CHAPTER

6 Direct Methods for Solving Linear Systems

6.1 Introduction

Systems of equations are used to represent physical problems that involve the interaction of various properties. The variables in the system represent the properties being studied, and the equations describe the interaction between the variables. The system is easiest to study when the equations are all linear. Often the number of equations is the same as the number of variables, for only in this case is it likely that a unique solution will exist.

Not all physical problems can be reasonably represented using a linear system with the same number of equations as unknowns, but the solutions to many problems either have this form or can be approximated by such a system. In fact, this is quite often the only approach that can give quantitative information about a physical problem.

In this chapter we consider *direct methods* for approximating the solution of a system of n linear equations in n unknowns. A direct method is one that gives the exact solution to the system, if it is assumed that all calculations can be performed without round-off error effects. However, we cannot generally avoid round-off error and we need to consider quite carefully the role of finite-digit arithmetic error in the approximation to the solution to the system, and how to arrange the calculations to minimize its effect.

6.2 Gaussian Elimination

If you have studied linear algebra or matrix theory, you probably have been introduced to Gaussian elimination, the most elementary method for systematically determining the solution of a system of linear equations. Variables are eliminated from the equations until one equation involves only one variable, a second equation involves only that variable and one other, a third has only these two and one additional, and so on. The solution is found by solving for the variable in the single equation, using this to reduce the second equation to one that now contains a single variable, and so on, until values for all the variables are found.

Three operations are permitted on a system of equations $E_1, E_2, \ldots, E_n$.

Operations on Systems of Equations

- 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 $(\lambda E_i) \rightarrow (E_i)$.
- 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 $(E_i + \lambda E_j) \rightarrow (E_i)$.
- Equations E_i and E_j can be transposed in order. This operation is denoted $(E_i) \leftrightarrow (E_j)$.

By a sequence of the operations just given, a linear system can be transformed to a more easily-solved linear system with the same solutions. The sequence of operations is shown in the next Illustration.

Illustration The four equations

$$\begin{aligned}
E_1:&\quad x_1 + x_2 + 3x_4 = 4,\\
E_2:&\quad 2x_1 + x_2 - x_3 + x_4 = 1,\\
E_3:&\quad 3x_1 - x_2 - x_3 + 2x_4 = -3,\\
E_4:&\quad -x_1 + 2x_2 + 3x_3 - x_4 = 4,
\end{aligned} \tag{6.1}$$

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

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

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 $(E_3 - 4E_2) \to (E_3)$ and $(E_4 + 3E_2) \to (E_4)$. This results in

$$\begin{aligned}
E_1:&\quad x_1 + x_2 + 3x_4 = 4,\\
E_2:&\quad - x_2 - x_3 - 5x_4 = -7,\\
E_3:&\quad 3x_3 + 13x_4 = 13,\\
E_4:&\quad - 13x_4 = -13.
\end{aligned} \tag{6.2}$$

The system of equations (6.2) 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 (6.2), and consequently to system (6.1), is therefore $x_1 = -1$, $x_2 = 2$, $x_3 = 0$, and $x_4 = 1$. □

Matrices and Vectors

When performing the calculations of the Illustration, we did 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 because they always remained in the same column. The only variation from system to system occurred 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**, a rectangular array of elements in which not only is the value of an element important, but also its position in the array. The matrix contains all the information about the system that is necessary to determine its solution in a compact form.

The notation for an $n \times m$ (n by 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. Its 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 subscript is omitted for vectors and a boldface lowercase letter is used for notation. So,

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

denotes a column vector, and $\mathbf{y} = [y_1\ y_2\ \cdots y_n]$ denotes a row vector.

A system of n linear equations in the n unknowns $x_1, x_2, \ldots, x_n$ has the form

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

An $n \times (n+1)$ matrix can be used to represent this linear system by first constructing

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

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

and then combining these matrices to form the *augmented matrix*:

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

where the vertical dotted line before the last column is used to separate the coefficients of the unknowns from the values on the right-hand side of the equations.

