# Notes on Mathematics - 102[1]

Peeyush Chandra, A. K. Lal, V. Raghavendra, G. Santhanam

[1] Supported by a grant from MHRD

# Contents

# Part I

# Linear Algebra

# Chapter 1

# Matrices

## 1.1 Definition of a Matrix

**Definition 1.1.1 (Matrix)** A rectangular array of numbers is called a matrix.

We shall mostly be concerned with matrices having real numbers as entries.

The horizontal arrays of a matrix are called its ROWS and the vertical arrays are called its COLUMNS. A matrix having $m$ rows and $n$ columns is said to have the order $m \times n$.

A matrix $A$ of ORDER $m \times n$ can be represented in the following form:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

where $a_{ij}$ is the entry at the intersection of the $i^{\text{th}}$ row and $j^{\text{th}}$ column.

In a more concise manner, we also denote the matrix $A$ by $[a_{ij}]$ by suppressing its order.

**Remark 1.1.2** *Some books also use* $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$ *to represent a matrix.*

Let $A = \begin{bmatrix} 1 & 3 & 7 \\ 4 & 5 & 6 \end{bmatrix}$. Then $a_{11} = 1, \ a_{12} = 3, \ a_{13} = 7, \ a_{21} = 4, \ a_{22} = 5, \text{ and } \ a_{23} = 6.$

A matrix having only one column is called a COLUMN VECTOR; and a matrix with only one row is called a ROW VECTOR.

WHENEVER A VECTOR IS USED, IT SHOULD BE UNDERSTOOD FROM THE CONTEXT WHETHER IT IS A ROW VECTOR OR A COLUMN VECTOR.

**Definition 1.1.3 (Equality of two Matrices)** Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ having the same order $m \times n$ are equal if $a_{ij} = b_{ij}$ for each $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

In other words, two matrices are said to be equal if they have the same order and their corresponding entries are equal.

**Example 1.1.4** The linear system of equations $2x + 3y = 5$ and $3x + 2y = 5$ can be identified with the matrix $\begin{bmatrix} 2 & 3 & : & 5 \\ 3 & 2 & : & 5 \end{bmatrix}$.

### 1.1.1 Special Matrices

**Definition 1.1.5** 1. A matrix in which each entry is zero is called a zero-matrix, denoted by $\mathbf{0}$. For example,

$$\mathbf{0}_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \mathbf{0}_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

2. A matrix having the number of rows equal to the number of columns is called a square matrix. Thus, its order is $m \times m$ (for some $m$) and is represented by $m$ only.

3. In a square matrix, $A = [a_{ij}]$, of order $n$, the entries $a_{11}, a_{22}, \ldots, a_{nn}$ are called the diagonal entries and form the principal diagonal of $A$.

4. A square matrix $A = [a_{ij}]$ is said to be a diagonal matrix if $a_{ij} = 0$ for $i \neq j$. In other words, the non-zero entries appear only on the principal diagonal. For example, the zero matrix $\mathbf{0}_n$ and $\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ are a few diagonal matrices.

   A diagonal matrix $D$ of order $n$ with the diagonal entries $d_1, d_2, \ldots, d_n$ is denoted by $D = \text{diag}(d_1, \ldots, d_n)$. If $d_i = d$ for all $i = 1, 2, \ldots, n$ then the diagonal matrix $D$ is called a **scalar matrix**.

5. A square matrix $A = [a_{ij}]$ with $a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ is called the identity matrix, denoted by $I_n$.

   For example, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

   The subscript $n$ is suppressed in case the order is clear from the context or if no confusion arises.

6. A square matrix $A = [a_{ij}]$ is said to be an upper triangular matrix if $a_{ij} = 0$ for $i > j$.

   A square matrix $A = [a_{ij}]$ is said to be an lower triangular matrix if $a_{ij} = 0$ for $i < j$.

   A square matrix $A$ is said to be triangular if it is an upper or a lower triangular matrix.

   For example $\begin{bmatrix} 2 & 1 & 4 \\ 0 & 3 & -1 \\ 0 & 0 & -2 \end{bmatrix}$ is an upper triangular matrix. An upper triangular matrix will be represented by $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$.

## 1.2 Operations on Matrices

**Definition 1.2.1 (Transpose of a Matrix)** The transpose of an $m \times n$ matrix $A = [a_{ij}]$ is defined as the $n \times m$ matrix $B = [b_{ij}]$, with $b_{ij} = a_{ji}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. The transpose of $A$ is denoted by $A^t$.

That is, by the transpose of an $m \times n$ matrix $A$, we mean a matrix of order $n \times m$ having the rows of $A$ as its columns and the columns of $A$ as its rows.

For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$ then $A^t = \begin{bmatrix} 1 & 0 \\ 4 & 1 \\ 5 & 2 \end{bmatrix}$.

Thus, the transpose of a row vector is a column vector and vice-versa.

**Theorem 1.2.2** For any matrix $A$, we have $(A^t)^t = A$.

Proof. Let $A = [a_{ij}]$, $A^t = [b_{ij}]$ and $(A^t)^t = [c_{ij}]$. Then, the definition of transpose gives

$$c_{ij} = b_{ji} = a_{ij} \quad \text{for all} \quad i, j$$

and the result follows. □

**Definition 1.2.3 (Addition of Matrices)** let $A = [a_{ij}]$ and $B = [b_{ij}]$ be are two $m \times n$ matrices. Then the sum $A + B$ is defined to be the matrix $C = [c_{ij}]$ with $c_{ij} = a_{ij} + b_{ij}$.

Note that, we define the sum of two matrices only when the order of the two matrices are same.

**Definition 1.2.4 (Multiplying a Scalar to a Matrix)** Let $A = [a_{ij}]$ be an $m \times n$ matrix. Then for any element $k \in \mathbb{R}$, we define $kA = [ka_{ij}]$.

For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$ and $k = 5$, then $5A = \begin{bmatrix} 5 & 20 & 25 \\ 0 & 5 & 10 \end{bmatrix}$.

**Theorem 1.2.5** Let $A, B$ and $C$ be matrices of order $m \times n$, and let $k, \ell \in \mathbb{R}$. Then

1. $A + B = B + A$ (commutativity).
2. $(A + B) + C = A + (B + C)$ (associativity).
3. $k(\ell A) = (k\ell)A$.
4. $(k + \ell)A = kA + \ell A$.

Proof. Part 1.
Let $A = [a_{ij}]$ and $B = [b_{ij}]$. Then

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = [b_{ij}] + [a_{ij}] = B + A$$

as real numbers commute.

The reader is required to prove the other parts as all the results follow from the properties of real numbers. □

**Exercise 1.2.6** 1. Suppose $A + B = A$. Then show that $B = \mathbf{0}$.

2. Suppose $A + B = \mathbf{0}$. Then show that $B = (-1)A = [-a_{ij}]$.

**Definition 1.2.7 (Additive Inverse)** Let $A$ be an $m \times n$ matrix.

1. Then there exists a matrix $B$ with $A + B = \mathbf{0}$. This matrix $B$ is called the additive inverse of $A$, and is denoted by $-A = (-1)A$.
2. Also, for the matrix $\mathbf{0}_{m \times n}$, $A + \mathbf{0} = \mathbf{0} + A = A$. Hence, the matrix $\mathbf{0}_{m \times n}$ is called the additive identity.

### 1.2.1 Multiplication of Matrices

**Definition 1.2.8 (Matrix Multiplication / Product)** Let $A = [a_{ij}]$ be an $m \times n$ matrix and $B = [b_{ij}]$ be an $n \times r$ matrix. The product $AB$ is a matrix $C = [c_{ij}]$ of order $m \times r$, with

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Observe that the product $AB$ is defined if and only if
THE NUMBER OF COLUMNS OF $A =$ THE NUMBER OF ROWS OF $B$.

For example, if $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 3 \\ 1 & 0 & 4 \end{bmatrix}$ then

$$AB = \begin{bmatrix} 1+0+3 & 2+0+0 & 1+6+12 \\ 2+0+1 & 4+0+0 & 2+12+4 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 19 \\ 3 & 4 & 18 \end{bmatrix}.$$

Note that in this example, while $AB$ is defined, the product $BA$ is not defined. However, for square matrices $A$ and $B$ of the same order, both the product $AB$ and $BA$ are defined.

**Definition 1.2.9** Two square matrices $A$ and $B$ are said to commute if $AB = BA$.

**Remark 1.2.10**

1. *Note that if $A$ is a square matrix of order $n$ then $AI_n = I_nA$. Also for any $d \in \mathbb{R}$, the matrix $dI_n$ commutes with every square matrix of order $n$. The matrices $dI_n$ for any $d \in \mathbb{R}$ are called* SCALAR *matrices.*

2. *In general, the matrix product is not commutative. For example, consider the following two matrices $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. Then check that the matrix product*

$$AB = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = BA.$$

**Theorem 1.2.11** Suppose that the matrices $A$, $B$ and $C$ are so chosen that the matrix multiplications are defined.

1. Then $(AB)C = A(BC)$. That is, the matrix multiplication is associative.

2. For any $k \in \mathbb{R}$, $(kA)B = k(AB) = A(kB)$.

3. Then $A(B + C) = AB + AC$. That is, multiplication distributes over addition.

4. If $A$ is an $n \times n$ matrix then $AI_n = I_nA = A$.

5. For any square matrix $A$ of order $n$ and $D = \text{diag}(d_1, d_2, \ldots, d_n)$, we have
    - the first row of $DA$ is $d_1$ times the first row of $A$;
    - for $1 \leq i \leq n$, the $i^{\text{th}}$ row of $DA$ is $d_i$ times the $i^{\text{th}}$ row of $A$.

   A similar statement holds for the columns of $A$ when $A$ is multiplied on the right by $D$.

PROOF. Part 1. Let $A = [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{n\times p}$ and $C = [c_{ij}]_{p\times q}$. Then

$$(BC)_{kj} = \sum_{\ell=1}^{p} b_{k\ell} c_{\ell j} \quad \text{and} \quad (AB)_{i\ell} = \sum_{k=1}^{n} a_{ik} b_{k\ell}.$$

Therefore,

$$\begin{aligned}
\big(A(BC)\big)_{ij} &= \sum_{k=1}^{n} a_{ik}\big(BC\big)_{kj} = \sum_{k=1}^{n} a_{ik}\big(\sum_{\ell=1}^{p} b_{k\ell}c_{\ell j}\big) \\
&= \sum_{k=1}^{n}\sum_{\ell=1}^{p} a_{ik}\big(b_{k\ell}c_{\ell j}\big) = \sum_{k=1}^{n}\sum_{\ell=1}^{p}\big(a_{ik}b_{k\ell}\big)c_{\ell j} \\
&= \sum_{\ell=1}^{p}\big(\sum_{k=1}^{n} a_{ik}b_{k\ell}\big)c_{\ell j} = \sum_{\ell=1}^{t}\big(AB\big)_{i\ell}c_{\ell j} \\
&= \big((AB)C\big)_{ij}.
\end{aligned}$$

Part 5. For all $j = 1, 2, \ldots, n$, we have

$$(DA)_{ij} = \sum_{k=1}^{n} d_{ik}a_{kj} = d_i a_{ij}$$

as $d_{ik} = 0$ whenever $i \neq k$. Hence, the required result follows.

The reader is required to prove the other parts. □

**Exercise 1.2.12** 1. Let $A$ and $B$ be two matrices. If the matrix addition $A + B$ is defined, then prove that $(A + B)^t = A^t + B^t$. Also, if the matrix product $AB$ is defined then prove that $(AB)^t = B^t A^t$.

2. Let $A = [a_1, a_2, \ldots, a_n]$ and $B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$. Compute the matrix products $AB$ and $BA$.

3. Let $n$ be a positive integer. Compute $A^n$ for the following matrices:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Can you guess a formula for $A^n$ and prove it by induction?

4. Find examples for the following statements.

(a) Suppose that the matrix product $AB$ is defined. Then the product $BA$ need not be defined.

(b) Suppose that the matrix products $AB$ and $BA$ are defined. Then the matrices $AB$ and $BA$ can have different orders.

(c) Suppose that the matrices $A$ and $B$ are square matrices of order $n$. Then $AB$ and $BA$ may or may not be equal.

## 1.3 Some More Special Matrices

**Definition 1.3.1** 1. A matrix $A$ over $\mathbb{R}$ is called symmetric if $A^t = A$ and skew-symmetric if $A^t = -A$.

2. A matrix $A$ is said to be orthogonal if $AA^t = A^tA = I$.

**Example 1.3.2** 1. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -1 \\ 3 & -1 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & -3 \\ -2 & 3 & 0 \end{bmatrix}$. Then $A$ is a symmetric matrix and $B$ is a skew-symmetric matrix.

2. Let $A = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix}$. Then $A$ is an orthogonal matrix.

3. Let $A = [a_{ij}]$ be an $n \times n$ matrix with $a_{ij} = \begin{cases} 1 & \text{if } i = j+1 \\ 0 & \text{otherwise} \end{cases}$. Then $A^n = \mathbf{0}$ and $A^\ell \neq \mathbf{0}$ for $1 \leq \ell \leq n-1$. The matrices $A$ for which a positive integer $k$ exists such that $A^k = \mathbf{0}$ are called NILPOTENT matrices. The least positive integer $k$ for which $A^k = \mathbf{0}$ is called the ORDER OF NILPOTENCY.

4. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Then $A^2 = A$. The matrices that satisfy the condition that $A^2 = A$ are called IDEMPOTENT matrices.

**Exercise 1.3.3** 1. Show that for any square matrix $A$, $S = \frac{1}{2}(A + A^t)$ is symmetric, $T = \frac{1}{2}(A - A^t)$ is skew-symmetric, and $A = S + T$.

2. Show that the product of two lower triangular matrices is a lower triangular matrix. A similar statement holds for upper triangular matrices.

3. Let $A$ and $B$ be symmetric matrices. Show that $AB$ is symmetric if and only if $AB = BA$.

4. Show that the diagonal entries of a skew-symmetric matrix are zero.

5. Let $A, B$ be skew-symmetric matrices with $AB = BA$. Is the matrix $AB$ symmetric or skew-symmetric?

6. Let $A$ be a symmetric matrix of order $n$ with $A^2 = \mathbf{0}$. Is it necessarily true that $A = \mathbf{0}$?

7. Let $A$ be a nilpotent matrix. Show that there exists a matrix $B$ such that $B(I + A) = I = (I + A)B$.

### 1.3.1 Submatrix of a Matrix

**Definition 1.3.4** A matrix obtained by deleting some of the rows and/or columns of a matrix is said to be a submatrix of the given matrix.

For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$, a few submatrices of $A$ are

$$[1], [2], \begin{bmatrix} 1 \\ 0 \end{bmatrix}, [1 \; 5], \begin{bmatrix} 1 & 5 \\ 0 & 2 \end{bmatrix}, \; A.$$

But the matrices $\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 4 \\ 0 & 2 \end{bmatrix}$ are not submatrices of $A$. (The reader is advised to give reasons.)

## Miscellaneous Exercises

**Exercise 1.3.5** 1. Complete the proofs of Theorems 1.2.5 and 1.2.11.

2. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Geometrically interpret $\mathbf{y} = A\mathbf{x}$ and $\mathbf{y} = B\mathbf{x}$.

3. Consider the two coordinate transformations

$$\begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2 \\ x_2 &= a_{21}y_1 + a_{22}y_2 \end{aligned} \quad \text{and} \quad \begin{aligned} y_1 &= b_{11}z_1 + b_{12}z_2 \\ y_2 &= b_{21}z_1 + b_{22}z_2 \end{aligned}.$$

(a) Compose the two transformations to express $x_1, x_2$ in terms of $z_1, z_2$.

(b) If $\mathbf{x}^t = [x_1,\ x_2]$, $\mathbf{y}^t = [y_1,\ y_2]$ and $\mathbf{z}^t = [z_1,\ z_2]$ then find matrices $A, B$ and $C$ such that $\mathbf{x} = A\mathbf{y}$, $\mathbf{y} = B\mathbf{z}$ and $\mathbf{x} = C\mathbf{z}$.

(c) Is $C = AB$?

4. For a square matrix $A$ of order $n$, we define **trace** of $A$, denoted by tr $(A)$ as

$$\text{tr } (A) = a_{11} + a_{22} + \cdots a_{nn}.$$

Then for two square matrices, $A$ and $B$ of the same order, show the following:

(a) tr $(A + B) =$ tr $(A) +$ tr $(B)$.

(b) tr $(AB) =$ tr $(BA)$.

5. Show that, there do not exist matrices $A$ and $B$ such that $AB - BA = cI_n$ for any $c \neq 0$.

6. Let $A$ and $B$ be two $m \times n$ matrices and let $\mathbf{x}$ be an $n \times 1$ column vector.

(a) Prove that if $A\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$, then $A$ is the zero matrix.

(b) Prove that if $A\mathbf{x} = B\mathbf{x}$ for all $\mathbf{x}$, then $A = B$.

7. Let $A$ be an $n \times n$ matrix such that $AB = BA$ for all $n \times n$ matrices $B$. Show that $A = \alpha I$ for some $\alpha \in \mathbb{R}$.

8. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$. Show that there exist infinitely many matrices $B$ such that $BA = I_2$. Also, show that there does not exist any matrix $C$ such that $AC = I_3$.

### 1.3.1 Block Matrices

Let $A$ be an $n \times m$ matrix and $B$ be an $m \times p$ matrix. Suppose $r < m$. Then, we can decompose the matrices $A$ and $B$ as $A = [P\ Q]$ and $B = \begin{bmatrix} H \\ K \end{bmatrix}$; where $P$ has order $n \times r$ and $H$ has order $r \times p$. That is, the matrices $P$ and $Q$ are submatrices of $A$ and $P$ consists of the first $r$ columns of $A$ and $Q$ consists of the last $m - r$ columns of $A$. Similarly, $H$ and $K$ are submatrices of $B$ and $H$ consists of the first $r$ rows of $B$ and $K$ consists of the last $m - r$ rows of $B$. We now prove the following important theorem.

**Theorem 1.3.6** Let $A = [a_{ij}] = [P\ Q]$ and $B = [b_{ij}] = \begin{bmatrix} H \\ K \end{bmatrix}$ be defined as above. Then

$$AB = PH + QK.$$

Proof. First note that the matrices $PH$ and $QK$ are each of order $n \times p$. The matrix products $PH$ and $QK$ are valid as the order of the matrices $P, H, Q$ and $K$ are respectively, $n \times r,\ r \times p,\ n \times (m - r)$ and $(m-r)\times p$. Let $P = [P_{ij}],\ Q = [Q_{ij}],\ H = [H_{ij}]$, and $K = [k_{ij}]$. Then, for $1 \leq i \leq n$ and $1 \leq j \leq p$, we have

$$\begin{aligned} (AB)_{ij} &= \sum_{k=1}^{m} a_{ik}b_{kj} = \sum_{k=1}^{r} a_{ik}b_{kj} + \sum_{k=r+1}^{m} a_{ik}b_{kj} \\ &= \sum_{k=1}^{r} P_{ik}H_{kj} + \sum_{k=r+1}^{m} Q_{ik}K_{kj} \\ &= (PH)_{ij} + (QK)_{ij} = (PH + QK)_{ij}. \end{aligned}$$

□

Theorem 1.3.6 is very useful due to the following reasons:

1. The order of the matrices $P, Q, H$ and $K$ are smaller than that of $A$ or $B$.
2. It may be possible to block the matrix in such a way that a few blocks are either identity matrices or zero matrices. In this case, it may be easy to handle the matrix product using the block form.
3. Or when we want to prove results using induction, then we may assume the result for $r \times r$ submatrices and then look for $(r+1) \times (r+1)$ submatrices, etc.

For example, if $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$, Then

$$AB = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} [e \; f] = \begin{bmatrix} a+2c & b+2d \\ 2a+5c & 2b+5d \end{bmatrix}.$$

If $A = \begin{bmatrix} 0 & -1 & 2 \\ 3 & 1 & 4 \\ -2 & 5 & -3 \end{bmatrix}$, then $A$ can be decomposed as follows:

$$A = \left[\begin{array}{c|cc} 0 & -1 & 2 \\ 3 & 1 & 4 \\ \hline -2 & 5 & -3 \end{array}\right], \text{ or } \quad A = \left[\begin{array}{cc|c} 0 & -1 & 2 \\ 3 & 1 & 4 \\ \hline -2 & 5 & -3 \end{array}\right], \text{ or}$$

$$A = \left[\begin{array}{cc|c} 0 & -1 & 2 \\ \hline 3 & 1 & 4 \\ -2 & 5 & -3 \end{array}\right] \text{ and so on.}$$

Suppose $A = \begin{array}{cc} & \begin{array}{cc} m_1 & m_2 \end{array} \\ \begin{array}{c} n_1 \\ n_2 \end{array} & \begin{bmatrix} P & Q \\ R & S \end{bmatrix} \end{array}$ and $B = \begin{array}{cc} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} r_1 \\ r_2 \end{array} & \begin{bmatrix} E & F \\ G & H \end{bmatrix} \end{array}$. Then the matrices $P$, $Q$, $R$, $S$ and $E$, $F$, $G$, $H$, are called the blocks of the matrices $A$ and $B$, respectively.

Even if $A+B$ is defined, the orders of $P$ and $E$ **may not be same** and hence, we may not be able to add $A$ and $B$ in the block form. But, if $A+B$ and $P+E$ is defined then $A+B = \begin{bmatrix} P+E & Q+F \\ R+G & S+H \end{bmatrix}$.

Similarly, if the product $AB$ is defined, the product $PE$ need not be defined. Therefore, we can talk of matrix product $AB$ as block product of matrices, if both the products $AB$ and $PE$ are defined. And in this case, we have $AB = \begin{bmatrix} PE+QG & PF+QH \\ RE+SG & RF+SH \end{bmatrix}$.

That is, **once a partition of $A$ is fixed, the partition of $B$ has to be properly chosen for purposes of block addition or multiplication.**

**Exercise 1.3.7** 1. Compute the matrix product $AB$ using the block matrix multiplication for the matrices

$$A = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ \hline 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \text{ and } B = \left[\begin{array}{cc|cc} 1 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 \\ \hline 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{array}\right].$$

2. Let $A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$. If $P, Q, R$ and $S$ are symmetric, what can you say about $A$? Are $P, Q, R$ and $S$ symmetric, when $A$ is symmetric?

3. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices. Suppose $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are the rows of $A$ and $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$ are the columns of $B$. If the product $AB$ is defined, then show that

$$AB = [A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_p] = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_n B \end{bmatrix}.$$

[That is, left multiplication by $A$, is same as multiplying each column of $B$ by $A$. Similarly, right multiplication by $B$, is same as multiplying each row of $A$ by $B$.]

## 1.4 Matrices over Complex Numbers

Here the entries of the matrix are complex numbers. All the definitions still hold. One just needs to look at the following additional definitions.

**Definition 1.4.1 (Conjugate Transpose of a Matrix)** 1. Let $A$ be an $m \times n$ matrix over $\mathbb{C}$. If $A = [a_{ij}]$ then the Conjugate of $A$, denoted by $\overline{A}$, is the matrix $B = [b_{ij}]$ with $b_{ij} = \overline{a_{ij}}$.

For example, Let $A = \begin{bmatrix} 1 & 4+3i & i \\ 0 & 1 & i-2 \end{bmatrix}$. Then

$$\overline{A} = \begin{bmatrix} 1 & 4-3i & -i \\ 0 & 1 & -i-2 \end{bmatrix}.$$

2. Let $A$ be an $m \times n$ matrix over $\mathbb{C}$. If $A = [a_{ij}]$ then the Conjugate Transpose of $A$, denoted by $A^*$, is the matrix $B = [b_{ij}]$ with $b_{ij} = \overline{a_{ji}}$.

For example, Let $A = \begin{bmatrix} 1 & 4+3i & i \\ 0 & 1 & i-2 \end{bmatrix}$. Then

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

3. A square matrix $A$ over $\mathbb{C}$ is called Hermitian if $A^* = A$.

4. A square matrix $A$ over $\mathbb{C}$ is called skew-Hermitian if $A^* = -A$.

5. A square matrix $A$ over $\mathbb{C}$ is called unitary if $A^*A = AA^* = I$.

6. A square matrix $A$ over $\mathbb{C}$ is called Normal if $AA^* = A^*A$.

**Remark 1.4.2** *If $A = [a_{ij}]$ with $a_{ij} \in \mathbb{R}$, then $A^* = A^t$.*

**Exercise 1.4.3** 1. Give examples of Hermitian, skew-Hermitian and unitary matrices that have entries with non-zero imaginary parts.

2. Restate the results on **transpose** in terms of **conjugate transpose**.

3. Show that for any square matrix $A$, $S = \frac{A+A^*}{2}$ is Hermitian, $T = \frac{A-A^*}{2}$ is skew-Hermitian, and $A = S + T$.

4. Show that if $A$ is a complex triangular matrix and $AA^* = A^*A$ then $A$ is a diagonal matrix.

# Chapter 2

# Linear System of Equations

## 2.1 Introduction

Let us look at some examples of linear systems.

1. Suppose $a, b \in \mathbb{R}$. Consider the system $ax = b$.

   (a) If $a \neq 0$ then the system has a UNIQUE SOLUTION $x = \frac{b}{a}$.

   (b) If $a = 0$ and

      i. $b \neq 0$ then the system has NO SOLUTION.

      ii. $b = 0$ then the system has INFINITE NUMBER OF SOLUTIONS, namely all $x \in \mathbb{R}$.

2. We now consider a system with 2 equations in 2 unknowns.
   Consider the equation $ax + by = c$. If one of the coefficients, $a$ or $b$ is non-zero, then this linear equation represents a line in $\mathbb{R}^2$. Thus for the system

   $$a_1x + b_1y = c_1 \text{ and } a_2x + b_2y = c_2,$$

   the set of solutions is given by the points of intersection of the two lines. There are three cases to be considered. Each case is illustrated by an example.

   (a) UNIQUE SOLUTION
   $x + 2y = 1$ and $x + 3y = 1$. The unique solution is $(x, y)^t = (1, 0)^t$.
   Observe that in this case, $a_1b_2 - a_2b_1 \neq 0$.

   (b) INFINITE NUMBER OF SOLUTIONS
   $x + 2y = 1$ and $2x + 4y = 2$. The set of solutions is $(x, y)^t = (1 - 2y, y)^t = (1, 0)^t + y(-2, 1)^t$ with $y$ arbitrary. In other words, both the equations represent the same line.
   Observe that in this case, $a_1b_2 - a_2b_1 = 0$, $a_1c_2 - a_2c_1 = 0$ and $b_1c_2 - b_2c_1 = 0$.

   (c) NO SOLUTION
   $x + 2y = 1$ and $2x + 4y = 3$. The equations represent a pair of parallel lines and hence there is no point of intersection.
   Observe that in this case, $a_1b_2 - a_2b_1 = 0$ but $a_1c_2 - a_2c_1 \neq 0$.

3. As a last example, consider 3 equations in 3 unknowns.
   A linear equation $ax + by + cz = d$ represent a plane in $\mathbb{R}^3$ provided $(a, b, c) \neq (0, 0, 0)$. As in the case of 2 equations in 2 unknowns, we have to look at the points of intersection of the given three planes. Here again, we have three cases. The three cases are illustrated by examples.

(a) UNIQUE SOLUTION
Consider the system $x+y+z=3,\;\; x+4y+2z=7$ and $4x+10y-z=13$. The unique solution to this system is $(x,y,z)^t=(1,1,1)^t$; *i.e.* THE THREE PLANES INTERSECT AT A POINT.

(b) INFINITE NUMBER OF SOLUTIONS
Consider the system $x+y+z=3,\;\; x+2y+2z=5$ and $3x+4y+4z=11$. The set of solutions to this system is $(x,y,z)^t=(1,2-z,z)^t=(1,2,0)^t+z(0,-1,1)^t$, with $z$ arbitrary: THE THREE PLANES INTERSECT ON A LINE.

(c) NO SOLUTION
The system $x+y+z=3,\;\; x+2y+2z=5$ and $3x+4y+4z=13$ has no solution. In this case, we get three parallel lines as intersections of the above planes taken two at a time.

The readers are advised to supply the proof.

## 2.2 Definition and a Solution Method

**Definition 2.2.1 (Linear System)** A linear system of $m$ equations in $n$ unknowns $x_1, x_2, \ldots, x_n$ is a set of equations of the form

$$\begin{array}{rcl} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n & = & b_1 \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n & = & b_2 \\ \vdots & & \vdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n & = & b_m \end{array} \tag{2.2.1}$$

where for $1\le i\le n$, and $1\le j\le m$; $a_{ij}, b_i\in\mathbb{R}$. Linear System (2.2.1) is called HOMOGENEOUS if $b_1=0=b_2=\cdots=b_m$ and NON-HOMOGENEOUS otherwise.

We rewrite the above equations in the form $A\mathbf{x}=\mathbf{b}$, where

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix},\quad \mathbf{x}=\begin{bmatrix} x_1\\ x_2\\ \vdots \\ x_n\end{bmatrix},\text{ and } \mathbf{b}=\begin{bmatrix} b_1\\ b_2\\ \vdots\\ b_m\end{bmatrix}$$

The matrix $A$ is called the COEFFICIENT matrix and the block matrix $[A\;\; \mathbf{b}]$, is the AUGMENTED matrix of the linear system (2.2.1).

**Remark 2.2.2** *Observe that the $i^{th}$ row of the augmented matrix $[A\;\; \mathbf{b}]$ represents the $i^{th}$ equation and the $j^{th}$ column of the coefficient matrix $A$ corresponds to coefficients of the $j^{th}$ variable $x_j$. That is, for $1\le i\le m$ and $1\le j\le n$, the entry $a_{ij}$ of the coefficient matrix $A$ corresponds to the $i^{th}$ equation and $j^{th}$ variable $x_j$..*

For a system of linear equations $A\mathbf{x}=\mathbf{b}$, the system $A\mathbf{x}=\mathbf{0}$ is called the ASSOCIATED HOMOGENEOUS SYSTEM.

**Definition 2.2.3 (Solution of a Linear System)** A solution of the linear system $A\mathbf{x}=\mathbf{b}$ is a column vector $\mathbf{y}$ with entries $y_1, y_2, \ldots, y_n$ such that the linear system (2.2.1) is satisfied by substituting $y_i$ in place of $x_i$.

That is, if $\mathbf{y}^t=[y_1,y_2,\ldots,y_n]$ then $A\mathbf{y}=\mathbf{b}$ holds.

**Note:** The zero $n$-tuple $\mathbf{x}=\mathbf{0}$ is always a solution of the system $A\mathbf{x}=\mathbf{0}$, and is called the TRIVIAL solution. A non-zero $n$-tuple $\mathbf{x}$, if it satisfies $A\mathbf{x}=\mathbf{0}$, is called a NON-TRIVIAL solution.

### 2.2.1 A Solution Method

**Example 2.2.4** Let us solve the linear system $x+7y+3z=11,\ x+y+z=3$, and $4x+10y-z=13$.
**Solution:**

1. The above linear system and the linear system

$$\begin{aligned} x+y+z &= 3 \quad \text{Interchange the first two equations.} \\ x+7y+3z &= 11 \\ 4x+10y-z &= 13 \end{aligned} \tag{2.2.2}$$

   have the same set of solutions. (why?)

2. Eliminating $x$ from $2^{\text{nd}}$ and $3^{\text{rd}}$ equation, we get the linear system

$$\begin{aligned} x+y+z &= 3 \\ 6y+2z &= 8 \quad \text{(obtained by subtracting the first} \\ & \qquad \text{equation from the second equation.)} \\ 6y-5z &= 1 \quad \text{(obtained by subtracting } 4 \text{ times the first equation} \\ & \qquad \text{from the third equation.)} \end{aligned} \tag{2.2.3}$$

   This system and the system (2.2.2) has the same set of solution. (why?)

3. Eliminating $y$ from the last two equations of system (2.2.3), we get the system

$$\begin{aligned} x+y+z &= 3 \\ 6y+2z &= 8 \\ 7z &= 7 \quad \text{obtained by subtracting the third equation} \\ & \qquad \text{from the second equation.} \end{aligned} \tag{2.2.4}$$

   which has the same set of solution as the system (2.2.3). (why?)

4. The system (2.2.4) and system

$$\begin{aligned} x+y+z &= 3 \\ 3y+z &= 4 \quad \text{divide the second equation by } 2 \\ z &= 1 \quad \text{divide the second equation by } 2 \end{aligned} \tag{2.2.5}$$

   has the same set of solution. (why?)

5. Now, $z=1$ implies $y=\dfrac{4-1}{3}=1$ and $x=3-(1+1)=1$. Or in terms of a vector, the set of solution is $\{\ (x,y,z)^t\ :(x,y,z)=(1,1,1)\}$.

## 2.3 Row Operations and Equivalent Systems

**Definition 2.3.1 (Elementary Operations)** The following operations 1, 2 and 3 are called elementary operations.

1. interchange of two equations, say "interchange the $i^{\text{th}}$ and $j^{\text{th}}$ equations";

   (compare the system (2.2.2) with the original system.)

2. multiply a non-zero constant throughout an equation, say "multiply the $k^{\text{th}}$ equation by $c \neq 0$";

   (compare the system (2.2.5) and the system (2.2.4).)

3. replace an equation by itself plus a constant multiple of another equation, say "replace the $k^{\text{th}}$ equation by $k^{\text{th}}$ equation plus $c$ times the $j^{\text{th}}$ equation".

   (compare the system (2.2.3) with (2.2.2) or the system (2.2.4) with (2.2.3).)

**Observations:**

1. In the above example, observe that the elementary operations helped us in getting a linear system (2.2.5), which was easily solvable.

2. Note that at Step 1, if we interchange the first and the second equation, we get back to the linear system from which we had started. This means the operation at Step 1, has an inverse operation. In other words, INVERSE OPERATION sends us back to the step where we had precisely started. It will be a useful exercise for the reader to IDENTIFY THE INVERSE OPERATIONS at each step in Example 2.2.4.

So, in Example 2.2.4, the application of a finite number of elementary operations helped us to obtain a simpler system whose solution can be obtained directly. That is, after applying a finite number of elementary operations, a simpler linear system is obtained which can be easily solved. Note that the three elementary operations defined above, have corresponding INVERSE operations, namely,

1. "interchange the $i^{\text{th}}$ and $j^{\text{th}}$ equations",

2. "divide the $k^{\text{th}}$ equation by $c \neq 0$";

3. "replace the $k^{\text{th}}$ equation by $k^{\text{th}}$ equation minus $c$ times the $j^{\text{th}}$ equation".

It will be a useful exercise for the reader to IDENTIFY THE INVERSE OPERATIONS at each step in Example 2.2.4.

**Definition 2.3.2** **(Equivalent Linear Systems)** Two linear systems are said to be equivalent if one can be obtained from the other by a finite number of elementary operations.

The linear systems at each step in Example 2.2.4 are equivalent to each other and also to the original linear system.

**Lemma 2.3.3** Let $C\mathbf{x} = \mathbf{d}$ be the linear system obtained from the linear system $A\mathbf{x} = \mathbf{b}$ by a single elementary operation. Then the linear systems $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ have the same set of solutions.

PROOF. We prove the result for the elementary operation "the $k^{\text{th}}$ equation is replaced by $k^{\text{th}}$ equation plus $c$ times the $j^{\text{th}}$ equation." The reader is advised to prove the result for other elementary operations.

In this case, the systems $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ vary only in the $k^{\text{th}}$ equation. Let $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ be a solution of the linear system $A\mathbf{x} = b$. Then substituting for $\alpha_i$'s in place of $x_i$'s in the $k^{\text{th}}$ and $j^{\text{th}}$ equations, we get

$$a_{k1}\alpha_1 + a_{k2}\alpha_2 + \cdots a_{kn}\alpha_n = b_k, \text{ and } a_{j1}\alpha_1 + a_{j2}\alpha_2 + \cdots a_{jn}\alpha_n = b_j.$$

Therefore,

$$(a_{k1} + ca_{j1})\alpha_1 + (a_{k2} + ca_{j2})\alpha_2 + \cdots + (a_{kn} + ca_{jn})\alpha_n = b_k + cb_j. \tag{2.3.1}$$

But then the $k^{\text{th}}$ equation of the linear system $C\mathbf{x} = \mathbf{d}$ is

$$(a_{k1} + ca_{j1})x_1 + (a_{k2} + ca_{j2})x_2 + \cdots + (a_{kn} + ca_{jn})x_n = b_k + cb_j. \tag{2.3.2}$$

Therefore, using Equation (2.3.1), $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is also a solution for the $k^{\text{th}}$ Equation (2.3.2).

