1

SYSTEMS OF LINEAR EQUATIONS

- Introduction
- Elimination Methods
- Decomposition Methods
- Matrix Inverse and Determinant
- Errors, Residuals and Condition Number
- Iteration Methods
- Incomplete and Redundant Systems

1.1 Introduction

- The system of *linear* equations is formed by the addition of the products of a variable with a coefficient, which is also a *constant*.
- The system of linear equation can be solved via *matrix approach*.
- The general form of a set of a linear equation having n linear equations and n unknowns is

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad \vdots \qquad \ddots \qquad \vdots \quad & \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \tag{1.1}$$

where $x_1, x_2, \ldots, x_n$ are variables or unknowns, a_{ij} and b_j are coefficient or constant (real or complex).

- Eq. (1.1) can be written in a more compact form:

$$[a_{ij}] \cdot \{x_j\} = \{b_i\} \qquad \mathbf{A} \cdot \mathbf{x} = \mathbf{b} \tag{1.2}$$

where **A** is a matrix $[a_{ij}]$ of size $n \times n$, **x** is a variable vector $\{x_j\}$ and **b** is a right-hand side vector $\{b_j\}$.

- The process of solving Eq. (1.2) yield *three* possible solutions:

1. **Unique solution** — e.g.:

$$\begin{aligned} 3x_1 + x_2 &= 1 \\ x_1 + 3x_2 &= 1 \end{aligned} \qquad x_1 = x_2 = \tfrac{1}{4}$$

2. **No solution** — e.g.:

$$\begin{aligned} -x_1 + x_2 &= 1 \\ x_1 - x_2 &= 1 \end{aligned}$$

3. **Infinite solutions** — e.g.:

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + 2x_2 &= 2 \end{aligned}$$

1.2 Elimination Methods

- The most popular method is the *Gauss elimination* method, which comprises of two steps:

  1. **Forward elimination** to form an *upper triangular* system via *row-based transformation* process,
  2. **Back substitution** to produce the solution of x_j.

- Consider the following system:

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

If $a_{11} \neq 0$, for $i = 2,3,\ldots,n$, substract the i-th equation with the product of a_{i1}/a_{11} with the first equation to produce the first transformed system:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ & & a_{22}^{(1)}x_2 & + & \cdots & + & a_{2n}^{(1)}x_n & = & b_2^{(1)} \\ & & \vdots & & \ddots & & \vdots & & \vdots \\ & & a_{n2}^{(1)}x_2 & + & \cdots & + & a_{nn}^{(1)}x_n & = & b_n^{(1)} \end{array}$$

where

$$a_{ij}^{(1)} = a_{ij} - \frac{a_{i1}}{a_{11}}a_{1j} \quad \text{for } i, j = 2,3,\ldots,n$$

$$b_i^{(1)} = b_i - \frac{a_{i1}}{a_{11}}b_1 \quad \text{for } i = 2,3,\ldots,n$$

The process can be repeated for $(n-1)$ times until the $(n-1)$-th transformed system is formed as followed, which completes the forward eliminations:

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ & & a_{22}^{(1)}x_2 & + & a_{23}^{(1)}x_3 & + & \cdots & + & a_{2n}^{(1)}x_n & = & b_2^{(1)} \\ & & & & a_{33}^{(2)}x_3 & + & \cdots & + & a_{3n}^{(2)}x_n & = & b_2^{(2)} \\ & & & & & & \ddots & & \vdots & & \vdots \\ & & & & & & & & a_{nn}^{(n-1)}x_n & = & b_n^{(n-1)} \end{array} \qquad (1.3)$$

where

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} a_{kj}^{(k-1)} \qquad \text{for } i, j = k+1,\ldots,n \tag{1.4a}$$

$$b_{i}^{(k)} = b_{i}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} b_{k}^{(k-1)} \qquad \text{for } i = k+1,\ldots,n \tag{1.4b}$$

Back substitutions can then be executed so that x_j are solved:

$$x_n = \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}} \tag{1.5a}$$

$$x_k = \frac{1}{a_{kk}^{(k-1)}}\left[b_k^{(k-1)} - \sum_{j=k+1}^{n} a_{kj}^{(k-1)} x_j \right] \quad \text{for } k = n-1,\ldots,1 \tag{1.5b}$$

- The above method can fail if $a_{kk} \to 0$, the row has to be interchanged, which is referred to as *pivoting*:

$$\begin{array}{ccccc} & & x_2 & = & 2 \\ x_1 & + & x_2 & = & 3 \end{array} \xrightarrow{\text{Pivoting}} \begin{array}{ccccc} x_1 & + & x_2 & = & 3 \\ & & x_2 & = & 2 \end{array}$$

where the new diagonal element a_{kk}^{*} is called a *pivot*, which can be selected among the maximum absolute value of a_{ik}.

