The augmented matrix is [A|B] and the system AX = B is represented as follows:

$$[\boldsymbol{A}|\boldsymbol{B}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1N} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2N} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} & b_N \end{array}\right]. \tag{7}$$

The system $\boldsymbol{AX} = \boldsymbol{B}$, with augmented matrix given in (7), can be solved by performing row operations on the augmented matrix $[\boldsymbol{A}|\boldsymbol{B}]$. The variables x_k are placeholders for the coefficients and can be omitted until the end of the calculation.

Theorem 3.8 (Elementary Row Operations). The following operations applied to the augmented matrix (7) yield an equivalent linear system.

(8) Interchanges: The order of two rows can be changed.

(9) Scaling: Multiplying a row by a nonzero constant.

(10) Replacement: The row can be replaced by the sum of that row and a nonzero multiple of any other row; that is:
$\text{row}_r = \text{row}_r - m_{rp} \times \text{row}_p$.

It is common to use (10) by replacing a row with the difference of that row and a multiple of another row.

Definition 3.3. The number a_{rr} in the coefficient matrix $\boldsymbol{A}$ that is used to eliminate a_{kr}, where $k = r + 1, r + 2, \ldots, N$, is called the rth ***pivotal element***, and the rth row is called the ***pivot row***. ▲

The following example illustrates how to use the operations in Theorem 3.8 to obtain an equivalent upper-triangular system $\boldsymbol{UX} = \boldsymbol{Y}$ from a linear system $\boldsymbol{AX} = \boldsymbol{B}$, where $\boldsymbol{A}$ is an $N \times N$ matrix.

Example 3.16. Express the following system in augmented matrix form and find an equivalent upper-triangular system and the solution.

$$\begin{aligned} x_1 + 2x_2 + x_3 + 4x_4 &= 13 \\ 2x_1 + 0x_2 + 4x_3 + 3x_4 &= 28 \\ 4x_1 + 2x_2 + 2x_3 + x_4 &= 20 \\ -3x_1 + x_2 + 3x_3 + 2x_4 &= 6. \end{aligned}$$

The augmented matrix is

$$\begin{array}{l} \text{pivot} \rightarrow \\ m_{21} = 2 \\ m_{31} = 4 \\ m_{41} = -3 \end{array} \left[\begin{array}{rrrr|r} \underline{1} & 2 & 1 & 4 & 13 \\ 2 & 0 & 4 & 3 & 28 \\ 4 & 2 & 2 & 1 & 20 \\ -3 & 1 & 3 & 2 & 6 \end{array}\right].$$

The first row is used to eliminate elements in the first column below the diagonal. We refer to the first row as the *pivotal row* and the element $a_{11} = 1$ is called the *pivotal element*. The values m_{k1} are the multiples of row 1 that are to be subtracted from row k for $k = 2, 3, 4$. The result after elimination is

$$\begin{array}{l} \\ \text{pivot} \rightarrow \\ m_{32} = 1.5 \\ m_{42} = -1.75 \end{array} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & 4 & 13 \\ 0 & \underline{-4} & 2 & -5 & 2 \\ 0 & -6 & -2 & -15 & -32 \\ 0 & 7 & 6 & 14 & 45 \end{array}\right].$$

The second row is used to eliminate elements in the second column that lie below the diagonal. The second row is the pivotal row and the values m_{k2} are the multiples of row 2 that are to be subtracted from row k for $k = 3, 4$. The result after elimination is

$$\begin{array}{l} \\ \\ \text{pivot} \rightarrow \\ m_{43} = -1.9 \end{array} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & 4 & 13 \\ 0 & -4 & 2 & -5 & 2 \\ 0 & 0 & \underline{-5} & -7.5 & -35 \\ 0 & 0 & 9.5 & 5.25 & 48.5 \end{array}\right].$$

Finally, the multiple $m_{43} = -1.9$ of the third row is subtracted from the fourth row, and the result is the upper-triangular system

$$\left[\begin{array}{rrrr|r} 1 & 2 & 1 & 4 & 13 \\ 0 & -4 & 2 & -5 & 2 \\ 0 & 0 & -5 & -7.5 & -35 \\ 0 & 0 & 0 & -9 & -18 \end{array}\right]. \tag{11}$$

The back-substitution algorithm can be used to solve (11), and we get

$$x_4 = 2, \quad x_3 = 4, \quad x_2 = -1, \quad x_1 = 3. \qquad \blacksquare$$

The process described above is called ***Gaussian elimination*** and must be modified so that it can be used in most circumstances. If $a_{kk} = 0$, row k cannot be used to eliminate the elements in column k, and row k must be interchanged with some row below the diagonal to obtain a nonzero pivot element. If this cannot be done, then the coefficient matrix of the system of linear equations is singular, and the system does not have a unique solution.