Use a similar argument to show that if $(\beta_1, \beta_2, \ldots, \beta_n)$ is a solution of the linear system $C\mathbf{x} = \mathbf{d}$ then it is also a solution of the linear system $A\mathbf{x} = \mathbf{b}$.

Hence, we have the proof in this case. □

Lemma 2.3.3 is now used as an induction step to prove the main result of this section (Theorem 2.3.4).

**Theorem 2.3.4** Two equivalent systems have the same set of solutions.

PROOF. Let $n$ be the number of elementary operations performed on $A\mathbf{x} = \mathbf{b}$ to get $C\mathbf{x} = \mathbf{d}$. We prove the theorem by induction on $n$.

If $n = 1$, Lemma 2.3.3 answers the question. If $n > 1$, assume that the theorem is true for $n = m$. Now, suppose $n = m + 1$. Apply the Lemma 2.3.3 again at the "last step" (that is, at the $(m+1)^{\text{th}}$ step from the $m^{\text{th}}$ step) to get the required result using induction. □

Let us formalise the above section which led to Theorem 2.3.4. For solving a linear system of equations, we applied elementary operations to equations. It is observed that in performing the elementary operations, the calculations were made on the COEFFICIENTS (numbers). The variables $x_1, x_2, \ldots, x_n$ and the sign of equality (that is, " $=$ ") are not disturbed. Therefore, in place of looking at the system of equations as a whole, we just need to work with the coefficients. These coefficients when arranged in a rectangular array gives us the augmented matrix $[A \;\; \mathbf{b}]$.

**Definition 2.3.5 (Elementary Row Operations)** The elementary row operations are defined as:

1. interchange of two rows, say "interchange the $i^{\text{th}}$ and $j^{\text{th}}$ rows", denoted $R_{ij}$;
2. multiply a non-zero constant throughout a row, say "multiply the $k^{\text{th}}$ row by $c \neq 0$", denoted $R_k(c)$;
3. replace a row by itself plus a constant multiple of another row, say "replace the $k^{\text{th}}$ row by $k^{\text{th}}$ row plus $c$ times the $j^{\text{th}}$ row", denoted $R_{kj}(c)$.

**Exercise 2.3.6** Find the INVERSE row operations corresponding to the elementary row operations that have been defined just above.

**Definition 2.3.7 (Row Equivalent Matrices)** Two matrices are said to be row-equivalent if one can be obtained from the other by a finite number of elementary row operations.

**Example 2.3.8** The three matrices given below are row equivalent.

$$\begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix} \overrightarrow{R_{12}} \begin{bmatrix} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix} \overrightarrow{R_1(1/2)} \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

Whereas the matrix $\begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$ is not row equivalent to the matrix $\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 2 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$.

### 2.3.1 Gauss Elimination Method

**Definition 2.3.9 (Forward/Gauss Elimination Method)** Gaussian elimination is a method of solving a linear system $A\mathbf{x} = \mathbf{b}$ (consisting of $m$ equations in $n$ unknowns) by bringing the augmented matrix

$$[A\ \ \mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

to an upper triangular form

$$\left[\begin{array}{cccc|c} c_{11} & c_{12} & \cdots & c_{1n} & d_1 \\ 0 & c_{22} & \cdots & c_{2n} & d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{mn} & d_m \end{array}\right].$$

This elimination process is also called the forward elimination method.

The following examples illustrate the Gauss elimination procedure.

**Example 2.3.10** Solve the linear system by Gauss elimination method.

$$\begin{aligned} y + z &= 2 \\ 2x + 3z &= 5 \\ x + y + z &= 3 \end{aligned}$$

**Solution:** In this case, the augmented matrix is $\begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$. The method proceeds along the following steps.

1. Interchange $1^{\text{st}}$ and $2^{\text{nd}}$ equation (or $R_{12}$).

$$\begin{aligned} 2x + 3z &= 5 \\ y + z &= 2 \\ x + y + z &= 3 \end{aligned} \qquad \begin{bmatrix} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

2. Divide the $1^{\text{st}}$ equation by 2 (or $R_1(1/2)$).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ x + y + z &= 3 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

3. Add $-1$ times the $1^{\text{st}}$ equation to the $3^{\text{rd}}$ equation (or $R_{31}(-1)$).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ y - \tfrac{1}{2}z &= \tfrac{1}{2} \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

4. Add $-1$ times the $2^{\text{nd}}$ equation to the $3^{\text{rd}}$ equation (or $R_{32}(-1)$).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ -\tfrac{3}{2}z &= -\tfrac{3}{2} \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -\frac{3}{2} & -\frac{3}{2} \end{bmatrix}.$$

5. Multiply the $3^{\text{rd}}$ equation by $\frac{-2}{3}$ (or $R_3(-\frac{2}{3})$).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ z &= 1 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

The last equation gives $z = 1$, the second equation now gives $y = 1$. Finally the first equation gives $x = 1$. Hence the set of solutions is $(x, y, z)^t = (1, 1, 1)^t$, A UNIQUE SOLUTION.

**Example 2.3.11** Solve the linear system by Gauss elimination method.

$$\begin{aligned} x + y + z &= 3 \\ x + 2y + 2z &= 5 \\ 3x + 4y + 4z &= 11 \end{aligned}$$

**Solution:** In this case, the augmented matrix is $\begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 2 & 2 & 5 \\ 3 & 4 & 4 & 11 \end{bmatrix}$ and the method proceeds as follows:

1. Add $-1$ times the first equation to the second equation.

$$\begin{aligned} x + y + z &= 3 \\ y + z &= 2 \\ 3x + 4y + 4z &= 11 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 3 & 4 & 4 & 11 \end{bmatrix}.$$

2. Add $-3$ times the first equation to the third equation.

$$\begin{aligned} x + y + z &= 3 \\ y + z &= 2 \\ y + z &= 2 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{bmatrix}.$$

3. Add $-1$ times the second equation to the third equation

$$\begin{aligned} x + y + z &= 3 \\ y + z &= 2 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, the set of solutions is $(x, y, z)^t = (1, 2 - z, z)^t = (1, 2, 0)^t + z(0, -1, 1)^t$, with $z$ arbitrary. In other words, the system has INFINITE NUMBER OF SOLUTIONS.

**Example 2.3.12** Solve the linear system by Gauss elimination method.

$$\begin{aligned} x + y + z &= 3 \\ x + 2y + 2z &= 5 \\ 3x + 4y + 4z &= 12 \end{aligned}$$

**Solution:** In this case, the augmented matrix is $\begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 2 & 2 & 5 \\ 3 & 4 & 4 & 12 \end{bmatrix}$ and the method proceeds as follows:

1. Add $-1$ times the first equation to the second equation.

$$\begin{aligned} x + y + z &= 3 \\ y + z &= 2 \\ 3x + 4y + 4z &= 12 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 3 & 4 & 4 & 12 \end{bmatrix}.$$

2. Add $-3$ times the first equation to the third equation.

$$\begin{aligned} x+y+z &= 3 \\ y+z &= 2 \\ y+z &= 3 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 3 \end{bmatrix}.$$

3. Add $-1$ times the second equation to the third equation

$$\begin{aligned} x+y+z &= 3 \\ y+z &= 2 \\ 0 &= 1 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The third equation in the last step is

$$0x + 0y + 0z = 1.$$

This can never hold for any value of $x, y, z$. Hence, the system has NO SOLUTION.

**Remark 2.3.13** *Note that to solve a linear system, $A\mathbf{x} = \mathbf{b}$, one needs to apply only the elementary row operations to the augmented matrix $[A \;\; \mathbf{b}]$.*

## 2.4 Row Reduced Echelon Form of a Matrix

**Definition 2.4.1 (Row Reduced Form of a Matrix)** A matrix $C$ is said to be in the row reduced form if

1. THE FIRST NON-ZERO ENTRY IN EACH ROW OF $C$ IS 1;

2. THE COLUMN CONTAINING THIS 1 HAS ALL ITS OTHER ENTRIES ZERO.

A matrix in the row reduced form is also called a ROW REDUCED MATRIX.

**Example 2.4.2** 1. One of the most important examples of a row reduced matrix is the $n \times n$ identity matrix, $I_n$. Recall that the $(i,j)^{\text{th}}$ entry of the identity matrix is

$$I_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}.$$

$\delta_{ij}$ is usually referred to as the Kronecker delta function.

2. The matrices $\begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ are also in row reduced form.

3. The matrix $\begin{bmatrix} 1 & 0 & 0 & 0 & 5 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ is not in the row reduced form. (why?)

**Definition 2.4.3 (Leading Term, Leading Column)** For a row-reduced matrix, the first non-zero entry of any row is called a LEADING TERM. The columns containing the leading terms are called the LEADING COLUMNS.

**Definition 2.4.4 (Basic, Free Variables)** Consider the linear system $A\mathbf{x} = \mathbf{b}$ in $n$ variables and $m$ equations. Let $[C \;\; \mathbf{d}]$ be the row-reduced matrix obtained by applying the Gauss elimination method to the augmented matrix $[A \;\; \mathbf{b}]$. Then the variables corresponding to the leading columns in the first $n$ columns of $[C \;\; \mathbf{d}]$ are called the BASIC variables. The variables which are not basic are called FREE variables.

The free variables are called so as they can be assigned arbitrary values and the value of the basic variables can then be written in terms of the free variables.

**Observation:** In Example 2.3.11, the solution set was given by

$$(x, y, z)^t = (1, 2 - z, z)^t = (1, 2, 0)^t + z(0, -1, 1)^t, \quad \text{with } z \; \text{ arbitrary.}$$

That is, we had two basic variables, $x$ and $y$, and $z$ as a free variable.

**Remark 2.4.5** *It is very important to observe that if there are $r$ non-zero rows in the row-reduced form of the matrix then there will be $r$ leading terms. That is, there will be $r$ leading columns. Therefore,* IF THERE ARE $r$ LEADING TERMS AND $n$ VARIABLES, THEN THERE WILL BE $r$ BASIC VARIABLES AND $n - r$ FREE VARIABLES.

### 2.4.1 Gauss-Jordan Elimination

We now start with Step 5 of Example 2.3.10 and apply the elementary operations once again. But this time, we start with the 3rd row.

I. Add $-1$ times the third equation to the second equation (or $R_{23}(-1)$).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y \quad &= 2 \\ z &= 1 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

II. Add $\frac{-3}{2}$ times the third equation to the first equation (or $R_{13}(-\frac{3}{2})$).

$$\begin{aligned} x &= 1 \\ y &= 1 \\ z &= 1 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

III. From the above matrix, we directly have the set of solution as $(x, y, z)^t = (1, 1, 1)^t$.

**Definition 2.4.6 (Row Reduced Echelon Form of a Matrix)** A matrix $C$ is said to be in the row reduced echelon form if

1. $C$ is already in the row reduced form;
2. The rows consisting of all zeros comes below all non-zero rows; and
3. the leading terms appear from left to right in successive rows. That is, for $1 \le \ell \le k$, let $i_\ell$ be the leading column of the $\ell^{\text{th}}$ row. Then $i_1 < i_2 < \cdots < i_k$.

**Example 2.4.7** Suppose $A = \begin{bmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ are in row reduced form. Then the corresponding matrices in the row reduced echelon form are respectively, $\begin{bmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

**Definition 2.4.8 (Row Reduced Echelon Matrix)** A matrix which is in the row reduced echelon form is also called a row reduced echelon matrix.

**Definition 2.4.9 (Back Substitution/Gauss-Jordan Method)** The procedure to get to Step II of Example 2.3.10 from Step 5 of Example 2.3.10 is called the back substitution.

The elimination process applied to obtain the row reduced echelon form of the augmented matrix is called the Gauss-Jordan elimination.

That is, the Gauss-Jordan elimination method consists of both the forward elimination and the backward substitution.

**Method to get the row-reduced echelon form of a given matrix $A$**

Let $A$ be an $m \times n$ matrix. Then the following method is used to obtain the row-reduced echelon form the matrix $A$.

**Step 1:** Consider the first column of the matrix $A$.

If all the entries in the first column are zero, move to the second column.

Else, find a row, say $i^{\text{th}}$ row, which contains a non-zero entry in the first column. Now, interchange the first row with the $i^{\text{th}}$ row. Suppose the non-zero entry in the $(1,1)$-position is $\alpha \neq 0$. Divide the whole row by $\alpha$ so that the $(1,1)$-entry of the new matrix is 1. Now, use the 1 to make all the entries below this 1 equal to 0.

**Step 2:** If all entries in the first column after the first step are zero, consider the right $m \times (n-1)$ submatrix of the matrix obtained in step 1 and proceed as in step 1.

Else, forget the first row and first column. Start with the lower $(m-1) \times (n-1)$ submatrix of the matrix obtained in the first step and proceed as in step 1.

**Step 3:** Keep repeating this process till we reach a stage where all the entries below a particular row, say $r$, are zero. Suppose at this stage we have obtained a matrix $C$. Then $C$ has the following form:

1. THE FIRST NON-ZERO ENTRY IN EACH ROW of $C$ is 1. These 1's are the leading terms of $C$ and the columns containing these leading terms are the leading columns.
2. THE ENTRIES OF $C$ BELOW THE LEADING TERM ARE ALL ZERO.

**Step 4:** Now use the leading term in the $r^{\text{th}}$ row to make all entries in the $r^{\text{th}}$ leading column equal to zero.

**Step 5:** Next, use the leading term in the $(r-1)^{\text{th}}$ row to make all entries in the $(r-1)^{\text{th}}$ leading column equal to zero and continue till we come to the first leading term or column.

The final matrix is the row-reduced echelon form of the matrix $A$.

**Remark 2.4.10** *Note that the row reduction involves only row operations and proceeds from* LEFT TO RIGHT. *Hence, if $A$ is a matrix consisting of first $s$ columns of a matrix $C$, then the row reduced form of $A$ will be the first $s$ columns of the row reduced form of $C$.*

The proof of the following theorem is beyond the scope of this book and is omitted.

**Theorem 2.4.11** The row reduced echelon form of a matrix is unique.

**Exercise 2.4.12** 1. Solve the following linear system.

(a) $x+y+z+w=0,\ x-y+z+w=0$ and $-x+y+3z+3w=0$.

(b) $x+2y+3z=1$ and $x+3y+2z=1$.

(c) $x+y+z=3,\ x+y-z=1$ and $x+y+7z=6$.

(d) $x+y+z=3,\ x+y-z=1$ and $x+y+4z=6$.

(e) $x+y+z=3,\ x+y-z=1,\ x+y+4z=6$ and $x+y-4z=-1$.

2. Find the row-reduced echelon form of the following matrices.

$$1.\ \begin{bmatrix} -1 & 1 & 3 & 5 \\ 1 & 3 & 5 & 7 \\ 9 & 11 & 13 & 15 \\ -3 & -1 & 13 & \end{bmatrix}, \qquad 2.\ \begin{bmatrix} 10 & 8 & 6 & 4 \\ 2 & 0 & -2 & -4 \\ -6 & -8 & -10 & -12 \\ -2 & -4 & -6 & -8 \end{bmatrix}$$

### 2.4.2 Elementary Matrices

**Definition 2.4.13** A square matrix $E$ of order $n$ is called an **elementary matrix** if it is obtained by applying exactly one elementary row operation to the identity matrix, $I_n$.

**Remark 2.4.14** *There are three types of elementary matrices.*

1. *$E_{ij}$, which is obtained by the application of the elementary row operation $R_{ij}$ to the identity matrix, $I_n$. Thus, the $(k,\ell)^{th}$ entry of $E_{ij}$ is* $(E_{ij})_{(k,\ell)} = \begin{cases} 1 & \text{if } k=\ell \text{ and } \ell \neq i,j \\ 1 & \text{if } (k,\ell)=(i,j) \text{ or } (k,\ell)=(j,i) \\ 0 & \text{otherwise} \end{cases}.$

2. *$E_k(c)$, which is obtained by the application of the elementary row operation $R_k(c)$ to the identity matrix, $I_n$. The $(i,j)^{th}$ entry of $E_k(c)$ is* $(E_k(c))_{(i,j)} = \begin{cases} 1 & \text{if } i=j \text{ and } i \neq k \\ c & \text{if } i=j=k \\ 0 & \text{otherwise} \end{cases}.$

3. *$E_{ij}(c)$, which is obtained by the application of the elementary row operation $R_{ij}(c)$ to the identity matrix, $I_n$. The $(k,\ell)^{th}$ entry of $E_{ij}(c)$ is* $(E_{ij})_{(k,\ell)} \begin{cases} 1 & \text{if } k=\ell \\ c & \text{if } (k,\ell)=(i,j) \\ 0 & \text{otherwise} \end{cases}.$

In particular,

$$E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},\ E_1(c) = \begin{bmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \text{and } E_{23}(c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}.$$

**Example 2.4.15** 1. Let $A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 3 & 4 \\ 3 & 4 & 5 & 6 \end{bmatrix}$. Then

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 3 & 4 \\ 3 & 4 & 5 & 6 \end{bmatrix} \xrightarrow{R_{23}} \begin{bmatrix} 1 & 2 & 3 & 0 \\ 3 & 4 & 5 & 6 \\ 2 & 0 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} A = E_{23}A.$$

That is, interchanging the two rows of the matrix $A$ is same as multiplying on the left by the corresponding elementary matrix. In other words, we see that the left multiplication of elementary matrices to a matrix results in elementary row operations.

2. Consider the augmented matrix $[A \; \mathbf{b}] = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$. Then the result of the steps given below is same as the matrix product

$$E_{23}(-1)E_{12}(-1)E_3(1/3)E_{32}(2)E_{23}E_{21}(-2)E_{13}[A \; \mathbf{b}].$$

$$\begin{aligned}
\begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix} \quad & \xrightarrow[R_{13}]{} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \end{bmatrix} \xrightarrow[R_{21}(-2)]{} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & -2 & 1 & -1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \xrightarrow[R_{23}]{} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & -2 & 1 & -1 \end{bmatrix} \\
& \xrightarrow[R_{32}(2)]{} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 3 & 3 \end{bmatrix} \xrightarrow[R_3(1/3)]{} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow[R_{12}(-1)]{} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix} \\
& \xrightarrow[R_{23}(-1)]{} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}
\end{aligned}$$

Now, consider an $m \times n$ matrix $A$ and an elementary matrix $E$ of order $n$. Then multiplying by $E$ on the right to $A$ corresponds to applying column transformation on the matrix $A$. Therefore, for each elementary matrix, there is a corresponding column transformation. We summarize:

**Definition 2.4.16** The column transformations obtained by right multiplication of elementary matrices are called **elementary column operations**.

**Example 2.4.17** Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \\ 3 & 4 & 5 \end{bmatrix}$ and consider the elementary column operation $f$ which interchanges the second and the third column of $A$. Then $f(A) = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 3 & 0 \\ 3 & 5 & 4 \end{bmatrix} = A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = AE_{23}$.

**Exercise 2.4.18**

1. Let $e$ be an elementary row operation and let $E = e(I)$ be the corresponding elementary matrix. That is, $E$ is the matrix obtained from $I$ by applying the elementary row operation $e$. Show that $e(A) = EA$.

2. Show that the Gauss elimination method is same as multiplying by a series of elementary matrices on the left to the augmented matrix.

   Does the Gauss-Jordan method also corresponds to multiplying by elementary matrices on the left? Give reasons.

3. Let $A$ and $B$ be two $m \times n$ matrices. Then prove that the two matrices $A, B$ are row-equivalent if and only if $B = PA$, where $P$ is product of elementary matrices. When is this $P$ unique?

## 2.5 Rank of a Matrix

In previous sections, we solved linear systems using Gauss elimination method or the Gauss-Jordan method. In the examples considered, we have encountered three possibilities, namely

1. existence of a unique solution,

2. existence of an infinite number of solutions, and

3. no solution.

Based on the above possibilities, we have the following definition.

**Definition 2.5.1 (Consistent, Inconsistent)** A linear system is called CONSISTENT if it admits a solution and is called INCONSISTENT if it admits no solution.

The question arises, as to whether there are conditions under which the linear system $A\mathbf{x} = \mathbf{b}$ is consistent. The answer to this question is in the affirmative. To proceed further, we need a few definitions and remarks.

Recall that the row reduced echelon form of a matrix is unique and therefore, the number of non-zero rows is a unique number. Also, note that the number of non-zero rows in either the row reduced form or the row reduced echelon form of a matrix are same.

**Definition 2.5.2 (Row rank of a Matrix)** The number of non-zero rows in the row reduced form of a matrix is called the row-rank of the matrix.

By the very definition, it is clear that row-equivalent matrices have the same row-rank. For a matrix $A$, we write 'row-rank $(A)$' to denote the row-rank of $A$.

**Example 2.5.3** 1. Determine the row-rank of $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}$.

**Solution:** To determine the row-rank of $A$, we proceed as follows.

(a) $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{R_{21}(-2), R_{31}(-1)} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{R_2(-1), R_{32}(1)} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$.

(c) $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} \xrightarrow{R_3(1/2), R_{12}(-2)} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

(d) $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_{23}(-1), R_{13}(1)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

The last matrix in Step 1d is the row reduced form of $A$ which has $3$ non-zero rows. Thus, row-rank$(A) = 3$. This result can also be easily deduced from the last matrix in Step 1b.

2. Determine the row-rank of $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

**Solution:** Here we have

(a) $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{R_{21}(-2), R_{31}(-1)} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & -1 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & -1 \end{bmatrix} \xrightarrow{R_2(-1), R_{32}(1)} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

From the last matrix in Step 2b, we deduce row-rank$(A) = 2$.

**Remark 2.5.4** *Let $A\mathbf{x} = \mathbf{b}$ be a linear system with $m$ equations and $n$ unknowns. Then the row-reduced echelon form of $A$ agrees with the first $n$ columns of $[A \;\; \mathbf{b}]$, and hence*

$$\textit{row-rank}(A) \leq \textit{row-rank}([A \;\; \mathbf{b}]).$$

*The reader is advised to supply a proof.*

**Remark 2.5.5** *Consider a matrix $A$. After application of a finite number of elementary column operations (see Definition 2.4.16) to the matrix $A$, we can have a matrix, say $B$, which has the following properties:*

1. *The first nonzero entry in each column is $1$.*
2. *A column containing only $0$'s comes after all columns with at least one non-zero entry.*
3. *The first non-zero entry (the leading term) in each non-zero column moves down in successive columns.*

Therefore, we can define **column-rank** of $A$ as the number of non-zero columns in $B$. It will be proved later that

$$\text{row-rank}(A) = \text{column-rank}(A).$$

Thus we are led to the following definition.

**Definition 2.5.6** The number of non-zero rows in the row reduced form of a matrix $A$ is called the **rank** of $A$, denoted rank $(A)$.

**Theorem 2.5.7** Let $A$ be a matrix of rank $r$. Then there exist elementary matrices $E_1, E_2, \ldots, E_s$ and $F_1, F_2, \ldots, F_\ell$ such that

$$E_1 E_2 \ldots E_s \; A \; F_1 F_2 \ldots F_\ell = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Proof. Let $C$ be the row reduced echelon matrix obtained by applying elementary row operations to the given matrix $A$. As rank$(A) = r$, the matrix $C$ will have the first $r$ rows as the non-zero rows. So by Remark 2.4.5, $C$ will have $r$ leading columns, say $i_1, i_2, \ldots, i_r$. Note that, for $1 \leq s \leq r$, the $i_s^{\text{th}}$ column will have 1 in the $s^{\text{th}}$ row and zero elsewhere.

We now apply column operations to the matrix $C$. Let $D$ be the matrix obtained from $C$ by successively interchanging the $s^{\text{th}}$ and $i_s^{\text{th}}$ column of $C$ for $1 \leq s \leq r$. Then the matrix $D$ can be written in the form $\begin{bmatrix} I_r & B \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $B$ is a matrix of appropriate size. As the $(1,1)$ block of $D$ is an identity matrix, the block $(1,2)$ can be made the zero matrix by application of column operations to $D$. This gives the required result. □

**Exercise 2.5.8**
1. Determine the ranks of the coefficient and the augmented matrices that appear in Part 1 and Part 2 of Exercise 2.4.12.
2. For any matrix $A$, prove that rank$(A) =$ rank$(A^t)$.
3. Let $A$ be an $n \times n$ matrix with rank$(A) = n$. Then prove that $A$ is row-equivalent to $I_n$.

## 2.6 Existence of Solution of $A\mathbf{x} = \mathbf{b}$

We try to understand the properties of the set of solutions of a linear system through an example, using the Gauss-Jordan method. Based on this observation, we arrive at the existence and uniqueness results for the linear system $A\mathbf{x} = \mathbf{b}$. This example is more or less a motivation.

### 2.6.1 Example

Consider a linear system $A\mathbf{x} = \mathbf{b}$ which after the application of the Gauss-Jordan method reduces to a matrix $[C \;\; \mathbf{d}]$ with

$$[C \;\; \mathbf{d}] = \begin{bmatrix} 1 & 0 & 2 & -1 & 0 & 0 & 2 & 8 \\ 0 & 1 & 1 & 3 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For this particular matrix $[C \;\; \mathbf{d}]$, we want to see the set of solutions. We start with some observations.
**Observations:**

1. The number of non-zero rows in $C$ is 4. This number is also equal to the number of non-zero rows in $[C \;\; \mathbf{d}]$.

2. The first non-zero entry in the non-zero rows appear in columns $1, 2, 5$ and $6$.

3. Thus, the respective variables $x_1, x_2, x_5$ and $x_6$ are the basic variables.

4. The remaining variables, $x_3, x_4$ and $x_7$ are free variables.

5. We assign arbitrary constants $k_1, k_2$ and $k_3$ to the free variables $x_3, x_4$ and $x_7$, respectively.

Hence, we have the set of solutions as

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} &= \begin{bmatrix} 8 - 2k_1 + k_2 - 2k_3 \\ 1 - k_1 - 3k_2 - 5k_3 \\ k_1 \\ k_2 \\ 2 + k_3 \\ 4 - k_3 \\ k_3 \end{bmatrix} \\ &= \begin{bmatrix} 8 \\ 1 \\ 0 \\ 0 \\ 2 \\ 4 \\ 0 \end{bmatrix} + k_1 \begin{bmatrix} -2 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 1 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + k_3 \begin{bmatrix} -2 \\ -5 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \end{aligned}$$

where $k_1, k_2$ and $k_3$ are arbitrary.

Let $\mathbf{u}_0 = \begin{bmatrix} 8\\1\\0\\0\\2\\4\\0 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} -2\\-1\\1\\0\\0\\0\\0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1\\-3\\0\\1\\0\\0\\0 \end{bmatrix}$ and $\mathbf{u}_3 = \begin{bmatrix} -2\\-5\\0\\0\\1\\-1\\1 \end{bmatrix}$.

Then it can easily be verified that $C\mathbf{u}_0 = \mathbf{d}$, and for $1 \leq i \leq 3$, $C\mathbf{u}_i = \mathbf{0}$.

A similar idea is used in the proof of the next theorem and is omitted. The interested readers can read the proof in Appendix 14.1.

### 2.6.2 Main Theorem

**Theorem 2.6.1** [Existence and Non-existence] Consider a linear system $A\mathbf{x} = \mathbf{b}$, where $A$ is a $m \times n$ matrix, and $\mathbf{x}$, $\mathbf{b}$ are vectors with orders $n \times 1$, and $m \times 1$, respectively. Suppose rank $(A) = r$ and rank$([A \;\; \mathbf{b}]) = r_a$. Then exactly one of the following statement holds:

1. if $r_a = r < n$, the set of solutions of the linear system is an infinite set and has the form
$$\{\mathbf{u}_0 + k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_{n-r}\mathbf{u}_{n-r} \;:\; k_i \in \mathbb{R},\ 1 \leq i \leq n-r\},$$
where $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-r}$ are $n \times 1$ vectors satisfying $A\mathbf{u}_0 = \mathbf{b}$ and $A\mathbf{u}_i = \mathbf{0}$ for $1 \leq i \leq n-r$.
2. if $r_a = r = n$, the solution set of the linear system has a unique $n \times 1$ vector $\mathbf{x}_0$ satisfying $A\mathbf{x}_0 = \mathbf{b}$.
3. If $r < r_a$, the linear system has no solution.

**Remark 2.6.2** *Let $A$ be an $m \times n$ matrix and consider the linear system $A\mathbf{x} = \mathbf{b}$. Then by Theorem 2.6.1, we see that the linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if*
$$\text{rank } (A) = \text{rank}([A \;\; \mathbf{b}]).$$

The following corollary of Theorem 2.6.1 is a very important result about the homogeneous linear system $A\mathbf{x} = \mathbf{0}$.

**Corollary 2.6.3** Let $A$ be an $m \times n$ matrix. Then the homogeneous system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution if and only if rank$(A) < n$.

Proof. Suppose the system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution, $\mathbf{x}_0$. That is, $A\mathbf{x}_0 = \mathbf{0}$ and $\mathbf{x}_0 \neq \mathbf{0}$. Under this assumption, we need to show that rank$(A) < n$. On the contrary, assume that rank$(A) = n$. So,
$$n = \text{rank}(A) = \text{rank}\big([A \;\; \mathbf{0}]\big) = r_a.$$
Also $A\mathbf{0} = \mathbf{0}$ implies that $\mathbf{0}$ is a solution of the linear system $A\mathbf{x} = \mathbf{0}$. Hence, by the uniqueness of the solution under the condition $r = r_a = n$ (see Theorem 2.6.1), we get $\mathbf{x}_0 = \mathbf{0}$. A contradiction to the fact that $\mathbf{x}_0$ was a given non-trivial solution.

Now, let us assume that rank$(A) < n$. Then
$$r_a = \text{rank}\big([A \;\; \mathbf{0}]\big) = \text{rank}(A) < n.$$
So, by Theorem 2.6.1, the solution set of the linear system $A\mathbf{x} = \mathbf{0}$ has infinite number of vectors $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{0}$. From this infinite set, we can choose any vector $\mathbf{x}_0$ that is different from $\mathbf{0}$. Thus, we have a solution $\mathbf{x}_0 \neq \mathbf{0}$. That is, we have obtained a non-trivial solution $\mathbf{x}_0$. □

We now state another important result whose proof is immediate from Theorem 2.6.1 and Corollary 2.6.3.

**Proposition 2.6.4** Consider the linear system $A\mathbf{x} = \mathbf{b}$. Then the two statements given below cannot hold together.

1. The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.
2. The system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.

**Remark 2.6.5** 1. *Suppose $\mathbf{x}_1, \mathbf{x}_2$ are two solutions of $A\mathbf{x} = \mathbf{0}$. Then $k_1\mathbf{x}_1 + k_2\mathbf{x}_2$ is also a solution of $A\mathbf{x} = \mathbf{0}$ for any $k_1, k_2 \in \mathbb{R}$.*

2. *If $\mathbf{u}, \mathbf{v}$ are two solutions of $A\mathbf{x} = \mathbf{b}$ then $\mathbf{u} - \mathbf{v}$ is a solution of the system $A\mathbf{x} = \mathbf{0}$. That is, $\mathbf{u} - \mathbf{v} = \mathbf{x}_h$ for some solution $\mathbf{x}_h$ of $A\mathbf{x} = \mathbf{0}$. That is, any two solutions of $A\mathbf{x} = \mathbf{b}$ differ by a solution of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$.*

   *In conclusion, for $\mathbf{b} \neq \mathbf{0}$, the set of solutions of the system $A\mathbf{x} = \mathbf{b}$ is of the form, $\{\mathbf{x}_0 + \mathbf{x}_h\}$; where $\mathbf{x}_0$ is a particular solution of $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x}_h$ is a solution $A\mathbf{x} = \mathbf{0}$.*

### 2.6.3 Exercises

**Exercise 2.6.6** 1. For what values of $c$ and $k$-the following systems have $i)$ no solution, $ii)$ a unique solution and $iii)$ infinite number of solutions.

   (a) $x + y + z = 3, \ \ x + 2y + cz = 4, \ \ 2x + 3y + 2cz = k.$

   (b) $x + y + z = 3, \ x + y + 2cz = 7, \ x + 2y + 3cz = k.$

   (c) $x + y + 2z = 3, \ x + 2y + cz = 5, \ x + 2y + 4z = k.$

   (d) $kx + y + z = 1, \ x + ky + z = 1, \ x + y + kz = 1.$

   (e) $x + 2y - z = 1, \ 2x + 3y + kz = 3, \ x + ky + 3z = 2.$

   (f) $x - 2y = 1, \ x - y + kz = 1, \ ky + 4z = 6.$

2. Find the condition on $a, b, c$ so that the linear system

$$x + 2y - 3z = a, \ 2x + 6y - 11z = b, \ x - 2y + 7z = c$$

   is consistent.

3. Let $A$ be an $n \times n$ matrix. If the system $A^2\mathbf{x} = \mathbf{0}$ has a non trivial solution then show that $A\mathbf{x} = \mathbf{0}$ also has a non trivial solution.

## 2.7 Invertible Matrices

### 2.7.1 Inverse of a Matrix

**Definition 2.7.1 (Inverse of a Matrix)** Let $A$ be a square matrix of order $n$.

1. A square matrix $B$ is said to be a LEFT INVERSE of $A$ if $BA = I_n$.
2. A square matrix $C$ is called a RIGHT INVERSE of $A$, if $AC = I_n$.
3. A matrix $A$ is said to be INVERTIBLE (or is said to have an INVERSE) if there exists a matrix $B$ such that $AB = BA = I_n$.

**Lemma 2.7.2** Let $A$ be an $n \times n$ matrix. Suppose that there exist $n \times n$ matrices $B$ and $C$ such that $AB = I_n$ and $CA = I_n$, then $B = C$.

PROOF. Note that

$$C = CI_n = C(AB) = (CA)B = I_nB = B.$$

□

**Remark 2.7.3** 1. *From the above lemma, we observe that if a matrix $A$ is invertible, then the inverse is unique.*

2. *As the inverse of a matrix $A$ is unique, we denote it by $A^{-1}$. That is, $AA^{-1} = A^{-1}A = I$.*

**Theorem 2.7.4** Let $A$ and $B$ be two matrices with inverses $A^{-1}$ and $B^{-1}$, respectively. Then

1. $(A^{-1})^{-1} = A$.
2. $(AB)^{-1} = B^{-1}A^{-1}$.
3. prove that $(A^t)^{-1} = (A^{-1})^t$.

PROOF. Proof of Part 1.
By definition $AA^{-1} = A^{-1}A = I$. Hence, if we denote $A^{-1}$ by $B$, then we get $AB = BA = I$. This again by definition, implies $B^{-1} = A$, or equivalently $(A^{-1})^{-1} = A$.

Proof of Part 2.
Verify that $(AB)(B^{-1}A^{-1}) = I = (B^{-1}A^{-1})(AB)$. Hence, the result follows by definition.

Proof of Part 3.
We know $AA^{-1} = A^{-1}A = I$. Taking transpose, we get

$$(AA^{-1})^t = (A^{-1}A)^t = I^t \Longleftrightarrow (A^{-1})^tA^t = A^t(A^{-1})^t = I.$$

Hence, by definition $(A^t)^{-1} = (A^{-1})^t$. □

**Exercise 2.7.5** 1. If $A$ is a symmetric matrix, is the matrix $A^{-1}$ symmetric?

2. Show that every elementary matrix is invertible. Is the inverse of an elementary matrix, also an elementary matrix?

3. Let $A_1, A_2, \ldots, A_r$ be invertible matrices. Prove that the product $A_1A_2\cdots A_r$ is also an invertible matrix.

4. If $P$ and $Q$ are invertible matrices and $PAQ$ is defined then show that rank $(PAQ) =$ rank $(A)$.

5. Find matrices $P$ and $Q$ which are product of elementary matrices such that $B = PAQ$ where $A = \begin{bmatrix} 2 & 4 & 8 \\ 1 & 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

6. Let $A$ and $B$ be two matrices. Show that

   (a) if $A + B$ is defined, then rank$(A + B) \leq$ rank$(A) +$ rank$(B)$,

   (b) if $AB$ is defined, then rank$(AB) \leq$ rank$(A)$ and rank$(AB) \leq$ rank$(B)$.

