Chapter 3

System of Linear Equations

Dr. Mohamed Abdelwahed
King Saud University

Definition 3.1 (Linear equation)

It is an equation in which the highest exponent in a variable term is no more than one. The graph of such equation is a straight line. •

For example, the equations

$$5x_1 + 2x_2 = 2, \qquad \frac{4}{5}x_1 + 2x_2 = 1, \qquad 2x_1 - 4x_2 = \pi,$$

Definition 3.2 (System of Linear Equations)

A system of linear equations (or linear system) is simply a finite set of linear equations.

For example,

$$\begin{array}{rcrcr} 4x_1 & - & 2x_2 & = & 5 \\ 3x_1 & + & 2x_2 & = & 4 \end{array}$$

is a system of two equations in two variables x_1 and x_2, and

$$\begin{array}{rcrcrcrcr} 2x_1 & + & x_2 & - & 5x_3 & + & 2x_4 & = & 9 \\ 4x_1 & + & 3x_2 & + & 2x_3 & + & 4x_4 & = & 3 \\ x_1 & + & 2x_2 & + & 3x_3 & + & 2x_4 & = & 11 \end{array}$$

In order to write a general system of m linear equations in the n variables $x_1, \ldots, x_n$, we have

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \cdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{3.1}$$

or, in compact form the system (3.1) can be written

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \qquad i = 1, 2, \ldots, m. \tag{3.2}$$

The system of linear equations (3.3) can be written as the single matrix equation

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}. \qquad \text{can be written} \qquad A\mathbf{x} = \mathbf{b},$$

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

3.4 Direct Numerical Methods for Linear Systems

3.4.1 Gaussian Elimination Method

$$\begin{array}{ccccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\
a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
a_{n1}x_1 & + & a_{n2}x_2 & & a_{n3}x_3 & + & \cdots & + & a_{nn}x_n & = & b_n
\end{array}$$

Simple Gaussian Elimination Method

Forward Elimination

Consider first equation of the given system (3.14)

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1,$$

as first pivotal equation with first pivot element a_{11}. Then the first equation times multiples $m_{i1} = (a_{i1}/a_{11})$, $i = 2, 3, \ldots, n$, is subtracted from the *ith* equation to eliminate first variable x_1, producing an equivalent system

$$\begin{array}{ccccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
 & & a_{22}^{(1)}x_2 & + & a_{23}^{(1)}x_3 & + & \cdots & + & a_{2n}^{(1)}x_n & = & b_2^{(1)} \\
 & & a_{32}^{(1)}x_2 & + & a_{33}^{(1)}x_3 & + & \cdots & + & a_{3n}^{(1)}x_n & = & b_3^{(1)} \\
 & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
 & & a_{n2}^{(1)}x_2 & + & a_{n3}^{(1)}x_3 & + & \cdots & + & a_{nn}^{(1)}x_n & = & b_n^{(1)}
\end{array} \qquad (3.16)$$

Now consider a second equation of the system (3.16), which is

$$a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \cdots + a_{2n}^{(1)}x_n = b_2^{(1)}, \tag{3.17}$$

as second pivotal equation with second pivot element $a_{22}^{(1)}$. Then the second equation times multiples $m_{i2} = (a_{i2}^{(1)}/a_{22}^{(1)}),\ i = 3, \ldots, n$, is subtracted from the ith equation to eliminate second variable x_2, producing an equivalent system

$$\begin{array}{cccccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
 & & a_{22}^{(1)}x_2 & + & a_{23}^{(1)}x_3 & + & \cdots & + & a_{2n}^{(1)}x_n & = & b_2^{(1)} \\
 & & & & a_{33}^{(2)}x_3 & + & \cdots & + & a_{3n}^{(2)}x_n & = & b_3^{(2)} \\
 & & & & \vdots & & \vdots & & \vdots & & \vdots \\
 & & & & a_{n3}^{(2)}x_3 & + & \cdots & + & a_{nn}^{(2)}x_n & = & b_n^{(2)}
\end{array} \tag{3.18}$$

Similarly, after *(n-1)th* steps, we have the *nth* pivotal equation which have only one unknown variable x_n, that is

$$\begin{array}{cccccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
 & + & a_{22}^{(1)}x_2 & + & a_{23}^{(1)}x_3 & + & \cdots & + & a_{2n}^{(1)}x_n & = & b_2^{(1)} \\
 & & & + & a_{33}^{(2)}x_3 & + & \cdots & + & a_{3n}^{(2)}x_n & = & b_3^{(2)} \\
 & & & & & & & & \vdots & & \vdots \\
 & & & & & & & & a_{nn}^{(n-1)}x_n & = & b_n^{(n-1)}
\end{array} \tag{3.20}$$

The Gaussian elimination can be carried out by writing only the coefficients and the right-hand side terms in a matrix form, which means the augmented matrix form.

$$\left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & b_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & b_n \end{array}\right)$$

Consequently system (3.20) is now written directly as follows:

$$\left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ & & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} & b_3^{(2)} \\ & & & \vdots & \vdots & \\ & & & & a_{nn}^{(n-1)} & b_n^{(n-1)} \end{array}\right),$$

Backward Substitution

$$\left.\begin{aligned}
x_n &= \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}} \\
x_{n-1} &= \frac{1}{a_{n-1n-1}^{(n-2)}}\left(b_{n-1}^{(n-2)} - a_{n-1n}^{(n-2)} x_n\right) \\
&\vdots \\
x_1 &= \frac{1}{a_{11}}\left(b_1 - \sum_{j=2}^{n} a_{1j} x_j\right)
\end{aligned}\right\}$$

Example 3.9 *Solve the following linear system using the simple Gaussian elimination method*

$$\begin{aligned}
x_1 + 2x_2 + x_3 &= 2 \\
2x_1 + 5x_2 + 2x_3 &= 1 \\
x_1 + 3x_2 + 4x_3 &= 5
\end{aligned}$$

solution. *The process begins with the augmented matrix form*

$$\begin{pmatrix} 1 & 2 & 1 & \vdots & 2 \\ 2 & 5 & 3 & \vdots & 1 \\ 1 & 3 & 4 & \vdots & 5 \end{pmatrix}.$$

Since $a_{11} = 1 \neq 0$, *In this case the multiples are* $m_{21} = \frac{2}{1} = 2$ *and* $m_{31} = \frac{1}{1} = 1$. *Hence*

$$\left(\begin{array}{ccccc} 1 & 2 & 1 & \vdots & 2 \\ 0 & 1 & 1 & \vdots & -3 \\ 0 & 1 & 3 & \vdots & 3 \end{array}\right).$$

As $a_{22}^{(1)} = 1 \neq 0$, *therefore, we eliminate entry in* $a_{32}^{(1)}$ *position by subtracting the multiple* $m_{32} = \frac{1}{1} = 1$ *of the second row from the third row, to get*

$$\left(\begin{array}{ccccc} 1 & 2 & 1 & \vdots & 2 \\ 0 & 1 & 1 & \vdots & -3 \\ 0 & 0 & 2 & \vdots & 6 \end{array}\right).$$

Now expressing the set in algebraic form yields

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & x_3 & = & 2 \\ & & x_2 & + & x_3 & = & -3 \\ & & & & 2x_3 & = & 6 \end{array}$$

Now using backward substitution, we get

$$\begin{array}{rcllll} 2x_3 & = & 6, & gives & x_3 = 3, \\ x_2 & = & -x_3 - 3 = -(3) - 3 = -6, & gives & x_2 = -6, \\ x_1 & = & 2 - 2x_2 - x_3 = 2 - 2(-6) - 3, & gives & x_1 = 11, \end{array}$$

Example 3.11 *Solve the linear system using the simple Gaussian elimination method*

$$\begin{array}{rcrcrcl} x_1 & + & x_2 & + & x_3 & = & 3 \\ 2x_1 & + & 2x_2 & + & 3x_3 & = & 7 \\ x_1 & + & 2x_2 & + & 3x_3 & = & 6 \end{array}$$