- The pivotal Gauss elimination gives a more accurate solutions, e.g. consider these systems (values to be rounded up to 3 significant figures):

Original Gauss elimination:

$$\left[\begin{array}{cc|c} 0.00126 & 0.417 & 0.418 \\ 1.34 & -0.708 & 0.632 \end{array}\right] \xrightarrow{(2)-\frac{1.34}{0.00126}(1)\Rightarrow(2)} \left[\begin{array}{cc|c} 0.00126 & 0.417 & 0.418 \\ 0.0 & -444.184 & -443.908 \end{array}\right]$$

$$x_2 = 0.999 \qquad x_1 = 1.125$$

Pivotal Gauss elimination:

$$\left[\begin{array}{cc|c} 1.34 & -0.708 & 0.632 \\ 0.00126 & 0.417 & 0.418 \end{array}\right] \xrightarrow{(2)-\frac{0.00126}{1.34}(1)\Rightarrow(2)} \left[\begin{array}{cc|c} 1.34 & -0.708 & 0.632 \\ 0.0 & 0.418 & 0.417 \end{array}\right]$$

$$x_2 = 0.998 \qquad x_1 = 0.999$$

Exact solution:

$$x_1 = 1 \qquad x_2 = 1$$

Example 1.1

Solve the following system using the Gauss elimination method:

$$\begin{aligned} 2x_1 + x_2 + 3x_3 &= 1 \\ 4x_1 + 4x_2 + 7x_3 &= 1 \\ 2x_1 + 5x_2 + 9x_3 &= 3 \end{aligned}$$

Solution

The system can be rewritten in matrix form as:

$$\begin{bmatrix} 2 & 1 & 3 \\ 4 & 4 & 7 \\ 2 & 5 & 9 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \\ 3 \end{Bmatrix} \quad \text{or} \quad [\mathbf{A} \mid \mathbf{b}] \equiv \left[\begin{array}{ccc|c} 2 & 1 & 3 & 1 \\ 4 & 4 & 7 & 1 \\ 2 & 5 & 9 & 3 \end{array}\right]$$

First step of forward elimination:

$$\xrightarrow[(3)-(1)\Rightarrow(3)]{(2)-2(1)\Rightarrow(2)} \left[\begin{array}{ccc|c} 2 & 1 & 3 & 1 \\ 0 & 2 & 1 & -1 \\ 0 & 4 & 6 & 2 \end{array}\right]$$

Second step of forward elimination:

$$\xrightarrow{(3)-2(2)\Rightarrow(3)} \left[\begin{array}{ccc|c} 2 & 1 & 3 & 1 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 4 & 4 \end{array}\right]$$

Hence, the transformed upper triangular system is:

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

Back substitutions are as follows

$$x_3 = 4/4 = 1$$

$$x_2 = \frac{-1 - x_3}{2} = -1$$

$$x_1 = \frac{1 - x_2 - 3x_3}{2} = -\tfrac{1}{2}$$

Example 1.2

Perform the pivotal Gauss elimination to the system given in Example 1.1.

Solution

The pivotal Gauss elimination can be performed as followed:

$$\left[\begin{array}{ccc|c} 2 & 1 & 3 & 1 \\ 4 & 4 & 7 & 1 \\ 2 & 5 & 9 & 3 \end{array}\right] \xrightarrow{(1)\Leftrightarrow(2)} \left[\begin{array}{ccc|c} 4 & 4 & 7 & 1 \\ 2 & 1 & 3 & 1 \\ 2 & 5 & 9 & 3 \end{array}\right] \xrightarrow[(3)-\frac{2}{4}(1)\Rightarrow(3)]{(2)-\frac{2}{4}(1)\Rightarrow(2)} \left[\begin{array}{ccc|c} 4 & 4 & 7 & 1 \\ 0 & -1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 3 & \frac{11}{2} & \frac{5}{2} \end{array}\right]$$

$$\xrightarrow{(2)\Leftrightarrow(3)} \left[\begin{array}{ccc|c} 4 & 4 & 7 & 1 \\ 0 & 3 & \frac{11}{2} & \frac{5}{2} \\ 0 & -1 & -\frac{1}{2} & \frac{1}{2} \end{array}\right] \xrightarrow{(3)-\frac{-1}{3}(2)\Rightarrow(3)} \left[\begin{array}{ccc|c} 4 & 4 & 7 & 1 \\ 0 & 3 & \frac{11}{2} & \frac{5}{2} \\ 0 & 0 & \frac{4}{3} & \frac{4}{3} \end{array}\right]$$

Hence, the upper triangular system is:

$$\begin{array}{rcrcrcl} 4x_1 & + & 4x_2 & + & 7x_3 & = & 1 \\ & & 3x_2 & + & \frac{11}{2}x_3 & = & \frac{5}{2} \\ & & & & \frac{4}{3}x_3 & = & \frac{4}{3} \end{array}$$

Then, back substitution can be performed:

$$x_3 = \tfrac{4}{3}\Big/\tfrac{4}{3} = 1,$$

$$x_2 = \frac{\frac{5}{2} - \frac{11}{2}(1)}{3} = -1,$$

$$x_1 = \frac{1 - 4(-1) - 7(1)}{4} = -\tfrac{1}{2}.$$

□

1.3 Decomposition Methods

- In some cases, the left-hand side matrix **A** is frequently used while the right-hand side vector **b** is changed depending on the case.

- The overall system can be transformed to an upper triangular form so that it can be used repeatedly for different **b**, thus matrix **A** has to be decomposed.