Repeating the operations involved in the Illustration on page 230 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

$$(E_2 - 2E_1) \to (E_2), \quad (E_3 - 3E_1) \to (E_3), \quad \text{and} \quad (E_4 + E_1) \to (E_4)$$

produces

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

Then

$$(E_3 - 4E_2) \to (E_3) \quad \text{and} \quad (E_4 + 3E_2) \to (E_4),$$

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 in approximately 200 BCE. 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 dwarf planet Ceres.

produces the final matrix

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

This final matrix can be transformed into its corresponding linear system and solutions for x_1, x_2, x_3, and x_4 obtained. The procedure involved in this process is called **Gaussian Elimination with Backward Substitution**.

The general Gaussian elimination procedure applied to the linear system

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

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

$$\tilde{A} = [A, \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & a_{1,n+1} \\ a_{21} & a_{22} & \cdots & a_{2n} & \vdots & a_{2,n+1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \vdots & a_{n,n+1} \end{bmatrix},$$

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

Suppose that $a_{11} \neq 0$. To convert the entries in the first column, below a_{11}, to zero, we perform the operations $(E_k - m_{k1}E_1) \to (E_k)$ for each $k = 2, 3 \ldots, n$ for an appropriate multiplier m_{k1}. We first designate the diagonal element in the column, a_{11} as the **pivot element**. The *multiplier* for the kth row is defined by $m_{k1} = a_{k1}/a_{11}$. Performing the operations $(E_k - m_{k1}E_1) \to (E_k)$ for each $k = 2, 3, \ldots, n$ eliminates (that is, changes to zero) the coefficient of x_1 in each of these rows:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & \vdots & b_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \vdots & b_n \end{bmatrix} \begin{matrix} E_2 - m_{21}E_1 \to E_2 \\ E_3 - m_{31}E_1 \to E_3 \\ \vdots \\ E_n - m_{n1}E_1 \to E_n \end{matrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & b_1 \\ 0 & a_{22} & \cdots & a_{2n} & \vdots & b_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} & \vdots & b_n \end{bmatrix}.$$

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}.

If the pivot element $a_{22} \neq 0$, we form the multipliers $m_{k2} = a_{k2}/a_{22}$ and perform the operations $(E_k - m_{k2}E_2) \to E_k$ for each $k = 3, \ldots, n$ obtaining

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & b_1 \\ 0 & a_{22} & \cdots & a_{2n} & \vdots & b_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} & \vdots & b_n \end{bmatrix} \begin{matrix} E_3 - m_{32}E_2 \to E_3 \\ \vdots \\ E_n - m_{n2}E_2 \to E_n \end{matrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & b_1 \\ 0 & a_{22} & \cdots & a_{2n} & \vdots & b_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn} & \vdots & b_n \end{bmatrix}$$

We then follow this sequential procedure for the rows $i = 3 \ldots, n-1$. Define the multiplier $m_{ki} = a_{ki}/a_{ii}$ and perform the operation

$$(E_k - m_{ki}E_i) \to (E_k)$$

for each $k = i+1, i+2, \ldots, n$, provided the pivot element a_{ii} is nonzero. This eliminates x_i in each row below the ith for all values of $i = 1, 2, \ldots, n-1$. The resulting matrix has the form

$$\tilde{\tilde{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & a_{1,n+1} \\ 0 & a_{22} & \cdots & a_{2n} & \vdots & a_{2,n+1} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{nn} & \vdots & a_{n,n+1} \end{bmatrix},$$

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, that is,

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

Backward substitution can be performed on this system. Solving the nth equation for x_n gives

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

Then 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

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

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

The procedure will fail if at the ith step the pivot element a_{ii} is zero, for then either the multipliers $m_{ki} = a_{ki}/a_{ii}$ are not defined (this occurs if $a_{ii} = 0$ for some $i < n$) or the backward substitution cannot be performed (if $a_{nn} = 0$). This does not necessarily mean that the system has no solution, but rather that the technique for finding the solution must be altered by interchanging rows when a pivot is 0.

The program GAUSEL61 implements Gaussian Elimination with Backward Substitution.