7. Let $A$ be any matrix of rank $r$. Then show that there exists invertible matrices $B_i, C_i$ such that $B_1A = \begin{bmatrix} R_1 & R_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $AC_1 = \begin{bmatrix} S_1 & \mathbf{0} \\ S_3 & \mathbf{0} \end{bmatrix}$, $B_2AC_2 = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, and $B_3AC_3 = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Also, prove that the matrix $A_1$ is an $r \times r$ invertible matrix.

8. Let $A$ be an $m \times n$ matrix of rank $r$. Then $A$ can be written as $A = BC$, where both $B$ and $C$ have rank $r$ and $B$ is a matrix of size $m \times r$ and $C$ is a matrix of size $r \times n$.

9. Let $A$ and $B$ be two matrices such that $AB$ is defined and rank $(A) =$ rank $(AB)$. Then show that $A = ABX$ for some matrix $X$. Similarly, if $BA$ is defined and rank $(A) =$ rank $(BA)$, then $A = YBA$ for some matrix $Y$. *[Hint: Choose non-singular matrices $P, Q$ and $R$ such that $PAQ = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and $P(AB)R = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Define $X = R \begin{bmatrix} C^{-1}A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} Q^{-1}$.]*

10. Let $A = [a_{ij}]$ be an invertible matrix and let $B = [p^{i-j} a_{ij}]$ for some nonzero real number $p$. Find the inverse of $B$.

11. If matrices $B$ and $C$ are invertible and the involved partitioned products are defined, then show that

$$\begin{bmatrix} A & B \\ C & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & C^{-1} \\ B^{-1} & -B^{-1}AC^{-1} \end{bmatrix}.$$

12. Suppose $A$ is the inverse of a matrix $B$. Partition $A$ and $B$ as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

If $A_{11}$ is invertible and $P = A_{22} - A_{21}(A_{11}^{-1}A_{12})$, then show that

$$B_{11} = A_{11}^{-1} + (A_{11}^{-1}A_{12})P^{-1}(A_{21}A_{11}^{-1}), \;\; B_{21} = -P^{-1}(A_{21}A_{11}^{-1}), \;\; B_{12} = -(A_{11}^{-1}A_{12})P^{-1},$$

and $B_{22} = P^{-1}$.

### 2.7.2 Equivalent conditions for Invertibility

**Definition 2.7.6** A square matrix $A$ or order $n$ is said to be of **full rank** if rank $(A) = n$.

**Theorem 2.7.7** For a square matrix $A$ of order $n$, the following statements are equivalent.

1. $A$ is invertible.
2. $A$ is of full rank.
3. $A$ is row-equivalent to the identity matrix.
4. $A$ is a product of elementary matrices.

Proof. $1 \Longrightarrow 2$

Let if possible rank$(A) = r < n$. Then there exists an invertible matrix $P$ (a product of elementary matrices) such that $PA = \begin{bmatrix} B_1 & B_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $B_1$ is an $r \times r$ matrix. Since $A$ is invertible, let $A^{-1} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$, where $C_1$ is an $r \times n$ matrix. Then

$$P = PI_n = P(AA^{-1}) = (PA)A^{-1} = \begin{bmatrix} B_1 & B_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} B_1C_1 + B_2C_2 \\ \mathbf{0} \end{bmatrix}. \tag{2.7.1}$$

Thus the matrix $P$ has $n - r$ rows as zero rows. Hence, $P$ cannot be invertible. A contradiction to $P$ being a product of invertible matrices. Thus, $A$ is of full rank.

$2 \Longrightarrow 3$

Suppose $A$ is of full rank. This implies, the row reduced echelon form of $A$ has all non-zero rows. But $A$ has as many columns as rows and therefore, the last row of the row reduced echelon form of $A$ will be $(0, 0, \ldots, 0, 1)$. Hence, the row reduced echelon form of $A$ is the identity matrix.

$3 \Longrightarrow 4$

Since $A$ is row-equivalent to the identity matrix there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $A = E_1 E_2 \cdots E_k I_n$. That is, $A$ is product of elementary matrices.

$4 \Longrightarrow 1$

Suppose $A = E_1 E_2 \cdots E_k$; where the $E_i$'s are elementary matrices. We know that elementary matrices are invertible and product of invertible matrices is also invertible, we get the required result. □

The ideas of Theorem 2.7.7 will be used in the next subsection to find the inverse of an invertible matrix. The idea used in the proof of the first part also gives the following important Theorem. We repeat the proof for the sake of clarity.

**Theorem 2.7.8** Let $A$ be a square matrix of order $n$.

1. Suppose there exists a matrix $B$ such that $AB = I_n$. Then $A^{-1}$ exists.

2. Suppose there exists a matrix $C$ such that $CA = I_n$. Then $A^{-1}$ exists.

PROOF. Suppose that $AB = I_n$. We will prove that the matrix $A$ is of full rank. That is, rank $(A) = n$.

Let if possible, rank$(A) = r < n$. Then there exists an invertible matrix $P$ (a product of elementary matrices) such that $PA = \begin{bmatrix} C_1 & C_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Let $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$, where $B_1$ is an $r \times n$ matrix. Then

$$P = PI_n = P(AB) = (PA)B = \begin{bmatrix} C_1 & C_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} C_1 B_1 + C_2 B_2 \\ \mathbf{0} \end{bmatrix}. \qquad (2.7.2)$$

Thus the matrix $P$ has $n - r$ rows as zero rows. So, $P$ cannot be invertible. A contradiction to $P$ being a product of invertible matrices. Thus, rank $(A) = n$. That is, $A$ is of full rank. Hence, using Theorem 2.7.7, $A$ is an invertible matrix. That is, $BA = I_n$ as well.

Using the first part, it is clear that the matrix $C$ in the second part, is invertible. Hence

$$AC = I_n = CA.$$

Thus, $A$ is invertible as well. □

**Remark 2.7.9** *This theorem implies the following: "if we want to show that a square matrix $A$ of order $n$ is invertible, it is enough to show the existence of*

1. *either a matrix $B$ such that $AB = I_n$*

2. *or a matrix $C$ such that $CA = I_n$.*

**Theorem 2.7.10** The following statements are equivalent for a square matrix $A$ of order $n$.

1. $A$ is invertible.

2. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

3. $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$ for every $\mathbf{b}$.

PROOF. $1 \Longrightarrow 2$

Since $A$ is invertible, by Theorem 2.7.7 $A$ is of full rank. That is, for the linear system $A\mathbf{x} = \mathbf{0}$, the number of unknowns is equal to the rank of the matrix $A$. Hence, by Theorem 2.6.1 the system $A\mathbf{x} = \mathbf{0}$ has a unique solution $\mathbf{x} = \mathbf{0}$.

$2 \Longrightarrow 1$

Let if possible $A$ be non-invertible. Then by Theorem 2.7.7, the matrix $A$ is not of full rank. Thus by Corollary 2.6.3, the linear system $A\mathbf{x} = \mathbf{0}$ has infinite number of solutions. This contradicts the assumption that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

$1 \Longrightarrow 3$

Since $A$ is invertible, for every $\mathbf{b}$, the system $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

$3 \Longrightarrow 1$

For $1 \leq i \leq n$, define $\mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{i\text{th position}}, 0, \ldots, 0)^t$, and consider the linear system $A\mathbf{x} = \mathbf{e}_i$. By assumption, this system has a solution $\mathbf{x}_i$ for each $i$, $1 \leq i \leq n$. Define a matrix $B = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$. That is, the $i^{\text{th}}$ column of $B$ is the solution of the system $A\mathbf{x} = \mathbf{e}_i$. Then

$$AB = A[\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_n] = [A\mathbf{x}_1, A\mathbf{x}_2 \ldots, A\mathbf{x}_n] = [\mathbf{e}_1, \mathbf{e}_2 \ldots, \mathbf{e}_n] = I_n.$$

Therefore, by Theorem 2.7.8, the matrix $A$ is invertible. □

**Exercise 2.7.11** 1. Show that a triangular matrix $A$ is invertible if and only if each diagonal entry of $A$ is non-zero.

2. Let $A$ be a $1 \times 2$ matrix and $B$ be a $2 \times 1$ matrix having positive entries. Which of $BA$ or $AB$ is invertible? Give reasons.

3. Let $A$ be an $n \times m$ matrix and $B$ be an $m \times n$ matrix. Prove that the matrix $I - BA$ is invertible if and only if the matrix $I - AB$ is invertible.

### 2.7.3 Inverse and Gauss-Jordan Method

We first give a consequence of Theorem 2.7.7 and then use it to find the inverse of an invertible matrix.

**Corollary 2.7.12** Let $A$ be an invertible $n \times n$ matrix. Suppose that a sequence of elementary row-operations reduces $A$ to the identity matrix. Then the same sequence of elementary row-operations when applied to the identity matrix yields $A^{-1}$.

PROOF. Let $A$ be a square matrix of order $n$. Also, let $E_1, E_2, \ldots, E_k$ be a sequence of elementary row operations such that $E_1 E_2 \cdots E_k A = I_n$. Then $E_1 E_2 \cdots E_k I_n = A^{-1}$. This implies $A^{-1} = E_1 E_2 \cdots E_k$. □

**Summary:** Let $A$ be an $n \times n$ matrix. Apply the Gauss-Jordan method to the matrix $[A \;\; I_n]$. Suppose the row reduced echelon form of the matrix $[A \;\; I_n]$ is $[B \;\; C]$. If $B = I_n$, then $A^{-1} = C$ or else $A$ is not invertible.

**Example 2.7.13** Find the inverse of the matrix $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ using Gauss-Jordan method.

**Solution:** Consider the matrix $\begin{bmatrix} 2 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 1 \end{bmatrix}$. A sequence of steps in the Gauss-Jordan method are:

1. $\begin{bmatrix}2&1&1&1&0&0\\1&2&1&0&1&0\\1&1&2&0&0&1\end{bmatrix}\underset{R_1(1/2)}{\longrightarrow}\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\1&2&1&0&1&0\\1&1&2&0&0&1\end{bmatrix}$

2. $\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\1&2&1&0&1&0\\1&1&2&0&0&1\end{bmatrix}\underset{\substack{R_{21}(-1)\\ \longrightarrow\\ R_{31}(-1)}}{\longrightarrow}\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&\frac{3}{2}&\frac{1}{2}&-\frac{1}{2}&1&0\\0&\frac{1}{2}&\frac{3}{2}&-\frac{1}{2}&0&1\end{bmatrix}$

3. $\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&\frac{3}{2}&\frac{1}{2}&-\frac{1}{2}&1&0\\0&\frac{1}{2}&\frac{3}{2}&-\frac{1}{2}&0&1\end{bmatrix}\underset{R_2(2/3)}{\longrightarrow}\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&1&\frac{1}{3}&-\frac{1}{3}&\frac{2}{3}&0\\0&\frac{1}{2}&\frac{3}{2}&-\frac{1}{2}&0&1\end{bmatrix}$

4. $\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&1&\frac{1}{3}&-\frac{1}{3}&\frac{2}{3}&0\\0&\frac{1}{2}&\frac{3}{2}&-\frac{1}{2}&0&1\end{bmatrix}\underset{R_{32}(-1/2)}{\longrightarrow}\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&1&\frac{1}{3}&-\frac{1}{3}&\frac{2}{3}&0\\0&0&\frac{4}{3}&-\frac{1}{3}&-\frac{1}{3}&1\end{bmatrix}$

5. $\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&1&\frac{1}{3}&-\frac{1}{3}&\frac{2}{3}&0\\0&0&\frac{4}{3}&-\frac{1}{3}&-\frac{1}{3}&1\end{bmatrix}\underset{R_3(3/4)}{\longrightarrow}\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&1&\frac{1}{3}&-\frac{1}{3}&\frac{2}{3}&0\\0&0&1&-\frac{1}{4}&-\frac{1}{4}&\frac{3}{4}\end{bmatrix}$

6. $\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&0&0\\0&1&\frac{1}{3}&\frac{-1}{3}&\frac{2}{3}&0\\0&0&1&\frac{-1}{4}&\frac{-1}{4}&\frac{3}{4}\end{bmatrix}\underset{\substack{R_{23}(-1/3)\\ \longrightarrow\\ R_{13}(-1/2)}}{\longrightarrow}\begin{bmatrix}1&\frac{1}{2}&0&\frac{5}{8}&\frac{1}{8}&\frac{-3}{8}\\0&1&0&\frac{-1}{4}&\frac{3}{4}&\frac{-1}{4}\\0&0&1&\frac{-1}{4}&\frac{-1}{4}&\frac{3}{4}\end{bmatrix}$

7. $\begin{bmatrix}1&\frac{1}{2}&0&\frac{5}{8}&\frac{1}{8}&\frac{-3}{8}\\0&1&0&\frac{-1}{4}&\frac{3}{4}&\frac{-1}{4}\\0&0&1&\frac{-1}{4}&\frac{-1}{4}&\frac{3}{4}\end{bmatrix}\underset{R_{12}(-1/2)}{\longrightarrow}\begin{bmatrix}1&0&0&\frac{3}{4}&\frac{-1}{4}&\frac{-1}{4}\\0&1&0&\frac{-1}{4}&\frac{3}{4}&\frac{-1}{4}\\0&0&1&\frac{-1}{4}&\frac{-1}{4}&\frac{3}{4}\end{bmatrix}.$

8. Thus, the inverse of the given matrix is $\begin{bmatrix}3/4&-1/4&-1/4\\-1/4&3/4&-1/4\\-1/4&-1/4&3/4\end{bmatrix}.$

**Exercise 2.7.14** Find the inverse of the following matrices using Gauss-Jordan method.
$(i)\ \begin{bmatrix}1&2&3\\1&3&2\\2&4&7\end{bmatrix},\quad (ii)\ \begin{bmatrix}1&3&3\\2&3&2\\2&4&7\end{bmatrix},\quad (iii)\ \begin{bmatrix}2&-1&3\\-1&3&-2\\2&4&1\end{bmatrix}.$

## 2.8 Determinant

Notation: For an $n \times n$ matrix $A$, by $A(\alpha|\beta)$, we mean the submatrix $B$ of $A$, which is obtained by deleting the $\alpha^{\text{th}}$ row and $\beta^{\text{th}}$ column.

**Example 2.8.1** Consider a matrix $A = \begin{bmatrix}1&2&3\\1&3&2\\2&4&7\end{bmatrix}$. Then $A(1|2) = \begin{bmatrix}1&2\\2&7\end{bmatrix}$, $A(1|3) = \begin{bmatrix}1&3\\2&4\end{bmatrix}$, and $A(1,2|1,3) = [4]$.

**Definition 2.8.2 (Determinant of a Square Matrix)** Let $A$ be a square matrix of order $n$. With $A$, we associate inductively (on $n$) a number, called the determinant of $A$, written $\det(A)$ (or $|A|$) by

$$\det(A) = \begin{cases} a & \text{if } A = [a]\ (n=1), \\ \sum\limits_{j=1}^{n} (-1)^{1+j} a_{1j} \det\big(A(1|j)\big), & \text{otherwise.} \end{cases}$$

**Definition 2.8.3 (Minor, Cofactor of a Matrix)** The number $\det\left(A(i|j)\right)$ is called the $(i,j)^{\text{th}}$ minor of $A$. We write $A_{ij} = \det\left(A(i|j)\right)$. The $(i,j)^{\text{th}}$ cofactor of $A$, denoted $C_{ij}$, is the number $(-1)^{i+j}A_{ij}$.

**Example 2.8.4** 1. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Then, $\det(A) = |A| = a_{11}A_{11} - a_{12}A_{12} = a_{11}a_{22} - a_{12}a_{21}$.

For example, for $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ $\det(A) = |A| = 1 - 2 \cdot 2 = -3$.

2. Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. Then,

$$\begin{aligned} \det(A) &= |A| = a_{11}A_{11} - a_{12}A_{12} + a_{13}A_{13} \\ &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \quad (2.8.1) \end{aligned}$$

For example, if $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$ then

$$\det(A) = |A| = 1 \cdot \begin{vmatrix} 3 & 1 \\ 2 & 2 \end{vmatrix} - 2 \cdot \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} + 3 \cdot \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = 4 - 2(3) + 3(1) = 1.$$

**Exercise 2.8.5** 1. Find the determinant of the following matrices.

$$i) \begin{bmatrix} 1 & 2 & 7 & 8 \\ 0 & 4 & 3 & 2 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix}, \quad ii) \begin{bmatrix} 3 & 5 & 2 & 1 \\ 0 & 2 & 0 & 5 \\ 6 & -7 & 1 & 0 \\ 2 & 0 & 3 & 0 \end{bmatrix}, \quad iii) \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}.$$

2. Show that the determinant of a triangular matrix is the product of its diagonal entries.

**Definition 2.8.6** A matrix $A$ is said to be a **singular** matrix if $\det(A) = 0$. It is called **non-singular** if $\det(A) \neq 0$.

The proof of the next theorem is omitted. The interested reader is advised to go through Appendix 14.3.

**Theorem 2.8.7** Let $A$ be an $n \times n$ matrix. Then

1. if $B$ is obtained from $A$ by interchanging two rows, then $\det(B) = -\det(A)$,

2. if $B$ is obtained from $A$ by multiplying a row by $c$ then $\det(B) = c\det(A)$,

3. if all the elements of one row or column are $0$ then $\det(A) = 0$,

4. if $B$ is obtained from $A$ by replacing the $j$th row by itself plus $k$ times the $i$th row, where $i \neq j$ then $\det(B) = \det(A)$,

5. if $A$ is a square matrix having two rows equal then $\det(A) = 0$.

**Remark 2.8.8** *1. Many authors define the determinant using "Permutations." It turns out that* THE WAY WE HAVE DEFINED DETERMINANT *is usually called the expansion of the determinant along the first row.*

*2. Part 1 of Lemma 2.8.7 implies that "one can also calculate the determinant by expanding along any row." Hence, for an $n \times n$ matrix $A$, for every $k$, $1 \le k \le n$, one also has*

$$\det(A) = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det\big(A(k|j)\big).$$

**Remark 2.8.9** *1. Let $\mathbf{u}^t = (u_1, u_2)$ and $\mathbf{v}^t = (v_1, v_2)$ be two vectors in $\mathbb{R}^2$. Then consider the parallelogram, $PQRS$, formed by the vertices $\{P = (0,0)^t, Q = \mathbf{u}, S = \mathbf{v}, R = \mathbf{u} + \mathbf{v}\}$. We*

$$\text{Claim:} \qquad \text{Area } (PQRS) = \left| \det \left( \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} \right) \right| = |u_1 v_2 - u_2 v_1|.$$

*Recall that the dot product, $\mathbf{u} \bullet \mathbf{v} = u_1 v_1 + u_2 v_2$, and $\sqrt{\mathbf{u} \bullet \mathbf{u}} = \sqrt{(u_1^2 + u_2^2)}$, is the length of the vector $\mathbf{u}$. We denote the length by $\ell(\mathbf{u})$. With the above notation, if $\theta$ is the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$, then*

$$\cos(\theta) = \frac{\mathbf{u} \bullet \mathbf{v}}{\ell(\mathbf{u})\ell(\mathbf{v})}.$$

*Which tells us,*

$$\begin{aligned} \text{Area}(PQRS) &= \ell(\mathbf{u})\ell(\mathbf{v}) \sin(\theta) = \ell(\mathbf{u})\ell(\mathbf{v}) \sqrt{1 - \left( \frac{\mathbf{u} \bullet \mathbf{v}}{\ell(\mathbf{u})\ell(\mathbf{v})} \right)^2} \\ &= \sqrt{\ell(\mathbf{u})^2 + \ell(v)^2 - (\mathbf{u} \bullet \mathbf{v})^2} = \sqrt{(u_1 v_2 - u_2 v_1)^2} \\ &= |u_1 v_2 - u_2 v_1|. \end{aligned}$$

*Hence, the claim holds. That is, in $\mathbb{R}^2$, the determinant is ± times the area of the parallelogram.*

*2. Let $\mathbf{u} = (u_1, u_2, u_3), \mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ be three elements of $\mathbb{R}^3$. Recall that the cross product of two vectors in $\mathbb{R}^3$ is,*

$$\mathbf{u} \times \mathbf{v} = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).$$

*Note here that if $A = [\mathbf{u}^t, \mathbf{v}^t, \mathbf{w}^t]$, then*

$$\det(A) = \begin{vmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{vmatrix} = \mathbf{u} \bullet (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \bullet (\mathbf{w} \times \mathbf{u}) = \mathbf{w} \bullet (\mathbf{u} \times \mathbf{v}).$$

*Let $P$ be the parallelopiped formed with $(0,0,0)$ as a vertex and the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ as adjacent vertices. Then observe that $\mathbf{u} \times \mathbf{v}$ is a vector perpendicular to the plane that contains the parallelogram formed by the vectors $\mathbf{u}$ and $\mathbf{v}$. So, to compute the volume of the parallelopiped $P$, we need to look at $\cos(\theta)$, where $\theta$ is the angle between the vector $\mathbf{w}$ and the normal vector to the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$. So,*

$$\text{volume } (P) = |\mathbf{w} \bullet (\mathbf{u} \times \mathbf{v})|.$$

*Hence, $|\det(A)| =$ volume $(P)$.*

*3. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathbb{R}^{n \times 1}$ and let $A = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ be an $n \times n$ matrix. Then the following properties of $\det(A)$ also hold for the volume of an $n$-dimensional parallelopiped formed with $\mathbf{0} \in \mathbb{R}^{n \times 1}$ as one vertex and the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ as adjacent vertices:*

(a) *If* $\mathbf{u}_1 = (1,0,\ldots,0)^t, \mathbf{u}_2 = (0,1,0,\ldots,0)^t, \ldots,$ *and* $\mathbf{u}_n = (0,\ldots,0,1)^t$, *then* $\det(A) = 1$. *Also, volume of a unit* $n$*-dimensional cube is* 1.

(b) *If we replace the vector* $\mathbf{u}_i$ *by* $\alpha\mathbf{u}_i$, *for some* $\alpha \in \mathbb{R}$, *then the determinant of the new matrix is* $\alpha \cdot \det(A)$. *This is also true for the volume, as the original volume gets multiplied by* $\alpha$.

(c) *If* $\mathbf{u}_1 = \mathbf{u}_i$ *for some* $i$, $2 \le i \le n$, *then the vectors* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ *will give rise to an* $(n-1)$*-dimensional parallelopiped. So, this parallelopiped lies on an* $(n-1)$*-dimensional hyperplane. Thus, its* $n$*-dimensional volume will be zero. Also,* $|\det(A)| = |0| = 0$.

*In general, for any* $n \times n$ *matrix* $A$, *it can be proved that* $|\det(A)|$ *is indeed equal to the volume of the* $n$*-dimensional parallelepiped. The actual proof is beyond the scope of this book.*

## 2.8.1 Adjoint of a Matrix

Recall that for a square matrix $A$, the notations $A_{ij}$ and $C_{ij} = (-1)^{i+j}A_{ij}$ were respectively used to denote the $(i,j)^{\text{th}}$ minor and the $(i,j)^{\text{th}}$ cofactor of $A$.

**Definition 2.8.10 (Adjoint of a Matrix)** Let $A$ be an $n \times n$ matrix. The matrix $B = [b_{ij}]$ with $b_{ij} = C_{ji}$, for $1 \le i, j \le n$ is called the Adjoint of $A$, denoted $Adj(A)$.

**Example 2.8.11** Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$. Then $Adj(A) = \begin{bmatrix} 4 & 2 & -7 \\ -3 & -1 & 5 \\ 1 & 0 & -1 \end{bmatrix}$;
as $C_{11} = (-1)^{1+1}A_{11} = 4, C_{12} = (-1)^{1+2}A_{12} = -3, C_{13} = (-1)^{1+3}A_{13} = 1$, and so on.

**Theorem 2.8.12** Let $A$ be an $n \times n$ matrix. Then

1. for $1 \le i \le n$, $\sum\limits_{j=1}^{n} a_{ij}\, C_{ij} = \sum\limits_{j=1}^{n} a_{ij}(-1)^{i+j}\, A_{ij} = \det(A)$,

2. for $i \neq \ell$, $\sum\limits_{j=1}^{n} a_{ij}\, C_{\ell j} = \sum\limits_{j=1}^{n} a_{ij}(-1)^{\ell+j}\, A_{\ell j} = 0$, and

3. $A(Adj(A)) = \det(A)I_n$. Thus,

$$\det(A) \neq 0 \Rightarrow A^{-1} = \frac{1}{\det(A)} Adj(A). \tag{2.8.2}$$

PROOF. Let $B = [b_{ij}]$ be a square matrix with

- the $\ell^{\text{th}}$ row of $B$ as the $i^{\text{th}}$ row of $A$,
- the other rows of $B$ are the same as that of $A$.

By the construction of $B$, two rows ($i^{\text{th}}$ and $\ell^{\text{th}}$) are equal. By Part 5 of Lemma 2.8.7, $\det(B) = 0$. By construction again, $\det\big(A(\ell|j)\big) = \det\big(B(\ell|j)\big)$ for $1 \le j \le n$. Thus, by Remark 2.8.8, we have

$$\begin{aligned} 0 = \det(B) &= \sum_{j=1}^{n}(-1)^{\ell+j} b_{\ell j} \det\big(B(\ell|j)\big) = \sum_{j=1}^{n}(-1)^{\ell+j} a_{ij} \det\big(B(\ell|j)\big) \\ &= \sum_{j=1}^{n}(-1)^{\ell+j} a_{ij} \det\big(A(\ell|j)\big) = \sum_{j=1}^{n} a_{ij} C_{\ell j}. \end{aligned}$$

Now,

$$\begin{aligned}\Big(A\big(\text{Adj}(A)\big)\Big)_{ij} &= \sum_{k=1}^{n} a_{ik}\big(\text{Adj}(A)\big)_{kj} = \sum_{k=1}^{n} a_{ik}C_{jk} \\ &= \begin{cases} 0 & \text{if } i \neq j \\ \det(A) & \text{if } i = j \end{cases}\end{aligned}$$

Thus, $A(Adj(A)) = \det(A)I_n$. Since, $\det(A) \neq 0$, $A\dfrac{1}{\det(A)}Adj(A) = I_n$. Therefore, $A$ has a right inverse. Hence, by Theorem 2.7.8 $A$ has an inverse and

$$A^{-1} = \frac{1}{\det(A)}Adj(A).$$

□

**Example 2.8.13** Let $A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$. Then

$$Adj(A) = \begin{bmatrix} -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -3 & 1 \end{bmatrix}$$

and $\det(A) = -2$. By Theorem 2.8.12.3, $A^{-1} = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ -1/2 & -1/2 & 1/2 \\ 1/2 & 3/2 & -1/2 \end{bmatrix}$.

The next corollary is an easy consequence of Theorem 2.8.12 (recall Theorem 2.7.8).

**Corollary 2.8.14** If $A$ is a non-singular matrix, then
$\big(Adj(A)\big)A = \det(A)I_n$ and $\sum\limits_{i=1}^{n} a_{ij}\, C_{ik} = \begin{cases} \det(A) & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$.

**Theorem 2.8.15** Let $A$ and $B$ be square matrices of order $n$. Then $\det(AB) = \det(A)\det(B)$.

Proof. **Step 1.** Let $\det(A) \neq 0$.
This means, $A$ is invertible. Therefore, either $A$ is an elementary matrix or is a product of elementary matrices (see Theorem 2.7.7). So, let $E_1, E_2, \ldots, E_k$ be elementary matrices such that $A = E_1E_2\cdots E_k$. Then, by using Parts 1, 2 and 4 of Lemma 2.8.7 repeatedly, we get

$$\begin{aligned}\det(AB) &= \det(E_1E_2\cdots E_kB) = \det(E_1)\det(E_2\cdots E_kB) \\ &= \det(E_1)\det(E_2)\det(E_3\cdots E_kB) \\ &= \det(E_1E_2)\det(E_3\cdots E_kB) \\ &= \vdots \\ &= \det(E_1E_2\cdots E_k)\det(B) \\ &= \det(A)\det(B).\end{aligned}$$

Thus, we get the required result in case $A$ is non-singular.

**Step 2.** Suppose $\det(A) = 0$.

Then $A$ is not invertible. Hence, there exists an invertible matrix $P$ such that $PA = C$, where $C = \begin{bmatrix} C_1 \\ \mathbf{0} \end{bmatrix}$. So, $A = P^{-1}C$, and therefore

$$\begin{aligned} \det(AB) &= \det((P^{-1}C)B) = \det(P^{-1}(CB)) = \det\left(P^{-1}\begin{bmatrix} C_1B \\ \mathbf{0} \end{bmatrix}\right) \\ &= \det(P^{-1}) \cdot \det\left(\begin{bmatrix} C_1B \\ \mathbf{0} \end{bmatrix}\right) \quad \text{as } P^{-1} \text{ is non-singular} \\ &= \det(P) \cdot 0 = 0 = 0 \cdot \det(B) = \det(A)\det(B). \end{aligned}$$

Thus, the proof of the theorem is complete. □

**Corollary 2.8.16** Let $A$ be a square matrix. Then $A$ is non-singular if and only if $A$ has an inverse.

Proof. Suppose $A$ is non-singular. Then $\det(A) \neq 0$ and therefore, $A^{-1} = \dfrac{1}{\det(A)} Adj(A)$. Thus, $A$ has an inverse.

Suppose $A$ has an inverse. Then there exists a matrix $B$ such that $AB = I = BA$. Taking determinant of both sides, we get

$$\det(A)\det(B) = \det(AB) = \det(I) = 1.$$

This implies that $\det(A) \neq 0$. Thus, $A$ is non-singular. □

**Theorem 2.8.17** Let $A$ be a square matrix. Then $\det(A) = \det(A^t)$.

Proof. If $A$ is a non-singular Corollary 2.8.14 gives $\det(A) = \det(A^t)$.

If $A$ is singular, then $\det(A) = 0$. Hence, by Corollary 2.8.16, $A$ doesn't have an inverse. Therefore, $A^t$ also doesn't have an inverse (for if $A^t$ has an inverse then $A^{-1} = \left((A^t)^{-1}\right)^t$). Thus again by Corollary 2.8.16, $\det(A^t) = 0$. Therefore, we again have $\det(A) = 0 = \det(A^t)$.

Hence, we have $\det(A) = \det(A^t)$. □

### 2.8.2 Cramer's Rule

Recall the following:

- The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ if and only if $A^{-1}$ exists.
- $A$ has an inverse if and only if $\det(A) \neq 0$.

Thus, $A\mathbf{x} = \mathbf{b}$ has a unique solution FOR EVERY $\mathbf{b}$ if and only if $\det(A) \neq 0$.

The following theorem gives a direct method of finding the solution of the linear system $A\mathbf{x} = \mathbf{b}$ when $\det(A) \neq 0$.

**Theorem 2.8.18 (Cramer's Rule)** Let $A\mathbf{x} = \mathbf{b}$ be a linear system with $n$ equations in $n$ unknowns. If $\det(A) \neq 0$, then the unique solution to this system is

$$x_j = \frac{\det(A_j)}{\det(A)}, \quad \text{for } j = 1, 2, \ldots, n,$$

where $A_j$ is the matrix obtained from $A$ by replacing the $j$th column of $A$ by the column vector $\mathbf{b}$.

PROOF. Since $\det(A) \neq 0$, $A^{-1} = \frac{1}{\det(A)} Adj(A)$. Thus, the linear system $A\mathbf{x} = \mathbf{b}$ has the solution $\mathbf{x} = \frac{1}{\det(A)} Adj(A)\mathbf{b}$. Hence, $x_j$, the $j$th coordinate of $\mathbf{x}$ is given by

$$x_j = \frac{b_1 C_{1j} + b_2 C_{2j} + \cdots + b_n C_{nj}}{\det(A)} = \frac{\det(A_j)}{\det(A)}.$$

□

The theorem implies that

$$x_1 = \frac{1}{\det(A)} \begin{vmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{vmatrix},$$

and in general

$$x_j = \frac{1}{\det(A)} \begin{vmatrix} a_{11} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ a_{12} & \cdots & a_{2j-1} & b_2 & a_{2j+1} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nj-1} & b_n & a_{nj+1} & \cdots & a_{nn} \end{vmatrix}$$

for $j = 2, 3, \ldots, n$.

**Example 2.8.19** Suppose that $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Use Cramer's rule to find a vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$.

**Solution:** Check that $\det(A) = 1$. Therefore $x_1 = \begin{vmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{vmatrix} = -1$, $x_2 = \begin{vmatrix} 1 & 1 & 3 \\ 2 & 1 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 1$, and $x_3 = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 2 & 1 \end{vmatrix} = 0$. That is, $\mathbf{x}^t = (-1, 1, 0)$.

## 2.9 Miscellaneous Exercises

**Exercise 2.9.1** 1. Let $A$ be an orthogonal matrix. Show that $\det A = \pm 1$.

2. If $A$ and $B$ are two $n \times n$ non-singular matrices, are the matrices $A + B$ and $A - B$ non-singular? Justify your answer.

3. For an $n \times n$ matrix $A$, prove that the following conditions are equivalent:

   (a) $A$ is singular ($A^{-1}$ doesn't exist).

   (b) rank$(A) \neq n$.

   (c) $\det(A) = 0$.

   (d) $A$ is not row-equivalent to $I_n$, the identity matrix of order $n$.

   (e) $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution for $\mathbf{x}$.

   (f) $A\mathbf{x} = b$ doesn't have a unique solution, i.e., it has no solutions or it has infinitely many solutions.

4. Let $A = \begin{bmatrix} 2 & 0 & 6 & 0 & 4 \\ 5 & 3 & 2 & 2 & 7 \\ 2 & 5 & 7 & 5 & 5 \\ 2 & 0 & 9 & 2 & 7 \\ 7 & 8 & 4 & 2 & 1 \end{bmatrix}$. We know that the numbers $20604, 53227$, $25755, 20927$ and $78421$ are all divisible by $17$. Does this imply $17$ divides $\det(A)$?

5. Let $A = [a_{ij}]_{n\times n}$ where $a_{ij} = x_i^{j-1}$. Show that $\det(A) = \prod\limits_{1\le i<j\le n} (x_j - x_i)$. [The matrix $A$ is usually called the Van-dermonde matrix.]

6. Let $A = [a_{ij}]$ with $a_{ij} = \max\{i, j\}$ be an $n \times n$ matrix. Compute $\det A$.

7. Let $A = [a_{ij}]$ with $a_{ij} = 1/(i + j)$ be an $n \times n$ matrix. Show that $A$ is invertible.

8. Solve the following system of equations by Cramer's rule.
   $i)$ $x + y + z - w = 1,\ x + y - z + w = 2,\ 2x + y + z - w = 7,\ x + y + z + w = 3.$
   $ii)$ $x - y + z - w = 1,\ x + y - z + w = 2,\ 2x + y - z - w = 7,\ x - y - z + w = 3.$

9. Suppose $A = [a_{ij}]$ and $B = [b_{ij}]$ are two $n \times n$ matrices such that $b_{ij} = p^{i-j}a_{ij}$ for $1 \le i, j \le n$ for some non-zero real number $p$. Then compute $\det(B)$ in terms of $\det(A)$.

10. The position of an element $a_{ij}$ of a determinant is called even or odd according as $i + j$ is even or odd. Show that

    (a) If all the entries in odd positions are multiplied with $-1$ then the value of the determinant doesn't change.

    (b) If all entries in even positions are multiplied with $-1$ then the determinant

        i. does not change if the matrix is of even order.

        ii. is multiplied by $-1$ if the matrix is of odd order.

11. Let $A$ be an $n \times n$ Hermitian matrix, that is, $A^* = A$. Show that $\det A$ is a real number. [$A$ is a matrix with complex entries and $A^* = \overline{A^t}$.]

12. Let $A$ be an $n \times n$ matrix. Then show that

$$A \text{ is invertible} \Longleftrightarrow Adj(A) \text{ is invertible.}$$

13. Let $A$ and $B$ be invertible matrices. Prove that $\text{Adj}(AB) = \text{Adj}(B)\text{Adj}(A)$.

14. Let $P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ be a rectangular matrix with $A$ a square matrix of order $n$ and $|A| \neq 0$. Then show that rank $(P) = n$ if and only if $D = CA^{-1}B$.