- For a general *non-symmetric* system, the popular method is the *Doolitle* or *LU decomposition*:

$$\mathbf{A} = \mathbf{LU} \tag{1.6}$$

where **L** and **U** are the lower and upper triangular matrices, respectively:

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

$$\equiv \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ l_{21} & u_{22} & u_{23} \\ l_{31} & l_{32} & u_{33} \end{bmatrix} \text{ (in memory)}$$

The solution steps of the system are as followed:

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} \quad \Rightarrow \quad \mathbf{LU} \cdot \mathbf{x} = \mathbf{b}$$

By taking an intermediate vector **y**:

$$\mathbf{U} \cdot \mathbf{x} = \mathbf{y} \tag{1.7}$$

Hence,

$$\mathbf{L} \cdot \mathbf{y} = \mathbf{b} \tag{1.8}$$

- The elements for **L** and **U** can be obtained from the Gauss elimination:

$$\mathbf{U} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} \\ 0 & 0 & a_{33}^{(2)} \end{bmatrix} \qquad \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ a_{21}/a_{11} & 1 & 0 \\ a_{31}/a_{11} & a_{32}^{(1)}/a_{22}^{(1)} & 1 \end{bmatrix}$$

- Another variation of the LU decomposition is the *Crout decomposition*, which maintains $u_{ii} = 1$ for $i = 1,2,\ldots,n$ in **U** instead of **L**:

  For the first row and column:

$$l_{i1} = a_{i1} \qquad \text{for } i = 1,2,\ldots,n \tag{1.9a}$$

$$u_{1j} = \frac{a_{1j}}{l_{11}} \qquad \text{for } j = 2,3,\ldots,n \tag{1.9b}$$

  For $j = 2,3,\ldots,n-1$:

$$l_{ij} = a_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj} \qquad \text{for } i = j, j+1,\ldots,n \tag{1.9c}$$

$$u_{jk} = \frac{a_{jk} - \sum_{i=1}^{j-1} l_{ji} u_{ik}}{l_{jj}} \qquad \text{for } k = j+1, j+2,\ldots,n \tag{1.9d}$$

  dan,

$$l_{nn} = a_{nn} - \sum_{k=1}^{n-1} l_{nk} u_{kn} \tag{1.9e}$$

- If the system is *symmetric*, the *Cholesky decomposition* can be used, where matrix **A** can be decomposed such that:

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathrm{T}} \tag{1.10}$$

  For the k-th row:

$$l_{ki} = \frac{a_{ki} - \sum_{j=1}^{i-1} l_{ij} l_{kj}}{l_{ii}} \qquad \text{for } i = 1,2,\ldots,k-1 \tag{1.11a}$$

$$l_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} l_{kj}^2} \tag{1.11b}$$

  This method optimises the use of computer memory in storing the decomposed form of **A**.

Example 1.5

Decompose the following matrix using the Doolittle LU decomposition:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 4 & 7 \\ 2 & 5 & 9 \end{bmatrix}$$

Solution

With reference to the matrix elements derived in Example 1.1:

$$\mathbf{U} = \begin{bmatrix} 2 & 1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 4 \end{bmatrix}, \qquad \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 4/2 & 1 & 0 \\ 2/2 & 4/2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

Example 1.7

Decompose the following matrix using the Cholesky decomposition:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 5 \\ 3 & 5 & 9 \end{bmatrix}$$

Solution

By using Eq. 1.11:

$$l_{11} = \sqrt{a_{11}} = \sqrt{2}, \qquad l_{21} = a_{21}/l_{11} = 1/\sqrt{2},$$
$$l_{31} = a_{31}/l_{11} = 3/\sqrt{2}, \qquad l_{22} = \sqrt{a_{22} - l_{21}^2} = \sqrt{7/2},$$
$$l_{32} = \frac{a_{32} - l_{21}l_{31}}{l_{22}} = \sqrt{7/2}, \qquad l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2} = 1.$$

Maka,

$$\mathbf{L} = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & \sqrt{7/2} & 0 \\ 3/\sqrt{2} & \sqrt{7/2} & 1 \end{bmatrix} = \begin{bmatrix} 1.41421 & 0 & 0 \\ 0.70712 & 1.87083 & 0 \\ 2.12132 & 1.87083 & 1 \end{bmatrix}$$

1.4 Matrix Inverse and Determinant

- The Gauss elimination can be used to generate the inverse of a square matrix **A** by replacing the left-hand side vector **b** with an identity matrix **I**.