Solution. *Writing the given system in the augmented matrix form*

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 2 & 2 & 3 & 7 \\ 1 & 2 & 3 & 6 \end{array}\right).$$

First elimination step is to eliminate the elements $a_{21} = 2$ *and* $a_{31} = 1$ *from second and third rows by subtracting multiples* $m_{21} = \frac{2}{1} = 2$ *and* $m_{31} = \frac{1}{1} = 1$ *of row 1 from row 2 and row 3 respectively, gives*

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{array}\right).$$

To start the second elimination step, since we note that the element $a_{22}^{(1)} = 0$,

so the simple Gaussian elimination cannot continue in its present form.
Therefore, we interchange the rows 2 and 3, to get

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

Using backward substitution, we get, $x_1 = 1,\ x_2 = 1,\quad x_3 = 1.$

Example 3.17 *Consider a linear system*

$$\begin{array}{rcrcl} 0.000100x_1 & + & x_2 & = & 1 \\ x_1 & + & x_2 & = & 2 \end{array}$$

which has exact solution $\mathbf{x} = [1.00010, 0.99990]^T$. *Now we solve this system by the simple Gaussian elimination. The first elimination step is to eliminate first variable* x_1 *from second equation by subtracting multiple* $m_{21} = 10000$ *of first equation from second equation, gives*

$$\begin{array}{rcrcr} 0.000100x_1 & + & x_2 & = & 1 \\ & - & 10000x_2 & = & -10000 \end{array}$$

Using backward substitution to get the solution $\mathbf{x}^* = [0, 1]^T$. *Thus a computational disaster has occurred. But if we interchange the equations, we obtain*

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & 2 \\ 0.000100x_1 & + & x_2 & = & 1 \end{array}$$

Apply the Gaussian elimination again, and we got the solution $\mathbf{x}^* = [1, 1]^T$. *This solution is as good as one would hope. So, we conclude from this example that it is not enough just to avoid zero pivot, one must also avoid relatively small one.* •

Here we need some pivoting strategies which help us to over come these difficulties facing during the process of simple Gaussian elimination.

3.4.2 Pivoting Strategies

Partial Pivoting

the basic approach is to use the largest (in absolute value) element on or below the diagonal in the column of current interest as the pivotal element for elimination in the rest of that column.

Example 3.18 *Solve the following linear system using the Gaussian elimination with partial pivoting*

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 1 \\ 2x_1 & + & 3x_2 & + & 4x_3 & = & 3 \\ 4x_1 & + & 9x_2 & + & 16x_3 & = & 11 \end{array}$$

Solution. *For the first elimination step, since 4 is the largest absolute coefficient of first variable* x_1*, therefore, the first row and the third row are interchange, giving us*

$$\begin{array}{rcrcrcr} 4x_1 & + & 9x_2 & + & 16x_3 & = & 11 \\ 2x_1 & + & 3x_2 & + & 4x_3 & = & 3 \\ x_1 & + & x_2 & + & x_3 & = & 1 \end{array}$$

the multiples $m_{21} = \dfrac{2}{4}$ *and* $m_{31} = \dfrac{1}{4}$ *of row 1 from row 2 and row 3 respectively, gives*

$$\begin{array}{rcrcrcr} 4x_1 & + & 9x_2 & + & 16x_3 & = & 11 \\ & - & 3/2x_2 & - & 4x_3 & = & -5/2 \\ & - & 5/4x_2 & - & x_3 & = & -7/5 \end{array}$$

For the second elimination step, $-3/2$ is the largest absolute coefficient of second variable x_2, so eliminate second variable x_2 from the third row by subtracting the multiple $m_{32} = \frac{5}{6}$ of row 2 from row 3, gives

$$\begin{array}{rcrcrcr} 4x_1 & + & 9x_2 & + & 16x_3 & = & 11 \\ & - & 3/2x_2 & - & 4x_3 & = & -5/2 \\ & & & & 1/3x_3 & = & 1/3 \end{array}$$

Obviously, the original set of equations has been transformed to an equivalent upper-triangular form. Now using backward substitution, gives, $x_1 = 1,\ x_2 = -1,\ x_3 = 1$, the required solution. •

3.4.4 LU Decomposition Method

Here we decompose or factorize the coefficient matrix A into the product of two triangular matrices in the form $A = LU$, where L is a lower-triangular matrix and U is the upper-triangular matrix.

$$L = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}, \quad U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{pmatrix},$$

such that $l_{ij} = 0$ for $i < j$, and $u_{ij} = 0$ for $i > j$.

Consider a linear system $A\mathbf{x} = \mathbf{b}$, $\Longleftrightarrow$ $LU\mathbf{x} = \mathbf{b}$,

which can be written as $L\mathbf{y} = \mathbf{b}$, where $\mathbf{y} = U\mathbf{x}$.

The solution of $A\mathbf{x} = \mathbf{b}$, can be computed in the following two steps:

1. Solve the lower-triangular system $L\mathbf{y} = \mathbf{b}$. By using the *forward elimination*,

$$\begin{aligned} y_1 &= b_1, \\ y_i &= b_i - \sum_{j=1}^{i-1} l_{ij} y_j, \qquad i = 2, 3, \ldots, n. \end{aligned}$$

2. Solve the upper-triangular system $U\mathbf{x} = \mathbf{y}$.

$$\begin{aligned} x_n &= \frac{y_n}{u_{nn}}, \\ x_i &= \frac{1}{u_{ii}} \left[y_i - \sum_{j=i+1}^{n} u_{ij} x_j \right], \quad i = n-1, n-2, \ldots, 1 \end{aligned}$$

Theorem 3.12 *If the 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 the matrices L and U are of the same size as A. •

Now we discuss the two possible variations of the LU decomposition to find the solution of the nonsingular linear system in the following.

Doolittle's Method

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

Crout's Method

The Crout's method, in which matrix U has unity on the main diagonal, is similar to Doolittle's method in all other aspects. The L and U matrices are obtained by expanding the matrix equation $A = LU$ term by term to determine the elements of the L and U matrices.

Example 3.20 *Construct the LU decomposition of the following matrix A by using the Gauss factorization (that is, the LU decomposition by Doolittle's method).*

Solution. *Applying the forward elimination step of Simple Gauss-elimination to the given matrix*

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 4 \end{pmatrix},$$

using the multiples $m_{21} = 2 = l_{21}$ *and* $m_{31} = 1 = l_{31}$, *we get* $\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 3 \end{pmatrix}$

Similarly, by using the multiple $m_{32} = 1 = l_{32}$, *we obtain* $\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix} = U.$

Hence we obtained the LU-decomposition of the given matrix as follows

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix},$$

where the unknown elements of matrix L are the used multiples and the matrix U is same as we obtained in forward elimination process. •

Example 3.23 *Solve the following linear system by LU decomposition using Doolittle's method*

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 3 \\ 2 & 2 & 2 \end{pmatrix} \quad and \quad \mathbf{b} = \begin{pmatrix} -2 \\ 3 \\ -6 \end{pmatrix}.$$

Solution. *Since*

$$A = LU = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}.$$

Performing the multiplication on the right-hand side, gives

$$\begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 3 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ l_{21}u_{11} & l_{21}u_{12} + u_{22} & l_{21}u_{13} + u_{23} \\ l_{31}u_{11} & l_{31}u_{12} + l_{32}u_{22} & l_{31}u_{13} + l_{32}u_{23} + u_{33} \end{pmatrix}.$$

Then equating elements of first column to obtain

$$\begin{aligned} 1 &= u_{11}, & u_{11} &= 1, \\ 1 &= l_{21}u_{11}, & l_{21} &= 1, \\ 2 &= l_{31}u_{11}, & l_{31} &= 2. \end{aligned}$$