# Chapter 3

# Finite Dimensional Vector Spaces

Consider the problem of finding the set of points of intersection of the two planes $2x + 3y + z + u = 0$ and $3x + y + 2z + u = 0$.
Let $V$ be the set of points of intersection of the two planes. Then $V$ has the following properties:

1. The point $(0, 0, 0, 0)$ is an element of $V$.
2. For the points $(-1, 0, 1, 1)$ and $(-5, 1, 7, 0)$ which belong to $V$; the point $(-6, 1, 8, 1) = (-1, 0, 1, 1) + (-5, 1, 7, 0) \in V$.
3. Let $\alpha \in \mathbb{R}$. Then the point $\alpha(-1, 0, 1, 1) = (-\alpha, 0, \alpha, \alpha)$ also belongs to $V$.

Similarly, for an $m \times n$ real matrix $A$, consider the set $V$, of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. This set satisfies the following properties:

1. If $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$, then $\mathbf{x}, \mathbf{y} \in V$. Then $\mathbf{x} + \mathbf{y} \in V$ as $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$. Also, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. It is clear that if $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ then $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
3. The vector $\mathbf{0} \in V$ as $A\mathbf{0} = \mathbf{0}$.
4. If $A\mathbf{x} = \mathbf{0}$ then $A(-\mathbf{x}) = -A\mathbf{x} = \mathbf{0}$. Hence, $-\mathbf{x} \in V$.
5. Let $\alpha \in \mathbb{R}$ and $\mathbf{x} \in V$. Then $\alpha\mathbf{x} \in V$ as $A(\alpha\mathbf{x}) = \alpha A\mathbf{x} = \mathbf{0}$.

Thus we are lead to the following.

## 3.1 Vector Spaces

### 3.1.1 Definition

**Definition 3.1.1 (Vector Space)** A vector space over $\mathbb{F}$, denoted $V(\mathbb{F})$, is a non-empty set, satisfying the following axioms:

1. Vector Addition: To every pair $\mathbf{u}, \mathbf{v} \in V$ there corresponds a unique element $\mathbf{u} \oplus \mathbf{v}$ in $V$ such that
   (a) $\mathbf{u} \oplus \mathbf{v} = \mathbf{v} \oplus \mathbf{u}$ (Commutative law).
   (b) $(\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w} = \mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w})$ (Associative law).
   (c) There is a unique element $\mathbf{0}$ in $V$ (the zero vector) such that $\mathbf{u} \oplus \mathbf{0} = \mathbf{u}$, for every $\mathbf{u} \in V$ (called **the additive identity**).

(d) For every $\mathbf{u} \in V$ there is a unique element $-\mathbf{u} \in V$ such that $\mathbf{u} \oplus (-\mathbf{u}) = \mathbf{0}$ (called **the additive inverse**).

$\oplus$ is called VECTOR ADDITION.

2. SCALAR MULTIPLICATION: For each $\mathbf{u} \in V$ and $\alpha \in \mathbb{F}$, there corresponds a unique element $\alpha \odot \mathbf{u}$ in $V$ such that

   (a) $\alpha \cdot (\beta \odot \mathbf{u}) = (\alpha\beta) \odot \mathbf{u}$ for every $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u} \in V$.

   (b) $1 \odot \mathbf{u} = \mathbf{u}$ for every $\mathbf{u} \in V$, where $1 \in \mathbb{R}$.

3. DISTRIBUTIVE LAWS: RELATING VECTOR ADDITION WITH SCALAR MULTIPLICATION
   For any $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$, the following distributive laws hold:

   (a) $\alpha \odot (\mathbf{u} \oplus \mathbf{v}) = (\alpha \odot \mathbf{u}) \ \oplus \ (\alpha \odot \mathbf{v})$.

   (b) $(\alpha + \beta) \odot \mathbf{u} = (\alpha \odot \mathbf{u}) \ \oplus \ (\beta \odot \mathbf{u})$.

**Note:** the number $0$ is the element of $\mathbb{F}$ whereas $\mathbf{0}$ is the zero vector.

**Remark 3.1.2** *The elements of $\mathbb{F}$ are called* SCALARS, *and that of $V$ are called* VECTORS. *If $\mathbb{F} = \mathbb{R}$, the vector space is called a* REAL VECTOR SPACE. *If $\mathbb{F} = \mathbb{C}$, the vector space is called a* COMPLEX VECTOR SPACE.

*We may sometimes write $V$ for a vector space if $\mathbb{F}$ is understood from the context.*

Some interesting consequences of Definition 3.1.1 is the following useful result. Intuitively, these results seem to be obvious but for better understanding of the axioms it is desirable to go through the proof.

**Theorem 3.1.3** Let $V$ be a vector space over $\mathbb{F}$. Then

1. $\mathbf{u} \oplus \mathbf{v} = \mathbf{u}$ implies $\mathbf{v} = \mathbf{0}$.

2. $\alpha \odot \mathbf{u} = \mathbf{0}$ if and only if either $\mathbf{u}$ is the zero vector or $\alpha = 0$.

3. $(-1) \odot \mathbf{u} = -\mathbf{u}$ for every $\mathbf{u} \in V$.

PROOF. Proof of Part 1.
For $\mathbf{u} \in V$, by Axiom 1d there exists $-\mathbf{u} \in V$ such that $-\mathbf{u} \oplus \mathbf{u} = \mathbf{0}$.

Hence, $\mathbf{u} \oplus \mathbf{v} = \mathbf{u}$ is equivalent to

$$-\mathbf{u} \oplus (\mathbf{u} \oplus \mathbf{v}) = -\mathbf{u} \oplus \mathbf{u} \Longleftrightarrow (-\mathbf{u} \oplus \mathbf{u}) \oplus \mathbf{v} = \mathbf{0} \Longleftrightarrow \mathbf{0} \oplus \mathbf{v} = \mathbf{0} \Longleftrightarrow \mathbf{v} = \mathbf{0}.$$

Proof of Part 2.
As $\mathbf{0} = \mathbf{0} \oplus \mathbf{0}$, using the distributive law, we have

$$\alpha \odot \mathbf{0} = \alpha \odot (\mathbf{0} \oplus \mathbf{0}) = (\alpha \odot \mathbf{0}) \ \oplus \ (\alpha \odot \mathbf{0}).$$

Thus, for any $\alpha \in \mathbb{F}$, the first part implies $\alpha \odot \mathbf{0} = \mathbf{0}$. In the same way,

$$0 \odot \mathbf{u} = (0 + 0) \odot \mathbf{u} = (0 \odot \mathbf{u}) \ \oplus (0 \odot \mathbf{u}).$$

Hence, using the first part, one has $0 \odot \mathbf{u} = \mathbf{0}$ for any $\mathbf{u} \in V$.

Now suppose $\alpha \odot \mathbf{u} = \mathbf{0}$. If $\alpha = 0$ then the proof is over. Therefore, let us assume $\alpha \neq 0$ (note that $\alpha$ is a real or complex number, hence $\dfrac{1}{\alpha}$ exists and

$$\mathbf{0} = \frac{1}{\alpha} \odot \mathbf{0} = \frac{1}{\alpha} \odot (\alpha \odot \mathbf{u}) = (\frac{1}{\alpha}\ \alpha) \odot \mathbf{u} = 1 \odot \mathbf{u} = \mathbf{u}$$

as $1 \odot \mathbf{u} = \mathbf{u}$ for every vector $\mathbf{u} \in V$.

Thus we have shown that if $\alpha \neq 0$ and $\alpha \odot \mathbf{u} = \mathbf{0}$ then $\mathbf{u} = \mathbf{0}$.

Proof of Part 3.

We have $\mathbf{0} = \mathbf{0u} = (1 + (-1))\mathbf{u} = \mathbf{u} + (-1)\mathbf{u}$ and hence $(-1)\mathbf{u} = -\mathbf{u}$. □

### 3.1.2 Examples

**Example 3.1.4** 1. The set $\mathbb{R}$ of real numbers, with the usual addition and multiplication (*i.e.*, $\oplus \equiv +$ and $\odot \equiv \cdot$) forms a vector space over $\mathbb{R}$.

2. Consider the set $\mathbb{R}^2 = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}$. For $x_1, x_2, y_1, y_2 \in \mathbb{R}$ and $\alpha \in \mathbb{R}$, define,

$$(x_1, x_2) \oplus (y_1, y_2) = (x_1 + y_1, x_2 + y_2) \text{ and } \alpha \odot (x_1, x_2) = (\alpha x_1, \alpha x_2).$$

Then $\mathbb{R}^2$ is a real vector space.

3. Let $\mathbb{R}^n = \{(a_1, a_2, \ldots, a_n) : a_i \in \mathbb{R}, 1 \leq i \leq n\}$, be the set of $n$-tuples of real numbers. For $\mathbf{u} = (a_1, \ldots, a_n)$, $\mathbf{v} = (b_1, \ldots, b_n)$ in $V$ and $\alpha \in \mathbb{R}$, we define

$$\mathbf{u} \oplus \mathbf{v} = (a_1 + b_1, \ldots, a_n + b_n) \text{ and } \alpha \odot \mathbf{u} = (\alpha a_1, \ldots, \alpha a_n)$$

(called component wise or coordinate wise operations). Then $V$ is a real vector space with addition and scalar multiplication defined as above. This vector space is denoted by $\mathbb{R}^n$, called **the real vector space of $n$-tuples**.

4. Let $V = \mathbb{R}^+$ (the set of positive real numbers). This is NOT A VECTOR SPACE under usual operations of addition and scalar multiplication (why?). We now define a new vector addition and scalar multiplication as

$$\mathbf{v}_1 \oplus \mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{v}_2 \text{ and } \alpha \odot \mathbf{v} = \mathbf{v}^\alpha$$

for all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v} \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}$. Then $\mathbb{R}^+$ is a real vector space with $1$ as the additive identity.

5. Let $V = \mathbb{R}^2$. Define $(x_1, x_2) \oplus (y_1, y_2) = (x_1 + y_1 + 1, x_2 + y_2 - 3)$, $\alpha \odot (x_1, x_2) = (\alpha x_1 + \alpha - 1, \alpha x_2 - 3\alpha + 3)$ for $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$. Then it can be easily verified that the vector $(-1, 3)$ is the additive identity and $V$ is indeed a real vector space.

Recall $\sqrt{-1}$ is denoted $i$.

6. Consider the set $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$ of complex numbers.

(a) For $x_1 + iy_1, x_2 + iy_2 \in \mathbb{C}$ and $\alpha \in \mathbb{R}$, define,

$$\begin{aligned} (x_1 + iy_1) \oplus (x_2 + iy_2) &= (x_1 + x_2) + i(y_1 + y_2) \text{ and} \\ \alpha \odot (x_1 + iy_1) &= (\alpha x_1) + i(\alpha y_1). \end{aligned}$$

Then $\mathbb{C}$ is a real vector space.

(b) For $x_1 + iy_1, x_2 + iy_2 \in \mathbb{C}$ and $\alpha + i\beta \in \mathbb{C}$, define,

$$\begin{aligned} (x_1 + iy_1) \oplus (x_2 + iy_2) &= (x_1 + x_2) + i(y_1 + y_2) \text{ and} \\ (\alpha + i\beta) \odot (x_1 + iy_1) &= (\alpha x_1 - \beta y_1) + i(\alpha y_1 + \beta x_1). \end{aligned}$$

Then $\mathbb{C}$ forms a complex vector space.

7. Consider the set $\mathbb{C}^n = \{(z_1, z_2, \ldots, z_n) : z_i \in \mathbb{C} \text{ for } 1 \le i \le n\}$. For $(z_1, \ldots, z_n), (w_1, \ldots, w_n) \in \mathbb{C}^n$ and $\alpha \in \mathbb{F}$, define,

$$\begin{aligned} (z_1, \ldots, z_n) \oplus (w_1, \ldots, w_n) &= (z_1 + w_1, \ldots, z_n + w_n) \text{ and} \\ \alpha \odot (z_1, \ldots, z_n) &= (\alpha z_1, \ldots, \alpha z_n). \end{aligned}$$

   (a) If the set $\mathbb{F}$ is the set $\mathbb{C}$ of complex numbers, then $\mathbb{C}^n$ is a complex vector space having $n$-tuple of complex numbers as its vectors.

   (b) If the set $\mathbb{F}$ is the set $\mathbb{R}$ of real numbers, then $\mathbb{C}^n$ is a real vector space having $n$-tuple of complex numbers as its vectors.

**Remark 3.1.5** *In Example 7a, the scalars are Complex numbers and hence $i(1,0) = (i,0)$. Whereas, in Example 7b, the scalars are Real Numbers and hence* WE CANNOT WRITE $i(1,0) = (i,0)$.

8. Fix a positive integer $n$ and let $M_n(\mathbb{R})$ denote the set of all $n \times n$ matrices with real entries. Then $M_n(\mathbb{R})$ is a real vector space with vector addition and scalar multiplication defined by

$$A \oplus B = [a_{ij}] \oplus [b_{ij}] = [a_{ij} + b_{ij}], \qquad \alpha \odot A = \alpha \odot [a_{ij}] = [\alpha a_{ij}].$$

9. Fix a positive integer $n$. Consider the set, $\mathcal{P}_n(\mathbb{R})$, of all polynomials of degree $\le n$ with coefficients from $\mathbb{R}$ in the indeterminate $x$. Algebraically,

$$\mathcal{P}_n(\mathbb{R}) = \{a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n : a_i \in \mathbb{R}, 0 \le i \le n\}.$$

   Let $f(x), g(x) \in \mathcal{P}_n(\mathbb{R})$. Then $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ and $g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n$ for some $a_i, b_i \in \mathbb{R}$, $0 \le i \le n$. It can be verified that $\mathcal{P}_n(\mathbb{R})$ is a real vector space with the addition and scalar multiplication defined by:

$$\begin{aligned} f(x) \oplus g(x) &= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n, \text{ and} \\ \alpha \odot f(x) &= \alpha a_0 + \alpha a_1 x + \cdots + \alpha a_n x^n \text{ for } \alpha \in \mathbb{R}. \end{aligned}$$

10. Consider the set $\mathcal{P}(\mathbb{R})$, of all polynomials with real coefficients. Let $f(x), g(x) \in \mathcal{P}(\mathbb{R})$. Observe that a polynomial of the form $a_0 + a_1 x + \cdots + a_m x^m$ can be written as $a_0 + a_1 x + \cdots + a_m x^m + 0 \cdot x^{m+1} + \cdots + 0 \cdot x^p$ for any $p > m$. Hence, we can assume $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_p x^p$ and $g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_p x^p$ for some $a_i, b_i \in \mathbb{R}$, $0 \le i \le p$, for some large positive integer $p$. We now define the vector addition and scalar multiplication as

$$\begin{aligned} f(x) \oplus g(x) &= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_p + b_p)x^p, \text{ and} \\ \alpha \odot f(x) &= \alpha a_0 + \alpha a_1 x + \cdots + \alpha a_p x^p \text{ for } \alpha \in \mathbb{R}. \end{aligned}$$

   Then $\mathcal{P}(\mathbb{R})$ forms a real vector space.

11. Let $C([-1,1])$ be the set of all real valued continuous functions on the interval $[-1,1]$. For $f, g \in C([-1,1])$ and $\alpha \in \mathbb{R}$, define

$$\begin{aligned} (f \oplus g)(x) &= f(x) + g(x), \text{ and} \\ (\alpha \odot f)(x) &= \alpha f(x), \text{ for all } x \in [-1,1]. \end{aligned}$$

   Then $C([-1,1])$ forms a real vector space. The operations defined above are called POINT WISE ADDITION AND SCALAR MULTIPLICATION.

12. Let $V$ and $W$ be real vector spaces with binary operations $(+, \bullet)$ and $(\oplus, \odot)$, respectively. Consider the following operations on the set $V \times W$ : for $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \in V \times W$ and $\alpha \in \mathbb{R}$, define
$$\begin{aligned} (\mathbf{x}_1, \mathbf{y}_1) \oplus' (\mathbf{x}_2, \mathbf{y}_2) &= (\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}_1 \oplus \mathbf{y}_2), \text{ and} \\ \alpha \circ (\mathbf{x}_1, \mathbf{y}_1) &= (\alpha \bullet \mathbf{x}_1, \alpha \odot \mathbf{y}_1). \end{aligned}$$
On the right hand side, we write $\mathbf{x}_1 + \mathbf{x}_2$ to mean the addition in $V$, while $\mathbf{y}_1 \oplus \mathbf{y}_2$ is the addition in $W$. Similarly, $\alpha \bullet \mathbf{x}_1$ and $\alpha \odot \mathbf{y}_1$ come from scalar multiplication in $V$ and $W$, respectively. With the above definitions, $V \times W$ also forms a real vector space.

The readers are advised to justify the statements made in the above examples.

From now on, we will use '$\mathbf{u} + \mathbf{v}$' in place of '$\mathbf{u} \oplus \mathbf{v}$' and '$\alpha \cdot \mathbf{u}$ or $\alpha\mathbf{u}$' in place of '$\alpha \odot \mathbf{u}$'.

### 3.1.3 Subspaces

**Definition 3.1.6 (Vector Subspace)** Let $S$ be a NON-EMPTY SUBSET of $V$. $S(\mathbb{F})$ is said to be a subspace of $V(\mathbb{F})$ if $\alpha\mathbf{u} + \beta\mathbf{v} \in S$ whenever $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in S$; where the vector addition and scalar multiplication are the same as that of $V(\mathbb{F})$.

**Remark 3.1.7** *Any subspace is a vector space in its own right with respect to the vector addition and scalar multiplication that is defined for $V(\mathbb{F})$.*

**Example 3.1.8** 1. Let $V(\mathbb{F})$ be a vector space. Then

(a) $S = \{\mathbf{0}\}$, the set consisting of the zero vector $\mathbf{0}$,

(b) $S = V$

are vector subspaces of $V$. These are called **trivial subspaces**.

2. Let $S = \{(x, y, z) \in \mathbb{R}^3 : x + y - z = 0\}$. Then $S$ is a subspace of $\mathbb{R}^3$. ($S$ is a plane in $\mathbb{R}^3$ passing through the origin.)

3. Let $S = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 3\}$. Then $S$ is not a subspace of $\mathbb{R}^3$. ($S$ is again a plane in $\mathbb{R}^3$ but it doesn't pass through the origin.)

4. Let $S = \{(x, y, z) \in \mathbb{R}^3 : z = x\}$. Then $S$ is a subspace of $\mathbb{R}^3$.

5. The vector space $\mathcal{P}_n(\mathbb{R})$ is a subspace of the vector space $\mathcal{P}(\mathbb{R})$.

**Exercise 3.1.9** 1. Which of the following are correct statements?

(a) Let $S = \{(x, y, z) \in \mathbb{R}^3 : z = x^2\}$. Then $S$ is a subspace of $\mathbb{R}^3$.

(b) Let $V(\mathbb{F})$ be a vector space. Let $\mathbf{x} \in V$. Then the set $\{\alpha\mathbf{x} : \alpha \in \mathbb{F}\}$ forms a vector subspace of $V$.

(c) Let $W = \{f \in C([-1, 1]) : f(1/2) = 0\}$. Then $W$ is a subspace of the real vector space, $C([-1, 1])$.

2. Which of the following are subspaces of $\mathbb{R}^n(\mathbb{R})$?

(a) $\{(x_1, x_2, \ldots, x_n) : x_1 \geq 0\}$.

(b) $\{(x_1, x_2, \ldots, x_n) : x_1 + 2x_2 = 4x_3\}$.

(c) $\{(x_1, x_2, \ldots, x_n) : x_1$is rational $\}$.

(d) $\{(x_1, x_2, \ldots, x_n) : x_1 = x_3^2\}$.

(e) $\{(x_1, x_2, \ldots, x_n) :$ either $x_1$ or $x_2$ or both is$0\}$.

(f) $\{(x_1, x_2, \ldots, x_n) : |x_1| \leq 1\}$.

3. Which of the following are subspaces of $i)\mathbb{C}^n(\mathbb{R})$ $ii)\mathbb{C}^n(\mathbb{C})$?

(a) $\{(z_1, z_2, \ldots, z_n) : z_1$is real $\}$.

(b) $\{(z_1, z_2, \ldots, z_n) : z_1 + z_2 = \overline{z_3}\}$.

(c) $\{(z_1, z_2, \ldots, z_n) :\mid z_1 \mid=\mid z_2 \mid\}$.

### 3.1.4 Linear Combinations

**Definition 3.1.10 (Linear Span)** Let $V(\mathbb{F})$ be a vector space and let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a non-empty subset of $V$. The linear span of $S$ is the set defined by

$$L(S) \quad = \quad \{\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_n \mathbf{u}_n : \alpha_i \in \mathbb{F}, 1 \leq i \leq n\}$$

If $S$ is an empty set we define $L(S) = \{\mathbf{0}\}$.

**Example 3.1.11** 1. Note that $(4,5,5)$ is a linear combination of $(1,0,0)$, $(1,1,0)$, and $(1,1,1)$ as $(4,5,5) = 5(1,1,1) - 1(1,0,0) + 0(1,1,0)$.

For each vector, the LINEAR COMBINATION IN TERMS OF THE VECTORS $(1,0,0), (1,1,0)$, AND $(1,1,1)$ IS UNIQUE.

2. Is $(4,5,5)$ a linear combination of $(1,2,3), (-1,1,4)$ and $(3,3,2)$?
**Solution:** We want to find $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ such that

$$\alpha_1(1,2,3) + \alpha_2(-1,1,4) + \alpha_3(3,3,2) = (4,5,5). \tag{3.1.1}$$

Check that $3(1,2,3)+(-1)(-1,1,4)+0(3,3,2) = (4,5,5)$. Also, in this case, the vector $(4,5,5)$ DOES NOT HAVE A UNIQUE EXPRESSION AS LINEAR COMBINATION OF VECTORS $(1,2,3), (-1,1,4)$ AND $(3,3,2)$.

3. Verify that $(4,5,5)$ is not a linear combination of the vectors $(1,2,1)$ and $(1,1,0)$?

4. The linear span of $S = \{(1,1,1), (2,1,3)\}$ over $\mathbb{R}$ is

$$\begin{aligned} L(S) &= \{\alpha(1,1,1) + \beta(2,1,3) : \alpha, \beta \in \mathbb{R}\} \\ &= \{(\alpha + 2\beta, \alpha + \beta, \alpha + 3\beta) : \alpha, \beta \in \mathbb{R}\} \\ &= \{(x,y,z) \in \mathbb{R}^3 : 2x - y = z\}. \end{aligned}$$

as $2(\alpha + 2\beta) - (\alpha + \beta) = \alpha + 3\beta$, and if $z = 2x - y$, take $\alpha = 2y - x$ and $\beta = x - y$.

**Lemma 3.1.12 (Linear Span is a subspace)** Let $V(\mathbb{F})$ be a vector space and let $S$ be a non-empty subset of $V$. Then $L(S)$ is a subspace of $V(\mathbb{F})$.

PROOF. By definition, $S \subset L(S)$ and hence $L(S)$ is non-empty subset of $V$. Let $\mathbf{u}, \mathbf{v} \in L(S)$. Then, for $1 \leq i \leq n$ there exist vectors $\mathbf{w}_i \in S$, and scalars $\alpha_i, \beta_i \in \mathbb{F}$ such that $\mathbf{u} = \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2 + \cdots + \alpha_n\mathbf{w}_n$ and $\mathbf{v} = \beta_1\mathbf{w}_1 + \beta_2\mathbf{w}_2 + \cdots + \beta_n\mathbf{w}_n$. Hence,

$$\mathbf{u} + \mathbf{v} = (\alpha_1 + \beta)\mathbf{w}_1 + \cdots + (\alpha_n + \beta_n)\mathbf{w}_n \in L(S).$$

Thus, $L(S)$ is a vector subspace of $V(\mathbb{F})$. □

**Remark 3.1.13** *Let $V(\mathbb{F})$ be a vector space and $W \subset V$ be a subspace. If $S \subset W$, then $L(S) \subset W$ is a subspace of $W$ as $W$ is a vector space in its own right.*

**Theorem 3.1.14** Let $S$ be a non-empty subset of a vector space $V$. Then $L(S)$ is the smallest subspace of $V$ containing $S$.

PROOF. For every $\mathbf{u} \in S$, $\mathbf{u} = 1.\mathbf{u} \in L(S)$ and therefore, $S \subseteq L(S)$. To show $L(S)$ is the smallest subspace of $V$ containing $S$, consider any subspace $W$ of $V$ containing $S$. Then by Proposition 3.1.13, $L(S) \subseteq W$ and hence the result follows. □

**Definition 3.1.15** Let $A$ be an $m \times n$ matrix with real entries. Then using the rows $\mathbf{a}_1^t, \mathbf{a}_2^t, \ldots, \mathbf{a}_m^t \in \mathbb{R}^n$ and columns $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n \in \mathbb{R}^m$, we define

1. $\mathbf{R}owSpace(A) = L(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$,
2. $\mathbf{C}olumnSpace(A) = L(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)$,
3. $\mathbf{N}ullSpace(A)$, denoted $\mathcal{N}(A)$ as $\{\mathbf{x}^t \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$.
4. $\mathbf{R}ange(A)$, denoted Im $(A) = \{\mathbf{y} : A\mathbf{x} = \mathbf{y}$ for some $\mathbf{x}^t \in \mathbb{R}^n\}$.

Note that the "column space" of a matrix $A$ consists of all $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ has a solution. Hence, $\mathbf{C}olumnSpace(A) = \mathbf{R}ange(A)$.

**Lemma 3.1.16** Let $A$ be a real $m \times n$ matrix. Suppose $B = EA$ for some elementary matrix $E$. Then Row Space$(A) =$ Row Space$(B)$.

PROOF. We prove the result for the elementary matrix $E_{ij}(c)$, where $c \neq 0$ and $i < j$. Let $\mathbf{a}_1^t, \mathbf{a}_2^t, \ldots, \mathbf{a}_m^t$ be the rows of the matrix $A$. Then $B = E_{ij}(c)A$ gives us

$$\begin{aligned}
\text{Row Space}(B) &= L(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_i + c\mathbf{a}_j, \ldots, \mathbf{a}_m) \\
&= \{\alpha_1\mathbf{a}_1 + \cdots + \alpha_{i-1}\mathbf{a}_{i-1} + \alpha_i(\mathbf{a}_i + c\mathbf{a}_j) + \cdots \\
&\qquad\qquad + \alpha_m\mathbf{a}_m : \ \alpha_\ell \in \mathbb{R}, 1 \le \ell \le m\} \\
&= \left\{\sum_{\ell=1}^{m} \alpha_\ell\mathbf{a}_\ell + \alpha_i \cdot c\mathbf{a}_j : \alpha_\ell \in \mathbb{R}, 1 \le \ell \le m\right\} \\
&= \left\{\sum_{\ell=1}^{m} \beta_\ell\mathbf{a}_\ell : \beta_\ell \in \mathbb{R}, 1 \le \ell \le m\right\} \\
&= L(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_i, \ldots, \mathbf{a}_m) \\
&= \text{Row Space}(A)
\end{aligned}$$

□

**Theorem 3.1.17** Let $A$ be an $m \times n$ matrix with real entries. Then

1. $\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$;
2. the non-zero row vectors of a matrix in row-reduced form, forms a basis for the row-space. Hence $\dim($ Row Space$(A)) =$ row rank of $(A)$.

PROOF. Part 1) can be easily proved. Let $A$ be an $m \times n$ matrix. For part 2), let $D$ be the row-reduced form of $A$ with non-zero rows $\mathbf{d}_1^t, \mathbf{d}_2^t, \ldots, \mathbf{d}_r^t$. Then $B = E_k E_{k-1} \cdots E_2 E_1 A$ for some elementary matrices $E_1, E_2, \ldots, E_k$. Then, a repeated application of Lemma 3.1.16 implies Row Space$(A) =$ Row Space$(B)$. That is, if the rows of the matrix $A$ are $\mathbf{a}_1^t, \mathbf{a}_2^t, \ldots, \mathbf{a}_m^t$, then

$$L(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m) = L(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r).$$

Hence the required result follows. □

**Exercise 3.1.18** 1. Show that any two row-equivalent matrices have the same row space. Give examples to show that the column space of two row-equivalent matrices need not be same.

2. Find all the vector subspaces of $\mathbb{R}^2$.

3. Let $P$ and $Q$ be two subspaces of a vector space $V$. Show that $P \cap Q$ is a subspace of $V$. Also show that $P \cup Q$ need not be a subspace of $V$. When is $P \cup Q$ a subspace of $V$?

4. Let $P$ and $Q$ be two subspaces of a vector space $V$. Define $P + Q = \{\mathbf{u} + \mathbf{v} : \mathbf{u} \in P, \mathbf{v} \in Q\}$. Show that $P + Q$ is a subspace of $V$. Also show that $L(P \cup Q) = P + Q$.

5. Let $S = \{x_1, x_2, x_3, x_4\}$ where $x_1 = (1,0,0,0)$, $x_2 = (1,1,0,0)$, $x_3 = (1,2,0,0)$, $x_4 = (1,1,1,0)$. Determine all $x_i$ such that $L(S) = L(S \setminus \{x_i\})$.

6. Let $C([-1,1])$ be the set of all continuous functions on the interval $[-1,1]$ (cf. Example 3.1.4.11). Let

$$\begin{aligned} W_1 &= \{f \in C([-1,1]) : f(0.2) = 0\}, \text{ and} \\ W_2 &= \{f \in C([-1,1]) : f'(\frac{1}{4}) \text{exists } \}. \end{aligned}$$

Are $W_1, W_2$ subspaces of $C([-1,1])$?

7. Let $V = \{(x,y) : x, y \in \mathbb{R}\}$ over $\mathbb{R}$. Define $(x,y) \oplus (x_1,y_1) = (x + x_1, 0)$ and $\alpha \odot (x,y) = (\alpha x, 0)$. Show that $V$ is not a vector space over $\mathbb{R}$.

8. Recall that $M_n(\mathbb{R})$ is the real vector space of all $n \times n$ real matrices. Prove that the following subsets are subspaces of $M_n(\mathbb{R})$.

   (a) $\mathsf{sl}_n = \{A \in M_n(\mathbb{R}) \ : \ \mathsf{trace}(A) = 0\}$

   (b) $\mathsf{Sym}_n = \{A \in M_n(\mathbb{R}) \ : \ A = A^t\}$

   (c) $\mathsf{Skew}_n = \{A \in M_n(\mathbb{R}) \ : \ A + A^t = \mathbf{0}\}$

9. Let $V = \mathbb{R}$. Define $x \oplus y = x - y$ and $\alpha \odot x = -\alpha x$. Which vector space axioms are not satisfied here?

In this section, we saw that a vector space has infinite number of vectors. Hence, one can start with any finite collection of vectors and obtain their span. It means that any vector space contains infinite number of other vector subspaces. Therefore, the following questions arise:

1. What are the conditions under which, the linear span of two distinct sets the same?

2. Is it possible to find/choose vectors so that the linear span of the chosen vectors is the whole vector space itself?

3. Suppose we are able to choose certain vectors whose linear span is the whole space. Can we find the minimum number of such vectors?

We try to answer these questions in the subsequent sections.

## 3.2 Linear Independence

**Definition 3.2.1 (Linear Independence and Dependence)** Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be any non-empty subset of $V$. If there exist some non-zero $\alpha_i$'s $1 \leq i \leq m$, such that

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0},$$

then the set $S$ is called a linearly dependent set. Otherwise, the set $S$ is called linearly independent.

**Example 3.2.2** 1. Let $S = \{(1,2,1),(2,1,4),(3,3,5)\}$. Then check that $1(1,2,1)+1(2,1,4)+(-1)(3,3,5) = (0,0,0)$. Since $\alpha_1 = 1, \alpha_2 = 1$ and $\alpha_3 = -1$ is a solution of (3.2.1), so the set $S$ is a linearly dependent subset of $\mathbb{R}^3$.

2. Let $S = \{(1,1,1),(1,1,0),(1,0,1)\}$. Suppose there exists $\alpha, \beta, \gamma \in \mathbb{R}$ such that $\alpha(1,1,1)+\beta(1,1,0)+\gamma(1,0,1) = (0,0,0)$. Then check that in this case we necessarily have $\alpha = \beta = \gamma = 0$ which shows that the set $S = \{(1,1,1),(1,1,0),(1,0,1)\}$ is a linearly independent subset of $\mathbb{R}^3$.

In other words, if $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is a non-empty subset of a vector space $V$, then to check whether the set $S$ is linearly dependent or independent, one needs to consider the equation

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0}. \qquad (3.2.1)$$

In case $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ is THE ONLY SOLUTION of (3.2.1), the set $S$ becomes a linearly independent subset of $V$. Otherwise, the set $S$ becomes a linearly dependent subset of $V$.

**Proposition 3.2.3** Let $V$ be a vector space.

1. Then the zero-vector cannot belong to a linearly independent set.

2. If $S$ is a linearly independent subset of $V$, then every subset of $S$ is also linearly independent.

3. If $S$ is a linearly dependent subset of $V$ then every set containing $S$ is also linearly dependent.

PROOF. We give the proof of the first part. The reader is required to supply the proof of other parts. Let $S = \{\mathbf{0} = \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a set consisting of the zero vector. Then for any $\gamma \neq o$, $\gamma \mathbf{u}_1 + o\mathbf{u}_2 + \cdots + 0\mathbf{u}_n = \mathbf{0}$. Hence, for the system $\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0}$, we have a non-zero solution $\alpha_1 = \gamma$ and $o = \alpha_2 = \cdots = \alpha_n$. Therefore, the set $S$ is linearly dependent. □

**Theorem 3.2.4** Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a linearly independent subset of a vector space $V$. Suppose there exists a vector $\mathbf{v}_{p+1} \in V$, such that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{v}_{p+1}\}$ is linearly dependent, then $\mathbf{v}_{p+1}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$.

PROOF. Since the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{v}_{p+1}\}$ is linearly dependent, there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_{p+1}$, NOT ALL ZERO such that

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p + \alpha_{p+1} \mathbf{v}_{p+1} = \mathbf{0}. \qquad (3.2.2)$$

CLAIM: $\alpha_{p+1} \neq 0$.

Let if possible $\alpha_{p+1} = 0$. Then equation (3.2.2) gives $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p = \mathbf{0}$ with not all $\alpha_i$, $1 \leq i \leq p$ zero. Hence, by the definition of linear independence, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is linearly dependent which is contradictory to our hypothesis. Thus, $\alpha_{p+1} \neq 0$ and we get

$$\mathbf{v}_{p+1} = -\frac{1}{\alpha_{p+1}}(\alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p).$$

Note that $\alpha_i \in \mathbb{F}$ for every $i, \; 1 \leq i \leq p+1$ and hence $-\frac{\alpha_i}{\alpha_{p+1}} \in \mathbb{F}$ for $1 \leq i \leq p$. Hence the result follows. □

We now state two important corollaries of the above theorem. We don't give their proofs as they are easy consequence of the above theorem.

**Corollary 3.2.5** Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a linearly dependent subset of a vector space $V$. Then there exists a smallest $k, \; 2 \leq k \leq n$ such that

$$L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}).$$

The next corollary follows immediately from Theorem 3.2.4 and Corollary 3.2.5.

**Corollary 3.2.6** Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a linearly independent subset of a vector space $V$. Suppose there exists a vector $\mathbf{v} \in V$, such that $\mathbf{v} \not\in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$. Then the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{v}\}$ is also linearly independent subset of $V$.