- By using the following identity:

$$\mathbf{A}\cdot\mathbf{A}^{-1}=\mathbf{I} \tag{1.12}$$

If all columns of $\mathbf{A}^{-1}$ are written as $\mathbf{x}^{(1)},\mathbf{x}^{(2)},\ldots,\mathbf{x}^{(n)}$ and the columns of the **I** as $\mathbf{e}^{(1)},\mathbf{e}^{(2)},\ldots,\mathbf{e}^{(n)}$, respectively, thus Eq. (1.12) can be rewritten as:

$$\mathbf{A}\cdot\left(\mathbf{x}^{(1)},\mathbf{x}^{(2)},\ldots,\mathbf{x}^{(n)}\right)=\left(\mathbf{e}^{(1)},\mathbf{e}^{(2)},\ldots,\mathbf{e}^{(n)}\right)$$

Then, a set of *n* linear systems can be assembled:

$$\begin{gathered}\mathbf{A}\cdot\mathbf{x}^{(1)}=\mathbf{e}^{(1)}\\ \mathbf{A}\cdot\mathbf{x}^{(2)}=\mathbf{e}^{(2)}\\ \vdots\\ \mathbf{A}\cdot\mathbf{x}^{(n)}=\mathbf{e}^{(n)}\end{gathered} \tag{1.13}$$

- Consequently, the determinant of matrix **A** can simply be calculated using:

$$\det(\mathbf{A})\equiv|\mathbf{A}|=(-1)^p\,a_{11}a_{22}^{(1)}a_{33}^{(2)}\ldots a_{nn}^{(n-1)}=(-1)^p\prod_{i=1}^{n}a_{ii}^{(i-1)} \tag{1.14}$$

where *p* is the number of row interchange operation during pivoting.

Example 1.8

Determine the inverse of the following matrix using the Gauss elimination:

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

Solution

The combination of **A** and **I** can be represented in an augmented form:

$$\left[\begin{array}{ccc|ccc} 4 & 2 & -1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 2 & -1 & -1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\text{Gauss forward elimination}} \left[\begin{array}{ccc|ccc} 4 & 2 & -1 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{5}{4} & -\frac{1}{4} & 1 & 0 \\ 0 & 0 & \frac{9}{2} & -\frac{3}{2} & 4 & 1 \end{array}\right]$$

Upon back substitution:

$$\mathbf{x}^{(1)} = \begin{bmatrix} 0 \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} \qquad \mathbf{x}^{(2)} = \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{9} \\ \frac{8}{9} \end{bmatrix} \qquad \mathbf{x}^{(3)} = \begin{bmatrix} \frac{1}{3} \\ -\frac{5}{9} \\ \frac{2}{9} \end{bmatrix}$$

Hence, the inverse of **A** is

$$\mathbf{A}^{-1} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{9} & -\frac{5}{9} \\ -\frac{1}{3} & \frac{8}{9} & \frac{2}{9} \end{bmatrix}$$

Example 1.9

Calculate the determinant of the matrix given in Example 1.8.

Penyelesaian

In Example 1.0, there is no row interchange performed, thus $p = 0$. Hence,

$$\det(\mathbf{A}) = \begin{vmatrix} 4 & 2 & -1 \\ 1 & 1 & 1 \\ 2 & -1 & -1 \end{vmatrix} = (-1)^0 \times \begin{vmatrix} 4 & 2 & -1 \\ 0 & \frac{1}{2} & \frac{5}{4} \\ 0 & 0 & \frac{9}{2} \end{vmatrix} = (4)\left(\tfrac{1}{2}\right)\left(\tfrac{9}{2}\right) = 9$$

1.5 Errors, Residuals and Condition Number

- If $\mathbf{x}^*$ is an approximate solution of a linear system $\mathbf{A}\cdot\mathbf{x}=\mathbf{b}$, then the system *error* is defined as

$$\mathbf{e}=\mathbf{x}-\mathbf{x}^* \tag{1.16}$$

- On the other hand, the system *residue* **r** is defined as

$$\mathbf{r}=\mathbf{A}\cdot\mathbf{e} \tag{1.17}$$

or,
$$\mathbf{r}=\mathbf{A}\cdot\mathbf{x}-\mathbf{A}\cdot\mathbf{x}^*=\mathbf{b}-\mathbf{A}\cdot\mathbf{x}^*$$

- For *a well-conditioned* system, the residue can represent the error.

- Moreover, for comparison, a matrix or vector can be expressed in form of a scalar known as *norm*.

- For a vector $\mathbf{x}=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$, the *p*-norm is defined as

$$\|\mathbf{x}\|_p=\left(|x_1|^p+|x_2|^p+\cdots+|x_n|^p\right)^{1/p}=\left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p} \tag{1.20}$$

If $p = 1$, it is known as 1-*norm*:

$$\|\mathbf{x}\|_1=|x_1|+|x_2|+\cdots+|x_n|=\sum_{i=1}^{n}|x_i| \tag{1.19}$$

If $p = 2$, it is known as *Euclidean norm*:

$$\|\mathbf{x}\|_e=\|\mathbf{x}\|_2=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}=\sqrt{\sum_{i=1}^{n}x_i^2} \tag{1.18}$$

If $p \to \infty$, it is known as a *maximum norm*:

$$\|\mathbf{x}\|_\infty=\max_{1\le i\le n}\left\{|x_1|,|x_2|,\ldots,|x_n|\right\}=\max_{1\le i\le n}|x_i| \tag{1.21}$$

- For a matrix $\mathbf{A}=[a_{ij}]$ of size $m\times n$, the *Frobenius* norm, which is equivalent to the Euclidean norm for vectors, is defined as

$$\|\mathbf{A}\|_e=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^2} \tag{1.22}$$

and, the equivalent 1-norm and maximum norm for a matrix are defined as

$$\|\mathbf{A}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}| = \text{maximum sum of columns} \tag{1.23}$$

$$\|\mathbf{A}\|_\infty = \max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}| = \text{maximum sum of rows} \tag{1.24}$$