Now equating elements of second column to obtain

$$\begin{aligned} 2 &= u_{12}, & u_{12} &= 2, \\ 3 &= l_{21}u_{12} + u_{22}, & u_{22} &= 3 - 2 = 1, \\ 2 &= l_{31}u_{12} + l_{32}u_{22}, & l_{32} &= 2 - 4 = -2. \end{aligned}$$

Finally, equating elements of third column to obtain

$$\begin{array}{rclrcl} 4 &=& u_{13}, & u_{13} &=& 4, \\ 3 &=& l_{21}u_{13} + u_{23}, & u_{23} &=& 3 - 4 = -1, \\ 2 &=& l_{31}u_{13} + l_{32}u_{23} + u_{33}, & u_{33} &=& 2 - 10 = -8. \end{array}$$

Thus we obtain

$$\begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 3 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & -8 \end{pmatrix},$$

Then solving the first system $L\mathbf{y} = \mathbf{b}$ *for unknown vector* $\mathbf{y}$*, that is*

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \\ -6 \end{pmatrix}.$$

Performing forward substitution yields

$$\begin{array}{rcrcrcrlrcr} y_1 & & & & & = & -2, & \textit{gives} & y_1 &=& -2, \\ y_1 &+& y_2 & & & = & 3, & \textit{gives} & y_2 &=& 5, \\ 2y_1 &-& 2y_2 &+& y_3 & = & -6, & \textit{gives} & y_3 &=& 8. \end{array}$$

Then solving the second system $U\mathbf{x} = \mathbf{y}$ *for unknown vector* $\mathbf{x}$, *that is*

$$\begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & -8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 \\ 5 \\ 8 \end{pmatrix}.$$

Performing backward substitution yields

$$\begin{array}{rcrcrcrlrcr} x_1 & + & 2x_2 & + & 4x_3 & = & -2, & \textit{gives} & x_1 & = & -6, \\ & & x_2 & - & x_3 & = & 5, & \textit{gives} & x_2 & = & 4, \\ & & & - & 8x_3 & = & 8, & \textit{gives} & x_3 & = & -1, \end{array}$$

Example 3.27 *Solve the following linear system by LU decomposition using Crout's method*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 2 & 5 & 6 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 1 \\ -1 \\ 5 \end{pmatrix}.$$

Solution. *Since*

$$A = LU = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{pmatrix}.$$

Performing the multiplication on the right-hand side, gives

$$\begin{pmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 2 & 5 & 6 \end{pmatrix} = \begin{pmatrix} l_{11} & l_{11}u_{12} & l_{11}u_{13} \\ l_{21} & l_{21}u_{12} + l_{22} & l_{21}u_{13} + l_{22}u_{23} \\ l_{31} & l_{31}u_{12} + l_{32} & l_{31}u_{13} + l_{32}u_{23} + l_{33} \end{pmatrix}.$$

Then equating elements of first column to obtain

$$\begin{aligned} 1 &= l_{11}, \\ 6 &= l_{21}, \\ 2 &= l_{31}. \end{aligned}$$

$$\begin{aligned} 2 &= l_{11}u_{12}, & u_{12} &= 2, \\ 5 &= l_{21}u_{12} + l_{22}, & l_{22} &= 5 - 12 = -7, \\ 5 &= l_{31}u_{12} + l_{32}, & l_{32} &= 5 - 4 = 1. \end{aligned}$$

Finally, then equating elements of third column to obtain

$$\begin{aligned} 3 &= l_{11}u_{13}, & u_{13} &= 3, \\ 4 &= l_{21}u_{13} + l_{22}u_{23}, & u_{23} &= (4 - 18)/-7 = 2, \\ 6 &= l_{31}u_{13} + l_{32}u_{23} + l_{33}, & l_{33} &= (6 - 6 - 2) = -2. \end{aligned}$$

Thus we get

$$\begin{pmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 2 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 6 & -7 & 0 \\ 2 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix},$$

Then solving the first system $L\mathbf{y} = \mathbf{b}$ *for unknown vector* $\mathbf{y}$, *that is*

$$\begin{pmatrix} 1 & 0 & 0 \\ 6 & -7 & 0 \\ 2 & 1 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 5 \end{pmatrix}.$$

Performing forward substitution yields

$$\begin{array}{rcrcrcrl} y_1 & & & & & = & 1, & \textit{gives} \;\; y_1 = 1, \\ 6y_1 & - & 7y_2 & & & = & -1, & \textit{gives} \;\; y_2 = 1, \\ 2y_1 & + & y_2 & - & 2y_3 & = & 5, & \textit{gives} \;\; y_3 = -1. \end{array}$$

Then solving the second system $U\mathbf{x} = \mathbf{y}$ *for unknown vector* $\mathbf{x}$, *that is*

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

Performing backward substitution yields

$$\begin{array}{rcrcrcrl} x_1 & + & 2x_2 & + & 3x_3 & = & 1, & \textit{gives} \;\; x_1 = -2, \\ & & x_2 & + & 2x_3 & = & 1, & \textit{gives} \;\; x_2 = 3, \\ & & & & x_3 & = & -1, & \textit{gives} \;\; x_3 = -1, \end{array}$$

and we obtained the approximate solution $\mathbf{x}^* = [-2, 3, -1]^T$.

Since we know that not every matrix has a direct LU decomposition. We define the following matrix which gives the sufficient condition for the LU decomposition of the matrix. It also, helps us for the convergence of the iterative methods for solving linear systems.

Definition 3.24 (Strictly Diagonally Dominant Matrix)

A square matrix is said to be strictly diagonally dominant (SDD) if the absolute value of each element on the main diagonal is greater than the sum of the absolute values of all the other elements in that row. Thus, strictly diagonally dominant matrix is defined as

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \qquad \text{for} \quad i = 1, 2, \ldots, n. \tag{3.35}$$

Example 3.30 *The matrix*

$$A = \begin{pmatrix} 7 & 3 & 1 \\ 1 & 6 & 3 \\ -2 & 4 & 8 \end{pmatrix},$$

is strictly diagonally dominant since

$$\begin{array}{rclcrcl} |7| & > & |3| + |1|, & \text{that is,} & 7 & > & 4, \\ |6| & > & |1| + |3|, & \text{that is} & 6 & > & 4, \\ |8| & > & |-2| + |4|, & \text{that is} & 8 & > & 6, \end{array}$$

but the following matrix

$$B = \begin{pmatrix} 6 & -3 & 4 \\ 3 & 7 & 3 \\ 5 & -4 & 10 \end{pmatrix},$$

is not strictly diagonally dominant since

$$|6| > |-3| + |4|, \quad \text{that is} \quad 6 > 7,$$

Theorem 3.14 *If a matrix A is strictly diagonally dominant, then:*

1. *Matrix A is nonsingular.*

2. *Gaussian elimination without row interchange can be performed on the linear system* $A\mathbf{x} = \mathbf{b}$.

