\ \alpha & 2 & 1 \\ -1 & 1 & 4 \end{bmatrix}$ is positive definite.

18. Find all α and $\beta > 0$ so that the matrix

$$A = \begin{bmatrix} 4 & \alpha & 1 \\ 2\beta & 5 & 4 \\ \beta & 2 & \alpha \end{bmatrix}$$

is strictly diagonally dominant.

19. Find all $\alpha > 0$ and $\beta > 0$ so that the matrix

$$A = \begin{bmatrix} 3 & 2 & \beta \\ \alpha & 5 & \beta \\ 2 & 1 & \alpha \end{bmatrix}$$

is strictly diagonally dominant.

20. Suppose that A and B are strictly diagonally dominant $n \times n$ matrices. Which of the following must be strictly diagonally dominant?

a. $-A$ **b.** A^t **c.** $A+B$ **d.** A^2 **e.** $A-B$

21. Suppose that A and B are positive definite $n \times n$ matrices. Which of the following must be positive definite?

a. $-A$ **b.** A^t **c.** $A+B$ **d.** A^2 **e.** $A-B$

22. Let

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

Find all values of α for which

a. A is singular.
b. A is strictly diagonally dominant.
c. A is symmetric.
d. A is positive definite.

23. Let

$$A = \begin{bmatrix} \alpha & 1 & 0 \\ \beta & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

Find all values of α and β for which

a. A is singular.
b. A is strictly diagonally dominant.
c. A is symmetric.
d. A is positive definite.

24. Suppose A and B commute, that is, $AB = BA$. Must A^t and B^t also commute?

25. Construct a matrix A that is nonsymmetric but for which $\mathbf{x}^t A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

26. Show that Gaussian elimination can be performed on A without row interchanges if and only if all leading principal submatrices of A are nonsingular. [*Hint:* Partition each matrix in the equation

$$A^{(k)} = M^{(k-1)}M^{(k-2)}\cdots M^{(1)}A$$

vertically between the kth and $(k+1)$st columns and horizontally between the kth and $(k+1)$st rows (see Exercise 14 of Section 6.3). Show that the nonsingularity of the leading principal submatrix of A is equivalent to $a^{(k)}_{k,k} \neq 0$.]

27. Tridiagonal matrices are usually labeled by using the notation

$$A = \begin{bmatrix} a_1 & c_1 & 0 & \cdots & 0 \\ b_2 & a_2 & c_2 & \ddots & \vdots \\ 0 & b_3 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & c_{n-1} \\ 0 & \cdots & 0 & b_n & a_n \end{bmatrix}$$

to emphasize that it is not necessary to consider all the matrix entries. Rewrite the Crout Factorization Algorithm using this notation, and change the notation of the l_{ij} and u_{ij} in a similar manner.

28. Prove Theorem 6.31. [*Hint:* Show that $|u_{i,i+1}| < 1$, for each $i = 1, 2, \ldots, n-1$, and that $|l_{ii}| > 0$, for each $i = 1, 2, \ldots, n$. Deduce that $\det A = \det L \cdot \det U \neq 0$.]

29. Suppose $V = 5.5$ volts in the lead example of this chapter. By reordering the equations, a tridiagonal linear system can be formed. Use the Crout Factorization Algorithm to find the solution of the modified system.

30. Construct the operation count for solving an $n \times n$ linear system using the Crout Factorization Algorithm.

31. In a paper by Dorn and Burdick [DoB], it is reported that the average wing length that resulted from mating three mutant varieties of fruit flies (*Drosophila melanogaster*) can be expressed in the symmetric matrix form

$$A = \begin{bmatrix} 1.59 & 1.69 & 2.13 \\ 1.69 & 1.31 & 1.72 \\ 2.13 & 1.72 & 1.85 \end{bmatrix},$$

where a_{ij} denotes the average wing length of an offspring resulting from the mating of a male of type i with a female of type j.

a. What physical significance is associated with the symmetry of this matrix?

b. Is this matrix positive definite? If so, prove it; if not, find a nonzero vector $\mathbf{x}$ for which $\mathbf{x}^t A\mathbf{x} \leq 0$.

32. Suppose that the positive definite matrix A has the Cholesky factorization $A = LL^t$ and also the factorization $A = \hat{L}D\hat{L}^t$, where D is the diagonal matrix with positive diagonal entries $d_{11}, d_{22}, \ldots, d_{nn}$. Let $D^{1/2}$ be the diagonal matrix with diagonal entries $\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}}$.

a. Show that $D = D^{1/2}D^{1/2}$.

b. Show that $L = \hat{L}D^{1/2}$.

6.7 Survey of Methods and Software

In this chapter we have looked at direct methods for solving linear systems. A linear system consists of n equations in n unknowns expressed in matrix notation as $A\mathbf{x} = \mathbf{b}$. These techniques use a finite sequence of arithmetic operations to determine the exact solution of the system subject only to round-off error. We found that the linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if A^{-1} exists, which is equivalent to $\det A \neq 0$. When A^{-1} is known, the solution of the linear system is the vector $\mathbf{x} = A^{-1}\mathbf{b}$.