- The properties of norms of a vector or matrix **A** are as followed:

1. $\|\mathbf{A}\| \ge 0$ and $\|\mathbf{A}\| = 0$ if, and only if, $\mathbf{A} = \mathbf{0}$.
2. $\|c\mathbf{A}\| = |c|\cdot\|\mathbf{A}\|$ where c is a scalar quantity.
3. $\|\mathbf{A}+\mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|$ *Triangular inequality*, where **B** is a vector or matrix of the same dimension of **A**.
4. $\|\mathbf{A}\cdot\mathbf{B}\| \le \|\mathbf{A}\|\cdot\|\mathbf{B}\|$ *Schwarz inequality*, where **B** is a vector or matrix which forms a valid product with **A**.

- The concept of norms can be used to calculate the *condition number* represents the 'health' of a linear system, either ill- or well-conditioned.

- If **e** is the error for the system $\mathbf{A}\cdot\mathbf{x} = \mathbf{b}$, from the relations $\mathbf{A}\cdot\mathbf{e} = \mathbf{r}$ and $\mathbf{e} = \mathbf{A}^{-1}\cdot\mathbf{r}$, the following inequality can be established:

$$\|\mathbf{A}\|\cdot\|\mathbf{e}\| \ge \|\mathbf{r}\| \quad \text{and} \quad \|\mathbf{e}\| \le \|\mathbf{A}^{-1}\|\cdot\|\mathbf{r}\| \quad \Rightarrow \quad \frac{\|\mathbf{r}\|}{\|\mathbf{A}\|} \le \|\mathbf{e}\| \le \|\mathbf{A}^{-1}\|\cdot\|\mathbf{r}\|$$

Also, from $\mathbf{A}\cdot\mathbf{x} = \mathbf{b}$ and $\mathbf{x} = \mathbf{A}^{-1}\cdot\mathbf{b}$:

$$\|\mathbf{A}\|\cdot\|\mathbf{x}\| \ge \|\mathbf{b}\| \quad \text{and} \quad \|\mathbf{x}\| \le \|\mathbf{A}^{-1}\|\cdot\|\mathbf{b}\| \quad \Rightarrow \quad \frac{\|\mathbf{b}\|}{\|\mathbf{A}\|} \le \|\mathbf{x}\| \le \|\mathbf{A}^{-1}\|\cdot\|\mathbf{b}\|$$

Thus, the combination of both inequality relations yields the range of the relative error $\|\mathbf{e}\|/\|\mathbf{x}\|$, i.e.

$$\frac{1}{\|\mathbf{A}\|\cdot\|\mathbf{A}^{-1}\|}\cdot\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} \le \frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} \le \left(\|\mathbf{A}\|\cdot\|\mathbf{A}^{-1}\|\right)\cdot\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$$

- Hence, the *condition number* is defined as

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\cdot\|\mathbf{A}^{-1}\| \tag{1.25}$$

where the range of the relative error is.

$$\frac{1}{\kappa(\mathbf{A})}\cdot\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}\le\frac{\|\mathbf{e}\|}{\|\mathbf{x}\|}\le\kappa(\mathbf{A})\cdot\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} \tag{1.26}$$

- The characteristics of the condition number are that:

  1. $\kappa(\mathbf{A})\ge 1$ — the smaller the better, and otherwise.
  2. If $\kappa(\mathbf{A})\to 1$, the relative residual $\|\mathbf{r}\|/\|\mathbf{b}\|$ can represent the relative errors $\|\mathbf{e}\|/\|\mathbf{x}\|$.

- If the error is solely contributed by matrix **A**, the inequality becomes:

$$\frac{\|\mathbf{e}\|}{\|\mathbf{x}^*\|}\le\kappa(\mathbf{A})\cdot\frac{\|\mathbf{E}_A\|}{\|\mathbf{A}\|} \tag{1.27}$$

- On the other hand, if the error is solely contributed by vector **b**, the inequality becomes:

$$\left\|\frac{\mathbf{e}}{\mathbf{x}}\right\|\le\kappa(\mathbf{A})\cdot\frac{\|\mathbf{e}_b\|}{\|\mathbf{b}\|} \tag{1.28}$$

- Therefore, from Eqs. (1.26-8), it can be seen that the condition number can determine the range of error and thus the health of a system.

1.7 Iteration Methods

- For large systems (size > 200), the elimination and decomposition methods are not efficient due to increasing number of arithmetic operations.

- The number of arithmetic operations can be reduced via iteration methods, such as the *Jacobi iteration* and the *Gauss-Seidel iteration* methods.