Program GAUSEL61 incorporates row interchanges when required. 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

$$\tilde{A} = \left[\begin{array}{cccc:c} 1 & -1 & 2 & -1 & -8 \\ 2 & -2 & 3 & -3 & -20 \\ 1 & 1 & 1 & 0 & -2 \\ 1 & -1 & 4 & 3 & 4 \end{array}\right].$$

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 the matrix

$$\left[\begin{array}{cccc:c} 1 & -1 & 2 & -1 & -8 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 2 & 4 & 12 \end{array}\right].$$

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

The new diagonal entry a_{22}, 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} and a_{42} for the first nonzero element. Since $a_{32} \neq 0$, the operation $(E_2) \leftrightarrow (E_3)$ is performed to obtain a new matrix,

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

Since x_2 is already eliminated from E_3 and E_4, the computations continue with the operation $(E_4 + 2E_3) \to (E_4)$, giving

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

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

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

■

To define the initial augmented matrix in MATLAB, which we will call AA, we enter the matrix row by row. A space is placed between each entry in a row, and the rows in AA are separated by a colon. So, for the matrix in Example 2 we have

```
AA = [1 -1 2 -1 -8; 2 -2 3 -3 -20; 1 1 1 0 -2; 1 -1 4 3 4]
```

MATLAB responds with

$$AA = \begin{bmatrix} 1 & -1 & 2 & -1 & -8 \\ 2 & -2 & 3 & -3 & -20 \\ 1 & 1 & 1 & 0 & -2 \\ 1 & -1 & 4 & 3 & 4 \end{bmatrix}$$

To perform the operation $(E_j + mE_i) \to (E_j)$ in MATLAB we use the command

```
AA(j,:) = AA(j,:) + m * AA(i,:)
```

The notation `AA(k,l)` refers to entry in the kth row and lth column. The use of `:` in MATLAB refers to multiplying an entire row or column. For example, multiplying the kth row by m is done with `m *AA(k,:)`. Similarly, multiplying the lth column by m would be done with `m*AA(:,l)`. So the next command subtracts twice the first row of AA from the second row.

```
AA(2,:) = AA(2,:)-2*AA(1,:)
```

which gives

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

We then subtract the first row of AA from the third row, followed by the subtraction of the first row from the fourth row with

```
AA(3,:) = AA(3,:)-AA(1,:)
```

and

```
AA(4,:) = AA(4,:)-AA(1,:)
```

This gives

$$AA = \begin{bmatrix} 1 & -1 & 2 & -1 & -8 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 2 & 4 & 12 \end{bmatrix}$$

The variable x_1 has now been eliminated from the rows corresponding to the second, third, and fourth equations. Since a_{22} is zero, we need to interchange rows to move a nonzero entry to a_{22}. To interchange rows 2 and 3, we store row 2 in a temporary row vector B, move row 3 to row 2, and then more the temporary row vector B to row 3. This is done with

```
B = AA(2,:)
AA(2,:) = AA(3,:)
AA(3,:) = B
```

The result is

$$AA = \begin{bmatrix} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 2 & 4 & 12 \end{bmatrix}$$

The final operation in Gaussian elimination for this matrix is to add 2 times the third row to the fourth row with

```
AA(4,:) = AA(4,:)+2*AA(3,:)
```

This produces

$$AA = \begin{bmatrix} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix}$$

To perform the backward substitution we need to define the vector **x** that will contain the solution. We initialize a vector **x** as the **0** vector and will replace these entries as we progress through the backward substitution.

```
x = [0 0 0 0]
```

Now we replace the 0 in the fourth column of **x** with

```
x(4) = AA(4,5)/AA(4,4)
```

which gives

$$x = 0\,0\,0\,2$$

Then

```
x(3) = (AA(3,5) - AA(3,4)*x(4))/AA(3,3)
```

gives

$$x = 0\,0\,2\,2$$

```
x(2) = (AA(2,5) - (AA(2,3)*x(3)+AA(2,4)*x(4)))/AA(2,2)
```

gives

$$x = 0\,3\,2\,2$$

and

```
x(1) = (AA(1,5) - (AA(1,2)*x(2)+AA(1,3)*x(3)+AA(1,4)*x(4)))/AA(1,1)
```

gives the final solution

$$x = -7\,3\,2\,2$$

which corresponds to $x_1 = -7$, $x_2 = 3$, $x_3 = 2$, and $x_4 = 2$.