**Exercise 3.2.7**

1. Consider the vector space $\mathbb{R}^2$. Let $\mathbf{u}_1 = (1, 0)$. Find all choices for the vector $\mathbf{u}_2$ such that the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linear independent subset of $\mathbb{R}^2$. Does there exist choices for vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ such that the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly independent subset of $\mathbb{R}^2$?
2. If none of the elements appearing along the principal diagonal of a lower triangular matrix is zero, show that the row vectors are linearly independent in $\mathbb{R}^n$. The same is true for column vectors.
3. Let $S = \{(1,1,1,1), (1,-1,1,2), (1,1,-1,1)\} \subset \mathbb{R}^4$. Determine whether or not the vector $(1,1,2,1) \in L(S)$?
4. Show that $S = \{(1,2,3), (-2,1,1), (8,6,10)\}$ is linearly dependent in $\mathbb{R}^3$.
5. Show that $S = \{(1,0,0), (1,1,0), (1,1,1)\}$ is a linearly independent set in $\mathbb{R}^3$. In general if $\{f_1, f_2, f_3\}$ is a linearly independent set then $\{f_1, f_1 + f_2, f_1 + f_2 + f_3\}$ is also a linearly independent set.
6. In $\mathbb{R}^3$, give an example of 3 vectors $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ such that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent but any set of 2 vectors from $\mathbf{u}, \mathbf{v}, \mathbf{w}$ is linearly independent.
7. What is the maximum number of linearly independent vectors in $\mathbb{R}^3$?
8. Show that any set of $k$ vectors in $\mathbb{R}^3$ is linearly dependent if $k \geq 4$.
9. Is the set of vectors $(1,0), (\; i, 0)$ linearly independent subset of $\mathbb{C}^2 \; (\mathbb{R})$?
10. Under what conditions on $\alpha$ are the vectors $(1+\alpha, 1-\alpha)$ and $(\alpha - 1, 1 + \alpha)$ in $\mathbb{C}^2(\mathbb{R})$ linearly independent?
11. Let $\mathbf{u}, \mathbf{v} \in V$ and $M$ be a subspace of $V$. Further, let $K$ be the subspace spanned by $M$ and $\mathbf{u}$ and $H$ be the subspace spanned by $M$ and $\mathbf{v}$. Show that if $\mathbf{v} \in K$ and $\mathbf{v} \not\in M$ then $\mathbf{u} \in H$.

## 3.3 Bases

**Definition 3.3.1 (Basis of a Vector Space)**

1. A non-empty subset $\mathcal{B}$ of a vector space $V$ is called a basis of $V$ if
   (a) $\mathcal{B}$ is a linearly independent set, and
   (b) $L(\mathcal{B}) = V$, i.e., every vector in $V$ can be expressed as a linear combination of the elements of $\mathcal{B}$.

2. A vector in $\mathcal{B}$ is called a basis vector.

**Remark 3.3.2** *Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a basis of a vector space $V(\mathbb{F})$. Then any $\mathbf{v} \in V$ is a* UNIQUE LINEAR COMBINATION *of the basis vectors, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$.*

*Observe that if there exists a $\mathbf{v} \in W$ such that $\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_p\mathbf{v}_p$ and $\mathbf{v} = \beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \cdots + \beta_p\mathbf{v}_p$ then*

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = (\alpha_1 - \beta_1)\mathbf{v}_1 + (\alpha_2 - \beta_2)\mathbf{v}_2 + \cdots + (\alpha_p - \beta_p)\mathbf{v}_p.$$

*But then the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is linearly independent and therefore the scalars $\alpha_i - \beta_i$ for $1 \le i \le p$ must all be equal to zero. Hence, for $1 \le i \le p$, $\alpha_i = \beta_i$ and we have the uniqueness.*

By convention, the linear span of an empty set is $\{\mathbf{0}\}$. Hence, the empty set is a basis of the vector space $\{\mathbf{0}\}$.

**Example 3.3.3**

1. Check that if $V = \{(x, y, 0) : x, y \in \mathbb{R}\} \subset \mathbb{R}^3$, then $\mathcal{B} = \{(1,0,0), (0,1,0)\}$ or $\mathcal{B} = \{(1,0,0), (1,1,0)\}$ or $\mathcal{B} = \{(2,0,0), (1,3,0)\}$ or $\cdots$ are bases of $V$.

2. For $1 \le i \le n$, let $\mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{i \text{ th place}}, 0, \ldots, 0) \in \mathbb{R}^n$. Then, the set $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ forms a basis of $\mathbb{R}^n$. This set is called the **standard basis** of $\mathbb{R}^n$.

   That is, if $n = 3$, then the set $\{(1,0,0), (0,1,0), (0,0,1)\}$ forms an standard basis of $\mathbb{R}^3$.

3. Let $V = \{(x, y, z) : x+y-z = 0,\ x, y, z \in \mathbb{R}\}$ be a vector subspace of $\mathbb{R}^3$. Then $S = \{(1,1,2), (2,1,3), (1,2,3)\} \subset V$. It can be easily verified that the vector $(3,2,5) \in V$ and

   $$(3,2,5) = (1,1,2) + (2,1,3) = 4(1,1,2) - (1,2,3).$$

   Then by Remark 3.3.2, $S$ cannot be a basis of $V$.

   A basis of $V$ can be obtained by the following method:
   The condition $x + y - z = 0$ is equivalent to $z = x + y$. we replace the value of $z$ with $x + y$ to get

   $$(x, y, z) = (x, y, x+y) = (x, 0, x) + (0, y, y) = x(1,0,1) + y(0,1,1).$$

   Hence, $\{(1,0,1), (0,1,1)\}$ forms a basis of $V$.

4. Let $V = \{a + ib : a, b \in \mathbb{R}\}$ and $\mathbb{F} = \mathbb{C}$. That is, $V$ is a complex vector space. Note that any element $a + ib \in V$ can be written as $a + ib = (a + ib)\mathbf{1}$. Hence, a basis of $V$ is $\{\mathbf{1}\}$.

5. Let $V = \{a + ib : a, b \in \mathbb{R}\}$ and $\mathbb{F} = \mathbb{R}$. That is, $V$ is a real vector space. Any element $a + ib \in V$ is expressible as $a \cdot \mathbf{1} + b \cdot \mathbf{i}$. Hence a basis of $V$ is $\{\mathbf{1}, \mathbf{i}\}$.

   Observe that $i$ is a vector in $\mathbb{C}$. Also, $i \notin \mathbb{R}$ and hence $i \cdot (1 + 0 \cdot i)$ is not defined.

6. Recall the vector space $\mathcal{P}(\mathbb{R})$, the vector space of all polynomials with real coefficients. A basis of this vector space is the set

   $$\{1, x, x^2, \ldots, x^n, \ldots\}.$$

   This basis has infinite number of vectors as the degree of the polynomial can be any positive integer.

**Definition 3.3.4 (Finite Dimensional Vector Space)** A vector space $V$ is said to be finite dimensional if there exists a basis consisting of finite number of elements. Otherwise, the vector space $V$ is called infinite dimensional.

In Example 3.3.3, the vector space of all polynomials is an example of an infinite dimensional vector space. All the other vector spaces are finite dimensional.

**Remark 3.3.5** *We can use the above results to obtain a basis of any finite dimensional vector space $V$ as follows:*

**Step 1:** *Choose a non-zero vector, say, $\mathbf{v}_1 \in V$. Then the set $\{\mathbf{v}_1\}$ is linearly independent.*

**Step 2:** *If $V = L(\mathbf{v}_1)$, we have got a basis of $V$. Else there exists a vector, say, $\mathbf{v}_2 \in V$ such that $\mathbf{v}_2 \notin L(\mathbf{v}_1)$. Then by Corollary 3.2.6, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.*

**Step 3:** *If $V = L(\mathbf{v}_1, \mathbf{v}_2)$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of $V$. Else there exists a vector, say, $\mathbf{v}_3 \in V$ such that $\mathbf{v}_3 \notin L(\mathbf{v}_1, \mathbf{v}_2)$. So, by Corollary 3.2.6, the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.*

*At the $i^{th}$ step, either $V = L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i)$, or $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) \neq V$.*

*In the first case, we have $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$ as a basis of $V$.*

*In the second case, $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) \subset V$. So, we choose a vector, say, $\mathbf{v}_{i+1} \in V$ such that $\mathbf{v}_{i+1} \notin L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i)$. Therefore, by Corollary 3.2.6, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i+1}\}$ is linearly independent.*

*This process will finally end as $V$ is a finite dimensional vector space.*

**Exercise 3.3.6**

1. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a subset of a vector space $V(\mathbb{F})$. Suppose $L(S) = V$ but $S$ is not a linearly independent set. Then prove that each vector in $V$ can be expressed in more than one way as a linear combination of vectors from $S$.

2. Show that the set $\{(1,0,1), (1,i,0), (1,1,1-i)\}$ is a basis of $\mathbb{C}^3(\mathbb{C})$.

3. Let $A$ be a matrix of rank $r$. Then show that the $r$ non-zero rows in the row-reduced echelon form of $A$ are linearly independent and they form a basis of the row space of $A$.

### 3.3.1 Important Results

**Theorem 3.3.7** Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of a given vector space $V$. If $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is a set of vectors from $V$ with $m > n$ then this set is linearly dependent.

PROOF. Since we want to find whether the set $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is linearly independent or not, we consider the linear system

$$\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2 + \cdots + \alpha_m\mathbf{w}_m = \mathbf{0} \tag{3.3.1}$$

with $\alpha_1, \alpha_2, \ldots, \alpha_m$ as the $m$ unknowns. If the solution set of this linear system of equations has more than one solution, then this set will be linearly dependent.

As $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of $V$ and $\mathbf{w}_i \in V$, for each $i$, $1 \leq i \leq m$, there exist scalars $a_{ij}$, $1 \leq i \leq n$, $1 \leq j \leq m$, such that

$$\begin{aligned}
\mathbf{w}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \cdots + a_{n1}\mathbf{v}_n \\
\mathbf{w}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \cdots + a_{n2}\mathbf{v}_n \\
\vdots &= \vdots \\
\mathbf{w}_m &= a_{1m}\mathbf{v}_1 + a_{2m}\mathbf{v}_2 + \cdots + a_{nm}\mathbf{v}_n.
\end{aligned}$$

The set of equations (3.3.1) can be rewritten as

$$\alpha_1\left(\sum_{j=1}^{n} a_{j1}\mathbf{v}_j\right) + \alpha_2\left(\sum_{j=1}^{n} a_{j2}\mathbf{v}_j\right) + \cdots + \alpha_m\left(\sum_{j=1}^{n} a_{jm}\mathbf{v}_j\right) = \mathbf{0}$$

$$\text{i.e.,} \quad \left(\sum_{i=1}^{m} \alpha_i a_{1i}\right)\mathbf{v}_1 + \left(\sum_{i=1}^{m} \alpha_i a_{2i}\right)\mathbf{v}_2 + \cdots + \left(\sum_{i=1}^{m} \alpha_i a_{ni}\right)\mathbf{v}_n = \mathbf{0}.$$

Since the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent, we have

$$\sum_{i=1}^{m} \alpha_i a_{1i} = \sum_{i=1}^{m} \alpha_i a_{2i} = \cdots = \sum_{i=1}^{m} \alpha_i a_{ni} = 0.$$

Therefore, finding $\alpha_i$'s satisfying equation (3.3.1) reduces to solving the system of homogeneous equations $A\alpha = \mathbf{0}$ where $\alpha^t = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$. Since $n < m$, i.e., THE NUMBER OF EQUATIONS is strictly less than THE NUMBER OF UNKNOWNS, Corollary 2.6.3 implies that the solution set consists of infinite number of elements. Therefore, the equation (3.3.1) has a solution with not all $\alpha_i$, $1 \le i \le m$, zero. Hence, the set $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is a linearly dependent set. □

**Remark 3.3.8** *Let $V$ be a vector subspace of $\mathbb{R}^n$ with spanning set $S$. We give a method of finding a basis of $V$ from $S$.*

1. *Construct a matrix $A$ whose rows are the vectors in $S$.*
2. *Use only the elementary row operations $R_i(c)$ and $R_{ij}(c)$ to get the row-reduced form $B$ of $A$ (in fact we just need to make as many zero-rows as possible).*
3. *Let $\mathcal{B}$ be the set of vectors in $S$ corresponding to the non-zero rows of $B$.*

*Then the set $\mathcal{B}$ is a basis of $L(S) = V$.*

**Example 3.3.9** Let $S = \{(1,1,1,1), (1,1,-1,1), (1,1,0,1), (1,-1,1,1)\}$ be a subset of $\mathbb{R}^4$. Find a basis of $L(S)$.

**Solution:** Here $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$. Applying row-reduction to $A$, we have

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix} \xrightarrow[R_{12}(-1), R_{13}(-1), R_{14}(-1)]{} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & -2 & 0 & 0 \end{bmatrix} \xrightarrow[R_{32(-2)}]{} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & -2 & 0 & 0 \end{bmatrix}.$$

Observe that the rows $1, 3$ and $4$ are non-zero. Hence, a basis of $L(S)$ consists of the first, third and fourth vectors of the set $S$. Thus, $\mathcal{B} = \{(1,1,1,1),\ (1,1,0,1),\ (1,-1,1,1)\}$ is a basis of $L(S)$.

Observe that at the last step, in place of the elementary row operation $R_{32}(-2)$, we can apply $R_{23}(-\frac{1}{2})$ to make the third row as the zero-row. In this case, we get $\{(1,1,1,1),\ (1,1,-1,1),\ (1,-1,1,1)\}$ as a basis of $L(S)$.

**Corollary 3.3.10** Let $V$ be a finite dimensional vector space. Then any two bases of $V$ have the same number of vectors.

PROOF. Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be two bases of $V$ with $m > n$. Then by the above theorem the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is linearly dependent if we take $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ as the basis of $V$. This contradicts the assumption that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is also a basis of $V$. Hence, we get $m = n$. □

**Definition 3.3.11 (Dimension of a Vector Space)** The dimension of a finite dimensional vector space $V$ is the number of vectors in a basis of $V$, denoted $\dim(V)$.

Note that the Corollary 3.2.6 can be used to generate a basis of ANY NON-TRIVIAL FINITE DIMENSIONAL VECTOR SPACE.

**Example 3.3.12** 1. Consider the complex vector space $\mathbb{C}^2(\mathbb{C})$. Then,

$$(a+ib, c+id) = (a+ib)(1,0) + (c+id)(0,1).$$

So, $\{(1,0),(0,1)\}$ is a basis of $\mathbb{C}^2(\mathbb{C})$ and thus $\dim(V) = 2$.

2. Consider the real vector space $\mathbb{C}^2(\mathbb{R})$. In this case, any vector

$$(a+ib, c+id) = a(1,0) + b(i,0) + c(0,1) + d(0,i).$$

Hence, the set $\{(1,0),(i,0),(0,1),(0,i)\}$ is a basis and $\dim(V) = 4$.

**Remark 3.3.13** *It is important to note that the dimension of a vector space may change if the underlying field (the set of scalars) is changed.*

**Example 3.3.14** Let $V$ be the set of all functions $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ with the property that $f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ and $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$. For $f, g \in V$, and $t \in \mathbb{R}$, define

$$\begin{aligned}(f \oplus g)(\mathbf{x}) &= f(\mathbf{x}) + g(\mathbf{x}) \quad \text{and} \\ (t \odot f)(\mathbf{x}) &= f(t\mathbf{x}).\end{aligned}$$

Then $V$ is a real vector space.

For $1 \le i \le n$, consider the functions

$$\mathbf{e}_i(\mathbf{x}) = \mathbf{e}_i\big((x_1, x_2, \ldots, x_n)\big) = x_i.$$

Then it can be easily verified that the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis of $V$ and hence $\dim(V) = n$.

The next theorem follows directly from Corollary 3.2.6 and Theorem 3.3.7. Hence, the proof is omitted.

**Theorem 3.3.15** Let $S$ be a linearly independent subset of a finite dimensional vector space $V$. Then the set $S$ can be extended to form a basis of $V$.

Theorem 3.3.15 is equivalent to the following statement:
Let $V$ be a vector space of dimension $n$. Suppose, we have found a linearly independent set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} \subset V$. Then there exist vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ in $V$ such that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of $V$.

**Corollary 3.3.16** Let $V$ be a vector space of dimension $n$. Then any set of $n$ linearly independent vectors forms a basis of $V$. Also, every set of $m$ vectors, $m > n$, is linearly dependent.

**Example 3.3.17** Let $V = \{(v,w,x,y,z) \in \mathbb{R}^5 \ : v + x - 3y + z = 0\}$ and $W = \{(v,w,x,y,z) \in \mathbb{R}^5 \ : w - x - z = 0, v = y\}$ be two subspaces of $\mathbb{R}^5$. Find bases of $V$ and $W$ containing a basis of $V \cap W$.
**Solution:** Let us find a basis of $V \cap W$. The solution set of the linear equations

$$v + x - 3y + z = 0, \ \ w - x - z = 0 \ \ \text{and} \ \ v = y$$

is given by

$$(v,w,x,y,z)^t = (y, 2y, x, y, 2y - x)^t = y(1,2,0,1,2)^t + x(0,0,1,0,-1)^t.$$

Thus, a basis of $V \cap W$ is

$$\{(1,2,0,1,2), (0,0,1,0,-1)\}.$$

To find a basis of $W$ containing a basis of $V \cap W$, we can proceed as follows:

1. Find a basis of $W$.

2. Take the basis of $V \cap W$ found above as the first two vectors and that of $W$ as the next set of vectors. Now use Remark 3.3.8 to get the required basis.

Heuristically, we can also find the basis in the following way:
A vector of $W$ has the form $(y, x+z, x, y, z)$ for $x, y, z \in \mathbb{R}$. Substituting $y = 1, x = 1$, and $z = 0$ in $(y, x+z, x, y, z)$ gives us the vector $(1,1,1,1,0) \in W$. It can be easily verified that a basis of $W$ is

$$\{(1,2,0,1,2), (0,0,1,0,-1), (1,1,1,1,0)\}.$$

Similarly, a vector of $V$ has the form $(v, w, x, y, 3y-v-x)$ for $v, w, x, y \in \mathbb{R}$. Substituting $v = 0, w = 1, x = 0$ and $y = 0$, gives a vector $(0,1,0,0,0) \in V$. Also, substituting $v = 0, w = 1, x = 1$ and $y = 1$, gives another vector $(0,1,1,1,2) \in V$. So, a basis of $V$ can be taken as

$$\{(1,2,0,1,2), (0,0,1,0,-1), (0,1,0,0,0), (0,1,1,1,2)\}.$$

Recall that for two vector subspaces $M$ and $N$ of a vector space $V(\mathbb{F})$, the vector subspace $M+N$ is defined by

$$M + N = \{\mathbf{u} + \mathbf{v} : \mathbf{u} \in M, \ \mathbf{v} \in N\}.$$

With this definition, we have the following very important theorem (for a proof, see Appendix 14.4.1).

**Theorem 3.3.18** Let $V(\mathbb{F})$ be a finite dimensional vector space and let $M$ and $N$ be two subspaces of $V$. Then

$$\dim(M) + \dim(N) = \dim(M+N) + \dim(M \cap N). \tag{3.3.2}$$

**Exercise 3.3.19**

1. Find a basis of the vector space $\mathcal{P}_n(\mathbb{R})$. Also, find $\dim(\mathcal{P}_n(\mathbb{R}))$. What can you say about the dimension of $\mathcal{P}(\mathbb{R})$?

2. Consider the real vector space, $C([0, 2\pi])$, of all real valued continuous functions. For each $n$ consider the vector $\mathbf{e}_n$ defined by $\mathbf{e}_n(x) = \sin(nx)$. Prove that the collection of vectors $\{\mathbf{e}_n : 1 \leq n < \infty\}$ is a linearly independent set.
[*Hint: On the contrary, assume that the set is linearly dependent. Then we have a finite set of vectors, say* $\{\mathbf{e}_{k_1}, \mathbf{e}_{k_2}, \ldots, \mathbf{e}_{k_\ell}\}$ *that are linearly dependent. That is, there exist scalars* $\alpha_i \in \mathbb{R}$ *for* $1 \leq i \leq \ell$ *not all zero such that*

$$\alpha_1 \sin(k_1 x) + \alpha_2 \sin(k_2 x) + \cdots + \alpha_\ell \sin(k_\ell x) = \mathbf{0} \quad \textit{for all} \quad x \in [0, 2\pi].$$

*Now for different values of $m$ integrate the function*

$$\int_0^{2\pi} \sin(mx) \left(\alpha_1 \sin(k_1 x) + \alpha_2 \sin(k_2 x) + \cdots + \alpha_\ell \sin(k_\ell x)\right) \, dx$$

*to get the required result.*]

3. Show that the set $\{(1,0,0), (1,1,0), (1,1,1)\}$ is a basis of $\mathbb{C}^3(\mathbb{C})$. Is it a basis of $\mathbb{C}^3(\mathbb{R})$ also?

4. Let $W = \{(x,y,z,w) \in \mathbb{R}^4 : x + y - z + w = 0\}$ be a subspace of $\mathbb{R}^4$. Find its basis and dimension.

5. Let $V = \{(x,y,z,w) \in \mathbb{R}^4 : x + y - z + w = 0, x + y + z + w = 0\}$ and $W = \{(x,y,z,w) \in \mathbb{R}^4 : x - y - z + w = 0, x + 2y - w = 0\}$ be two subspaces of $\mathbb{R}^4$. Find bases and dimensions of $V$, $W$, $V \cap W$ and $V + W$.

6. Let $V$ be the set of all real symmetric $n \times n$ matrices. Find its basis and dimension. What if $V$ is the complex vector space of all $n \times n$ Hermitian matrices?

7. If $M$ and $N$ are 4-dimensional subspaces of a vector space $V$ of dimension 7 then show that $M$ and $N$ have at least one vector in common other than the zero vector.

8. Let $P = L\{(1,0,0),(1,1,0)\}$ and $Q = L\{(1,1,1)\}$ be vector subspaces of $\mathbb{R}^3$. Show that $P+Q = \mathbb{R}^3$ and $P \cap Q = \{\mathbf{0}\}$. If $\mathbf{u} \in \mathbb{R}^3$, determine $\mathbf{u}_P, \mathbf{u}_Q$ such that $\mathbf{u} = \mathbf{u}_P + \mathbf{u}_Q$ where $\mathbf{u}_P \in P$ and $\mathbf{u}_Q \in Q$. Is it necessary that $\mathbf{u}_P$ and $\mathbf{u}_Q$ are unique?

9. Let $W_1$ be a $k$-dimensional subspace of an $n$-dimensional vector space $V(\mathbb{F})$ where $k \geq 1$. Prove that there exists an $(n-k)$-dimensional subspace $W_2$ of $V$ such that $W_1 \cap W_2 = \{\mathbf{0}\}$ and $W_1 + W_2 = V$.

10. Let $P$ and $Q$ be subspaces of $\mathbb{R}^n$ such that $P + Q = \mathbb{R}^n$ and $P \cap Q = \{\mathbf{0}\}$. Then show that each $\mathbf{u} \in \mathbb{R}^n$ can be uniquely expressed as $\mathbf{u} = \mathbf{u}_P + \mathbf{u}_Q$ where $\mathbf{u}_P \in P$ and $\mathbf{u}_Q \in Q$.

11. Let $P = L\{(1,-1,0),(1,1,0)\}$ and $Q = L\{(1,1,1),(1,2,1)\}$ be vector subspaces of $\mathbb{R}^3$. Show that $P+Q = \mathbb{R}^3$ and $P \cap Q \neq \{\mathbf{0}\}$. Show that there exists a vector $\mathbf{u} \in \mathbb{R}^3$ such that $\mathbf{u}$ cannot be written uniquely in the form $\mathbf{u} = \mathbf{u}_P + \mathbf{u}_Q$ where $\mathbf{u}_P \in P$ and $\mathbf{u}_Q \in Q$.

12. Recall the vector space $\mathcal{P}_4(\mathbb{R})$. Is the set,
$$W = \{p(x) \in \mathcal{P}_4(\mathbb{R}) \ : \ p(-1) = p(1) = 0\}$$
a subspace of $\mathcal{P}_4(\mathbb{R})$? If yes, find its dimension.

13. Let $V$ be the set of all $2 \times 2$ matrices with complex entries and $a_{11} + a_{22} = 0$. Show that $V$ is a real vector space. Find its basis. Also let $W = \{A \in V : a_{21} = \overline{-a_{12}}\}$. Show $W$ is a vector subspace of $V$, and find its dimension.

14. Let $A = \begin{bmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 2 & 2 & 2 & 4 \\ 2 & -2 & 4 & 0 & 8 \\ 4 & 2 & 5 & 6 & 10 \end{bmatrix}$, and $B = \begin{bmatrix} 2 & 4 & 0 & 6 \\ -1 & 0 & -2 & 5 \\ -3 & -5 & 1 & -4 \\ -1 & -1 & 1 & 2 \end{bmatrix}$ be two matrices. For $A$ and $B$ find the following:

    (a) their row-reduced echelon forms.

    (b) the matrices $P_1$ and $P_2$ such that $P_1 A$ and $P_2 B$ are in row-reduced form.

    (c) a basis each for the row spaces of $A$ and $B$.

    (d) a basis each for the range spaces of $A$ and $B$.

    (e) bases of the null spaces of $A$ and $B$.

    (f) the dimensions of all the vector subspaces so obtained.

15. Let $M(n,\mathbb{R})$ denote the space of all $n \times n$ real matrices. For the sets given below, check that they are subspaces of $M(n,\mathbb{R})$ and also find their dimension.

    (a) $sl(n,\mathbb{R}) = \{A \in M(n,\mathbb{R}) \ : \ \text{tr}(A) = 0\}$, where recall that tr$(A)$ stands for trace of $A$.

    (b) $S(n,\mathbb{R}) = \{A \in M(n,\mathbb{R}) \ : \ A = A^t\}$.

    (c) $A(n,\mathbb{R}) = \{A \in M(n,\mathbb{R}) \ : \ A + A^t = \mathbf{0}\}$.

Before going to the next section, we prove that for any matrix $A$ of order $m \times n$
$$\text{Row rank}(A) = \text{Column rank}(A).$$

**Proposition 3.3.20** Let $A$ be an $m \times n$ real matrix. Then

$$\text{Row rank}(A) = \text{Column rank}(A).$$

Proof. Let $R_1, R_2, \ldots, R_m$ be the rows of $A$ and $C_1, C_2, \ldots, C_n$ be the columns of $A$. Note that Row rank$(A) = r$, means that

$$\dim\big(L(R_1, R_2, \ldots, R_m)\big) = r.$$

Hence, there exists vectors

$$\mathbf{u}_1 = (u_{11}, \ldots, u_{1n}), \mathbf{u}_2 = (u_{21}, \ldots, u_{2n}), \ldots, \mathbf{u}_r = (u_{r1}, \ldots, u_{rn}) \in \mathbb{R}^n$$

with

$$R_i \in L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r) \in \mathbb{R}^n, \quad \text{for all } \; i, 1 \le i \le m.$$

Therefore, there exist real numbers $\alpha_{ij}, \; 1 \le i \le m, \; 1 \le j \le r$ such that

$$R_1 = \alpha_{11}\mathbf{u}_1 + \alpha_{12}\mathbf{u}_2 + \cdots + \alpha_{1r}\mathbf{u}_r = (\sum_{i=1}^{r} \alpha_{1i}u_{i1}, \sum_{i=1}^{r} \alpha_{1i}u_{i2}, \ldots, \sum_{i=1}^{r} \alpha_{1i}u_{in}),$$

$$R_2 = \alpha_{21}\mathbf{u}_1 + \alpha_{22}\mathbf{u}_2 + \cdots + \alpha_{2r}\mathbf{u}_r = (\sum_{i=1}^{r} \alpha_{2i}u_{i1}, \sum_{i=1}^{r} \alpha_{2i}u_{i2}, \ldots, \sum_{i=1}^{r} \alpha_{2i}u_{in}),$$

and so on, till

$$R_m = \alpha_{m1}\mathbf{u}_1 + \cdots + \alpha_{mr}\mathbf{u}_r = (\sum_{i=1}^{r} \alpha_{mi}u_{i1}, \sum_{i=1}^{r} \alpha_{mi}u_{i2}, \ldots, \sum_{i=1}^{r} \alpha_{mi}u_{in}).$$

So,

$$C_1 = \begin{bmatrix} \sum\limits_{i=1}^{r} \alpha_{1i}u_{i1} \\ \sum\limits_{i=1}^{r} \alpha_{2i}u_{i1} \\ \vdots \\ \sum\limits_{i=1}^{r} \alpha_{mi}u_{i1} \end{bmatrix} = u_{11} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{bmatrix} + u_{21} \begin{bmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{bmatrix} + \cdots + u_{r1} \begin{bmatrix} \alpha_{1r} \\ \alpha_{2r} \\ \vdots \\ \alpha_{mr} \end{bmatrix}.$$

In general, for $1 \le j \le n$, we have

$$C_j = \begin{bmatrix} \sum\limits_{i=1}^{r} \alpha_{1i}u_{ij} \\ \sum\limits_{i=1}^{r} \alpha_{2i}u_{ij} \\ \vdots \\ \sum\limits_{i=1}^{r} \alpha_{mi}u_{ij} \end{bmatrix} = u_{1j} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{bmatrix} + u_{2j} \begin{bmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{bmatrix} + \cdots + u_{rj} \begin{bmatrix} \alpha_{1r} \\ \alpha_{2r} \\ \vdots \\ \alpha_{mr} \end{bmatrix}.$$

Therefore, we observe that the columns $C_1, C_2, \ldots, C_n$ are linear combination of the $r$ vectors

$$(\alpha_{11}, \alpha_{21}, \ldots, \alpha_{m1})^t, (\alpha_{12}, \alpha_{22}, \ldots, \alpha_{m2})^t, \ldots, (\alpha_{1r}, \alpha_{2r}, \ldots, \alpha_{mr})^t.$$

Therefore,

$$\text{Column rank}(A) = \dim\big(L(C_1, C_2, \ldots, C_n)\big) =\le r = \text{Row rank}(A).$$

A similar argument gives

$$\text{Row rank}(A) \le \text{Column rank}(A).$$

Thus, we have the required result. □

## 3.4 Ordered Bases

Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis of a vector space $V(\mathbb{F})$. As $\mathcal{B}$ is a set, there is no ordering of its elements. In this section, we want to associate an order among the vectors in any basis of $V$.

**Definition 3.4.1 (Ordered Basis)** An ordered basis for a vector space $V(\mathbb{F})$ of dimension $n$, is a basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ together with a one-to-one correspondence between the sets $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\{1, 2, 3, \ldots, n\}$.

If the ordered basis has $\mathbf{u}_1$ as the first vector, $\mathbf{u}_2$ as the second vector and so on, then we denote this ordered basis by

$$(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n).$$

**Example 3.4.2** Consider $\mathcal{P}_2(\mathbb{R})$, the vector space of all polynomials of degree less than or equal to 2 with coefficients from $\mathbb{R}$. The set $\{1-x, 1+x, x^2\}$ is a basis of $\mathcal{P}_2(\mathbb{R})$.

For any element $a_0 + a_1x + a_2x^2 \in \mathcal{P}_2(\mathbb{R})$, we have

$$a_0 + a_1x + a_2x^2 = \frac{a_0 - a_1}{2}(1-x) + \frac{a_0 + a_1}{2}(1+x) + a_2x^2.$$

If $(1-x, 1+x, x^2)$ is an ordered basis, then $\frac{a_0 - a_1}{2}$ is the first component, $\frac{a_0 + a_1}{2}$ is the second component, and $a_2$ is the third component of the vector $a_0 + a_1x + a_2x^2$.

If we take $(1+x, 1-x, x^2)$ as an ordered basis, then $\frac{a_0 + a_1}{2}$ is the first component, $\frac{a_0 - a_1}{2}$ is the second component, and $a_2$ is the third component of the vector $a_0 + a_1x + a_2x^2$.

That is, as ordered bases $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$, $(\mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n, \mathbf{u}_1)$, and $(\mathbf{u}_n, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1})$ are different even though they have the same set of vectors as elements.

**Definition 3.4.3 (Coordinates of a Vector)** Let $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis of a vector space $V(\mathbb{F})$ and let $\mathbf{v} \in V$. If

$$\mathbf{v} = \beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \cdots + \beta_n\mathbf{v}_n$$

then the tuple $(\beta_1, \beta_2, \ldots, \beta_n)$ is called the coordinate of the vector $\mathbf{v}$ with respect to the ordered basis $\mathcal{B}$.

Mathematically, we denote it by $[\mathbf{v}]_{\mathcal{B}} = (\beta_1, \ldots, \beta_n)^t$, A COLUMN VECTOR.

Suppose $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{u}_n, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1})$ are two ordered bases of $V$. Then for any $\mathbf{x} \in V$ there exists unique scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that

$$\mathbf{x} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n = \alpha_n\mathbf{u}_n + \alpha_1\mathbf{u}_1 + \cdots + \alpha_{n-1}\mathbf{u}_{n-1}.$$

Therefore,

$$[\mathbf{x}]_{\mathcal{B}_1} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t \quad \text{and} \quad [\mathbf{x}]_{\mathcal{B}_2} = (\alpha_n, \alpha_1, \alpha_2, \ldots, \alpha_{n-1})^t.$$

Note that $\mathbf{x}$ is uniquely written as $\sum\limits_{i=1}^{n} \alpha_i\mathbf{u}_i$ and hence the coordinates with respect to an ordered basis are unique.

Suppose that the ordered basis $\mathcal{B}_1$ is changed to the ordered basis $\mathcal{B}_3 = (\mathbf{u}_2, \mathbf{u}_1, \mathbf{u}_3, \ldots, \mathbf{u}_n)$. Then $[\mathbf{x}]_{\mathcal{B}_3} = (\alpha_2, \alpha_1, \alpha_3, \ldots, \alpha_n)^t$. So, the coordinates of a vector depend on the ordered basis chosen.

**Example 3.4.4** Let $V = \mathbb{R}^3$. Consider the ordered bases $\mathcal{B}_1 = \big((1,0,0), (0,1,0), (0,0,1)\big)$, $\mathcal{B}_2 = \big((1,0,0), (1,1,0), (1,1,1)\big)$ and $\mathcal{B}_3 = \big((1,1,1), (1,1,0), (1,0,0)\big)$ of $V$. Then, with respect to the above bases we have

$$\begin{aligned}
(1,-1,1) &= 1\cdot(1,0,0) + (-1)\cdot(0,1,0) + 1\cdot(0,0,1). \\
&= 2\cdot(1,0,0) + (-2)\cdot(1,1,0) + 1\cdot(1,1,1). \\
&= 1\cdot(1,1,1) + (-2)\cdot(1,1,0) + 2\cdot(1,0,0).
\end{aligned}$$

Therefore, if we write $\mathbf{u} = (1, -1, 1)$, then

$$[\mathbf{u}]_{\mathcal{B}_1} = (1, -1, 1)^t, \; [\mathbf{u}]_{\mathcal{B}_2} = (2, -2, 1)^t, \; [\mathbf{u}]_{\mathcal{B}_3} = (1, -2, 2)^t.$$

In general, let $V$ be an $n$-dimensional vector space with ordered bases $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. Since, $\mathcal{B}_1$ is a basis of $V$, there exists unique scalars $a_{ij}, \; 1 \le i, j \le n$ such that

$$\mathbf{v}_i = \sum_{l=1}^{n} a_{li} \mathbf{u}_l \qquad \text{for } 1 \le i \le n.$$

That is, for each $i, \; 1 \le i \le n, \; [\mathbf{v}_i]_{\mathcal{B}_1} = (a_{1i}, a_{2i}, \ldots, a_{ni})^t$.

Let $\mathbf{v} \in V$ with $[\mathbf{v}]_{\mathcal{B}_2} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t$. As $\mathcal{B}_2$ as ordered basis $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$, we have

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{i=1}^{n} \alpha_i \left( \sum_{j=1}^{n} a_{ji} \mathbf{u}_j \right) = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} a_{ji} \alpha_i \right) \mathbf{u}_j.$$

Since $\mathcal{B}_1$ is a basis this representation of $\mathbf{v}$ in terms of $\mathbf{u}_i$'s is unique. So,

$$\begin{aligned} [\mathbf{v}]_{\mathcal{B}_1} &= \left( \sum_{i=1}^{n} a_{1i}\alpha_i, \sum_{i=1}^{n} a_{2i}\alpha_i, \ldots, \sum_{i=1}^{n} a_{ni}\alpha_i \right)^t \\ &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \\ &= A[\mathbf{v}]_{\mathcal{B}_2}. \end{aligned}$$

Note that the $i^{th}$ column of the matrix $A$ is equal to $[\mathbf{v}_i]_{\mathcal{B}_1}$, i.e., the $i^{th}$ column of $A$ is the coordinate of the $i^{th}$ vector $\mathbf{v}_i$ of $\mathcal{B}_2$ with respect to the ordered basis $\mathcal{B}_1$. Hence, we have proved the following theorem.