3. *Matrix A has LU factorization.* $\bullet$

3.5 Norms of Vectors and Matrices

Vector Norms

It is sometimes useful to have a scalar measure of the magnitude of a vector. Such a measure is called a *vector norm* and for a vector $\mathbf{x}$ is written as $\|\mathbf{x}\|$.
A vector norm on $\mathbf{R}^n$ is a function, from $\mathbf{R}^n$ to $\mathbf{R}$ satisfying:

1. $\|\mathbf{x}\| > 0 \quad \text{for all} \quad \mathbf{x} \in \mathbf{R}^n.$
2. $\|\mathbf{x}\| = 0 \quad \text{if and only if} \quad \mathbf{x} = \mathbf{0}.$
3. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|, \quad \text{for all} \quad \alpha \in \mathbf{R}, \quad \mathbf{x} \in \mathbf{R}^n.$
4. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbf{R}^n.$

There are three norms in $\mathbf{R}^n$ that are most commonly used in applications, called l_1-norm, l_2-norm, and l_∞-norm, and are defined for the given vectors $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ as

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|, \qquad \|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}, \qquad \|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Example 3.32 *Compute l_p-norms ($p = 1, 2, \infty$) of the vector $\mathbf{x} = [-5, 3, -2]^T$ in $\mathbf{R}^3$.*

Solution. *These l_p-norms ($p = 1, 2, \infty$) of the given vector are:*

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + |x_3| = |-5| + |3| + |-2| = 10.$$

$$\|\mathbf{x}\|_2 = (x_1^2 + x_2^2 + x_3^2)^{1/2} = \left[(-5)^2 + (3)^2 + (-2)^2\right]^{1/2} \approx 6.16.$$

$$\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, |x_3|\} = \max\{|-5|, |3|, |-2|\} = 5.$$

Matrix Norms

a matrix norm on the set of all $n \times n$ matrices is a real-valued function, $\|.\|$, defined on this set, satisfying for all $n \times n$ matrices A and B and all real number α as follows:

1. $\|A\| > 0, \qquad A \neq \mathbf{0}.$
2. $\|A\| = 0, \qquad A = \mathbf{0}.$
3. $\|I\| = 1, \qquad I \; is \; the \; identity \; matrix.$
4. $\|\alpha A\| = |\alpha| \|A\|, \quad$ for scalar $\quad \alpha \in \mathbf{R}.$
5. $\|A + B\| \leq \|A\| + \|B\|.$
6. $\|AB\| \leq \|A\| \|B\|$
7. $\|A - B\| \geq \Big| \|A\| - \|B\| \Big|.$

Several norms for matrices have been defined, we shall use the following three natural norms l_1, l_2, and l_∞ for a square matrix of order n:

$$\|A\|_1 = \max_j \left(\sum_{i=1}^{n} |a_{ij}| \right) = \text{maximum column-sum.}$$

$$\|A\|_2 = \max_{\|x\|_2 = 1} \|A\mathbf{x}\|_2 = \text{spectral norm.}$$

$$\|A\|_\infty = \max_i \left(\sum_{j=1}^{n} |a_{ij}| \right) = \text{row-sum norm.}$$

For $m \times n$ matrix, we can paraphrase the *Frobenius norm* (or *Euclidean norm*, $\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$

Example 3.33 *Compute l_p-norms $(p = 1, \infty, F)$ of the following matrix*

$$A = \begin{pmatrix} 4 & 2 & -1 \\ 3 & 5 & -2 \\ 1 & -2 & 7 \end{pmatrix}.$$

Solution. *These norms are:*

$$\sum_{i=1}^{3}|a_{i1}| = |4| + |3| + |1| = 8, \quad \sum_{i=1}^{3}|a_{i2}| = |2| + |5| + |-2| = 9, \quad \sum_{i=1}^{3}|a_{i3}| = |-1| + |-2| + |7| = 10,$$

so $\|A\|_1 = max\{8, 9, 10\} = 10.$

Also, $\sum_{j=1}^{3}|a_{1j}| = |4| + |2| + |-1| = 7, \quad \sum_{j=1}^{3}|a_{2j}| = |3| + |5| + |-2| = 10, \quad \sum_{j=1}^{3}|a_{3j}| = |1| + |-2| + |7| = 10,$

so $\|A\|_\infty = max\{7, 10, 10\} = 10.$

Finally, we have

$$\|A\|_F = (16 + 4 + 1 + 9 + 25 + 4 + 1 + 4 + 49)^{1/2} \approx 10.6301,$$

3.6 Iterative Methods for Solving Linear Systems

3.6.1 Jacobi Iterative Method

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

$$\Longleftrightarrow \quad \begin{aligned} x_1 &= \frac{1}{a_{11}}\left[b_1 - a_{12}x_2 - a_{13}x_3\right] \\ x_2 &= \frac{1}{a_{22}}\left[b_2 - a_{21}x_1 - a_{23}x_3\right] \\ x_3 &= \frac{1}{a_{33}}\left[b_3 - a_{31}x_1 - a_{32}x_2\right] \end{aligned}$$

Let $\mathbf{x}^{(k)} = \left[x_1^{(k)}, x_2^{(k)}, x_3^{(k)}\right]^T$ then we define an iterative sequence called the Jacobi formula

$$\left\{ \begin{aligned} x_1^{(k+1)} &= \frac{1}{a_{11}}\left[b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)}\right] \\ x_2^{(k+1)} &= \frac{1}{a_{22}}\left[b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)}\right] \\ x_3^{(k+1)} &= \frac{1}{a_{33}}\left[b_3 - a_{31}x_1^{(k)} - a_{32}x_2^{(k)}\right] \end{aligned} \right.$$

For a general system of n linear equations, the Jacobi method is

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right] \qquad i = 1,2,\dots,n, \quad k = 0,1,2,\dots,$$

provided that the diagonal elements $a_{ii} \neq 0$ for each $i = 1,2,\dots,n$.

it is conventional to start with $x_i^{(0)} = \mathbf{0}$ for all i.

The iterations defined by (3.37) are stopped when $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < \epsilon$, or $\dfrac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\|\mathbf{x}^{(k+1)}\|} < \epsilon$,

Example 3.34 *Solve the following system of equations using the Jacobi iterative method, using $\epsilon = 10^{-6}$ in the l_∞-norm.*

$$\begin{array}{rcrcrcr} 5x_1 & - & x_2 & + & x_3 & = & 10 \\ 2x_1 & + & 8x_2 & - & x_3 & = & 11 \\ -x_1 & + & x_2 & + & 4x_3 & = & 3 \end{array}$$

Start with the initial solution $\mathbf{x}^{(0)} = [0,0,0]^T$.

The Jacobi iterative method for the given system has the form

$$x_1^{(k+1)} = \frac{1}{5}\left[10 + x_2^{(k)} - x_3^{(k)}\right]$$

$$x_2^{(k+1)} = \frac{1}{8}\left[11 - 2x_1^{(k)} + x_3^{(k)}\right]$$

$$x_3^{(k+1)} = \frac{1}{4}\left[3 + x_1^{(k)} - x_2^{(k)}\right]$$

and starting with initial approximation $x_1^{(0)} = 0, x_2^{(0)} = 0, x_3^{(0)} = 0$*, then for* $k = 0$*, we obtain*

$$\begin{aligned} x_1^{(1)} &= \frac{1}{5}\Big[10 + x_2^{(0)} - x_3^{(0)}\Big] = \frac{1}{5}\Big[10 + 0 - 0\Big] = 2, \\ x_2^{(1)} &= \frac{1}{8}\Big[11 - 2x_1^{(0)} + x_3^{(0)}\Big] = \frac{1}{8}\Big[11 - 0 + 0\Big] = 1.375, \\ x_3^{(1)} &= \frac{1}{4}\Big[3 + x_1^{(0)} - x_2^{(0)}\Big] = \frac{1}{4}\Big[3 + 0 - 0\Big] = 0.75. \end{aligned}$$

The first and subsequent iterations are listed in Table 3.1.

Table 3.1: Solution of the Example 3.34

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.000000	0.000000	0.000000
1	2.000000	1.375000	0.750000
2	2.125000	0.968750	0.906250
⋮	⋮	⋮	⋮
15	2.000000	0.999999	1.000000
16	2.000000	1.000000	1.000000

3.6.2 Gauss-Seidel Iterative Method

the Gauss-Seidel formula for the system of three equations

$$
\begin{aligned}
x_1^{(k+1)} &= \frac{1}{a_{11}}\left[b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)}\right] \\
x_2^{(k+1)} &= \frac{1}{a_{22}}\left[b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)}\right] \\
x_3^{(k+1)} &= \frac{1}{a_{33}}\left[b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)}\right]
\end{aligned}
$$

For a general system of n linear equations, the Gauss-Seidel iterative method defined as

$$
x_i^{(k+1)} = \frac{1}{a_{ii}}\left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right] \qquad i = 1,2,\ldots,n, \quad k = 0,1,2,\ldots
$$

Example 3.36 *Solve the following system of equations using the Gauss-Seidel iterative method, with $\epsilon = 10^{-6}$ in l_∞-norm.*

$$
\begin{aligned}
5x_1 - x_2 + x_3 &= 10 \\
2x_1 + 8x_2 - x_3 &= 11 \\
-x_1 + x_2 + 4x_3 &= 3
\end{aligned}
$$

Start with the initial solution $\mathbf{x}^{(0)} = [0,0,0]^T$.