- In the Jacobi iteration, Eq. (1.1) can be written for x_i from the i-th equation:

$$\begin{aligned} x_1 &= -\frac{1}{a_{11}}\left(a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n - b_1\right) \\ x_2 &= -\frac{1}{a_{22}}\left(a_{21}x_1 + a_{23}x_3 + \cdots + a_{2n}x_n - b_2\right) \\ &\vdots \\ x_n &= -\frac{1}{a_{nn}}\left(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{n,n-1}x_{n-1} - b_n\right) \end{aligned} \tag{1.29}$$

Eq. (1.29) needs initial values $\mathbf{x}^{(0)} = \left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}\right)^{\mathrm{T}}$, which yield $\mathbf{x}^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}\right)^{\mathrm{T}}$, and the computation continues as followed:

$$\begin{aligned} x_1^{(k+1)} &= -\frac{1}{a_{11}}\left(a_{12}x_2^{(k)} + a_{13}x_3^{(k)} + \cdots + a_{1n}x_n^{(k)} - b_1\right) \\ x_2^{(k+1)} &= -\frac{1}{a_{22}}\left(a_{21}x_1^{(k)} + a_{23}x_3^{(k)} + \cdots + a_{2n}x_n^{(k)} - b_2\right) \\ &\vdots \\ x_n^{(k+1)} &= -\frac{1}{a_{nn}}\left(a_{n1}x_1^{(k)} + a_{n3}x_3^{(k)} + \cdots + a_{n,n-1}x_{n-1}^{(k)} - b_n\right) \end{aligned} \tag{1.30}$$

For $k \to \infty$, vector $\mathbf{x}^{(k)}$ converges to its exact solution if the *diagonal domain condition* is followed, i.e.

$$\left|a_{ii}\right| > \sum_{\substack{j=1 \\ j\neq i}}^{n} \left|a_{ij}\right| \qquad \text{for } i = 1, 2, \ldots, n \tag{1.31}$$

and the matrix which follows this condition is called a *diagonal domain matrix*.

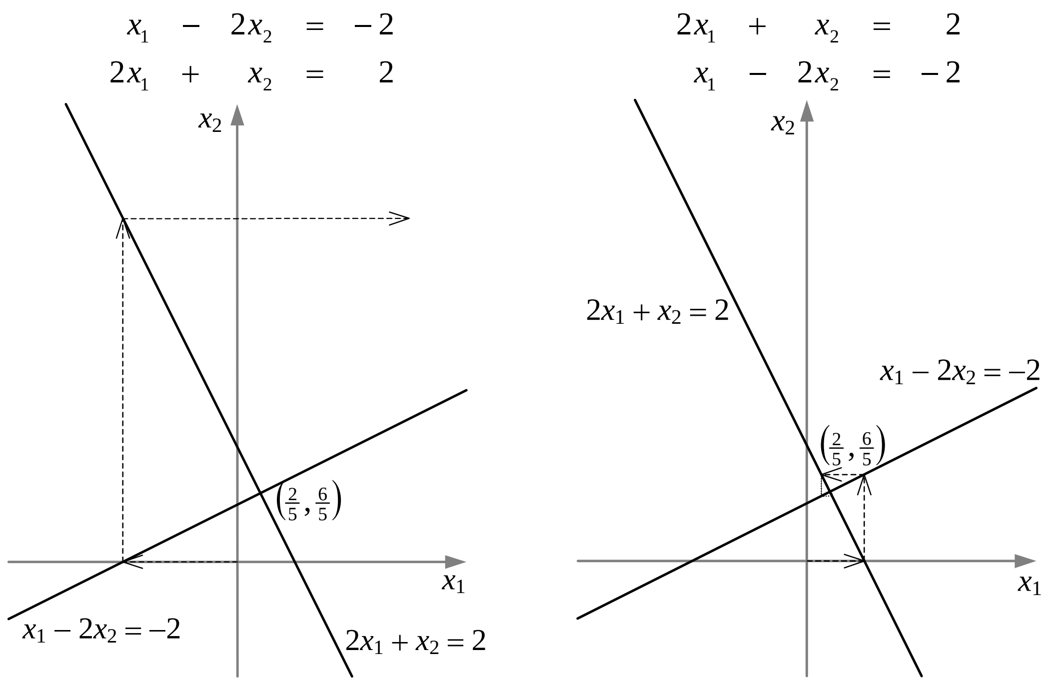
- To terminate the iteration process, a *convergence* or *termination criterion* can be specified, i.e.

$$\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right\| < \varepsilon \tag{1.32}$$

- The Gauss-Siedel iteration method uses the most current known solution after each arithmetic operation in order to speed up convergence:

$$\begin{aligned}
x_1^{(k+1)} &= -\frac{1}{a_{11}}\left(a_{12}x_2^{(k)} + a_{13}x_3^{(k)} + \cdots + a_{1n}x_n^{(k)} - b_1\right)\\
x_2^{(k+1)} &= -\frac{1}{a_{22}}\left(a_{21}x_1^{(k+1)} + a_{23}x_3^{(k)} + \cdots + a_{2n}x_n^{(k)} - b_2\right)\\
&\vdots\\
x_n^{(k+1)} &= -\frac{1}{a_{nn}}\left(a_{n1}x_1^{(k+1)} + a_{n3}x_3^{(k+1)} + \cdots + a_{n,n-1}x_{n-1}^{(k+1)} - b_n\right)
\end{aligned} \tag{1.33}$$

- As of the Jacobi method, the Gauss-Siedel method must also observe the diagonal domain condition for convergence to be possible (see Fig. 1.1).