**Theorem 3.4.5** Let $V$ be an $n$-dimensional vector space with ordered bases $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. Let

$$A = \left[ [\mathbf{v}_1]_{\mathcal{B}_1}, [\mathbf{v}_2]_{\mathcal{B}_1}, \ldots, [\mathbf{v}_n]_{\mathcal{B}_1} \right].$$

Then for any $\mathbf{v} \in V$,

$$[v]_{\mathcal{B}_1} = A[\mathbf{v}]_{\mathcal{B}_2}.$$

**Example 3.4.6** Consider two bases $\mathcal{B}_1 = \big((1,0,0),(1,1,0),(1,1,1)\big)$ and $\mathcal{B}_2 = \big((1,1,1),(1,-1,1),(1,1,0)\big)$ of $\mathbb{R}^3$.

1. Then

$$\begin{aligned} [(x,y,z)]_{\mathcal{B}_1} &= (x-y)\cdot(1,0,0) + (y-z)\cdot(1,1,0) + z\cdot(1,1,1) \\ &= (x-y, y-z, z)^t \end{aligned}$$

and

$$\begin{aligned} [(x,y,z)]_{\mathcal{B}_2} &= (\frac{y-x}{2} + z)\cdot(1,1,1) + \frac{x-y}{2}\cdot(1,-1,1) \\ &\quad +(x-z)\cdot(1,1,0) \\ &= (\frac{y-x}{2} + z, \frac{x-y}{2}, x-z)^t. \end{aligned}$$

2. Let $A = [a_{ij}] = \begin{bmatrix} 0 & 2 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. The columns of the matrix $A$ are obtained by the following rule:

$$[(1,1,1)]_{\mathcal{B}_1} = 0 \cdot (1,0,0) + 0 \cdot (1,1,0) + 1 \cdot (1,1,1) = (0,0,1)^t,$$

$$[(1,-1,1)]_{\mathcal{B}_1} = 2 \cdot (1,0,0) + (-2) \cdot (1,1,0) + 1 \cdot (1,1,1) = (2,-2,1)^t$$

and

$$[(1,1,0)]_{\mathcal{B}_1} = 0 \cdot (1,0,0) + 1 \cdot (1,1,0) + 0 \cdot (1,1,1) = (0,1,0)^t.$$

That is, the elements of $\mathcal{B}_2 = \big((1,1,1),(1,-1,1),(1,1,0)\big)$ are expressed in terms of the ordered basis $\mathcal{B}_1$.

3. Note that for any $(x,y,z) \in \mathbb{R}^3$,

$$[(x,y,z)]_{\mathcal{B}_1} = \begin{bmatrix} x-y \\ y-z \\ z \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{y-x}{2} + z \\ \frac{x-y}{2} \\ x-z \end{bmatrix} = A \;\; [(x,y,z)]_{\mathcal{B}_2}.$$

4. The matrix $A$ is invertible and hence $[(x,y,z)]_{\mathcal{B}_2} = A^{-1} \;\; [(x,y,z)]_{\mathcal{B}_1}$.

In the next chapter, we try to understand Theorem 3.4.5 again using the ideas of 'linear transformations / functions'.

**Exercise 3.4.7** 1. Determine the coordinates of the vectors $(1,2,1)$ and $(4,-2,2)$ with respect to the basis $\mathcal{B} = \big((2,1,0),(2,1,1),(2,2,1)\big)$ of $\mathbb{R}^3$.

2. Consider the vector space $\mathcal{P}_3(\mathbb{R})$.

   (a) Show that $\mathcal{B}_1 = (1-x, 1+x^2, 1-x^3, 3+x^2-x^3)$ and $\mathcal{B}_2 = (1, 1-x, 1+x^2, 1-x^3)$ are bases of $\mathcal{P}_3(\mathbb{R})$.

   (b) Find the coordinates of the vector $\mathbf{u} = 1 + x + x^2 + x^3$ with respect to the ordered basis $\mathcal{B}_1$ and $\mathcal{B}_2$.

   (c) Find the matrix $A$ such that $[\mathbf{u}]_{\mathcal{B}_2} = A[\mathbf{u}]_{\mathcal{B}_1}$.

   (d) Let $\mathbf{v} = a_0 + a_1 x + a_2 x^2 + a_3 x^3$. Then verify the following:

$$\begin{aligned} [\mathbf{v}]_{\mathcal{B}_1} &= \begin{bmatrix} -a_1 \\ -a_0 - a_1 + 2a_2 - a_3 \\ -a_0 - a_1 + a_2 - 2a_3 \\ a_0 + a_1 - a_2 + a_3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 + a_1 - a_2 + a_3 \\ -a_1 \\ a_2 \\ -a_3 \end{bmatrix} \\ &= [\mathbf{v}]_{\mathcal{B}_2}. \end{aligned}$$

# Chapter 4

# Linear Transformations

## 4.1 Definitions and Basic Properties

Throughout this chapter, the scalar field $\mathbb{F}$ is either always the set $\mathbb{R}$ or always the set $\mathbb{C}$.

**Definition 4.1.1 (Linear Transformation)** Let $V$ and $W$ be vector spaces over $\mathbb{F}$. A map $T : V \longrightarrow W$ is called a linear transformation if

$$T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}), \qquad \text{for all } \alpha, \beta \in \mathbb{F}, \text{ and } \mathbf{u}, \mathbf{v} \in V.$$

We now give a few examples of linear transformations.

**Example 4.1.2** 1. Define $T : \mathbb{R} \longrightarrow \mathbb{R}^2$ by $T(x) = (x, 3x)$ for all $x \in \mathbb{R}$. Then $T$ is a linear transformation as

$$T(x + y) = (x + y, 3(x + y)) = (x, 3x) + (y, 3y) = T(x) + T(y).$$

2. Verify that the maps given below from $\mathbb{R}^n$ to $\mathbb{R}$ are linear transformations. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

   (a) Define $T(\mathbf{x}) = \sum\limits_{i=1}^{n} x_i$.

   (b) For any $i$, $1 \leq i \leq n$, define $T_i(\mathbf{x}) = x_i$.

   (c) For a fixed vector $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$, define $T(\mathbf{x}) = \sum\limits_{i=1}^{n} a_i x_i$. Note that examples $(a)$ and $(b)$ can be obtained by assigning particular values for the vector $\mathbf{a}$.

3. Define $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ by $T((x, y)) = (x + y, 2x - y, x + 3y)$.
   Then $T$ is a linear transformation with $T((1, 0)) = (1, 2, 1)$ and $T((0, 1)) = (1, -1, 3)$.

4. Let $A$ be an $m \times n$ real matrix. Define a map $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by

   $$T_A(\mathbf{x}) = A\mathbf{x} \quad \text{for every} \quad \mathbf{x}^t = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n.$$

   Then $T_A$ is a linear transformation. That is, every $m \times n$ real matrix defines a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

5. Recall that $\mathcal{P}_n(\mathbb{R})$ is the set of all polynomials of degree less than or equal to $n$ with real coefficients. Define $T : \mathbb{R}^{n+1} \longrightarrow \mathcal{P}_n(\mathbb{R})$ by

   $$T((a_1, a_2, \ldots, a_{n+1})) = a_1 + a_2 x + \cdots + a_{n+1} x^n$$

   for $(a_1, a_2, \ldots, a_{n+1}) \in \mathbb{R}^{n+1}$. Then $T$ is a linear transformation.

**Proposition 4.1.3** Let $T : V \longrightarrow W$ be a linear transformation. Suppose that $\mathbf{0}_V$ is the zero vector in $V$ and $\mathbf{0}_W$ is the zero vector of $W$. Then $T(\mathbf{0}_V) = \mathbf{0}_W$.

Proof. Since $\mathbf{0}_V = \mathbf{0}_V + \mathbf{0}_V$, we have

$$T(\mathbf{0}_V) = T(\mathbf{0}_V + \mathbf{0}_V) = T(\mathbf{0}_V) + T(\mathbf{0}_V).$$

So, $T(\mathbf{0}_V) = \mathbf{0}_W$ as $T(\mathbf{0}_V) \in W$. □

From now on, we write $\mathbf{0}$ for both the zero vector of the domain space and the zero vector of the range space.

**Definition 4.1.4 (Zero Transformation)** Let $V$ be a vector space and let $T : V \longrightarrow W$ be the map defined by

$$T(\mathbf{v}) = \mathbf{0} \text{ for every } \mathbf{v} \in V.$$

Then $T$ is a linear transformation. Such a linear transformation is called the zero transformation and is denoted by $\mathbf{0}$.

**Definition 4.1.5 (Identity Transformation)** Let $V$ be a vector space and let $T : V \longrightarrow V$ be the map defined by

$$T(\mathbf{v}) = \mathbf{v} \text{ for every } \mathbf{v} \in V.$$

Then $T$ is a linear transformation. Such a linear transformation is called the Identity transformation and is denoted by $I$.

We now prove a result that relates a linear transformation $T$ with its value on a basis of the domain space.

**Theorem 4.1.6** Let $T : V \longrightarrow W$ be a linear transformation and $\mathcal{B} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ be an ordered basis of $V$. Then the linear transformation $T$ is a linear combination of the vectors $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$.

In other words, $T$ is determined by $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$.

Proof. Since $\mathcal{B}$ is a basis of $V$, for any $\mathbf{x} \in V$, there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that $\mathbf{x} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_n \mathbf{u}_n$. So, by the definition of a linear transformation

$$T(\mathbf{x}) = T(\alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n) = \alpha_1 T(\mathbf{u}_1) + \cdots + \alpha_n T(\mathbf{u}_n).$$

Observe that, given $\mathbf{x} \in V$, we know the scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$. Therefore, to know $T(\mathbf{x})$, we just need to know the vectors $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$ in $W$.

That is, for every $\mathbf{x} \in V$, $T(\mathbf{x})$ is determined by the coordinates $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ of $\mathbf{x}$ with respect to the ordered basis $\mathcal{B}$ and the vectors $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n) \in W$. □

**Exercise 4.1.7** 1. Which of the following are linear transformations $T : V \longrightarrow W$? Justify your answers.

(a) Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$ with $T\big( (x, y) \big) = (x + y + 1, 2x - y, x + 3y)$

(b) Let $V = W = \mathbb{R}^2$ with $T\big( (x, y) \big) = (x - y, x^2 - y^2)$

(c) Let $V = W = \mathbb{R}^2$ with $T\big( (x, y) \big) = (x - y, |x|)$

(d) Let $V = \mathbb{R}^2$ and $W = \longrightarrow \mathbb{R}^4$ with $T\big( (x, y) \big) = (x + y, x - y, 2x + y, 3x - 4y)$

(e) Let $V = W = \mathbb{R}^4$ with $T\big( (x, y, z, w) \big) = (z, x, w, y)$

2. Recall that $M_2(\mathbb{R})$ is the space of all $2 \times 2$ matrices with real entries. Then, which of the following are linear transformations $T : M_2(\mathbb{R}) \longrightarrow M_2(\mathbb{R})$?

$(a)$ $T(A) = A^t$ $(b)$ $T(A) = I + A$ $(c)$ $T(A) = A^2$

$(d)$ $T(A) = BAB^{-1}$, where $B$ is some fixed $2 \times 2$ matrix.

3. Let $T : \mathbb{R} \longrightarrow \mathbb{R}$ be a map. Then $T$ is a linear transformation if and only if there exists a unique $c \in \mathbb{R}$ such that $T(\mathbf{x}) = c\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}$.

4. Let $A$ be an $n \times n$ real matrix. Consider the linear transformation

$$T_A(\mathbf{x}) = A\mathbf{x} \quad \text{for every} \quad \mathbf{x} \in \mathbb{R}^n.$$

Then prove that $T^2(\mathbf{x}) := T(T(\mathbf{x})) = A^2\mathbf{x}$. In general, for $k \in \mathbb{N}$, prove that $T^k(\mathbf{x}) = A^k\mathbf{x}$.

5. Use the ideas of matrices to give examples of linear transformations $T, S : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ that satisfy:

   (a) $T \neq \mathbf{0}, \;\; T^2 \neq \mathbf{0}, \;\; T^3 = \mathbf{0}$.

   (b) $T \neq \mathbf{0}, \;\; S \neq \mathbf{0}, \;\; S \circ T \neq \mathbf{0}, \;\; T \circ S = \mathbf{0}$; where $T \circ S(\mathbf{x}) = T\big(S(\mathbf{x})\big)$.

   (c) $S^2 = T^2, \;\; S \neq T$.

   (d) $T^2 = I, \;\; T \neq I$.

6. Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear transformation such that $T \neq \mathbf{0}$ and $T^2 = \mathbf{0}$. Let $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) \neq \mathbf{0}$. Then prove that the set $\{\mathbf{x}, T(\mathbf{x})\}$ is linearly independent. In general, if $T^k \neq \mathbf{0}$ for $1 \leq k \leq p$ and $T^{p+1} = \mathbf{0}$, then for any vector $\mathbf{x} \in \mathbb{R}^n$ with $T^p(\mathbf{x}) \neq \mathbf{0}$ prove that the set $\{\mathbf{x}, T(\mathbf{x}), \ldots, T^p(\mathbf{x})\}$ is linearly independent.

7. Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation, and let $\mathbf{x}_0 \in \mathbb{R}^n$ with $T(\mathbf{x}_0) = y$. Consider the sets

$$S = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{y}\} \quad \text{and} \quad N = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{0}\}.$$

Show that for every $\mathbf{x} \in S$ there exists $\mathbf{z} \in N$ such that $\mathbf{x} = \mathbf{x}_0 + z$.

8. Define a map $T : \mathbb{C} \longrightarrow \mathbb{C}$ by $T(z) = \overline{z}$, the complex conjugate of $z$. Is $T$ linear on

   (a) $\mathbb{C}$ over $\mathcal{R}$ (b) $\mathbb{C}$ over $\mathbb{C}$.

9. Find all functions $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ that satisfy the conditions

   (a) $f(\ (x, x)\ ) = (x, x)$ and

   (b) $f(\ (x, y)\ ) = (y, x)$ for all $(x, y) \in \mathbb{R}^2$.

That is, $f$ fixes the line $y = x$ and sends the point $(x_1, y_1)$ for $x_1 \neq y_1$ to its mirror image along the line $y = x$.

Is this function a linear transformation? Justify your answer.

**Theorem 4.1.8** Let $T : V \longrightarrow W$ be a linear transformation. For $\mathbf{w} \in W$, define the set

$$T^{-1}(\mathbf{w}) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{w}\}.$$

Suppose that the map $T$ is one-one and onto.

1. Then for each $\mathbf{w} \in W$, the set $T^{-1}(\mathbf{w})$ is a set consisting of a single element.

2. The map $T^{-1} : W \longrightarrow V$ defined by

$$T^{-1}(\mathbf{w}) = \mathbf{v} \;\; \text{whenever} \;\; T(\mathbf{v}) = \mathbf{w}.$$

is a linear transformation.

PROOF. Since $T$ is onto, for each $\mathbf{w} \in W$ there exists a vector $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. So, the set $T^{-1}(\mathbf{w})$ is non-empty.

Suppose there exist vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_1) = T(\mathbf{v}_2)$. But by assumption, $T$ is one-one and therefore $\mathbf{v}_1 = \mathbf{v}_2$. This completes the proof of Part 1.

We now show that $T^{-1}$ as defined above is a linear transformation. Let $\mathbf{w}_1, \mathbf{w}_2 \in W$. Then by Part 1, there exist unique vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T^{-1}(\mathbf{w}_1) = \mathbf{v}_1$ and $T^{-1}(\mathbf{w}_2) = \mathbf{v}_2$. Or equivalently, $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. So, for any $\alpha_1, \alpha_2 \in \mathbb{F}$, we have $T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2$.

Thus for any $\alpha_1, \alpha_2 \in \mathbb{F}$,

$$T^{-1}(\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \alpha_1 T^{-1}(\mathbf{w}_1) + \alpha_2 T^{-1}(\mathbf{w}_2).$$

Hence $T^{-1} : W \longrightarrow V$, defined as above, is a linear transformation. □

**Definition 4.1.9 (Inverse Linear Transformation)** Let $T : V \longrightarrow W$ be a linear transformation. If the map $T$ is one-one and onto, then the map $T^{-1} : W \longrightarrow V$ defined by

$$T^{-1}(\mathbf{w}) = \mathbf{v} \quad \text{whenever } T(\mathbf{v}) = \mathbf{w}$$

is called the inverse of the linear transformation $T$.

**Example 4.1.10**

1. Define $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ by $T((x, y)) = (x + y, x - y)$. Then $T^{-1} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is defined by

$$T^{-1}((x, y)) = (\frac{x+y}{2}, \frac{x-y}{2}).$$

Note that

$$\begin{aligned} T \circ T^{-1}((x, y)) &= T(T^{-1}((x, y))) = T((\frac{x+y}{2}, \frac{x-y}{2})) \\ &= (\frac{x+y}{2} + \frac{x-y}{2}, \frac{x+y}{2} - \frac{x-y}{2}) \\ &= (x, y). \end{aligned}$$

Hence, $T \circ T^{-1} = I$, the identity transformation. Verify that $T^{-1} \circ T = I$. Thus, the map $T^{-1}$ is indeed the inverse of the linear transformation $T$.

2. Recall the vector space $\mathcal{P}_n(\mathbb{R})$ and the linear transformation $T : \mathbb{R}^{n+1} \longrightarrow \mathcal{P}_n(\mathbb{R})$ defined by

$$T((a_1, a_2, \ldots, a_{n+1})) = a_1 + a_2 x + \cdots + a_{n+1}x^n$$

for $(a_1, a_2, \ldots, a_{n+1}) \in \mathbb{R}^{n+1}$. Then $T^{-1} : \mathcal{P}_n(\mathbb{R}) \longrightarrow \mathbb{R}^{n+1}$ is defined as

$$T^{-1}(a_1 + a_2 x + \cdots + a_{n+1}x^n) = (a_1, a_2, \ldots, a_{n+1})$$

for $a_1 + a_2 x + \cdots + a_{n+1}x^n \in \mathcal{P}_n(\mathbb{R})$. Verify that $T \circ T^{-1} = T^{-1} \circ T = I$. Hence, conclude that the map $T^{-1}$ is indeed the inverse of the linear transformation $T$.

## 4.2 Matrix of a linear transformation

In this section, we relate linear transformation over finite dimensional vector spaces with matrices. For this, we ask the reader to recall the results on ordered basis, studied in Section 3.4.

Let $V$ and $W$ be finite dimensional vector spaces over the set $\mathbb{F}$ with respective dimensions $m$ and $n$. Also, let $T : V \longrightarrow W$ be a linear transformation. Suppose $\mathcal{B}_1 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is an ORDERED BASIS of

$V$. In the last section, we saw that a linear transformation is determined by its image on a basis of the domain space. We therefore look at the images of the vectors $\mathbf{v}_j \in \mathcal{B}_1$ for $1 \le j \le n$.

Now for each $j,\ 1 \le j \le n$, the vectors $T(\mathbf{v}_j) \in W$. We now express these vectors in terms of an ordered basis $\mathcal{B}_2 = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ of $W$. So, for each $j,\ 1 \le j \le n$, there exist unique scalars $a_{1j}, a_{2j}, \ldots, a_{mj} \in \mathbb{F}$ such that

$$\begin{aligned} T(\mathbf{v}_1) &= a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + \cdots + a_{m1}\mathbf{w}_m \\ T(\mathbf{v}_2) &= a_{12}\mathbf{w}_1 + a_{22}\mathbf{w}_2 + \cdots + a_{m2}\mathbf{w}_m \\ &\vdots \\ T(\mathbf{v}_n) &= a_{1n}\mathbf{w}_1 + a_{2n}\mathbf{w}_2 + \cdots + a_{mn}\mathbf{w}_m. \end{aligned}$$

Or in short, $T(\mathbf{v}_j) = \sum\limits_{i=1}^{m} a_{ij}\mathbf{w}_i$ for $1 \le j \le n$. In other words, for each $j,\ 1 \le j \le n$, the coordinates of $T(\mathbf{v}_j)$ with respect to the ordered basis $\mathcal{B}_2$ is the column vector $[a_{1j}, a_{2j}, \ldots, a_{mj}]^t$. Equivalently,

$$[T(\mathbf{v}_j)]_{\mathcal{B}_2} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

Let $[\mathbf{x}]_{\mathcal{B}_1} = [x_1, x_2, \ldots, x_n]^t$ be the coordinates of a vector $\mathbf{x} \in V$. Then

$$\begin{aligned} T(\mathbf{x}) &= T(\sum_{j=1}^{n} x_j \mathbf{v}_j) = \sum_{j=1}^{n} x_j T(\mathbf{v}_j) \\ &= \sum_{j=1}^{n} x_j (\sum_{i=1}^{m} a_{ij}\mathbf{w}_i) \\ &= \sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij} x_j)\mathbf{w}_i. \end{aligned}$$

Define a matrix $A$ by $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$. Then the coordinates of the vector $T(\mathbf{x})$ with respect to the ordered basis $\mathcal{B}_2$ is

$$\begin{aligned} [T(\mathbf{x})]_{\mathcal{B}_2} &= \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= A\,[\mathbf{x}]_{\mathcal{B}_1}. \end{aligned}$$

The matrix $A$ is called the matrix of the linear transformation $T$ with respect to the ordered bases $\mathcal{B}_1$ and $\mathcal{B}_2$, and is denoted by $T[\mathcal{B}_1, \mathcal{B}_2]$.

We thus have the following theorem.

**Theorem 4.2.1** Let $V$ and $W$ be finite dimensional vector spaces with dimensions $n$ and $m$, respectively. Let $T : V \longrightarrow W$ be a linear transformation. If $\mathcal{B}_1$ is an ordered basis of $V$ and $\mathcal{B}_2$ is an ordered basis of $W$, then there exists an $m \times n$ matrix $A = T[\mathcal{B}_1, \mathcal{B}_2]$ such that

$$[T(\mathbf{x})]_{\mathcal{B}_2} = A\ \ [x]_{\mathcal{B}_1}.$$

**Remark 4.2.2** *Let $\mathcal{B}_1 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis of $V$ and $\mathcal{B}_2 = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ be an ordered basis of $W$. Let $T : V \longrightarrow W$ be a linear transformation with $A = T[\mathcal{B}_1, \mathcal{B}_2]$. Then the first column of $A$ is the coordinate of the vector $T(\mathbf{v}_1)$ in the basis $\mathcal{B}_2$. In general, the $i^{th}$ column of $A$ is the coordinate of the vector $T(\mathbf{v}_i)$ in the basis $\mathcal{B}_2$.*

We now give a few examples to understand the above discussion and the theorem.

**Example 4.2.3** 1. Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a linear transformation, given by

$$T(\ (x,y)\ ) = (x+y, x-y).$$

We obtain $T[\mathcal{B}_1, \mathcal{B}_2]$, the matrix of the linear transformation $T$ with respect to the ordered bases

$$\mathcal{B}_1 = \big((1,0),(0,1)\big) \quad \text{and} \quad \mathcal{B}_2 = \big((1,1),(1,-1)\big) \quad \text{of} \quad \mathbb{R}^2.$$

For any vector

$$(x,y) \in \mathbb{R}^2, \ \ [(x,y)]_{\mathcal{B}_1} = \begin{bmatrix} x \\ y \end{bmatrix}$$

as $(x,y) = x(1,0) + y(0,1)$. Also, by definition of the linear transformation $T$, we have

$$T(\ (1,0)\ ) = (1,1) = 1 \cdot (1,1) + 0 \cdot (1,-1). \ \text{So,} \ [T(\ (1,0)\ )]_{\mathcal{B}_2} = (1,0)^t$$

and

$$T(\ (0,1)\ ) = (1,-1) = 0 \cdot (1,1) + 1 \cdot (1,-1).$$

That is, $[T(\ (0,1)\ )]_{\mathcal{B}_2} = (0,1)^t$. So the $T[\mathcal{B}_1, \mathcal{B}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Observe that in this case,

$$[T(\ (x,y)\ )]_{\mathcal{B}_2} = [(x+y, x-y)]_{\mathcal{B}_2} = x(1,1) + y(1,-1) = \begin{bmatrix} x \\ y \end{bmatrix}, \ \text{and}$$

$$T[\mathcal{B}_1, \mathcal{B}_2]\ [(x,y)]_{\mathcal{B}_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} = [T(\ (x,y)\ )]_{\mathcal{B}_2}.$$

2. Let $\mathcal{B}_1 = \big((1,0,0),(0,1,0),(0,0,1)\big)$, $\mathcal{B}_2 = \big((1,0,0),(1,1,0),(1,1,1)\big)$ be two ordered bases of $\mathbb{R}^3$. Define

$$T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3 \ \text{by} \ T(\mathbf{x}) = \mathbf{x}.$$

Then

$$\begin{aligned} T((1,0,0)) &= 1 \cdot (1,0,0) + 0 \cdot (1,1,0) + 0 \cdot (1,1,1), \\ T((0,1,0)) &= -1 \cdot (1,0,0) + 1 \cdot (1,1,0) + 0 \cdot (1,1,1), \ \text{and} \\ T((0,0,1)) &= 0 \cdot (1,0,0) + (-1) \cdot (1,1,0) + 1 \cdot (1,1,1). \end{aligned}$$

Thus, we have

$$\begin{aligned} T[\mathcal{B}_1, \mathcal{B}_2] &= [[T((1,0,0))]_{\mathcal{B}_2}, \ [T((0,1,0))]_{\mathcal{B}_2}, \ [T((0,0,1))]_{\mathcal{B}_2}] \\ &= [(1,0,0)^t, \ (-1,1,0)^t, \ (0,-1,1)^t] \\ &= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Similarly check that $T[\mathcal{B}_1, \mathcal{B}_1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

3. Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ be define by $T((x,y,z)) = (x+y-z, x+z)$. Let $\mathcal{B}_1 = \big((1,0,0),(0,1,0),(0,0,1)\big)$ and $\mathcal{B}_2 = \big((1,0),(0,1)\big)$ be the ordered bases of the domain and range space, respectively. Then
$$T[\mathcal{B}_1, \mathcal{B}_2] = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$
Check that that $[T(x,y,z)]_{\mathcal{B}_2} = T[\mathcal{B}_1, \mathcal{B}_2] \ [(x,y,z)]_{\mathcal{B}_1}$.

**Exercise 4.2.4** Recall the space $\mathcal{P}_n(\mathbb{R})$ ( the vector space of all polynomials of degree less than or equal to $n$). We define a linear transformation $D : \mathcal{P}_n(\mathbb{R}) \longrightarrow \mathcal{P}_n(\mathbb{R})$ by
$$D(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1}.$$
Find the matrix of the linear transformation $D$.

However, note that the image of the linear transformation is contained in $\mathcal{P}_{n-1}(\mathbb{R})$.

**Remark 4.2.5** *1. Observe that*
$$T[\mathcal{B}_1, \mathcal{B}_2] = [[T(\mathbf{v}_1)]_{\mathcal{B}_2}, [T(\mathbf{v}_2)]_{\mathcal{B}_2}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}_2}].$$

*2. It is important to note that*
$$[T(\mathbf{x})]_{\mathcal{B}_2} = T[\mathcal{B}_1, \mathcal{B}_2] \ [\mathbf{x}]_{\mathcal{B}_1}.$$
*That is, we multiply the matrix of the linear transformation with the coordinates $[\mathbf{x}]_{\mathcal{B}_1}$, of the vector $\mathbf{x} \in V$ to obtain the coordinates of the vector $T(\mathbf{x}) \in W$.*

*3. If $A$ is an $m \times n$ matrix, then $A$ induces a linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, defined by*
$$T_A(\mathbf{x}) = A\mathbf{x}.$$
*We sometimes write $A$ for $T_A$. Suppose that the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ are the ordered bases $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. Then observe that*
$$T[\mathcal{B}_1, \mathcal{B}_2] = A.$$

## 4.3 Rank-Nullity Theorem

**Definition 4.3.1 (Range and Null Space)** Let $V, W$ be finite dimensional vector spaces over the same set of scalars and $T : V \longrightarrow W$ be a linear transformation. We define

1. $\mathcal{R}(T) = \{T(\mathbf{x}) : \mathbf{x} \in V\}$, and
2. $\mathcal{N}(T) = \{\mathbf{x} \in V : T(\mathbf{x}) = \mathbf{0}\}$.

**Proposition 4.3.2** Let $V$ and $W$ be finite dimensional vector spaces and let $T : V \longrightarrow W$ be a linear transformation. Suppose that $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is an ordered basis of $V$. Then

1. (a) $\mathcal{R}(T)$ is a subspace of $W$.
   (b) $\mathcal{R}(T) = L(T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n))$.
   (c) $\dim(\mathcal{R}(T)) \leq \dim(W)$.
2. (a) $\mathcal{N}(T)$ is a subspace of $V$.
   (b) $\dim(\mathcal{N}(T)) \leq \dim(V)$.

3. $T$ is one-one $\Longleftrightarrow$ $\mathcal{N}(T) = \{\mathbf{0}\}$ is the zero subspace of $V \Longleftrightarrow$ $\{T(u_i) : 1 \leq i \leq n\}$ is a basis of $\mathcal{R}(T)$.

4. $\dim(\mathcal{R}(T)) = \dim(V)$ if and only if $\mathcal{N}(T) = \{\mathbf{0}\}$.

Proof. The results about $\mathcal{R}(T)$ and $\mathcal{N}(T)$ can be easily proved. We thus leave the proof for the readers.
We now assume that $T$ is one-one. We need to show that $\mathcal{N}(T) = \{\mathbf{0}\}$.
Let $\mathbf{u} \in \mathcal{N}(T)$. Then by definition, $T(\mathbf{u}) = \mathbf{0}$. Also for any linear transformation (see Proposition 4.1.3), $T(\mathbf{0}) = \mathbf{0}$. Thus $T(\mathbf{u}) = T(\mathbf{0})$. So, $T$ is one-one implies $\mathbf{u} = \mathbf{0}$. That is, $\mathcal{N}(T) = \{\mathbf{0}\}$.

Let $\mathcal{N}(T) = \{\mathbf{0}\}$. We need to show that $T$ is one-one. So, let us assume that for some $\mathbf{u}, \mathbf{v} \in V$, $T(\mathbf{u}) = T(\mathbf{v})$. Then, by linearity of $T$, $T(\mathbf{u}-\mathbf{v}) = \mathbf{0}$. This implies, $\mathbf{u}-\mathbf{v} \in \mathcal{N}(T) = \{\mathbf{0}\}$. This in turn implies $\mathbf{u} = \mathbf{v}$. Hence, $T$ is one-one.

The other parts can be similarly proved. □

**Remark 4.3.3** *1. The space $\mathcal{R}(T)$ is called the* RANGE SPACE *of $T$ and $\mathcal{N}(T)$ is called the* NULL SPACE *of $T$.*

*2. We write $\rho(T) = \dim(\mathcal{R}(T))$ and $\nu(T) = \dim(\mathcal{N}(T))$.*

*3. $\rho(T)$ is called the rank of the linear transformation $T$ and $\nu(T)$ is called the nullity of $T$.*

**Example 4.3.4** Determine the range and null space of the linear transformation

$$T : \mathbb{R}^3 \longrightarrow \mathbb{R}^4 \quad \text{with} \quad T(x,y,z) = (x-y+z, y-z, x, 2x-5y+5z).$$

**Solution:** By Definition $\mathcal{R}(T) = L(T(1,0,0), T(0,1,0), T(0,0,1))$. We therefore have

$$\begin{aligned}
\mathcal{R}(T) &= L\big((1,0,1,2), (-1,1,0,-5), (1,-1,0,5)\big) \\
&= L\big((1,0,1,2), (1,-1,0,5)\big) \\
&= \{\alpha(1,0,1,2) + \beta(1,-1,0,5) \ : \alpha, \beta \in \mathbb{R}\} \\
&= \{(\alpha+\beta, -\beta, \alpha, 2\alpha+5\beta) \ : \alpha, \beta \in \mathbb{R}\} \\
&= \{(x,y,z,w) \in \mathbb{R}^4 \ : x+y-z = 0, 5y-2z+w = 0\}.
\end{aligned}$$

Also, by definition

$$\begin{aligned}
\mathcal{N}(T) &= \{(x,y,z) \in \mathbb{R}^3 \ : T(x,y,z) = \mathbf{0}\} \\
&= \{(x,y,z) \in \mathbb{R}^3 \ : (x-y+z, y-z, x, 2x-5y+5z) = \mathbf{0}\} \\
&= \{(x,y,z) \in \mathbb{R}^3 \ : x-y+z = 0, y-z = 0, \\
&\qquad\qquad x = 0, 2x-5y+5z = 0\} \\
&= \{(x,y,z) \in \mathbb{R}^3 \ : y-z = 0, x = 0\} \\
&= \{(x,y,z) \in \mathbb{R}^3 \ : y = z, x = 0\} \\
&= \{(0,y,y) \in \mathbb{R}^3 \ : y \text{ arbitrary}\} \\
&= L((0,1,1))
\end{aligned}$$

**Exercise 4.3.5** 1. Let $T : V \longrightarrow W$ be a linear transformation and let $\{T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)\}$ be linearly independent in $\mathcal{R}(T)$. Prove that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset V$ is linearly independent.

2. Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ be defined by
$$T\big((1,0)\big) = (1,0,0), \; T\big((0,1)\big) = (1,0,0).$$
Then the vectors $(1,0)$ and $(0,1)$ are linearly independent whereas $T\big((1,0)\big)$ and $T\big((0,1)\big)$ are linearly dependent.

3. Is there a linear transformation
$$T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2 \text{ such that } T(1,-1,1) = (1,2), \quad \text{and} \quad T(-1,1,2) = (1,0)?$$

4. Recall the vector space $\mathcal{P}_n(\mathbb{R})$. Define a linear transformation
$$D : \mathcal{P}_n(\mathbb{R}) \longrightarrow \mathcal{P}_n(\mathbb{R})$$
by
$$D(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1}.$$
Describe the null space and range space of $D$. Note that the range space is contained in the space $\mathcal{P}_{n-1}(\mathbb{R})$.

5. Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be defined by
$$T(1,0,0) = (0,0,1), \; T(1,1,0) = (1,1,1) \text{ and } T(1,1,1) = (1,1,0).$$
   (a) Find $T(x,y,z)$ for $x,y,z \in \mathbb{R}$,

   (b) Find $\mathcal{R}(T)$ and $\mathcal{N}(T)$. Also calculate $\rho(T)$ and $\nu(T)$.

   (c) Show that $T^3 = T$ and find the matrix of the linear transformation with respect to the standard basis.

6. Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a linear transformation with
$$T((3,4)) = (0,1), \; T((-1,1)) = (2,3).$$
Find the matrix representation $T[\mathcal{B},\mathcal{B}]$ of $T$ with respect to the ordered basis $\mathcal{B} = \big((1,0),(1,1)\big)$ of $\mathbb{R}^2$.

7. Determine a linear transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ whose range space is $L\{(1,2,0),(0,1,1),(1,3,1)\}$.

8. Suppose the following chain of matrices is given.
$$A \longrightarrow B_1 \longrightarrow B_1 \longrightarrow B_2 \cdots \longrightarrow B_{k-1} \longrightarrow B_k \longrightarrow B.$$
If row space of $B$ is in the row space of $B_k$ and the row space of $B_l$ is in the row space of $B_{l-1}$ for $2 \le l \le k$ then show that the row space of $B$ is in the row space of $A$.

We now state and prove the rank-nullity Theorem. This result also follows from Proposition 4.3.2.

**Theorem 4.3.6 (Rank Nullity Theorem)** Let $T : V \longrightarrow W$ be a linear transformation and $V$ be a finite dimensional vector space. Then
$$\dim(\mathcal{R}(T)) + \dim(\mathcal{N}(T)) = \dim(V),$$
or equivalently $\rho(T) + \nu(T) = \dim(V)$.