Theorem 3.9 (Gaussian Elimination with Back Substitution). If $\boldsymbol{A}$ is an $N \times N$ nonsingular matrix, then there exists a system $\boldsymbol{UX} = \boldsymbol{Y}$, equivalent to $\boldsymbol{AX} = \boldsymbol{B}$, where $\boldsymbol{U}$ is an upper-triangular matrix with $u_{kk} \neq 0$. After $\boldsymbol{U}$ and $\boldsymbol{Y}$ are constructed, back substitution can be used to solve $\boldsymbol{UX} = \boldsymbol{Y}$ for $\boldsymbol{X}$.

Proof. We will use the augmented matrix with $\boldsymbol{B}$ stored in column $N+1$:

$$\boldsymbol{AX} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1N}^{(1)} \\ a_{21}^{(1)} & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2N}^{(1)} \\ a_{31}^{(1)} & a_{32}^{(1)} & a_{33}^{(1)} & \cdots & a_{3N}^{(1)} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{N1}^{(1)} & a_{N2}^{(1)} & a_{N3}^{(1)} & \cdots & a_{NN}^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} a_{1\,N+1}^{(1)} \\ a_{2\,N+1}^{(1)} \\ a_{3\,N+1}^{(1)} \\ \vdots \\ a_{N\,N+1}^{(1)} \end{bmatrix} = \boldsymbol{B}.$$

Then we will construct an equivalent upper-triangular system $\boldsymbol{UX} = \boldsymbol{Y}$:

$$\boldsymbol{UX} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1N}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2N}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3N}^{(3)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{NN}^{(N)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} a_{1\,N+1}^{(1)} \\ a_{2\,N+1}^{(2)} \\ a_{3\,N+1}^{(3)} \\ \vdots \\ a_{N\,N+1}^{(N)} \end{bmatrix} = \boldsymbol{Y}.$$

Step 1. Store the coefficients in the augmented matrix. The superscript on $a_{rc}^{(1)}$ means that this is the first time that a number is stored in location (r, c):

$$\left[\begin{array}{ccccc|c} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1N}^{(1)} & a_{1\,N+1}^{(1)} \\ a_{21}^{(1)} & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2N}^{(1)} & a_{2\,N+1}^{(1)} \\ a_{31}^{(1)} & a_{32}^{(1)} & a_{33}^{(1)} & \cdots & a_{3N}^{(1)} & a_{3\,N+1}^{(1)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ a_{N1}^{(1)} & a_{N2}^{(1)} & a_{N3}^{(1)} & \cdots & a_{NN}^{(1)} & a_{N\,N+1}^{(1)} \end{array}\right].$$

Step 2. If necessary, switch rows so that $a_{11}^{(1)} \neq 0$; then eliminate x_1 in rows 2 through N. In this process, m_{r1} is the multiple of row 1 that is subtracted from row r.

$$\begin{array}{l} \text{for } r = 2 : N \\ \quad m_{r1} = a_{r1}^{(1)} / a_{11}^{(1)}; \\ \quad a_{r1}^{(2)} = 0; \\ \quad \text{for } c = 2 : N+1 \\ \quad\quad a_{rc}^{(2)} = a_{rc}^{(1)} - m_{r1} * a_{1c}^{(1)}; \\ \quad \text{end} \\ \text{end} \end{array}$$

The new elements are written $a_{rc}^{(2)}$ to indicate that this is the second time that a number has been stored in the matrix at location (r, c). The result after step 2 is

$$\left[\begin{array}{ccccc|c} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1N}^{(1)} & a_{1\,N+1}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2N}^{(2)} & a_{2\,N+1}^{(2)} \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & \cdots & a_{3N}^{(2)} & a_{3\,N+1}^{(2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & a_{N2}^{(2)} & a_{N3}^{(2)} & \cdots & a_{NN}^{(2)} & a_{N\,N+1}^{(2)} \end{array}\right].$$

Step 3. If necessary, switch the second row with some row below it so that $a_{22}^{(2)} \neq 0$; then eliminate x_2 in rows 3 through N. In this process, m_{r2} is the multiple of row 2 that is subtracted from row r.