(a) The off-diagonal domain system (b) The diagonal domain system

FIG. 1.1 Divergence and convergence in the Gauss-Seidel method

Example 1.10

Use the Jacobi iteration method to solve the following system up to 5 decimal points:

$$\begin{aligned} 64x_1 - 3x_2 - x_3 &= 14 \\ x_1 + x_2 + 40x_3 &= 20 \\ 2x_1 - 90x_2 + x_3 &= -5 \end{aligned}$$

Solution

First of all, form a diagonal domain system:

$$\begin{aligned} 64x_1 - 3x_2 - x_3 &= 14 \\ 2x_1 - 90x_2 + x_3 &= -5 \\ x_1 + x_2 + 40x_3 &= 20 \end{aligned}$$

Then, rewrite the system according to Eq. (1.30):

$$x_1^{(k+1)} = -\frac{1}{64}\left(-3x_2^{(k)} - x_3^{(k)} - 14\right)$$

$$x_2^{(k+1)} = +\frac{1}{90}\left(+2x_1^{(k)} + x_3^{(k)} + 5\right)$$

$$x_3^{(k+1)} = -\frac{1}{40}\left(+x_1^{(k)} + x_2^{(k)} - 20\right)$$

By taking an initial values $\mathbf{x}^{(0)} = (0,\ 0,\ 0)^{\mathrm{T}}$, thus the method converges within **5** iterations:

Iteration no. 1: $\mathbf{x}^{(1)} = (0.21875, 0.05556, 0.50000)^{\mathrm{T}}$,

Iteration no. 2: $\mathbf{x}^{(2)} = (0.22917, 0.06597, 0.49592)^{\mathrm{T}}$,

Iteration no. 3: $\mathbf{x}^{(3)} = (0.22955, 0.06613, 0.49262)^{\mathrm{T}}$,

Iteration no. 4: $\mathbf{x}^{(4)} = (0.22955, 0.06613, 0.49261)^{\mathrm{T}}$,

Iteration no. 5: $\mathbf{x}^{(5)} = (0.22955, 0.06613, 0.49261)^{\mathrm{T}}$.

Example 1.11

Repeat problem given in Example 1.10 using the Gauss-Seidel iteration method.

Solution

First of all, form a diagonal domain system:

$$\begin{aligned} 64x_1 &- 3x_2 - x_3 = 14 \\ 2x_1 &- 90x_2 + x_3 = -5 \\ x_1 &+ x_2 + 40x_3 = 20 \end{aligned}$$

By taking an initial values $\mathbf{x}^{(0)} = (0, 0, 0)^T$, the first solution in the first iteration:

$$x_1^{(1)} = -\frac{1}{64}\left[-3(0) - 0 - 14\right] = 0.21875$$

Use $x_1^{(1)}$ to calculate $x_2^{(1)}$ and so on, i.e.

$$x_2^{(1)} = \frac{1}{90}\left[2(0.21875) + 0 + 5\right] = 0.06042$$

$$x_3^{(1)} = -\frac{1}{40}\left(0.21875 + 0.06042 - 20\right) = 0.49302$$

Hence, the method converges within **4** iterations:

$$\mathbf{x}^{(1)} = (0.21875, 0.06042, 0.49302)^T,$$
$$\mathbf{x}^{(2)} = (0.22929, 0.06613, 0.49262)^T,$$
$$\mathbf{x}^{(3)} = (0.22955, 0.06613, 0.49261)^T,$$
$$\mathbf{x}^{(4)} = (0.22955, 0.06613, 0.49261)^T.$$

1.8 Incomplete and Redundant Systems

- If $m \neq n$, there will be two situations:

  1. $m < n$ — incomplete system.
  2. $m > n$ — redundant system.

- For incomplete system, no solution is possible since additional $(n - m)$ equations from other independent sources are required until $m = n$.

- For redundant system, a unique solution is not possible, and the system has to be optimised via *least square method* (also known as *linear regression*):

$$\begin{aligned} S &= \|\mathbf{e}\|_e^2 = \mathbf{e}^{\mathrm{T}}\mathbf{e}, \\ &= (\mathbf{b} - \mathbf{A}\cdot\mathbf{x})^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\cdot\mathbf{x}), \\ &= \mathbf{b}^{\mathrm{T}}\mathbf{b} - (\mathbf{A}\cdot\mathbf{x})^{\mathrm{T}}\mathbf{b} - \mathbf{b}^{\mathrm{T}}(\mathbf{A}\cdot\mathbf{x}) + (\mathbf{A}\cdot\mathbf{x})^{\mathrm{T}}(\mathbf{A}\cdot\mathbf{x}), \\ &= (\mathbf{A}\cdot\mathbf{x})^{\mathrm{T}}(\mathbf{A}\cdot\mathbf{x}) - (\mathbf{A}\cdot\mathbf{x})^{\mathrm{T}}\mathbf{b}. \end{aligned}$$

Using the identity $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$:

$$S = \mathbf{x}^{\mathrm{T}} \cdot \mathbf{A}^{\mathrm{T}}\mathbf{A}\cdot\mathbf{x} - \mathbf{x}^{\mathrm{T}} \cdot \mathbf{A}^{\mathrm{T}} \cdot \mathbf{b}$$

Minimising S:

$$\frac{\partial S}{\partial \mathbf{x}^{\mathrm{T}}} = 0 = \mathbf{A}^{\mathrm{T}}\mathbf{A}\cdot\mathbf{x} - \mathbf{A}^{\mathrm{T}} \cdot \mathbf{b}$$

forms an approximate system of n equations, i.e.