PROOF. Let $\dim(V) = n$ and $\dim(\mathcal{N}(T)) = r$. Suppose $\{u_1, u_2, \ldots, u_r\}$ is a basis of $\mathcal{N}(T)$. Since $\{u_1, u_2, \ldots, u_r\}$ is a linearly independent set in $V$, we can extend it to form a basis of $V$ (see Corollary 3.3.15). So, there exist vectors $\{u_{r+1}, u_{r+2}, \ldots, u_n\}$ such that $\{u_1, \ldots, u_r, u_{r+1}, \ldots, u_n\}$ is a basis of $V$. Therefore, by Proposition 4.3.2

$$\begin{aligned}\mathcal{R}(T) &= L(T(u_1), T(u_2), \ldots, T(u_n)) \\ &= L(\mathbf{0}, \ldots, \mathbf{0}, T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)) \\ &= L(T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)).\end{aligned}$$

We now prove that the set $\{T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)\}$ is linearly independent. Suppose the set is not linearly independent. Then, there exists scalars, $\alpha_{r+1}, \alpha_{r+2}, \ldots, \alpha_n$, not all zero such that

$$\alpha_{r+1}T(u_{r+1}) + \alpha_{r+2}T(u_{r+2}) + \cdots + \alpha_n T(u_n) = \mathbf{0}.$$

That is,

$$T(\alpha_{r+1}u_{r+1} + \alpha_{r+2}u_{r+2} + \cdots + \alpha_n u_n) = \mathbf{0}.$$

So, by definition of $\mathcal{N}(T)$,

$$\alpha_{r+1}u_{r+1} + \alpha_{r+2}u_{r+2} + \cdots + \alpha_n u_n \in \mathcal{N}(T) = L(u_1, \ldots, u_r).$$

Hence, there exists scalars $\alpha_i,\ 1 \le i \le r$ such that

$$\alpha_{r+1}u_{r+1} + \alpha_{r+2}u_{r+2} + \cdots + \alpha_n u_n = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r.$$

That is,

$$\alpha_1 u_1 + + \cdots + \alpha_r u_r - \alpha_{r+1}u_{r+1} - \cdots - \alpha_n u_n = \mathbf{0}.$$

But the set $\{u_1, u_2, \ldots, u_n\}$ is a basis of $V$ and so linearly independent. Thus by definition of linear independence

$$\alpha_i = 0 \text{ for all } i,\ 1 \le i \le n.$$

In other words, we have shown that $\{T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)\}$ is a basis of $\mathcal{R}(T)$. Hence,

$$\dim(\mathcal{R}(T)) + \dim(\mathcal{N}(T)) = (n - r) + r = n = \dim(V).$$

□

Using the Rank-nullity theorem, we give a short proof of the following result.

**Corollary 4.3.7** Let $T : V \longrightarrow V$ be a linear transformation on a finite dimensional vector space $V$. Then

$$T \text{ is one-one} \iff T \text{ is onto} \iff T \text{ is invertible.}$$

PROOF. By Proposition 4.3.2, $T$ is one-one if and only if $\mathcal{N}(T) = \{\mathbf{0}\}$. By the rank-nullity Theorem 4.3.6 $\mathcal{N}(T) = \{\mathbf{0}\}$ is equivalent to the condition $\dim(\mathcal{R}(T)) = \dim(V)$. Or equivalently $T$ is onto.

By definition, $T$ is invertible if $T$ is one-one and onto. But we have shown that $T$ is one-one if and only if $T$ is onto. Thus, we have the last equivalent condition. □

**Remark 4.3.8** *Let $V$ be a finite dimensional vector space and let $T : V \longrightarrow V$ be a linear transformation. If either $T$ is one-one or $T$ is onto, then $T$ is invertible.*

The following are some of the consequences of the rank-nullity theorem. The proof is left as an exercise for the reader.

**Corollary 4.3.9** The following are equivalent for an $m \times n$ real matrix $A$.

1. Rank $(A) = k$.
2. There exist exactly $k$ rows of $A$ that are linearly independent.
3. There exist exactly $k$ columns of $A$ that are linearly independent.
4. There is a $k \times k$ submatrix of $A$ with non-zero determinant and every $(k+1) \times (k+1)$ submatrix of $A$ has zero determinant.
5. The dimension of the range space of $A$ is $k$.
6. There is a subset of $\mathbb{R}^m$ consisting of exactly $k$ linearly independent vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k$ such that the system $A\mathbf{x} = \mathbf{b}_i$ for $1 \leq i \leq k$ is consistent.
7. The dimension of the null space of $A = n - k$.

**Exercise 4.3.10** 1. Let $T : V \longrightarrow W$ be a linear transformation.

(a) If $V$ is finite dimensional then show that the null space and the range space of $T$ are also finite dimensional.

(b) If $V$ and $W$ are both finite dimensional then show that

i. if $\dim(V) < \dim(W)$ then $T$ is onto.

ii. if $\dim(V) > \dim(W)$ then $T$ is not one-one.

2. Let $A$ be an $m \times n$ real matrix. Then

(a) if $n > m$, then the system $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions,

(b) if $n < m$, then there exists a non-zero vector $\mathbf{b} = (b_1, b_2, \ldots, b_m)^t$ such that the system $A\mathbf{x} = \mathbf{b}$ does not have any solution.

3. Let $A$ be an $m \times n$ matrix. Prove that
Row Rank $(A) =$ Column Rank $(A)$.
*[Hint: Define $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $T_A(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. Let Row Rank $(A) = r$. Use Theorem 2.6.1 to show, $A\mathbf{x} = \mathbf{0}$ has $n - r$ linearly independent solutions. This implies,*
$\nu(T_A) = \dim(\{\mathbf{v} \in \mathbb{R}^n : T_A(\mathbf{v}) = \mathbf{0}\}) = \dim(\{\mathbf{v} \in \mathbb{R}^n : A\mathbf{v} = \mathbf{0}\}) = n - r$.
*Now observe that $\mathcal{R}(T_A)$ is the linear span of columns of $A$ and use the rank-nullity Theorem 4.3.6 to get the required result.]*

4. Prove Theorem 2.6.1.
*[Hint: Consider the linear system of equation $A\mathbf{x} = \mathbf{b}$ with the orders of $A, \mathbf{x}$ and $\mathbf{b}$, respectively as $m \times n, n \times 1$ and $m \times 1$. Define a linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $T(\mathbf{v}) = A\mathbf{v}$. First observe that if the solution exists then $\mathbf{b}$ is a linear combination of the columns of $A$ and the linear span of the columns of $A$ give us $\mathcal{R}(T)$. Note that $\rho(A) =$ column rank$(A) = \dim(\mathcal{R}(T)) = \ell$(say). Then for part $i)$ one can proceed as follows.*
*$i)$ Let $C_{i_1}, C_{i_2}, \ldots, C_{i_\ell}$ be the linearly independent columns of $A$. Then rank$(A) <$ rank$([A \; b])$ implies that $\{C_{i_1}, C_{i_2}, \ldots, C_{i_\ell}, \mathbf{b}\}$ is linearly independent. Hence $\mathbf{b} \notin L(C_{i_1}, C_{i_2}, \ldots, C_{i_\ell})$. Hence, the system doesn't have any solution.*

*On similar lines prove the other two parts.]*

5. Let $T, S : V \longrightarrow V$ be linear transformations with $\dim(V) = n$.

(a) Show that $\mathcal{R}(T+S) \subset \mathcal{R}(T) + \mathcal{R}(S)$. Deduce that $\rho(T+S) \le \rho(T) + \rho(S)$.
*Hint: For two subspaces $M, N$ of a vector space $V$, recall the definition of the vector subspace $M+N$.*

(b) Use the above and the rank-nullity Theorem 4.3.6 to prove $\nu(T+S) \ge \nu(T) + \nu(S) - n$.

6. Let $V$ be the complex vector space of all complex polynomials of degree at most $n$. Given $k$ distinct complex numbers $z_1, z_2, \ldots, z_k$, we define a linear transformation

$$T : V \longrightarrow \mathbb{C}^k \quad \text{by} \quad T\big(P(z)\big) = \big(P(z_1), P(z_2), \ldots, P(z_k)\big).$$

For each $k \ge 1$, determine the dimension of the range space of $T$.

7. Let $A$ be an $n \times n$ real matrix with $A^2 = A$. Consider the linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, defined by $T_A(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. Prove that

(a) $T_A \circ T_A = T_A$ (use the condition $A^2 = A$).

(b) $\mathcal{N}(T_A) \cap \mathcal{R}(T_A) = \{\mathbf{0}\}$.
*Hint: Let $\mathbf{x} \in \mathcal{N}(T_A) \cap \mathcal{R}(T_A)$. This implies $T_A(\mathbf{x}) = \mathbf{0}$ and $\mathbf{x} = T_A(\mathbf{y})$ for some $\mathbf{y} \in \mathbb{R}^n$. So,*

$$\mathbf{x} = T_A(\mathbf{y}) = (T_A \circ T_A)(\mathbf{y}) = T_A\big(T_A(\mathbf{y})\big) = T_A(\mathbf{x}) = \mathbf{0}.$$

(c) $\mathbb{R}^n = \mathcal{N}(T_A) + \mathcal{R}(T_A)$.
*Hint: Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $\mathcal{N}(T_A)$. Extend it to get a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$. Then by Rank-nullity Theorem 4.3.6, $\{T_A(\mathbf{v}_{k+1}), \ldots, T_A(\mathbf{v}_n)\}$ is a basis of $\mathcal{R}(T_A)$.*

## 4.4 Similarity of Matrices

In the last few sections, the following has been discussed in detail:
Given a finite dimensional vector space $V$ of dimension $n$, we fixed an ordered basis $\mathcal{B}$. For any $\mathbf{v} \in V$, we calculated the column vector $[\mathbf{v}]_{\mathcal{B}}$, to obtain the coordinates of $\mathbf{v}$ with respect to the ordered basis $\mathcal{B}$. Also, for any linear transformation $T : V \longrightarrow V$, we got an $n \times n$ matrix $T[\mathcal{B}, \mathcal{B}]$, the matrix of $T$ with respect to the ordered basis $\mathcal{B}$. That is, once an ordered basis of $V$ is fixed, every linear transformation is represented by a matrix with entries from the scalars.

In this section, we understand the matrix representation of $T$ in terms of different bases $\mathcal{B}_1$ and $\mathcal{B}_2$ of $V$. That is, we relate the two $n \times n$ matrices $T[\mathcal{B}_1, \mathcal{B}_1]$ and $T[\mathcal{B}_2, \mathcal{B}_2]$. We start with the following important theorem. This theorem also enables us to understand WHY THE MATRIX PRODUCT IS DEFINED SOMEWHAT DIFFERENTLY.

**Theorem 4.4.1 (Composition of Linear Transformations)** Let $V$, $W$ and $Z$ be finite dimensional vector spaces with ordered bases $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$, respectively. Also, let $T : V \longrightarrow W$ and $S : W \longrightarrow Z$ be linear transformations. Then the composition map $S \circ T : V \longrightarrow Z$ is a linear transformation and

$$(S \circ T)\,[\mathcal{B}_1, \mathcal{B}_3] = S[\mathcal{B}_2, \mathcal{B}_3]\;\; T[\mathcal{B}_1, \mathcal{B}_2].$$

PROOF. Let $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$, $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ and $\mathcal{B}_3 = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p)$ be ordered bases of $V, W$ and $Z$, respectively. Then

$$(S \circ T)\,[\mathcal{B}_1, \mathcal{B}_3] = [[S \circ T(\mathbf{u}_1)]_{\mathcal{B}_3},\; [S \circ T(\mathbf{u}_2)]_{\mathcal{B}_3}, \ldots, [S \circ T(\mathbf{u}_n)]_{\mathcal{B}_3}].$$

Now for $1 \leq t \leq n$,

$$\begin{aligned}
(S \circ T)\ (\mathbf{u}_t) &= S(T(\mathbf{u}_t)) = S\Big(\sum_{j=1}^{m} (T[\mathcal{B}_1, \mathcal{B}_2])_{jt} \mathbf{v}_j\Big) = \sum_{j=1}^{m} (T[\mathcal{B}_1, \mathcal{B}_2])_{jt} S(\mathbf{v}_j) \\
&= \sum_{j=1}^{m} (T[\mathcal{B}_1, \mathcal{B}_2])_{jt} \sum_{k=1}^{p} (S[\mathcal{B}_2, \mathcal{B}_3])_{kj} \mathbf{w}_k \\
&= \sum_{k=1}^{p} (\sum_{j=1}^{m} (S[\mathcal{B}_2, \mathcal{B}_3])_{kj} (T[\mathcal{B}_1, \mathcal{B}_2])_{jt}) \mathbf{w}_k \\
&= \sum_{k=1}^{p} (S[\mathcal{B}_2, \mathcal{B}_3]\ \ T[\mathcal{B}_1, \mathcal{B}_2])_{kt} \mathbf{w}_k.
\end{aligned}$$

So,

$$[(S \circ T)\ (\mathbf{u}_t)]_{\mathcal{B}_3} = ((S[\mathcal{B}_2, \mathcal{B}_3]\ T[\mathcal{B}_1, \mathcal{B}_2])_{1t}, \ldots, (S[\mathcal{B}_2, \mathcal{B}_3]\ T[\mathcal{B}_1, \mathcal{B}_2])_{pt})^t.$$

Hence,

$$(S \circ T)\ [\mathcal{B}_1, \mathcal{B}_3] = \big[\ [(S \circ T)\ (u_1)]_{\mathcal{B}_3}, \ldots, [(S \circ T)\ (u_n)]_{\mathcal{B}_3}\big] = S[\mathcal{B}_2, \mathcal{B}_3]\ \ T[\mathcal{B}_1, \mathcal{B}_2].$$

This completes the proof. □

**Proposition 4.4.2** Let $V$ be a finite dimensional vector space and let $T, S : V \longrightarrow V$ be a linear transformations. Then

$$\nu(T) + \nu(S) \geq \nu(T \circ S) \geq \max\{\nu(T), \nu(S)\}.$$

PROOF. We first prove the second inequality.
Suppose that $\mathbf{v} \in \mathcal{N}(S)$. Then $T \circ S(\mathbf{v}) = T(S(\mathbf{v})) = T(\mathbf{0}) = \mathbf{0}$. So, $\mathcal{N}(S) \subset \mathcal{N}(T \circ S)$. Therefore, $\nu(S) \leq \nu(T \circ S)$.

Suppose $\dim(V) = n$. Then using the rank-nullity theorem, observe that

$$\nu(T \circ S) \geq \nu(T) \Longleftrightarrow n - \nu(T \circ S) \leq n - \nu(T) \Longleftrightarrow \rho(T \circ S) \leq \rho(T).$$

So, to complete the proof of the second inequality, we need to show that $\mathcal{R}(T \circ S) \subset \mathcal{R}(T)$. This is true as $\mathcal{R}(S) \subset V$.

We now prove the first inequality.
Let $k = \nu(S)$ and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a basis of $\mathcal{N}(S)$. Clearly, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subset \mathcal{N}(T \circ S)$ as $T(\mathbf{0}) = \mathbf{0}$. We extend it to get a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell\}$ of $\mathcal{N}(T \circ S)$.

**Claim:** The set $\{S(\mathbf{u}_1), S(\mathbf{u}_2), \ldots, S(\mathbf{u}_\ell)\}$ is linearly independent subset of $\mathcal{N}(T)$.

As $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell \in \mathcal{N}(T \circ S)$, the set $\{S(\mathbf{u}_1), S(\mathbf{u}_2), \ldots, S(\mathbf{u}_\ell)\}$ is a subset of $\mathcal{N}(T)$. Let if possible the given set be linearly dependent. Then there exist non-zero scalars $c_1, c_2, \ldots, c_\ell$ such that

$$c_1 S(\mathbf{u}_1) + c_2 S(\mathbf{u}_2) + \cdots + c_\ell S(\mathbf{u}_\ell) = \mathbf{0}.$$

So, the vector $\sum\limits_{i=1}^{\ell} c_i \mathbf{u}_i \in \mathcal{N}(S)$ and is a linear combination of the basis vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ of $\mathcal{N}(S)$. Therefore, there exist scalars $\alpha_1, \alpha_2, \alpha_k$ such that

$$\sum_{i=1}^{\ell} c_i \mathbf{u}_i = \sum_{i=1}^{k} \alpha_i \mathbf{v}_i.$$

Or equivalently

$$\sum_{i=1}^{\ell} c_i \mathbf{u}_i + \sum_{i=1}^{k} (-\alpha_i) \mathbf{v}_i = \mathbf{0}.$$

That is, the $\mathbf{0}$ vector is a non-trivial linear combination of the basis vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell$ of $\mathcal{N}(T \circ S)$. A contradiction.

Thus, the set $\{S(\mathbf{u}_1), S(\mathbf{u}_2), \ldots, S(\mathbf{u}_\ell)\}$ is a linearly independent subset of $\mathcal{N}(T)$ and so $\nu(T) \geq \ell$. Hence,

$$\nu(T \circ S) = k + \ell \leq \nu(S) + \nu(T).$$

□

Recall from Theorem 4.1.8 that if $T$ is an invertible linear Transformation, then $T^{-1} : V \longrightarrow V$ is a linear transformation defined by $T^{-1}(\mathbf{u}) = \mathbf{v}$ whenever $T(\mathbf{v}) = \mathbf{u}$. We now state an important result about inverse of a linear transformation. The reader is required to supply the proof (use Theorem 4.4.1).

**Theorem 4.4.3 (Inverse of a Linear Transformation)** Let $V$ be a finite dimensional vector space with ordered bases $\mathcal{B}_1$ and $\mathcal{B}_2$. Also let $T : V \longrightarrow V$ be an invertible linear transformation. Then the matrix of $T$ and $T^{-1}$ are related by

$$T[\mathcal{B}_1, \mathcal{B}_2]^{-1} = T^{-1}[\mathcal{B}_2, \mathcal{B}_1].$$

**Exercise 4.4.4** For the linear transformations given below, find the matrix $T[\mathcal{B}, \mathcal{B}]$.

1. Let $\mathcal{B} = \big((1,1,1), (1,-1,1), (1,1,-1)\big)$ be an ordered basis of $\mathbb{R}^3$. Define $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ by $T(1,1,1) = (1,-1,1)$, $T(1,-1,1) = (1,1,-1)$, and $T(1,1,-1) = (1,1,1)$. Is $T$ an invertible linear transformation? Give reasons.

2. Let $\mathcal{B} = \big(1, x, x^2, x^3)\big)$ be an ordered basis of $\mathcal{P}_3(\mathbb{R})$. Define $T : \mathcal{P}_3(\mathbb{R}) \longrightarrow \mathcal{P}_3(\mathbb{R})$ by

$$T(1) = 1, T(x) = 1 + x, T(x^2) = (1+x)^2, \text{ and } T(x^3) = (1+x)^3.$$

Prove that $T$ is an invertible linear transformation. Also, find $T^{-1}[\mathcal{B}, \mathcal{B}]$.

Let $V$ be a vector space with $\dim(V) = n$. Let $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be two ordered bases of $V$. Recall from Definition 4.1.5 that $I : V \longrightarrow V$ is the identity linear transformation defined by $I(\mathbf{x}) = \mathbf{x}$ for every $\mathbf{x} \in V$. Suppose $\mathbf{x} \in V$ with $[\mathbf{x}]_{\mathcal{B}_1} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t$ and $[\mathbf{x}]_{\mathcal{B}_2} = (\beta_1, \beta_2, \ldots, \beta_n)^t$.

We now express each vector in $\mathcal{B}_2$ as a linear combination of the vectors from $\mathcal{B}_1$. Since $\mathbf{v}_i \in V$, for $1 \leq i \leq n$, and $\mathcal{B}_1$ is a basis of $V$, we can find scalars $a_{ij}, 1 \leq i, j \leq n$ such that

$$\mathbf{v}_i = I(\mathbf{v}_i) = \sum_{j=1}^{n} a_{ji}\mathbf{u}_j \quad \text{for all } \; i, 1 \leq i \leq n.$$

Hence, $[I(\mathbf{v}_i)]_{\mathcal{B}_1} = [\mathbf{v}_i]_{\mathcal{B}_1} = (a_{1i}, a_{2i}, \cdots, a_{ni})^t$ and

$$\begin{aligned} I[\mathcal{B}_2, \mathcal{B}_1] &= [[I(\mathbf{v}_1)]_{\mathcal{B}_1}, [I(\mathbf{v}_2)]_{\mathcal{B}_1}, \ldots, [I(\mathbf{v}_n)]_{\mathcal{B}_1}] \\ &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}. \end{aligned}$$

Thus, we have proved the following result.

**Theorem 4.4.5 (Change of Basis Theorem)** Let $V$ be a finite dimensional vector space with ordered bases $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Suppose $\mathbf{x} \in V$ with $[\mathbf{x}]_{\mathcal{B}_1} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t$ and $[\mathbf{x}]_{\mathcal{B}_2} = (\beta_1, \beta_2, \ldots, \beta_n)^t$. Then

$$[\mathbf{x}]_{\mathcal{B}_1} = I[\mathcal{B}_2, \mathcal{B}_1] \;\; [\mathbf{x}]_{\mathcal{B}_2}.$$

Equivalently,

$$\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix} \begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{bmatrix}.$$

**Note:** Observe that the identity linear transformation $I : V \longrightarrow V$ defined by $I(\mathbf{x}) = \mathbf{x}$ for every $\mathbf{x} \in V$ is invertible and

$$I[\mathcal{B}_2, \mathcal{B}_1]^{-1} = I^{-1}[\mathcal{B}_1, \mathcal{B}_2] = I[\mathcal{B}_1, \mathcal{B}_2].$$

Therefore, we also have

$$[\mathbf{x}]_{\mathcal{B}_2} = I[\mathcal{B}_1, \mathcal{B}_2]\ [\mathbf{x}]_{\mathcal{B}_1}.$$

Let $V$ be a finite dimensional vector space and let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two ordered bases of $V$. Let $T : V \longrightarrow V$ be a linear transformation. We are now in a position to relate the two matrices $T[\mathcal{B}_1, \mathcal{B}_1]$ and $T[\mathcal{B}_2, \mathcal{B}_2]$.

**Theorem 4.4.6** Let $V$ be a finite dimensional vector space and let $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be two ordered bases of $V$. Let $T : V \longrightarrow V$ be a linear transformation with $B = T[\mathcal{B}_1, \mathcal{B}_1]$ and $C = T[\mathcal{B}_2, \mathcal{B}_2]$ as matrix representations of $T$ in bases $\mathcal{B}_1$ and $\mathcal{B}_2$.

Also, let $A = [a_{ij}] = I[\mathcal{B}_2, \mathcal{B}_1]$, be the matrix of the identity linear transformation with respect to the bases $\mathcal{B}_1$ and $\mathcal{B}_2$. Then $BA = AC$. Equivalently $B = ACA^{-1}$.

Proof. For any $\mathbf{x} \in V$, we represent $[T(\mathbf{x})]_{\mathcal{B}_2}$ in two ways. Using Theorem 4.2.1, the first expression is

$$[T(\mathbf{x})]_{\mathcal{B}_2} = T[\mathcal{B}_2, \mathcal{B}_2]\ [\mathbf{x}]_{\mathcal{B}_2}. \tag{4.4.1}$$

Using Theorem 4.4.5, the other expression is

$$\begin{aligned}[T(\mathbf{x})]_{\mathcal{B}_2} &= I[\mathcal{B}_1, \mathcal{B}_2]\ [T(\mathbf{x})]_{\mathcal{B}_1}\\ &= I[\mathcal{B}_1, \mathcal{B}_2]\ T[\mathcal{B}_1, \mathcal{B}_1]\ [\mathbf{x}]_{\mathcal{B}_1}\\ &= I[\mathcal{B}_1, \mathcal{B}_2]\ T[\mathcal{B}_1, \mathcal{B}_1]\ I[\mathcal{B}_2, \mathcal{B}_1]\ [\mathbf{x}]_{\mathcal{B}_2}. \end{aligned} \tag{4.4.2}$$

Hence, using (4.4.1) and (4.4.2), we see that for every $\mathbf{x} \in V$,

$$I[\mathcal{B}_1, \mathcal{B}_2]\ T[\mathcal{B}_1, \mathcal{B}_1]\ I[\mathcal{B}_2, \mathcal{B}_1]\ [\mathbf{x}]_{\mathcal{B}_2} = T[\mathcal{B}_2, \mathcal{B}_2]\ [\mathbf{x}]_{\mathcal{B}_2}.$$

Since the result is true for all $\mathbf{x} \in V$, we get

$$I[\mathcal{B}_1, \mathcal{B}_2]\ T[\mathcal{B}_1, \mathcal{B}_1]\ I[\mathcal{B}_2, \mathcal{B}_1] = T[\mathcal{B}_2, \mathcal{B}_2]. \tag{4.4.3}$$

That is, $A^{-1}BA = C$ or equivalently $ACA^{-1} = B$. □

**Another Proof:**

Let $B = [b_{ij}]$ and $C = [c_{ij}]$. Then for $1 \le i \le n$,

$$T(\mathbf{u}_i) = \sum_{j=1}^{n} b_{ji}\mathbf{u}_j \ \text{ and } \ T(\mathbf{v}_i) = \sum_{j=1}^{n} c_{ji}\mathbf{v}_j.$$

So, for each $j, 1 \le j \le n$,

$$\begin{aligned} T(\mathbf{v}_j) &= T(I(\mathbf{v}_j)) = T(\sum_{k=1}^{n} a_{kj}\mathbf{u}_k) = \sum_{k=1}^{n} a_{kj}T(\mathbf{u}_k)\\ &= \sum_{k=1}^{n} a_{kj}(\sum_{\ell=1}^{n} b_{\ell k}\mathbf{u}_\ell) = \sum_{\ell=1}^{n}(\sum_{k=1}^{n} b_{\ell k}a_{kj})\mathbf{u}_\ell \end{aligned}$$

and therefore,

$$[T(\mathbf{v}_j)]_{\mathcal{B}_1} = \begin{bmatrix} \sum\limits_{k=1}^{n} b_{1k}a_{kj} \\ \sum\limits_{k=1}^{n} b_{2k}a_{kj} \\ \vdots \\ \sum\limits_{k=1}^{n} b_{nk}a_{kj} \end{bmatrix} = B \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}.$$

Hence $T[\mathcal{B}_2, \mathcal{B}_1] = BA$.

Also, for each $j, 1 \leq j \leq n$,

$$\begin{aligned} T(\mathbf{v}_j) &= \sum_{k=1}^{n} c_{kj}\mathbf{v}_k = \sum_{k=1}^{n} c_{kj} I(\mathbf{v}_k) = \sum_{k=1}^{n} c_{kj} (\sum_{\ell=1}^{n} a_{\ell k}\mathbf{u}_\ell) \\ &= \sum_{\ell=1}^{n} (\sum_{k=1}^{n} a_{\ell k} c_{kj})\mathbf{u}_\ell \end{aligned}$$

and so

$$[T(\mathbf{v}_j)]_{\mathcal{B}_1} = \begin{bmatrix} \sum\limits_{k=1}^{n} a_{1k}c_{kj} \\ \sum\limits_{k=1}^{n} a_{2k}c_{kj} \\ \vdots \\ \sum\limits_{k=1}^{n} a_{nk}c_{kj} \end{bmatrix} = A \begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{nj} \end{bmatrix}.$$

This gives us $T[\mathcal{B}_2, \mathcal{B}_1] = AC$. We thus have $AC = T[\mathcal{B}_2, \mathcal{B}_1] = BA$. ■

Let $V$ be a vector space with $\dim(V) = n$, and let $T : V \longrightarrow V$ be a linear transformation. Then for each ordered basis $\mathcal{B}$ of $V$, we get an $n \times n$ matrix $T[\mathcal{B}, \mathcal{B}]$. Also, we know that for any vector space we have infinite number of choices for an ordered basis. So, as we change an ordered basis, the matrix of the linear transformation changes. Theorem 4.4.6 tells us that all these matrices are related.

Now, let $A$ and $B$ be two $n \times n$ matrices such that $P^{-1}AP = B$ for some invertible matrix $P$. Recall the linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by $T_A(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. Then we have seen that if the standard basis of $\mathbb{R}^n$ is the ordered basis $\mathcal{B}$, then $A = T_A[\mathcal{B}, \mathcal{B}]$. Since $P$ is an invertible matrix, its columns are linearly independent and hence we can take its columns as an ordered basis $\mathcal{B}_1$. Then note that $B = T_A[\mathcal{B}_1, \mathcal{B}_1]$. The above observations lead to the following remark and the definition.

**Remark 4.4.7** *The identity (4.4.3) shows how the matrix representation of a linear transformation $T$ changes if the ordered basis used to compute the matrix representation is changed. Hence, the matrix $I[\mathcal{B}_1, \mathcal{B}_2]$ is called the $\mathcal{B}_1 : \mathcal{B}_2$ change of basis matrix.*

**Definition 4.4.8 (Similar Matrices)** Two square matrices $B$ and $C$ of the same order are said to be similar if there exists a non-singular matrix $P$ such that $B = PCP^{-1}$ or equivalently $BP = PC$.

**Remark 4.4.9** *Observe that if $A = T[\mathcal{B}, \mathcal{B}]$ then*

$$\{S^{-1}AS : S \text{ is } n \times n \text{ invertible matrix } \}$$

*is the set of all matrices that are similar to the given matrix $A$. Therefore, similar matrices are just different matrix representations of a single linear transformation.*

**Example 4.4.10** 1. Consider $\mathcal{P}_2(\mathbb{R})$, with ordered bases

$$\mathcal{B}_1 = \big(1, 1+x, 1+x+x^2\big) \quad \text{and} \quad \mathcal{B}_2 = \big(1+x-x^2, 1+2x+x^2, 2+x+x^2\big).$$

Then

$$[1+x-x^2]_{\mathcal{B}_1} = 0\cdot 1 + 2\cdot(1+x) + (-1)\cdot(1+x+x^2) = (0,2,-1)^t,$$
$$[1+2x+x^2]_{\mathcal{B}_1} = (-1)\cdot 1 + 1\cdot(1+x) + 1\cdot(1+x+x^2) = (-1,1,1)^t, \text{ and}$$
$$[2+x+x^2]_{\mathcal{B}_1} = 1\cdot 1 + 0\cdot(1+x) + 1\cdot(1+x+x^2) = (1,0,1)^t.$$

Therefore,

$$\begin{aligned} I[\mathcal{B}_2,\mathcal{B}_1] &= [[I(1+x-x^2)]_{\mathcal{B}_1}, [I(1+2x+x^2)]_{\mathcal{B}_1}, [I(2+x+x^2)]_{\mathcal{B}_1}] \\ &= [[1+x-x^2]_{\mathcal{B}_1}, [1+2x+x^2]_{\mathcal{B}_1}, [2+x+x^2]_{\mathcal{B}_1}] \\ &= \begin{bmatrix} 0 & -1 & 1 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}. \end{aligned}$$

Find the matrices $T[\mathcal{B}_1,\mathcal{B}_1]$ and $T[\mathcal{B}_2,\mathcal{B}_2]$. Also verify that

$$\begin{aligned} T[\mathcal{B}_2,\mathcal{B}_2] &= I[\mathcal{B}_1,\mathcal{B}_2]\, T[\mathcal{B}_1,\mathcal{B}_1]\, I[\mathcal{B}_2,\mathcal{B}_1] \\ &= I^{-1}[\mathcal{B}_2,\mathcal{B}_1]\, T[\mathcal{B}_1,\mathcal{B}_1]\, I[\mathcal{B}_2,\mathcal{B}_1]. \end{aligned}$$

2. Consider two bases $\mathcal{B}_1 = \big((1,0,0),(1,1,0),(1,1,1)\big)$ and $\mathcal{B}_2 = \big((1,1,-1),(1,2,1),(2,1,1)\big)$ of $\mathbb{R}^3$. Suppose $T:\mathbb{R}^3\longrightarrow\mathbb{R}^3$ is a linear transformation defined by

$$T((x,y,z)) = (x+y, x+y+2z, y-z).$$

Then

$$T[\mathcal{B}_1,\mathcal{B}_1] = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 1 & 4 \\ 0 & 1 & 0 \end{bmatrix}, \quad \text{and} \quad T[\mathcal{B}_2,\mathcal{B}_2] = \begin{bmatrix} -4/5 & 1 & 8/5 \\ -2/5 & 2 & 9/5 \\ 8/5 & 0 & -1/5 \end{bmatrix}.$$

Find $I[\mathcal{B}_1,\mathcal{B}_2]$ and verify,

$$I[\mathcal{B}_1,\mathcal{B}_2]\, T[\mathcal{B}_1,\mathcal{B}_1]\, I[\mathcal{B}_2,\mathcal{B}_1] = T[\mathcal{B}_2,\mathcal{B}_2].$$

Check that,

$$T[\mathcal{B}_1,\mathcal{B}_1]\, I[\mathcal{B}_2,\mathcal{B}_1] = I[\mathcal{B}_2,\mathcal{B}_1]\, T[\mathcal{B}_2,\mathcal{B}_2] = \begin{bmatrix} 2 & -2 & -2 \\ -2 & 4 & 5 \\ 2 & 1 & 0 \end{bmatrix}.$$

**Exercise 4.4.11** 1. Let $V$ be an $n$-dimensional vector space and let $T: V\longrightarrow V$ be a linear transformation. Suppose $T$ has the property that $T^{n-1} \neq \mathbf{0}$ but $T^n = \mathbf{0}$.

(a) Then prove that there exists a vector $\mathbf{u} \in V$ such that the set

$$\{\mathbf{u}, T(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u})\}$$

is a basis of $V$.

(b) Let $\mathcal{B} = (\mathbf{u}, T(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u}))$. Then prove that

$$T[\mathcal{B},\mathcal{B}] = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

(c) Let $A$ be an $n \times n$ matrix with the property that $A^{n-1} \neq \mathbf{0}$ but $A^n = \mathbf{0}$. Then prove that $A$ is similar to the matrix given above.

2. Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be a linear transformation given by

$$T((x, y, z)) = (x + y + 2z, x - y - 3z, 2x + 3y + z).$$

Let $\mathcal{B}$ be the standard basis and $\mathcal{B}_1 = \big((1, 1, 1), (1, -1, 1), (1, 1, 2)\big)$ be another ordered basis.

(a) Find the matrices $T[\mathcal{B}, \mathcal{B}]$ and $T[\mathcal{B}_1, \mathcal{B}_1]$.

(b) Find the matrix $P$ such that $P^{-1}T[\mathcal{B}, \mathcal{B}]\, P = T[\mathcal{B}_1, \mathcal{B}_1]$.

3. Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be a linear transformation given by

$$T((x, y, z)) = (x, x + y, x + y + z).$$

Let $\mathcal{B}$ be the standard basis and $\mathcal{B}_1 = \big((1, 0, 0), (1, 1, 0), (1, 1, 1)\big)$ be another ordered basis.

(a) Find the matrices $T[\mathcal{B}, \mathcal{B}]$ and $T[\mathcal{B}_1, \mathcal{B}_1]$.

(b) Find the matrix $P$ such that $P^{-1}T[\mathcal{B}, \mathcal{B}]\, P = T[\mathcal{B}_1, \mathcal{B}_1]$.

4. Let $\mathcal{B}_1 = \big((1, 2, 0), (1, 3, 2), (0, 1, 3)\big)$ and $\mathcal{B}_2 = \big((1, 2, 1), (0, 1, 2), (1, 4, 6)\big)$ be two ordered bases of $\mathbb{R}^3$.

(a) Find the change of basis matrix $P$ from $\mathcal{B}_1$ to $\mathcal{B}_2$.

(b) Find the change of basis matrix $Q$ from $\mathcal{B}_2$ to $\mathcal{B}_1$.

(c) Verify that $PQ = I = QP$.

(d) Find the change of basis matrix from the standard basis of $\mathbb{R}^3$ to $\mathcal{B}_1$. What do you notice?

# Chapter 5

# Inner Product Spaces

We had learned that given vectors $\vec{i}$ and $\vec{j}$ (which are at an angle of 90°) in a plane, any vector in the plane is a linear combination of the vectors $\vec{i}$ and $\vec{j}$. In this section, we investigate a method by which any basis of a finite dimensional vector can be transferred to another basis in such a way that the vectors in the new basis are at an angle of 90° to each other. To do this, we start by defining a notion of INNER PRODUCT (dot product) in a vector space. This helps us in finding out whether two vectors are at 90° or not.

## 5.1 Definition and Basic Properties

In $\mathbb{R}^2$, given two vectors $\mathbf{x} = (x_1, x_2),\ \mathbf{y} = (y_1, y_2)$, we know the inner product $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2$. Note that for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$, this inner product satisfies the conditions

$$\mathbf{x} \cdot (\mathbf{y} + \alpha \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \alpha \mathbf{x} \cdot \mathbf{z},\ \mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}, \text{ and } \mathbf{x} \cdot \mathbf{x} \geq 0$$

and $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$. Thus, we are motivated to define an inner product on an arbitrary vector space.