Example 2 illustrates what is done if one of the pivot elements is zero. If the ith pivot element is zero, the ith column of the matrix is searched from the ith row downward for the first nonzero entry, and a row interchange is performed to obtain the new matrix. Then the procedure continues as before. If no nonzero entry is found the procedure stops, and the linear system does not have a unique solution. It might have no solution or an infinite number of solutions.

Operation Counts

The computations in the program are performed using only one $n \times (n+1)$ array for storage. This is done by replacing, at each step, the previous value of a_{ij} by the new one. In addition, the multipliers are stored in the locations of a_{ki} known to have zero values—that is, when $i < n$ and $k = i+1, i+2, \ldots, n$. Thus, the original matrix A is overwritten by the multipliers below the main diagonal and by the nonzero entries of the final reduced matrix on and above the main diagonal. We will see in Section 6.5 that these values can be used to solve other linear systems involving the original matrix A.

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. Even though the actual differences in execution time depend on the particular computing system being used, the count of the additions/subtractions are kept separate from the count of the multiplications/divisions because of the time differential. The total number of arithmetic operations depends on the size n, as follows:

$$\text{Multiplications/divisions:} \quad \frac{n^3}{3} + n^2 - \frac{n}{3}.$$

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

For large n, the total number of multiplications and divisions is approximately $n^3/3$, that is, $O(n^3)$, as is the total number of additions and subtractions. The amount of computation, and the time required to perform it, increases with n in approximate proportion to $n^3/3$, 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 6.2

1. Obtain a solution by graphical methods of the following linear systems, if possible.

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

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

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

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

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

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

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

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

2. Use Gaussian elimination 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. $4x_1 - x_2 + x_3 = 8,$
   $2x_1 + 5x_2 + 2x_3 = 3,$
   $x_1 + 2x_2 + 4x_3 = 11.$

   b. $4x_1 + x_2 + 2x_3 = 9,$
   $2x_1 + 4x_2 - x_3 = -5,$
   $x_1 + x_2 - 3x_3 = -9.$

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

   a. $x_1 - x_2 + 3x_3 = 2,$
   $3x_1 - 3x_2 + x_3 = -1,$
   $x_1 + x_2 \quad\quad = 3.$

   b. $2x_1 - 1.5x_2 + 3x_3 = 1,$
   $-x_1 \quad\quad + 2x_3 = 3,$
   $4x_1 - 4.5x_2 + 5x_3 = 1.$

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

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

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

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

4. Use MATLAB with `long format` and Gaussian elimination 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. $$\begin{aligned} 2x_1 + x_2 - x_3 + x_4 - 3x_5 &= 7,\\ x_1 \qquad + 2x_3 - x_4 + x_5 &= 2,\\ -2x_2 - x_3 + x_4 - x_5 &= -5,\\ 3x_1 + x_2 - 4x_3 \qquad + 5x_5 &= 6,\\ x_1 - x_2 - x_3 - x_4 + x_5 &= 3. \end{aligned}$$

5. 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.

6. 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.

7. 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 $j = 1, \ldots, n$; b_i represent the available daily

supply of the ith food; and a_{ij} represent the amount of the ith food consumed on 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

$$A = [a_{ij}] = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

$\mathbf{x} = (x_j) = [1000, 500, 350, 400]$, 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?

8. 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

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

are solved for $u(x_0), u(x_1), \ldots, u(x_m)$. The integrals are approximated using quadrature formulas based on the nodes $x_0, \ldots, x_m$. 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 (b) using the Composite Simpson's rule.

6.3 Pivoting Strategies

If all the calculations could be done using exact arithmetic, we could almost end the chapter with the previous section. We now know how many calculations are needed to perform Gaussian elimination on a system, and from this we should be able to determine whether