$$\begin{array}{l} \text{for } r = 3 : N \\ \quad m_{r2} = a_{r2}^{(2)}/a_{22}^{(2)}; \\ \quad a_{r2}^{(3)} = 0; \\ \quad \text{for } c = 3 : N + 1 \\ \qquad a_{rc}^{(3)} = a_{rc}^{(2)} - m_{r2} * a_{2c}^{(2)}; \\ \quad \text{end} \\ \text{end} \end{array}$$

The new elements are written $a_{rc}^{(3)}$ to indicate that this is the third time that a number has been stored in the matrix at location (r, c). The result after step 3 is

$$\left[\begin{array}{ccccc|c} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1N}^{(1)} & a_{1\,N+1}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2N}^{(2)} & a_{2\,N+1}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3N}^{(3)} & a_{3\,N+1}^{(3)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & a_{N3}^{(3)} & \cdots & a_{NN}^{(3)} & a_{N\,N+1}^{(3)} \end{array}\right].$$

Step $p + 1$. This is the general step. If necessary, switch row p with some row beneath it so that $a_{pp}^{(p)} \neq 0$; then eliminate x_p in rows $p + 1$ through N. Here m_{rp} is the multiple of row p that is subtracted from row r.

$$\begin{array}{l} \text{for } r = p + 1 : N \\ \quad m_{rp} = a_{rp}^{(p)}/a_{pp}^{(p)}; \\ \quad a_{rp}^{(p+1)} = 0; \end{array}$$

```
        for c = p + 1 : N + 1
            a_rc^(p+1) = a_rc^(p) - m_rp * a_pc^(p);
        end
    end
```

The final result after x_{N-1} has been eliminated from row N is

$$\left[\begin{array}{ccccc|c} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1N}^{(1)} & a_{1\,N+1}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2N}^{(2)} & a_{2\,N+1}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3N}^{(3)} & a_{3\,N+1}^{(3)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{NN}^{(N)} & a_{N\,N+1}^{(N)} \end{array}\right].$$

The upper-triangularization process is now complete.

Since $\boldsymbol{A}$ is nonsingular, when row operations are performed the successive matrices are also nonsingular. This guarantees that $a_{kk}^{(k)} \neq 0$ for all k in the construction process. Hence back substitution can be used to solve $\boldsymbol{UX} = \boldsymbol{Y}$ for $\boldsymbol{X}$, and the theorem is proved. •

Pivoting to Avoid $a_{pp}^{(p)} = 0$

If $a_{pp}^{(p)} = 0$, row p cannot be used to eliminate the elements in column p below the main diagonal. It is necessary to find row k, where $a_{kp}^{(p)} \neq 0$ and $k > p$, and then interchange row p and row k so that a nonzero pivot element is obtained. This process is called *pivoting*, and the criterion for deciding which row to choose is called a *pivoting strategy*. The ***trivial pivoting*** strategy is as follows. If $a_{pp}^{(p)} \neq 0$, do not switch rows. If $a_{pp}^{(p)} = 0$, locate the first row below p in which $a_{kp}^{(p)} \neq 0$ and switch rows k and p. This will result in a new element $a_{pp}^{(p)} \neq 0$, which is a nonzero pivot element.

Pivoting to Reduce Error

Because the computer uses fixed-precision arithmetic, it is possible that a small error will be introduced each time that an arithmetic operation is performed. The following example illustrates how use of the trivial pivoting strategy in Gaussian elimination can lead to significant error in the solution of a linear system of equations.

Example 3.17. The values $x_1 = x_2 = 1.000$ are the solutions to

$$\begin{aligned} 1.133x_1 + 5.281x_2 &= 6.414 \\ 24.14x_1 - 1.210x_2 &= 22.93. \end{aligned} \tag{12}$$

Use four-digit arithmetic (see Exercises 6 and 7 in Section 1.3) and Gaussian elimination with trivial pivoting to find a computed approximate solution to the system.