Solution. *The Gauss-Seidel iteration for the given system is*

$$
\begin{aligned}
x_1^{(k+1)} &= \frac{1}{5}\Big[10 + x_2^{(k)} - x_3^{(k)}\Big] \\
x_2^{(k+1)} &= \frac{1}{8}\Big[11 - 2x_1^{(k+1)} + x_3^{(k)}\Big] \\
x_3^{(k+1)} &= \frac{1}{4}\Big[3 + x_1^{(k+1)} - x_2^{(k+1)}\Big]
\end{aligned}
$$

and starting with initial approximation $x_1^{(0)} = 0, x_2^{(0)} = 0, x_3^{(0)} = 0$, *then for* $k = 0$, *we obtain*

$$
\begin{aligned}
x_1^{(1)} &= \frac{1}{5}\Big[10 + x_2^{(0)} - x_3^{(0)}\Big] = \frac{1}{5}\Big[10 + 0 - 0\Big] = 2, \\
x_2^{(1)} &= \frac{1}{8}\Big[11 - 2x_1^{(1)} + x_3^{(0)}\Big] = \frac{1}{8}\Big[11 - 4 + 0\Big] = 0.875, \\
x_3^{(1)} &= \frac{1}{4}\Big[3 + x_1^{(1)} - x_2^{(1)}\Big] = \frac{1}{4}\Big[3 + 2 - 0.875\Big] = 1.03125.
\end{aligned}
$$

The first and subsequent iterations are listed in Table 3.3.

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.000000	0.000000	0.000000
1	2.000000	0.875000	1.031250
2	1.968750	1.011719	0.989258
3	2.004492	0.997534	1.001740
⋮	⋮	⋮	⋮
9	2.000000	0.999999	1.000000
10	2.000000	1.000000	1.000000

From the Examples 3.34 and (3.36), we noted that the solution by the Gauss-Seidel method converges more quickly than the Jacobi method. In general, we may state that **if both the Jacobi method and the Gauss-Seidel method are converge, then the Gauss-Seidel method converges more quickly**. This is generally the case but not always true. In fact, there are some linear systems for which the Jacobi method converges but the Gauss-Seidel method does not, and others for which the Gauss-Seidel method converges but the Jacobi method does not.

3.6.3 Matrix Forms of Iterative Methods for Linear System

The iterative methods to solve the system of linear equations $A\mathbf{x} = \mathbf{b}$, start with an initial approximation $\mathbf{x}^{(0)} \in \mathbf{R}$ and generates a sequence of vectors $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ that converges to $\mathbf{x}$.

$$A\mathbf{x} = \mathbf{b}, \quad \text{equivalent} \quad \mathbf{x} = T\mathbf{x} + \mathbf{c},$$

the sequence of approximate solutions vector is generated by computing

$$\mathbf{x}^{(k+1)} = T\mathbf{x}^{(k)} + \mathbf{c}, \qquad \text{for} \quad k = 0, 1, 2, \ldots$$

Let a matrix A can be written $\quad A = L + D + U,$

$$L = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ a_{21} & 0 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 0 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & 0 & a_{23} & \cdots & a_{2n} \\ 0 & 0 & 0 & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad D = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

$$\mathbf{x} = T\mathbf{x} + \mathbf{c}, \quad \Longleftrightarrow \quad (L + D + U)\mathbf{x} = \mathbf{b}. \qquad (3.47)$$

Jacobi Iterative Method

The equation (3.47) can be written as $D\mathbf{x} = -(L + U)\mathbf{x} + \mathbf{b}.$

Since matrix D is nonsingular, $\mathbf{x} = -D^{-1}(L + U)\mathbf{x} + D^{-1}\mathbf{b},$

which can be put in the form $\mathbf{x}^{(k+1)} = T_J\mathbf{x}^{(k)} + \mathbf{c}_j, \quad \text{for} \quad k = 0, 1, 2, \ldots$

$$T_J = -D^{-1}(L + U) \quad \text{and} \quad \mathbf{c}_j = D^{-1}\mathbf{b},$$

Gauss-Seidel Iterative Method

The equation (3.47) can also be written as $(L + D)\mathbf{x} = -U\mathbf{x} + \mathbf{b}.$

Since lower-triangular matrix $(L + D)$ is nonsingular, $\mathbf{x} = -(L + D)^{-1}U\mathbf{x} + (L + D)^{-1}\mathbf{b},$

which can be put in the form $\mathbf{x}^{(k+1)} = T_G\mathbf{x}^{(k)} + \mathbf{c}_G, \quad \text{for} \quad k = 0, 1, 2, \ldots$

$$T_G = -(L + D)^{-1}U \quad \text{and} \quad \mathbf{c}_G = (L + D)^{-1}\mathbf{b},$$

Example 3.38 *Consider the following system*

$$\begin{array}{rcrcrcr} 6x_1 & + & 2x_2 & & & = & 1 \\ x_1 & + & 7x_2 & - & 2x_3 & = & 2 \\ 3x_1 & - & 2x_2 & + & 9x_3 & = & -1 \end{array}$$

(a) *Find the matrix form of iterative (Jacobi and Gauss-Seidel) methods.*

Solution. *Since the given matrix A is*

$$A = \begin{pmatrix} 6 & 2 & 0 \\ 1 & 7 & -2 \\ 3 & -2 & 9 \end{pmatrix},$$

and so

$$A = L + U + D = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 3 & -2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 6 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

Jacobi Iterative Method

(a) *Since the matrix form of the Jacobi iterative method can be written as*

$$\mathbf{x}^{(k+1)} = T_J\mathbf{x}^{(k)} + \mathbf{c}, \qquad k = 0, 1, 2, \dots,$$

where

$$T_J = -D^{-1}(L+U) \qquad and \qquad \mathbf{c} = D^{-1}\mathbf{b}.$$

One can easily compute the Jacobi iteration matrix T_J *and the vector* $\mathbf{c}$ *as follows:*

$$T_J = \begin{pmatrix} 0 & -\frac{2}{6} & 0 \\ -\frac{1}{7} & 0 & \frac{2}{7} \\ -\frac{3}{9} & \frac{2}{9} & 0 \end{pmatrix} \quad and \quad \mathbf{c} = \begin{pmatrix} \frac{1}{6} \\ \frac{2}{7} \\ -\frac{1}{9} \end{pmatrix}.$$

Thus the matrix form of Jacobi iterative method is

$$\mathbf{x}^{(k+1)} = \begin{pmatrix} 0 & -\frac{2}{6} & 0 \\ -\frac{1}{7} & 0 & \frac{2}{7} \\ -\frac{3}{9} & \frac{2}{9} & 0 \end{pmatrix} \mathbf{x}^{(k)} + \begin{pmatrix} \frac{1}{6} \\ \frac{2}{7} \\ -\frac{1}{9} \end{pmatrix}, \quad k = 0, 1, 2.$$

Gauss-Seidel Iterative Method

(a) *Now by using Gauss-Seidel method, first we compute the Gauss-Seidel iteration matrix T_G and the vector* $\mathbf{c}$ *as follows:*

$$T_G = \begin{pmatrix} 0 & -\frac{1}{3} & 0 \\ 0 & \frac{1}{21} & \frac{2}{7} \\ 0 & \frac{23}{189} & \frac{4}{63} \end{pmatrix} \quad \textit{and} \quad \mathbf{c} = \begin{pmatrix} \frac{1}{6} \\ \frac{11}{42} \\ -\frac{41}{378} \end{pmatrix}.$$

Thus the matrix form of Gauss-Seidel iterative method is

$$\mathbf{x}^{(k+1)} = \begin{pmatrix} 0 & -\frac{1}{3} & 0 \\ 0 & \frac{1}{21} & \frac{2}{7} \\ 0 & \frac{23}{189} & \frac{4}{63} \end{pmatrix} \mathbf{x}^{(k)} + \begin{pmatrix} \frac{1}{6} \\ \frac{11}{42} \\ -\frac{41}{378} \end{pmatrix}, \quad k = 0, 1, 2.$$

3.6.4 Convergence Criteria of Iterative Methods

Since we noted that the Jacobi method and the Gauss-Seidel method do not always converge to the solution of the given system of linear equations. Here we need some conditions which make the both methods converge. The sufficient conditions for the convergence of both iterative methods are discussed in the following theorems.