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\cdot\mathbf{x} = \mathbf{A}^{\mathrm{T}} \cdot \mathbf{b} \tag{1.34}$$

where the left-hand side matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is *symmetry* and the standard deviation σ can be calculated from the Euclidean norm of $\mathbf{e}$, i.e.:

$$\sigma = \sqrt{\frac{S}{m-n}} = \frac{\|\mathbf{e}\|_e}{\sqrt{m-n}} \tag{1.35}$$

Example 1.12

Calculate the best approximate solution for the following system:

$$
\begin{array}{rcrcrcr}
2x_1 & + & x_2 & + & 3x_3 & = & 1 \\
4x_1 & + & 4x_2 & + & 7x_3 & = & 1 \\
2x_1 & + & 5x_2 & + & 9x_3 & = & 3 \\
5x_1 & + & 5x_2 & + & 9x_3 & = & 2 \\
7x_1 & + & 10x_2 & + & 15x_3 & = & 4
\end{array}
$$

Also, calculate the resulting standard deviation.

Solution

The above system can be rewritted in form of $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ as:

$$
\begin{bmatrix} 2 & 1 & 3 \\ 4 & 4 & 7 \\ 2 & 5 & 9 \\ 5 & 5 & 9 \\ 7 & 10 & 15 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \\ 3 \\ 2 \\ 4 \end{Bmatrix}
$$

By using Eq. (1.34):

$$
\left(\begin{bmatrix} 2 & 4 & 2 & 5 & 7 \\ 1 & 4 & 5 & 5 & 10 \\ 3 & 7 & 9 & 9 & 15 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 4 & 4 & 7 \\ 2 & 5 & 9 \\ 5 & 5 & 9 \\ 7 & 10 & 15 \end{bmatrix} \right) \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{bmatrix} 2 & 4 & 2 & 5 & 7 \\ 1 & 4 & 5 & 5 & 10 \\ 3 & 7 & 9 & 9 & 15 \end{bmatrix} \begin{Bmatrix} 1 \\ 1 \\ 3 \\ 2 \\ 4 \end{Bmatrix}
$$

$$
\begin{bmatrix} 98 & 123 & 202 \\ 123 & 167 & 271 \\ 202 & 271 & 445 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 50 \\ 70 \\ 100 \end{Bmatrix}
$$

where its solutions are

$$
x_1 = -0.34930 , \quad x_2 = -0.01996 , \quad x_3 = -0.42914 .
$$

The standard deviation can be obtained from the Euclidean norm of the error **e**:

$$\mathbf{e} = \begin{Bmatrix} 1 \\ 1 \\ 3 \\ 2 \\ 4 \end{Bmatrix} - \begin{bmatrix} 2 & 1 & 3 \\ 4 & 4 & 7 \\ 2 & 5 & 9 \\ 5 & 5 & 9 \\ 7 & 10 & 15 \end{bmatrix} \begin{Bmatrix} -0.34930 \\ -0.01996 \\ 0.42914 \end{Bmatrix} = \begin{Bmatrix} 0.43114 \\ -0.52695 \\ -0.06387 \\ -0.01597 \\ 0.20758 \end{Bmatrix}$$

$$\|\mathbf{e}\|_e = \sqrt{0.43114^2 + (-0.52695)^2 + (-0.06387)^2 + (-0.01597)^2 + 0.20758^2},$$
$$= 0.71483.$$

Therefore,

$$\sigma = \frac{0.71483}{\sqrt{5-3}} = 0.50546$$

◳

Exercises

1. Consider the following system:

$$\begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 12 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{Bmatrix} = \begin{Bmatrix} 1.2 \\ -0.2 \\ 0.8 \\ 1.5 \end{Bmatrix}$$

   a. Use the Gauss elimination method to obtain the solution of x_i.
   b. Calculate the determinant for the left-hand side matrix.
   c. Generate the lower and upper triangular matrices using the Doolittle factorisation.

2. Consider the following system of 2 complex equations:

$$\begin{bmatrix} 2+2i & -1+2i \\ -3i & 3-2i \end{bmatrix} \begin{Bmatrix} z_1 \\ z_2 \end{Bmatrix} = \begin{Bmatrix} 1-4i \\ 2+4i \end{Bmatrix}$$

   By writing $z_k = x_k + y_k i$, solve the equation using the Gauss-Siedel iteration method using Microsoft Excel until it converges up to 5 decimal points.

3. Consider the following set of redundant equations:

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

   a. Derive an approximate system of linear equations and solve it via the Gauss elimination.
   b. Calculate the corresponding standard deviation.