The multiple $m_{21} = 24.14/1.133 = 21.31$ of row 1 is to be subtracted from row 2 to obtain the upper-triangular system. Using four digits in the calculations, we obtain the new coefficients

$$\begin{aligned} a_{22}^{(2)} &= -1.210 - 21.31(5.281) = -1.210 - 112.5 = -113.7 \\ a_{23}^{(2)} &= 22.93 - 21.31(6.414) = 22.93 - 136.7 = -113.8. \end{aligned}$$

The computed upper-triangular system is

$$\begin{aligned} 1.133x_1 + 5.281x_2 &= 6.414 \\ -113.7x_2 &= -113.8. \end{aligned}$$

Back substitution is used to compute $x_2 = -113.8/(-113.7) = 1.001$, and $x_1 = (6.414 - 5.281(1.001))/(1.133) = (6.414 - 5.286)/1.133 = 0.9956$. ∎

The error in the solution of the linear system (12) is due to the magnitude of the multiplier $m_{21} = 21.31$. In the next example the magnitude of the multiplier m_{21} is reduced by first interchanging the first and second equations in the linear system (12) and then using the trivial pivoting strategy in Gaussian elimination to solve the system.

Example 3.18. Use four-digit arithmetic and Gaussian elimination with trivial pivoting to solve the linear system

$$\begin{aligned} 24.14x_1 - 1.210x_2 &= 22.93 \\ 1.133x_1 + 5.281x_2 &= 6.414. \end{aligned}$$

This time $m_{21} = 1.133/24.14 = 0.04693$ is the multiple of row 1 that is to be subtracted from row 2. The new coefficients are

$$\begin{aligned} a_{22}^{(2)} &= 5.281 - 0.04693(-1.210) = 5.281 + 0.05679 = 5.338 \\ a_{23}^{(2)} &= 6.414 - 0.04693(22.93) \quad = 6.414 - 1.076 \quad = 5.338. \end{aligned}$$

The computed upper-triangular system is

$$\begin{aligned} 24.14x_1 - 1.210x_2 &= 22.93 \\ 5.338x_2 &= 5.338. \end{aligned}$$

Back substitution is used to compute $x_2 = 5.338/5.338 = 1.000$, and $x_1 = (22.93 + 1.210(1.000))/24.14 = 1.000$. ∎

The purpose of a pivoting strategy is to move the entry of greatest magnitude to the main diagonal and then use it to eliminate the remaining entries in the column. If there is more than one nonzero element in column p that lies on or below the main diagonal, then there is a choice to determine which rows to interchange. The ***partial pivoting*** strategy, illustrated in Example 3.18, is the most common one and is used in Program 3.2. To reduce the propagation of error, it is suggested that one check the magnitude of all the elements in column p that lie on or below the main diagonal. Locate row k in which the element that has the largest absolute value lies, that is,

$$|a_{kp}| = \max\{|a_{pp}|, |a_{p+1\,p}|, \ldots, |a_{N-1\,p}|, |a_{Np}|\},$$

and then switch row p with row k if $k > p$. Now, each of the multipliers m_{rp} for $r = p+1, \ldots, N$ will be less than or equal to 1 in absolute value. This process will usually keep the relative magnitudes of the elements of the matrix $\boldsymbol{U}$ in Theorem 3.9 the same as those in the original coefficient matrix $\boldsymbol{A}$. Usually, the choice of the larger pivot element in partial pivoting will result in a smaller error being propagated.

In Section 3.5 we will find that it takes a total of $(4N^3 + 9N^2 - 7N)/6$ arithmetic operations to solve an $N \times N$ system. When $N = 20$, the total number of arithmetic operations that must be performed is 5910, and the propagation of error in the computations could result in an erroneous answer. The technique of ***scaled partial pivoting*** or equilibrating can be used to further reduce the effect of error propagation. In scaled partial pivoting we search all the elements in column p that lie on or below the main diagonal for the one that is largest relative to the entries in its row. First search rows p through N for the largest element in magnitude in each row, say s_r:

$$(13) \qquad s_r = \max\{|a_{rp}|, |a_{rp+1}|, \ldots, |a_{rN}|\} \qquad \text{for } r = p,\ p+1,\ \ldots,\ N.$$

The pivotal row k is determined by finding

$$(14) \qquad \frac{|a_{kp}|}{s_k} = \max\left\{\frac{|a_{pp}|}{s_p}, \frac{|a_{p+1\,p}|}{s_{p+1}}, \ldots, \frac{|a_{Np}|}{s_N}\right\}.$$

Now interchange row p and k, unless $p = k$. Again, this pivoting process is designed to keep the relative magnitudes of the elements in the matrix $\boldsymbol{U}$ in Theorem 3.9 the same as those in the original coefficient matrix $\boldsymbol{A}$.

Numerical Methods Using Matlab, 4th Edition, 2004

John H. Mathews and Kurtis K. Fink

ISBN: 0-13-065248-2

Prentice-Hall Inc.
Upper Saddle River, New Jersey, USA
http://vig.prenhall.com/