Theorem 3.15 (Sufficient Condition for Convergence)

If the matrix A is strictly diagonally dominant (SDD), then for any choice of initial approximation $\mathbf{x}^{(0)} \in \mathbf{R}$ both the Jacobi method and the Gauss-Seidel method give sequence $\{x^{(k)}\}_{k=0}^{\infty}$ of approximations that converge to the solution of the linear system. •

Example 3.39 *Rearrange the following linear system of equations*

$$\begin{array}{rcrcrcl} x_1 & + & 6x_2 & - & 3x_3 & = & 4 \\ 2x_1 & + & 2x_2 & + & 4x_3 & = & 7 \\ 5x_1 & + & 2x_2 & - & x_3 & = & 6 \end{array}$$

such that the convergence of both iterative methods (Jacobi and Gauss-Seidel) is guaranteed. Use initial solution $\mathbf{x}^{(0)} = [0, 0, 0]^T$, compute approximation solution within accuracy 0.5×10^{-2}.

Solution. *For the guarantee convergence of iterative methods, the system must be SDD form, so rearrange the given system in the following form*

$$\begin{array}{rcrcrcl} 5x_1 & + & 2x_2 & - & x_3 & = & 6 \\ x_1 & + & 6x_2 & - & 3x_3 & = & 4 \\ 2x_1 & + & 2x_2 & + & 4x_3 & = & 7 \end{array}$$

Jacobi Iterative Method

$$x_1^{(k+1)} = \frac{1}{5}\Big[6 - 2x_2^{(k)} + x_3^{(k)}\Big]$$

$$x_2^{(k+1)} = \frac{1}{6}\Big[4 - x_1^{(k)} + 3x_3^{(k)}\Big]$$

$$x_3^{(k+1)} = \frac{1}{4}\Big[7 - 2x_1^{(k)} - 2x_2^{(k)}\Big]$$

Starting with $\mathbf{x}^{(0)} = [0, 0, 0]^T$, *the first and subsequent iterations are listed in Table 3.5.*

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.000	0.000	0.000
1	1.200	0.667	1.750
2	1.283	1.342	0.983
3	0.860	0.944	0.773
$\vdots$	$\vdots$	$\vdots$	$\vdots$
10	0.997	0.998	0.999
11	1.001	1.000	1.002

Gauss-Seidel Iterative Method

$$\begin{aligned} x_1^{(k+1)} &= \frac{1}{5}\Big[6 - 2x_2^{(k)} + x_3^{(k)}\Big] \\ x_2^{(k+1)} &= \frac{1}{6}\Big[4 - x_1^{(k+1)} + 3x_3^{(k)}\Big] \\ x_3^{(k+1)} &= \frac{1}{4}\Big[7 - 2x_1^{(k+1)} - 2x_2^{(k+1)}\Big] \end{aligned}$$

Starting with initial approximation $\mathbf{x}^{(0)} = [0,0,0]^T$, *the first and subsequent iterations are listed in Table 3.6. Note that Gauss-Seidel iterative method converges faster than Jacobi iterative method.*

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.000	0.000	0.000
1	1.200	0.467	1.033
2	1.220	0.980	0.895
3	0.987	0.950	1.019
⋮	⋮	⋮	⋮
6	1.003	1.002	0.998
7	0.999	0.999	1.001

Theorem 3.16 (Sufficient Condition for Convergence)

For any initial approximation $\mathbf{x}^{(0)} \in \mathbf{R}$, *the sequence* $\{x^{(k)}\}_{k=0}^{\infty}$ *of approximations defined by*

$$\mathbf{x}^{(k+1)} = T\mathbf{x}^{(k)} + \mathbf{c}, \quad \textit{for each} \quad k \geq 0, \quad \textit{and} \quad \mathbf{c} \neq 0, \tag{3.52}$$

converges to the unique solution of $\mathbf{x} = T\mathbf{x} + \mathbf{c}$ *if* $\|T\| < 1$ *for any natural matrix norm, and the following* **error bounds** *hold:*

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^{(k)}\| &\leq \|T\|^k \|\mathbf{x}^{(0)} - \mathbf{x}\|, \\ \|\mathbf{x} - \mathbf{x}^{(k)}\| &\leq \frac{\|T\|^k}{1 - \|T\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|. \end{aligned} \tag{3.53}$$

Note that smaller the value of the $\|T\|$, faster the convergence of iterative methods.

Example 3.41 *Consider the following linear system of equations*

$$\begin{array}{rcrcrcr} 4x_1 & - & x_2 & + & x_3 & = & 12 \\ -x_1 & + & 3x_2 & + & x_3 & = & 1 \\ x_1 & + & x_2 & + & 5x_3 & = & -14 \end{array}$$

(a) Show that both iterative methods (Jacobi and Gauss-Seidel) will converge by using $\|T\|_\infty < 1$.

(b) Find second approximation $\mathbf{x}^{(2)}$ *when the initial solution is* $\mathbf{x}^{(0)} = [4, 3, -3]^T$.

(c) Compute the error bounds for your approximations.

(d) How many iterations needed to get an accuracy within 10^{-4}.

Solution. *From (3.46), we have*

$$\begin{aligned} A &= \begin{pmatrix} 4 & -1 & 1 \\ -1 & 3 & 1 \\ 1 & 1 & 5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \\ &= L + U + D. \end{aligned}$$

Jacobi Method

(a) Since the Jacobi iteration matrix is defined as

$$T_J = -D^{-1}(L + U),$$

and by using the given information, we have

$$T_J = -\begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{3} & 0 & -\frac{1}{3} \\ -\frac{1}{5} & -\frac{1}{5} & 0 \end{pmatrix}.$$

Then the l_∞ norm of the matrix T_J is

$$\|T_J\|_\infty = \max\left\{\frac{2}{4}, \frac{2}{3}, \frac{2}{5}\right\} = \frac{2}{3} < 1.$$

Thus the Jacobi method will converge for the given linear system.

(b) The Jacobi method for the given system is

$$\begin{aligned} x_1^{(k+1)} &= \frac{1}{4}\Big[12 + x_2^{(k)} - x_3^{(k)}\Big] \\ x_2^{(k+1)} &= \frac{1}{3}\Big[1 + x_1^{(k)} - x_3^{(k)}\Big] \\ x_3^{(k+1)} &= \frac{1}{5}\Big[-14 - x_1^{(k)} - x_2^{(k)}\Big] \end{aligned}$$

Starting with initial approximation $x_1^{(0)} = 4, x_2^{(0)} = 3, x_3^{(0)} = -3$, *and for* $k = 0, 1$, *we obtain the first and the second approximations as*

$$\mathbf{x}^{(1)} = [4.5, 2.6667, -4.2]^T \quad and \quad \mathbf{x}^{(2)} = [4.7167, 3.2333, -4.2333]^T.$$

(c) Using the error bound formula (3.53), we obtain

$$\|\mathbf{x} - \mathbf{x}^{(2)}\| \leq \frac{(2/3)^2}{1 - 2/3} \left\| \begin{pmatrix} 4.5 \\ 2.6667 \\ -4.2 \end{pmatrix} - \begin{pmatrix} 4 \\ 3 \\ -3 \end{pmatrix} \right\| \leq \frac{4}{3}(1.2) = 1.6.$$

(d) To find the number of iterations, we use the formula (3.53) as

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|T_J\|^k}{1 - \|T_J\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| \leq 10^{-4}.$$

It gives

$$\frac{(2/3)^k}{1/3}(1.2) \leq 10^{-4}, \quad or \quad (2/3)^k \leq \frac{10^{-4}}{3.6}.$$

Taking ln *on both sides, we obtain*

$$k \ln(2/3) \leq \ln\left(\frac{10^{-4}}{3.6}\right), \quad gives \quad k \geq 25.8789, \quad or \quad k = 26,$$

which is the required number of iterations.