Pivoting techniques were introduced to minimize the effects of round-off error, which can dominate the solution when using direct methods. We studied partial pivoting, scaled partial pivoting, and briefly discussed complete pivoting. We recommend the partial or scaled partial pivoting methods for most problems because these decrease the effects of round-off error without adding much extra computation. Complete pivoting should be used if round-off error is suspected to be large. In Section 5 of Chapter 7 we will see some procedures for estimating this round-off error.

Gaussian elimination with minor modifications was shown to yield a factorization of the matrix A into LU, where L is lower triangular with 1s on the diagonal and U is upper triangular. This process is called Doolittle factorization. Not all nonsingular matrices can be factored this way, but a permutation of the rows will always give a factorization of the form $PA = LU$, where P is the permutation matrix used to rearrange the rows of A. The advantage of the factorization is that the work is significantly reduced when solving linear systems $A\mathbf{x} = \mathbf{b}$ with the same coefficient matrix A and different vectors $\mathbf{b}$.

Factorizations take a simpler form when the matrix A is positive definite. For example, the Choleski factorization has the form $A = LL^t$, where L is lower triangular. A symmetric matrix that has an LU factorization can also be factored in the form $A = LDL^t$, where D is diagonal and L is lower triangular with 1s on the diagonal. With these factorizations, manipulations involving A can be simplified. If A is tridiagonal, the LU factorization takes a particularly simple form, with U having 1s on the main diagonal and 0s elsewhere, except on the diagonal immediately above the main diagonal. In addition, L has its only nonzero entries on the main diagonal and one diagonal below. Another important method of matrix factorization is considered in Section 6 of Chapter 9.

The direct methods are the methods of choice for most linear systems. For tridiagonal, banded, and positive definite matrices, the special methods are recommended. For the general case, Gaussian elimination or LU factorization methods, which allow pivoting, are recommended. In these cases, the effects of round-off error should be monitored. In Section 7.5 we discuss estimating errors in direct methods.

Large linear systems with primarily 0 entries occurring in regular patterns can be solved efficiently using an iterative procedure such as those discussed in Chapter 7. Systems of this type arise naturally, for example, when finite-difference techniques are used to solve boundary-value problems, a common application in the numerical solution of partial-differential equations.

It can be very difficult to solve a large linear system that has primarily nonzero entries or one where the 0 entries are not in a predictable pattern. The matrix associated with the system can be placed in secondary storage in partitioned form and portions read into main memory only as needed for calculation. Methods that require secondary storage can be either iterative or direct, but they generally require techniques from the fields of data structures and graph theory. The reader is referred to [BuR] and [RW] for a discussion of the current techniques.

The software for matrix operations and the direct solution of linear systems implemented in IMSL and NAG is based on LAPACK, a subroutine package in the public domain. There is excellent documentation available with it and from the books written about it. We will focus on several of the subroutines that are available in all three sources.

Accompanying LAPACK is a set of lower-level operations called Basic Linear Algebra Subprograms (BLAS). Level 1 of BLAS generally consists of vector-vector operations such as vector additions with input data and operation counts of $O(n)$. Level 2 consists of the matrix-vector operations such as the product of a matrix and a vector with input data and operation counts of $O(n^2)$. Level 3 consists of the matrix-matrix operations such as matrix products with input data and operation counts of $O(n^3)$.

The subroutines in LAPACK for solving linear systems first factor the matrix A. The factorization depends on the type of matrix in the following way:

1. General matrix $PA = LU$;
2. Positive definite matrix $A = LL^t$;
3. Symmetric matrix $A = LDL^t$;
4. Tridiagonal matrix $A = LU$ (in banded form).

In addition, inverses and determinants can be computed.

Many of the subroutines in LINPACK, and its successor LAPACK, can be implemented using MATLAB. A nonsingular matrix A can be factored into the form $PA = LU$, where P is the permutation matrix defined by performing partial pivoting to solve a linear system involving A. A system of the form $A\mathbf{x} = \mathbf{b}$ is found by solving a lower triangular system followed by the solution to an upper triangular system.

Other MATLAB commands include computing the inverse, transpose, and determinant of matrix A by issuing the commands $\text{inv}(A)$, A', and $\det(A)$, respectively.

The IMSL Library includes counterparts to almost all the LAPACK subroutines and some extensions as well. The NAG Library has numerous subroutines for direct methods of solving linear systems similar to those in LAPACK and IMSL.

Further information on the numerical solution of linear systems and matrices can be found in Golub and Van Loan [GV], Forsythe and Moler [FM], and Stewart [Stew1]. The use of direct techniques for solving large sparse systems is discussed in detail in George and Liu [GL] and in Pissanetzky [Pi]. Coleman and Van Loan [CV] consider the use of BLAS, LINPACK, and MATLAB.