Gauss-Seidel Method

(a) Since the Gauss-Seidel iteration matrix is defined as

$$T_G = -(D+L)^{-1}U,$$

and by using the given information, we have

$$T_G = -\begin{pmatrix} \frac{1}{4} & 0 & 0 \\ \frac{1}{12} & \frac{1}{3} & 0 \\ -\frac{4}{60} & -\frac{1}{15} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{4} & -\frac{1}{4} \\ 0 & \frac{1}{12} & -\frac{5}{12} \\ 0 & -\frac{4}{60} & \frac{8}{60} \end{pmatrix}.$$

Then the l_∞ norm of the matrix T_G is

$$\|T_G\|_\infty = \max\left\{\frac{2}{4}, \frac{6}{12}, \frac{12}{60}\right\} = \frac{1}{2} < 1.$$

Thus the Gauss-Seidel method will converge for the given linear system.

(b) The Gauss-Seidel method for the given system is

$$
\begin{aligned}
x_1^{(k+1)} &= \frac{1}{4}\Big[12 + x_2^{(k)} - x_3^{(k)}\Big] \\
x_2^{(k+1)} &= \frac{1}{3}\Big[1 + x_1^{(k+1)} - x_3^{(k)}\Big] \\
x_3^{(k+1)} &= \frac{1}{5}\Big[-12 - x_1^{(k+1)} - x_2^{(k+1)}\Big]
\end{aligned}
$$

Starting with initial approximation $x_1^{(0)} = 4, x_2^{(0)} = 3, x_3^{(0)} = -3$, *and for* $k = 0, 1$, *we obtain the first and the second approximations as*

$$\mathbf{x}^{(1)} = [4.5, 2.8333, -4.2667]^T \quad \textit{and} \quad \mathbf{x}^{(2)} = [4.775, 3.3472, -4.4244]^T.$$

(c) Using the error bound formula (3.53), we obtain

$$\|\mathbf{x} - \mathbf{x}^{(2)}\| \leq \frac{(1/2)^2}{1 - 1/2} \left\| \begin{pmatrix} 4.5 \\ 2.8333 \\ -4.2667 \end{pmatrix} - \begin{pmatrix} 4 \\ 3 \\ -3 \end{pmatrix} \right\| \leq \frac{1}{2}(1.2667) = 0.6334.$$

(d) To find the number of iterations, we use the formula (3.53) as

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|T_J\|^k}{1 - \|T_J\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| \leq 10^{-4}.$$

It gives

$$\frac{(1/2)^k}{1/2}(1.2667) \leq 10^{-4}, \quad or \quad (1/2)^k \leq \frac{10^{-4}}{2.5334}.$$

Taking ln *on both sides, we obtain*

$$k \ln(1/2) \leq \ln\left(\frac{10^{-4}}{2.5334}\right), \quad gives \quad k \geq 14.6084 \quad or \quad k = 15,$$

which is the required number of iterations.

Remark

Since $\|T_G\|_\infty < \|T_J\|_\infty$, *which shows that Gauss-Seidel method will converge faster than Jacobi method for the given linear system.* •

3.7 Errors in Solving Linear Systems

if $\mathbf{x}^*$ is an approximate solution of the given system $A\mathbf{x} = \mathbf{b}$, we compute a vector

$\mathbf{r} = \mathbf{b} - A\mathbf{x}^*$, which is called the *residual vector* and can be easily calculated.

The quantity $\dfrac{\|\mathbf{r}\|}{\|\mathbf{b}\|} = \dfrac{\|\mathbf{b} - A\mathbf{x}^*\|}{\|\mathbf{b}\|}$, is called the *relative residual.*

The smallness of the residual then provides a measure of the goodness of the approximate solution $\mathbf{x}^*$. If every component of vector $\mathbf{r}$ vanishes, then $\mathbf{x}^*$ is the exact solution. If $\mathbf{x}^*$ is a good approximation then we would expect each component of $\mathbf{r}$ to be small, at least in a relative sense. For example, the following linear system

$$\begin{aligned} x_1 &+ 2x_2 = 3 \\ 1.0001x_1 &+ 2x_2 = 3.0001 \end{aligned}$$

has the exact solution $\mathbf{x} = [1, 1]^T$ but has a poor approximate solution $\mathbf{x}^* = [3, 0]^T$. To see ho good this solution is, we compute the residual, $\mathbf{r} = [0, -0.0002]^T$, and so $\|\mathbf{r}\|_\infty = 0.0002$. Althoug the norm of the residual vector is small, the approximate solution $\mathbf{x}^* = [3, 0]^T$ is obviously quit poor; in fact $\|\mathbf{x} - \mathbf{x}^*\|_\infty = 2$.

Intuitively it would seem reasonable to assume that when $\|\mathbf{r}\|$ is small for a given vector norm, then the error $\|\mathbf{x} - \mathbf{x}^*\|$ would be small as well. In fact this is true for some systems. However, there are systems of equations which do not satisfy this property. Such systems are said to be *ill-conditioned.*

3.7.1 Conditioning of Linear Systems

These are systems in which small changes in the coefficients of the system

lead to large changes in the solution. For example,

$$\begin{aligned} x_1 &+ x_2 &= 2 \\ x_1 &+ 1.01x_2 &= 2.01 \end{aligned}$$

The exact solution is easily verified to be $x_1 = x_2 = 1$. On the other hand, the system

$$\begin{aligned} x_1 &+ x_2 &= 2 \\ 1.001x_1 &+ x_2 &= 2.01 \end{aligned}$$

has the solution $x_1 = 10, x_2 = -8$. Thus change of 1 percent in the coefficients has changed the solution by a factor of 10. If in the above given system, we substitute $x_1 = 10, x_2 = 8$, we find that the residual are $r_1 = 0, r_2 = 0.09$, so that this solution looks a reasonable although it is grossly in error. In practical problems we can expect the coefficients in the system to be subject to small errors, either because of round-off or because of physical measurement. If the system is ill-conditioned.

We have seen that for ill-conditioned systems the residual is not necessarily a good measure of the accuracy of a solution. How then can we tell when a system is ill-conditioned ? In the following we discuss the some possible indicators of ill-conditioned system.

Definition 3.25 (Condition Number of a Matrix)

The number $\|A\|\|A^{-1}\|$ *is called the condition number of a nonsingular matrix* A *and is denoted by* $K(A)$*, that is*

$$cond(A) = K(A) = \|A\|\|A^{-1}\|. \tag{3.55}$$

Note that the condition number $K(A)$ for A depends on the matrix norm used and can, for some matrices, vary considerably as the matrix norm is changed. Since

$$1 = \|I\| = \|AA^{-1}\| \leq \|A\|\|A^{-1}\| = K(A),$$

therefore, the condition number is always in the range $1 \leq K(A) \leq \infty$ regardless of any natural norm. The lower limit is attained for identity matrices and $K(A) = \infty$ if A is singular. So the matrix A is *well-behaved*(or well-conditioned) if $K(A)$ is close to 1 and is increasingly *ill-conditioned* when $K(A)$ is significantly greater than 1, that is, $K(A) \to \infty$.

Example 3.44 *Compute the condition number of the following matrix using the* l_∞*-norm*

$$A = \begin{pmatrix} 2 & -1 & 0 \\ 2 & -4 & -1 \\ -1 & 0 & 2 \end{pmatrix}.$$

Solution. *Since the condition number of a matrix is defined as*

$$K(A) = \|A\|_\infty \|A^{-1}\|_\infty.$$

First we calculate the inverse of the given matrix which is

$$A^{-1} = \begin{pmatrix} \frac{8}{13} & -\frac{2}{13} & -\frac{1}{13} \\ \frac{3}{13} & -\frac{4}{13} & -\frac{2}{13} \\ \frac{4}{13} & -\frac{1}{13} & \frac{6}{13} \end{pmatrix}.$$

Now we calculate the l_∞-norm of both the matrices A and A^{-1}. Since the l_∞-norm of a matrix is the maximum of the absolute row sums, we have

$$\|A\|_\infty = max\{|2| + |-1| + |0|, |2| + |-4| + |-1|, |-1| + |0| + |2|\} = 7,$$

and

$$\|A^{-1}\|_\infty = max\Big\{\Big|\frac{8}{13}\Big| + \Big|\frac{-2}{13}\Big| + \Big|\frac{-1}{13}\Big|, \Big|\frac{3}{13}\Big| + \Big|\frac{-4}{13}\Big| + \Big|\frac{-2}{13}\Big|, \Big|\frac{4}{13}\Big| + \Big|\frac{-1}{13}\Big| + \Big|\frac{6}{13}\Big|\Big\},$$

which gives

$$\|A^{-1}\|_\infty = \frac{11}{13}.$$

Therefore,

$$K(A) = \|A\|_\infty \|A^{-1}\|_\infty = (7)\left(\frac{11}{13}\right) \approx 5.9231.$$

Depending on the application, we might consider this number to be reasonably small and conclude that the given matrix A is reasonably well-conditioned. •

Some matrices are notoriously ill-conditioned. For example, consider the 4×4 Hilbert matrix

$$H = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix},$$

The inverse of the matrix H can be obtained as

$$H^{-1} = \begin{pmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{pmatrix}.$$

Then the condition number of the Hilbert matrix is

$$K(H) = \|H\|_\infty \|H^{-1}\|_\infty = (2.0833)(13620) \approx 28375,$$

which is quite large. Note that the condition number of Hilbert matrices increase rapidly as the size of the matrices increases. Therefore, large Hilbert matrices are considered to be extremely ill-conditioned.

Theorem 3.18 (Error in Linear Systems)

Suppose that $\mathbf{x}^*$ *is an approximation to the solution* $\mathbf{x}$ *of the linear system* $A\mathbf{x} = \mathbf{b}$ *and* A *is a nonsingular matrix and* $\mathbf{r}$ *is the residual vector for* $\mathbf{x}^*$*. Then for any natural norm, the error is*

$$\|\mathbf{x} - \mathbf{x}^*\| \leq \|\mathbf{r}\| \|A^{-1}\| \tag{3.56}$$

and the relative error is

$$\frac{\|\mathbf{x} - \mathbf{x}^*\|}{\|\mathbf{x}\|} \leq K(A) \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}, \quad \textit{provided} \quad \mathbf{x} \neq 0, \ \mathbf{b} \neq 0. \tag{3.57}$$

Proof. Since $\mathbf{r} = \mathbf{b} - A\mathbf{x}^*$ and A is nonsingular, then

$$A\mathbf{x} - A\mathbf{x}^* = \mathbf{b} - (\mathbf{b} - \mathbf{r}) = \mathbf{r},$$

which implies that

$$A(\mathbf{x} - \mathbf{x}^*) = \mathbf{r}, \quad \text{or} \quad \mathbf{x} - \mathbf{x}^* = A^{-1}\mathbf{r}. \tag{3.58}$$

Taking norm on both side, gives

$$\|\mathbf{x} - \mathbf{x}^*\| = \|A^{-1}\mathbf{r}\| \leq \|A^{-1}\| \|\mathbf{r}\|.$$

Moreover, since $\mathbf{b} = A\mathbf{x}$, then

$$\|\mathbf{b}\| \leq \|A\| \|\mathbf{x}\|, \qquad \text{or} \quad \|\mathbf{x}\| \geq \frac{\|\mathbf{b}\|}{\|A\|}.$$

Hence

$$\frac{\|\mathbf{x}-\mathbf{x}^*\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\|\|\mathbf{r}\|}{\|\mathbf{b}\|/\|A\|} \leq K(A)\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}.$$

The inequalities (3.56) and (3.57) imply that the quantities $\|A^{-1}\|$ and $K(A)$ can be used to give an indication of the connection between the residual vector and the accuracy of the approximation. If the quantity $K(A) \approx 1$, the relative error will be fairly close to the relative residual. But if $K(A) >> 1$, then the relative error could be many times larger than the relative residual. •

Example 3.47 *Consider a following linear system*

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & - & x_3 & = & 1 \\ x_1 & + & 2x_2 & - & 2x_3 & = & 0 \\ -2x_1 & + & x_2 & + & x_3 & = & -1 \end{array}$$

(a) Discuss the ill-conditioning of the given linear system.

(b) If $\mathbf{x}^* = [2.01, 1.01, 1.98]^T$ be an approximate solution of the given system, then find the residual vector $\mathbf{r}$ and its norm $\|\mathbf{r}\|_\infty$.

(c) Estimate the relative error using (3.57).

(d) Use the simple Gaussian elimination method to find approximate error using (3.58).

Solution. *(a) Given the matrix*

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

and whose inverse can be computed as

$$A^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ 1.5 & -0.5 & 0.5 \\ 2.5 & -1.5 & 0.5 \end{pmatrix}.$$

Then the l_∞-norm of both matrices are

$$\|A\|_\infty = 5 \quad and \quad \|A^{-1}\|_\infty = 4.5.$$

Using the values of both matrices norms, we can find the value of the condition number of A as follows:

$$K(A) = \|A\|_\infty \| |A^{-1}\|_\infty = 22.5 >> 1,$$

which shows that the matrix is ill-conditioned. Thus the given system is ill-conditioned.

(b) The residual vector can be calculated as

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}^* = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -2 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2.01 \\ 1.01 \\ 1.98 \end{pmatrix} = \begin{pmatrix} -0.04 \\ -0.07 \\ 0.03 \end{pmatrix},$$

and it gives

$$\|\mathbf{r}\|_\infty = 0.07.$$

(c) From (3.57), we have

$$\frac{\|\mathbf{x} - \mathbf{x}^*\|}{\|\mathbf{x}\|} \leq K(A) \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}.$$

By using above parts (a) and (b) and the value $\|\mathbf{b}\|_\infty = 1$, we obtain

$$\frac{\|\mathbf{x} - \mathbf{x}^*\|}{\|\mathbf{x}\|} \leq (22.5) \frac{(0.07)}{1} = 1.575$$

(d) To solve the linear system $A\mathbf{e} = \mathbf{r}$, where

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -2 \\ -2 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{r} = \begin{pmatrix} -0.04 \\ -0.07 \\ 0.03 \end{pmatrix},$$

and $\mathbf{e} = \mathbf{x} - \mathbf{x}^*$. *Writing the above system in the augmented matrix form*

$$\left(\begin{array}{rrrcr} 1 & 1 & -1 & \vdots & -0.04 \\ 1 & 2 & -2 & \vdots & -0.07 \\ -2 & 1 & 1 & \vdots & 0.03 \end{array}\right).$$

After applying forward elimination step of the simple Gauss elimination method, we obtain

$$\left(\begin{array}{rrrcr} 1 & 1 & -1 & \vdots & -0.04 \\ 0 & 1 & -1 & \vdots & -0.03 \\ 0 & 0 & 2 & \vdots & 0.04 \end{array}\right).$$

Now by using the backward substitution, we obtain the solution

$$\mathbf{e}^* = [-0.01, -0.01, 0.02]^T,$$

which is the required approximation of the exact error. •