**Definition 5.1.1 (Inner Product)** Let $V(\mathbb{F})$ be a vector space over $\mathbb{F}$. An inner product over $V(\mathbb{F})$, denoted by $\langle\ ,\ \rangle$, is a map,

$$\langle\ ,\ \rangle : V \times V \longrightarrow \mathbb{F}$$

such that for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a, b \in \mathbb{F}$

1. $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$,
2. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$, the complex conjugate of $\langle \mathbf{u}, \mathbf{v} \rangle$, and
3. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ for all $\mathbf{u} \in V$ and equality holds if and only if $\mathbf{u} = \mathbf{0}$.

**Definition 5.1.2 (Inner Product Space)** Let $V$ be a vector space with an inner product $\langle\ ,\ \rangle$. Then $(V, \langle\ ,\ \rangle)$ is called an inner product space, in short denoted by IPS.

**Example 5.1.3** The first two examples given below are called the STANDARD INNER PRODUCT or the DOT PRODUCT on $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively..

1. Let $V = \mathbb{R}^n$ be the real vector space of dimension $n$. Given two vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ of $V$, we define

$$\langle u, v \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \mathbf{u}\mathbf{v}^t.$$

   Verify $\langle\ ,\ \rangle$ is an inner product.

2. Let $V = \mathbb{C}^n$ be a complex vector space of dimension $n$. Then for $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in $V$, check that
$$\langle u, v\rangle = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n} = \mathbf{u}\mathbf{v}^*$$
is an inner product.

3. Let $V = \mathbb{R}^2$ and let $A = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$. Define $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}A\mathbf{y}^t$. Check that $\langle\ ,\ \rangle$ is an inner product. *Hint: Note that* $\mathbf{x}A\mathbf{y}^t = 4x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$.

4. let $\mathbf{x} = (x_1, x_2, x_3),\ \mathbf{y} = (y_1, y_2, y_3) \in \mathbb{R}^3$., Show that $\langle \mathbf{x}, \mathbf{y}\rangle = 10x_1y_1 + 3x_1y_2 + 3x_2y_1 + 2x_2y_2 + x_2y_3 + x_3y_2 + x_3y_3$ is an inner product in $\mathbb{R}^3(\mathbb{R})$.

5. Consider the real vector space $\mathbb{R}^2$. In this example, we define three products that satisfy two conditions out of the three conditions for an inner product. Hence the three products are not inner products.

   (a) Define $\langle \mathbf{x}, \mathbf{y}\rangle = \langle (x_1, x_2), (y_1, y_2)\rangle = x_1y_1$. Then it is easy to verify that the third condition is not valid whereas the first two conditions are valid.

   (b) Define $\langle \mathbf{x}, \mathbf{y}\rangle = \langle (x_1, x_2), (y_1, y_2)\rangle = x_1^2 + y_1^2 + x_2^2 + y_2^2$. Then it is easy to verify that the first condition is not valid whereas the second and third conditions are valid.

   (c) Define $\langle \mathbf{x}, \mathbf{y}\rangle = \langle (x_1, x_2), (y_1, y_2)\rangle = x_1y_1^3 + x_2y_2^3$. Then it is easy to verify that the second condition is not valid whereas the first and third conditions are valid.

**Remark 5.1.4** *Note that in parts 1 and 2 of Example 5.1.3, the inner products are* $\mathbf{u}\mathbf{v}^t$ *and* $\mathbf{u}\mathbf{v}^*$*, respectively. This occurs because the vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *are row vectors. In general,* $\mathbf{u}$ *and* $\mathbf{v}$ *are taken as column vectors and hence one uses the notation* $\mathbf{u}^t\mathbf{v}$ *or* $\mathbf{u}^*\mathbf{v}$.

**Exercise 5.1.5** Verify that inner products defined in parts 3 and 4 of Example 5.1.3, are indeed inner products.

**Definition 5.1.6 (Length/Norm of a Vector)** For $\mathbf{u} \in V$, we define the length (norm) of $\mathbf{u}$, denoted $\|\mathbf{u}\|$, by $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}$, the positive square root.

A very useful and a fundamental inequality concerning the inner product is due to Cauchy and Schwartz. The next theorem gives the statement and a proof of this inequality.

**Theorem 5.1.7 (Cauchy-Schwartz inequality)** Let $V(\mathbb{F})$ be an inner product space. Then for any $\mathbf{u}, \mathbf{v} \in V$
$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\ \|\mathbf{v}\|.$$
The equality holds if and only if the vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. Further, if $\mathbf{u} \ne \mathbf{0}$, then $\mathbf{v} = \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$.

PROOF. If $\mathbf{u} = \mathbf{0}$, then the inequality holds. Let $\mathbf{u} \ne \mathbf{0}$. Note that $\langle \lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v}\rangle \ge 0$ for all $\lambda \in \mathbb{F}$. In particular, for $\lambda = -\frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}$, we get
$$\begin{aligned} 0 &\le \langle \lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v}\rangle \\ &= \lambda\overline{\lambda}\|\mathbf{u}\|^2 + \lambda\langle \mathbf{u}, \mathbf{v}\rangle + \overline{\lambda}\langle \mathbf{v}, \mathbf{u}\rangle + \|\mathbf{v}\|^2 \\ &= \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\frac{\overline{\langle \mathbf{v}, \mathbf{u}\rangle}}{\|\mathbf{u}\|^2}\|\mathbf{u}\|^2 - \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\langle \mathbf{u}, \mathbf{v}\rangle - \frac{\overline{\langle \mathbf{v}, \mathbf{u}\rangle}}{\|\mathbf{u}\|^2}\langle \mathbf{v}, \mathbf{u}\rangle + \|\mathbf{v}\|^2 \\ &= \|\mathbf{v}\|^2 - \frac{|\langle \mathbf{v}, \mathbf{u}\rangle|^2}{\|\mathbf{u}\|^2}. \end{aligned}$$

Or, in other words

$$|\langle \mathbf{v}, \mathbf{u}\rangle|^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$$

and the proof of the inequality is over.

Observe that if $\mathbf{u} \neq \mathbf{0}$ then the equality holds if and only of $\lambda \mathbf{u} + \mathbf{v} = \mathbf{0}$ for $\lambda = -\dfrac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}$. That is, $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. We leave it for the reader to prove

$$\mathbf{v} = \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

□

**Definition 5.1.8 (Angle between two vectors)** Let $V$ be a real vector space. Then for every $\mathbf{u}, \mathbf{v} \in V$, by the Cauchy-Schwartz inequality, we have

$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\| \; \|\mathbf{v}\|} \leq 1.$$

We know that $\cos : [0, \pi] \longrightarrow [-1, \ 1]$ is an one-one and onto function. Therefore, for every real number $\dfrac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\| \; \|\mathbf{v}\|}$, there exists a unique $\theta, \ 0 \leq \theta \leq \pi$, such that

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\| \; \|\mathbf{v}\|}.$$

1. The real number $\theta$ with $0 \leq \theta \leq \pi$ and satisfying $\cos\theta = \dfrac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\| \; \|\mathbf{v}\|}$ is called the angle between the two vectors $\mathbf{u}$ and $\mathbf{v}$ in $V$.
2. The vectors $\mathbf{u}$ and $\mathbf{v}$ in $V$ are said to be orthogonal if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.
3. A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is called mutually orthogonal if $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = 0$ for all $1 \leq i \neq j \leq n$.

**Exercise 5.1.9** 1. Let $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$. Then prove that with respect to the standard inner product on $\mathbb{R}^n$, the vectors $\mathbf{e}_i$ satisfy the following:

   (a) $\|\mathbf{e}_i\| = 1$ for $1 \leq i \leq n$.

   (b) $\langle \mathbf{e}_i, \mathbf{e}_j\rangle = 0$ for $1 \leq i \neq j \leq n$.

2. Recall the following inner product on $\mathbb{R}^2$ : for $\mathbf{x} = (x_1, x_2)^t$ and $\mathbf{y} = (y_1, y_2)^t$,

$$\langle \mathbf{x}, \mathbf{y}\rangle = 4x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2.$$

   (a) Find the angle between the vectors $e_1 = (1, 0)^t$ and $e_2 = (0, 1)^t$.

   (b) Let $\mathbf{u} = (1, 0)^t$. Find $\mathbf{v} \in \mathbb{R}^2$ such that $\langle \mathbf{v}, \mathbf{u}\rangle = 0$.

   (c) Find two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, such that $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ and $\langle \mathbf{x}, \mathbf{y}\rangle = 0$.

3. Find an inner product in $\mathbb{R}^2$ such that the following conditions hold:

$$\|(1, 2)\| = \|(2, -1)\| = 1, \quad \text{and} \quad \langle (1, 2), \ (2, -1)\rangle = 0.$$

*[Hint: Consider a symmetric matrix* $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. *Define* $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{y}^t A\mathbf{x}$ *and solve a system of* 3 *equations for the unknowns* $a, b, c$.*]*

4. Let $V$ be a complex vector space with $\dim(V) = n$. Fix an ordered basis $\mathcal{B} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$. Define a map
$$\langle \; , \; \rangle : V \times V \longrightarrow \mathbb{C} \text{ by } \langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} a_i \overline{b_i}$$
whenever $[\mathbf{u}]_{\mathcal{B}} = (a_1, a_2, \ldots, a_n)^t$ and $[\mathbf{v}]_{\mathcal{B}} = (b_1, b_2, \ldots, b_n)^t$. Show that the above defined map is indeed an inner product.

5. Let $\mathbf{x} = (x_1, x_2, x_3), \; \mathbf{y} = (y_1, y_2, y_3) \in \mathbb{R}^3$. Show that
$$\langle \mathbf{x}, \mathbf{y} \rangle = 10x_1y_1 + 3x_1y_2 + 3x_2y_1 + 2x_2y_2 + x_2y_3 + x_3y_2 + x_3y_3$$
is an inner product in $\mathbb{R}^3(\mathbb{R})$. With respect to this inner product, find the angle between the vectors $(1, 1, 1)$ and $(2, -5, 2)$.

6. Consider the set $M_{n\times n}(\mathbb{R})$ of all real square matrices of order $n$. For $A, B \in M_{n\times n}(\mathbb{R})$ we define $\langle A, B\rangle = tr(AB^t)$. Then
$$\langle A + B, C\rangle = tr\big((A + B)C^t\big) = tr(AC^t) + tr(BC^t) = \langle A, C\rangle + \langle B, C\rangle.$$
$$\langle A, B\rangle = tr(AB^t) = tr( \; (AB^t)^t \; ) = tr(BA^t) = \langle B, A\rangle.$$
Let $A = (a_{ij})$. Then
$$\langle A, A\rangle = tr(AA^t) = \sum_{i=1}^{n}(AA^t)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$$
and therefore, $\langle A, A\rangle > 0$ for all non-zero matrices $A$. So, it is clear that $\langle A, B\rangle$ is an inner product on $M_{n\times n}(\mathbb{R})$.

7. Let $V$ be the real vector space of all continuous functions with domain $[-2\pi, 2\pi]$. That is, $V = C[-2\pi, \; 2\pi]$. Then show that $V$ is an inner product space with inner product $\int_{-1}^{1} f(x)g(x)dx$.

   For different values of $m$ and $n$, find the angle between the functions $\cos(mx)$ and $\sin(nx)$.

8. Let $V$ be an inner product space. Prove that
$$\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\| \quad \text{for every} \quad \mathbf{u}, \mathbf{v} \in V.$$
This inequality is called the TRIANGLE INEQUALITY.

9. Let $z_1, z_2, \ldots, z_n \in \mathbb{C}$. Use the Cauchy-Schwartz inequality to prove that
$$|z_1 + z_2 + \cdots + z_n| \le \sqrt{n(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2)}.$$
When does the equality hold?

10. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Observe that $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$. Hence or otherwise prove the following:

    (a) $\langle \mathbf{x}, \mathbf{y}\rangle = 0 \Longleftrightarrow \|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, (This is called PYTHAGORAS THEOREM).

    (b) $\|\mathbf{x}\| = \|\mathbf{y}\| \Longleftrightarrow \langle \mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle = 0$, ($\mathbf{x}$ and $\mathbf{y}$ form adjacent sides of a rhombus as the diagonals $\mathbf{x} + \mathbf{y}$ and $\mathbf{x} - \mathbf{y}$ are orthogonal).

    (c) $\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$, (This is called the PARALLELOGRAM LAW).

    (d) $4\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2$ (This is called the POLARISATION IDENTITY).

**Remark 5.1.10** *i. Suppose the norm of a vector is given. Then, the polarisation identity can be used to define an inner product.*

ii. *Observe that if* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ *then the parallelogram spanned by the vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *is a rectangle. The above equality tells us that the lengths of the two diagonals are equal.*

Are these results true if $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n(\mathbb{C})$?

11. Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n(\mathbb{C})$. Prove that

    (a) $4\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 + i\|\mathbf{x}+i\mathbf{y}\|^2 - i\|\mathbf{x}-i\mathbf{y}\|^2$.

    (b) If $\mathbf{x} \neq \mathbf{0}$ then $\|\mathbf{x}+i\mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|i\mathbf{x}\|^2$, even though $\langle \mathbf{x}, i\mathbf{x} \rangle \neq 0$.

    (c) If $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ and $\|\mathbf{x}+i\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|i\mathbf{y}\|^2$ then show that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

12. Let $V$ be an $n$-dimensional inner product space, with an inner product $\langle\ ,\ \rangle$. Let $\mathbf{u} \in V$ be a fixed vector with $\|\mathbf{u}\| = 1$. Then give reasons for the following statements.

    (a) Let $S^\perp = \{\mathbf{v} \in V\ :\ \langle \mathbf{v}, \mathbf{u} \rangle = 0\}$. Then $S$ is a subspace of $V$ of dimension $n-1$.

    (b) Let $0 \neq \alpha \in \mathbb{F}$ and let $S = \{\mathbf{v} \in V\ :\ \langle \mathbf{v}, \mathbf{u} \rangle = \alpha\}$. Then $S$ is not a subspace of $V$.

    (c) For any $\mathbf{v} \in S$, there exists a vector $\mathbf{v}_0 \in S^\perp$, such that $\mathbf{v} = \mathbf{v}_0 + \alpha\mathbf{u}$.

**Theorem 5.1.11** Let $V$ be an inner product space. Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a set of non-zero, mutually orthogonal vectors of $V$.

1. Then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is linearly independent.
2. $\| \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i \|^2 = \sum\limits_{i=1}^{n} |\alpha_i|^2 \|\mathbf{u}_i\|^2$;
3. Let $\dim(V) = n$ and also let $\|\mathbf{u}_i\| = 1$ for $i = 1, 2, \ldots, n$. Then for any $\mathbf{v} \in V$,
   $$\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{u}_i \rangle \mathbf{u}_i.$$
   In particular, $\langle \mathbf{v}, \mathbf{u}_i \rangle = 0$ for all $i = 1, 2, \ldots, n$ if and only if $\mathbf{v} = \mathbf{0}$.

PROOF. Consider the set of non-zero, mutually orthogonal vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$. Suppose there exist scalars $c_1, c_2, \ldots, c_n$ not all zero, such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = \mathbf{0}.$$

Then for $1 \leq i \leq n$, we have

$$0 = \langle \mathbf{0}, \mathbf{u}_i \rangle = \langle c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n, \mathbf{u}_i \rangle = \sum_{j=1}^{n} c_j \langle \mathbf{u}_j, \mathbf{u}_i \rangle = c_i$$

as $\langle \mathbf{u}_j, \mathbf{u}_i \rangle = 0$ for all $j \neq i$ and $\langle \mathbf{u}_i, \mathbf{u}_i \rangle = 1$. This gives a contradiction to our assumption that some of the $c_i$'s are non-zero. This establishes the linear independence of a set of non-zero, mutually orthogonal vectors.

For the second part, using $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ \|\mathbf{u}_i\|^2 & \text{if } i = j \end{cases}$ for $1 \leq i, j \leq n$, we have

$$\begin{aligned} \|\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\|^2 &= \langle \sum_{i=1}^{n} \alpha_i \mathbf{u}_i, \sum_{i=1}^{n} \alpha_i \mathbf{u}_i \rangle = \sum_{i=1}^{n} \alpha_i \langle \mathbf{u}_i, \sum_{j=1}^{n} \alpha_j \mathbf{u}_j \rangle \\ &= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \overline{\alpha_j} \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \sum_{i=1}^{n} \alpha_i \overline{\alpha_i} \langle \mathbf{u}_i, \mathbf{u}_i \rangle \\ &= \sum_{i=1}^{n} |\alpha_i|^2 \|\mathbf{u}_i\|^2. \end{aligned}$$

For the third part, observe from the first part, the linear independence of the non-zero mutually orthogonal vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. Since $\dim(V) = n$, they form a basis of $V$. Thus, for every vector $\mathbf{v} \in V$, there exist scalars $\alpha_i$, $1 \le i \le n$, such that $\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{u}_n$. Hence,

$$\langle \mathbf{v}, \mathbf{u}_j \rangle = \langle \sum_{i=1}^n \alpha_i \mathbf{u}_i, \mathbf{u}_j \rangle = \sum_{i=1}^n \alpha_i \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \alpha_j.$$

Therefore, we have obtained the required result. □

**Definition 5.1.12 (Orthonormal Set)** Let $V$ be an inner product space. A set of non-zero, mutually orthogonal vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in $V$ is called an orthonormal set if $\|\mathbf{v}_i\| = 1$ for $i = 1, 2, \ldots, n$.

If the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is also a basis of $V$, then the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is called an orthonormal basis of $V$.

**Example 5.1.13**

1. Consider the vector space $\mathbb{R}^2$ with the standard inner product. Then the standard ordered basis $\mathcal{B} = \big((1,0), (0,1)\big)$ is an orthonormal set. Also, the basis $\mathcal{B}_1 = \big(\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1)\big)$ is an orthonormal set.
2. Let $\mathbb{R}^n$ be endowed with the standard inner product. Then by Exercise 5.1.9.1, the standard ordered basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ is an orthonormal set.

In view of Theorem 5.1.11, we inquire into the question of extracting an orthonormal basis from a given basis. In the next section, we describe a process (called the Gram-Schmidt Orthogonalisation process) that generates an orthonormal set from a given set containing finitely many vectors.

**Remark 5.1.14** *The last part of the above theorem can be rephrased as "suppose $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of an inner product space V. Then for each $\mathbf{u} \in V$ the numbers $\langle \mathbf{u}, \mathbf{v}_i \rangle$ for $1 \le i \le n$ are the coordinates of $\mathbf{u}$ with respect to the above basis".*

*That is, let $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis. Then for any $\mathbf{u} \in V$,*

$$[\mathbf{u}]_{\mathcal{B}} = (\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \ldots, \langle \mathbf{u}, \mathbf{v}_n \rangle)^t.$$

## 5.2 Gram-Schmidt Orthogonalisation Process

Let $V$ be a finite dimensional inner product space. Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ is a linearly independent subset of $V$. Then the Gram-Schmidt orthogonalisation process uses the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ to construct new vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ such that $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $i \neq j$, $\|\mathbf{v}_i\| = 1$ and Span $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i\} =$ Span $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$ for $i = 1, 2, \ldots, n$. This process proceeds with the following idea.

Suppose we are given two vectors $\mathbf{u}$ and $\mathbf{v}$ in a plane. If we want to get vectors $\mathbf{z}$ and $\mathbf{y}$ such that $\mathbf{z}$ is a unit vector in the direction of $\mathbf{u}$ and $\mathbf{y}$ is a unit vector perpendicular to $\mathbf{z}$, then they can be obtained in the following way:

Take the first vector $\mathbf{z} = \dfrac{\mathbf{u}}{\|\mathbf{u}\|}$. Let $\theta$ be the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$. Then $\cos(\theta) = \dfrac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|u\| \; \|v\|}$. Defined $\alpha = \|\mathbf{v}\| \; \cos(\theta) = \dfrac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|} = \langle \mathbf{z}, \mathbf{v} \rangle$. Then $\mathbf{w} = \mathbf{v} - \alpha \; \mathbf{z}$ is a vector perpendicular to the unit vector $\mathbf{z}$, as we have removed the component of $\mathbf{z}$ from $\mathbf{v}$. So, the vectors that we are interested in are $\mathbf{z}$ and $\mathbf{y} = \dfrac{\mathbf{w}}{\|\mathbf{w}\|}$.

This idea is used to give the Gram-Schmidt Orthogonalization process which we now describe.

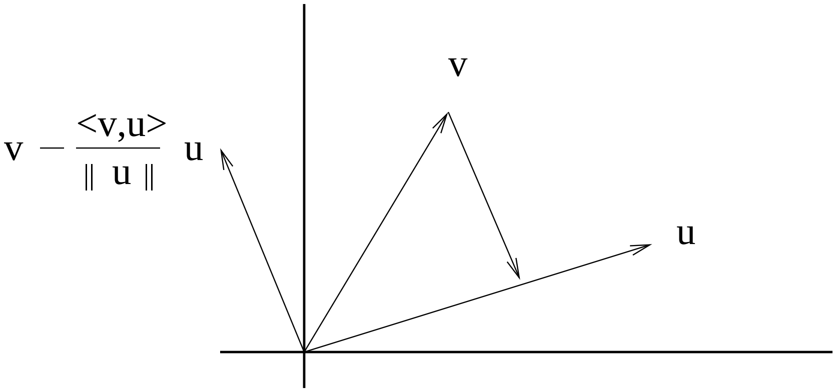


Figure 5.1: Gram-Schmidt Process

**Theorem 5.2.1 (Gram-Schmidt Orthogonalization Process)** Let $V$ be an inner product space. Suppose $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a set of linearly independent vectors of $V$. Then there exists a set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of vectors of $V$ satisfying the following:

1. $\|\mathbf{v}_i\| = 1$ for $1 \leq i \leq n$,

2. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $1 \leq i, j \leq n, \; i \neq j$ and

3. $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i)$ for $1 \leq i \leq n$.

Proof. We successively define the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ as follows.

$\mathbf{v}_1 = \dfrac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$.

Calculate $\mathbf{w}_2 = \mathbf{u}_2 - \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1$, and let $\mathbf{v}_2 = \dfrac{\mathbf{w}_2}{\|\mathbf{w}_2\|}$.

Obtain $\mathbf{w}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2$, and let $\mathbf{v}_3 = \dfrac{\mathbf{w}_3}{\|\mathbf{w}_3\|}$.

In general, if $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \ldots, \mathbf{v}_{i-1}$ are already obtained, we compute

$$\mathbf{w}_i = \mathbf{u}_i - \langle \mathbf{u}_i, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_i, \mathbf{v}_2 \rangle \mathbf{v}_2 - \cdots - \langle \mathbf{u}_i, \mathbf{v}_{i-1} \rangle \mathbf{v}_{i-1}, \tag{5.2.1}$$

and define

$$\mathbf{v}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}.$$

We prove the theorem by induction on $n$, the number of linearly independent vectors.

For $n = 1$, we have $\mathbf{v}_1 = \dfrac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$. Since $\mathbf{u}_1 \neq \mathbf{0}, \; \mathbf{v}_1 \neq \mathbf{0}$ and

$$\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \langle \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \rangle = \frac{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2} = 1.$$

Hence, the result holds for $n = 1$.

Let the result hold for all $k \leq n - 1$. That is, suppose we are given any set of $k, \; 1 \leq k \leq n - 1$ linearly independent vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of $V$. Then by the inductive assumption, there exists a set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of vectors satisfying the following:

1. $\|\mathbf{v}_i\| = 1$ for $1 \leq i \leq k$,

2. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $1 \leq i \neq j \leq k$, and

3. $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i)$ for $1 \leq i \leq k$.

Now, let us assume that we are given a set of $n$ linearly independent vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of $V$. Then by the inductive assumption, we already have vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$ satisfying

1. $\|\mathbf{v}_i\| = 1$ for $1 \leq i \leq n-1$,

2. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $1 \leq i \neq j \leq n-1$, and

3. $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i)$ for $1 \leq i \leq n-1$.

Using (5.2.1), we define

$$\mathbf{w}_n = \mathbf{u}_n - \langle \mathbf{u}_n, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_n, \mathbf{v}_2 \rangle \mathbf{v}_2 - \cdots - \langle \mathbf{u}_n, \mathbf{v}_{n-1} \rangle \mathbf{v}_{n-1}. \tag{5.2.2}$$

We first show that $\mathbf{w}_n \notin L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$. This will also imply that $\mathbf{w}_n \neq \mathbf{0}$ and hence $\mathbf{v}_n = \dfrac{\mathbf{w}_n}{\|\mathbf{w}_n\|}$ is well defined.

On the contrary, assume that $\mathbf{w}_n \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$. Then there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ such that

$$\mathbf{w}_n = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_{n-1} \mathbf{v}_{n-1}.$$

So, by (5.2.2)

$$\mathbf{u}_n = \big(\alpha_1 + \langle \mathbf{u}_n, \mathbf{v}_1 \rangle\big) \mathbf{v}_1 + \big(\alpha_2 + \langle \mathbf{u}_n, \mathbf{v}_2 \rangle\big) \mathbf{v}_2 + \cdots + \big(\big(\alpha_{n-1} + \langle \mathbf{u}_n, \mathbf{v}_{n-1} \rangle\big) \mathbf{v}_{n-1}.$$

Thus, by the third induction assumption,

$$\mathbf{u}_n \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}).$$

This gives a contradiction to the given assumption that the set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is linear independent.

So, $\mathbf{w}_n \neq \mathbf{0}$. Define $\mathbf{v}_n = \dfrac{\mathbf{w}_n}{\|\mathbf{w}_n\|}$. Then $\|\mathbf{v}_n\| = 1$. Also, it can be easily verified that $\langle \mathbf{v}_n, \mathbf{v}_i \rangle = 0$ for $1 \leq i \leq n-1$. Hence, by the principle of mathematical induction, the proof of the theorem is complete. □

We illustrate the Gram-Schmidt process by the following example.

**Example 5.2.2** Let $\{(1,-1,1,1), (1,0,1,0), (0,1,0,1)\}$ be a linearly independent set in $\mathbb{R}^4(\mathbb{R})$. Find an orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ such that $L(\ (1,-1,1,1), (1,0,1,0), (0,1,0,1)\ ) = L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$.

**Solution:** Let $\mathbf{u}_1 = (1,0,1,0)$. Define $\mathbf{v}_1 = \dfrac{(1,0,1,0)}{\sqrt{2}}$. Let $\mathbf{u}_2 = (0,1,0,1)$. Then

$$\mathbf{w}_2 = (0,1,0,1) - \langle (0,1,0,1), \frac{(1,0,1,0)}{\sqrt{2}} \rangle \mathbf{v}_1 = (0,1,0,1).$$

Hence, $\mathbf{v}_2 = \dfrac{(0,1,0,1)}{\sqrt{2}}$. Let $\mathbf{u}_3 = (1,-1,1,1)$. Then

$$\begin{aligned} \mathbf{w}_3 &= (1,-1,1,1) - \langle (1,-1,1,1), \frac{(1,0,1,0)}{\sqrt{2}} \rangle \mathbf{v}_1 - \langle (1,-1,1,1), \frac{(0,1,0,1)}{\sqrt{2}} \rangle \mathbf{v}_2 \\ &= (0,-1,0,1) \end{aligned}$$

and $\mathbf{v}_3 = \dfrac{(0,-1,0,1)}{\sqrt{2}}$.

**Remark 5.2.3** 1. *Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be any basis of a $k$-dimensional subspace $W$ of $\mathbb{R}^n$. Then by Gram-Schmidt orthogonalisation process, we get an orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$ with $W = L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$, and for $1 \leq i \leq k$,*

$$L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i).$$

2. *Suppose we are given a set of $n$ vectors, $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of $V$ that are linearly dependent. Then by Corollary 3.2.5, there exists a smallest $k$, $2 \leq k \leq n$ such that*

$$L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}).$$

*We claim that in this case, $\mathbf{w}_k = \mathbf{0}$.*

*Since, we have chosen the smallest $k$ satisfying*

$$L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{i-1}),$$

*for $2 \leq i \leq n$, the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}\}$ is linearly independent (use Corollary 3.2.5). So, by Theorem 5.2.1, there exists an orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}\}$ such that*

$$L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}) = L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}).$$

*As $\mathbf{u}_k \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1})$, by Remark 5.1.14*

$$\mathbf{u}_k = \langle \mathbf{u}_k, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}_k, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{u}_k, \mathbf{v}_{k-1} \rangle \mathbf{v}_{n-1}.$$

*So, by definition of $\mathbf{w}_k$, $\mathbf{w}_k = \mathbf{0}$.*

*Therefore, in this case, we can continue with the Gram-Schmidt process by replacing $\mathbf{u}_k$ by $\mathbf{u}_{k+1}$.*

3. *Let $S$ be a countably infinite set of linearly independent vectors. Then one can apply the Gram-Schmidt process to get a countably infinite orthonormal set.*

4. *Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be an orthonormal subset of $\mathbb{R}^n$. Let $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ be the standard ordered basis of $\mathbb{R}^n$. Then there exist real numbers $\alpha_{ij}$, $1 \leq i \leq k$, $1 \leq j \leq n$ such that*

$$[\mathbf{v}_i]_{\mathcal{B}} = (\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_{ni})^t.$$

*Let $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$. Then in the ordered basis $\mathcal{B}$, we have*

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nk} \end{bmatrix}$$

*is an $n \times k$ matrix.*

*Also, observe that the conditions $\|\mathbf{v}_i\| = 1$ and $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $1 \leq i \neq j \leq n$, implies that*

$$\left.\begin{aligned} 1 = \|\mathbf{v}_i\| = \|\mathbf{v}_i\|^2 = \langle \mathbf{v}_i, \mathbf{v}_i \rangle &= \sum_{j=1}^{n} \alpha_{ji}^2, \\ \text{and} \quad 0 = \langle \mathbf{v}_i, \mathbf{v}_j \rangle &= \sum_{s=1}^{n} \alpha_{si}\alpha_{sj}. \end{aligned}\right\} \qquad (5.2.3)$$

*Note that,*

$$\begin{aligned} A^tA &= \begin{bmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_k^t \end{bmatrix} [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k] = \begin{bmatrix} \|\mathbf{v}_1\|^2 & \langle \mathbf{v}_1, \mathbf{v}_2\rangle & \cdots & \langle \mathbf{v}_1, \mathbf{v}_k\rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1\rangle & \|\mathbf{v}_2\|^2 & \cdots & \langle \mathbf{v}_2, \mathbf{v}_k\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{v}_k, \mathbf{v}_1\rangle & \langle \mathbf{v}_k, \mathbf{v}_2\rangle & \cdots & \|\mathbf{v}_k\|^2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_k. \end{aligned}$$

*Or using (5.2.3), in the language of matrices, we get*

$$A^tA = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1k} & \alpha_{2k} & \cdots & \alpha_{nk} \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nk} \end{bmatrix} = I_k.$$

Perhaps the readers must have noticed that the inverse of $A$ is its transpose. Such matrices are called orthogonal matrices and they have a special role to play.

**Definition 5.2.4 (Orthogonal Matrix)** A $n \times n$ real matrix $A$ is said to be an orthogonal matrix if $A\ A^t = A^tA = I_n$.

It is worthwhile to solve the following exercises.

**Exercise 5.2.5**

1. Let $A$ and $B$ be two $n \times n$ orthogonal matrices. Then prove that $AB$ and $BA$ are both orthogonal matrices.

2. Let $A$ be an $n \times n$ orthogonal matrix. Then prove that

   (a) the rows of $A$ form an orthonormal basis of $\mathbb{R}^n$.

   (b) the columns of $A$ form an orthonormal basis of $\mathbb{R}^n$.

   (c) for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n\times 1}$, $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$.

   (d) for any vector $\mathbf{x} \in \mathbb{R}^{n\times 1}$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$.

3. Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthonormal basis of $\mathbb{R}^n$. Let $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ be the standard basis of $\mathbb{R}^n$. Construct an $n \times n$ matrix $A$ by

   $$A = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

   where

   $$\mathbf{u}_i = \sum_{j=1}^{n} a_{ji}\mathbf{e}_j, \text{ for } 1 \le i \le n.$$

   Prove that $A^tA = I_n$. Hence deduce that $A$ is an orthogonal matrix.

4. Let $A$ be an $n \times n$ upper triangular matrix. If $A$ is also an orthogonal matrix, then prove that $A = I_n$.

**Theorem 5.2.6 (QR Decomposition)** Let $A$ be a square matrix of order $n$. Then there exist matrices $Q$ and $R$ such that $Q$ is orthogonal and $R$ is upper triangular with $A = QR$.

In case, $A$ is non-singular, the diagonal entries of $R$ can be chosen to be positive. Also, in this case, the decomposition is unique.

Proof. We prove the theorem when $A$ is non-singular. The proof for the singular case is left as an exercise.

Let the columns of $A$ be $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. The Gram-Schmidt orthogonalisation process applied to the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ gives the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ satisfying

$$\left.\begin{array}{c} L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i) = L(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i), \\ \|\mathbf{u}_i\| = 1, \ \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0, \end{array}\right\} \text{ for } 1 \leq i \neq j \leq n. \tag{5.2.4}$$

Now, consider the ordered basis $\mathcal{B} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$. From (5.2.4), for $1 \leq i \leq n$, we have $L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i) = L(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i)$. So, we can find scalars $\alpha_{ji}, 1 \leq j \leq i$ such that

$$\mathbf{x}_i = \alpha_{1i}\mathbf{u}_1 + \alpha_{2i}\mathbf{u}_2 + \cdots + \alpha_{ii}\mathbf{u}_i = \left[(\alpha_{1i}, \ldots, \alpha_{ii}, 0 \ldots, 0)^t\right]_{\mathcal{B}}. \tag{5.2.5}$$

Let $Q = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$. Then by Exercise 5.2.5.3, $Q$ is an orthogonal matrix. We now define an $n \times n$ upper triangular matrix $R$ by

$$R = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ 0 & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_{nn} \end{bmatrix}.$$

By using (5.2.5), we get

$$\begin{aligned} QR &= [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ 0 & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_{nn} \end{bmatrix} \\ &= \left[\alpha_{11}\mathbf{u}_1, \ \alpha_{12}\mathbf{u}_1 + \alpha_{22}\mathbf{u}_2, \ldots, \sum_{i=1}^{n} \alpha_{in}\mathbf{u}_i\right] \\ &= [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] = A. \end{aligned}$$

Thus, we see that $A = QR$, where $Q$ is an orthogonal matrix (see Remark 5.2.3.4) and $R$ is an upper triangular matrix.

The proof doesn't guarantee that for $1 \leq i \leq n$, $\alpha_{ii}$ is positive. But this can be achieved by replacing the vector $\mathbf{u}_i$ by $-\mathbf{u}_i$ whenever $\alpha_{ii}$ is negative.

**Uniqueness:** suppose $Q_1R_1 = Q_2R_2$ then $Q_2^{-1}Q_1 = R_2R_1^{-1}$. Observe the following properties of upper triangular matrices.

1. The inverse of an upper triangular matrix is also an upper triangular matrix, and
2. product of upper triangular matrices is also upper triangular.

Thus the matrix $R_2R_1^{-1}$ is an upper triangular matrix. Also, by Exercise 5.2.5.1, the matrix $Q_2^{-1}Q_1$ is an orthogonal matrix. Hence, by Exercise 5.2.5.4, $R_2R_1^{-1} = I_n$. So, $R_2 = R_1$ and therefore $Q_2 = Q_1$. □

Suppose we have matrix $A = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]$ of dimension $n \times k$ with rank $(A) = r$. Then by Remark 5.2.3.2, the application of the Gram-Schmidt orthogonalisation process yields a set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ of orthonormal vectors of $\mathbb{R}^n$. In this case, for each $i$, $1 \le i \le r$, we have

$$L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i) = L(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j), \text{ for some } j, \; i \le j \le k.$$

Hence, proceeding on the lines of the above theorem, we have the following result.

**Theorem 5.2.7 (Generalised QR Decomposition)** Let $A$ be an $n \times k$ matrix of rank $r$. Then $A = QR$, where

1. $Q$ is an $n \times r$ matrix with $Q^t Q = I_r$. That is, the columns of $Q$ form an orthonormal set,
2. If $Q = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]$, then $L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r) = L(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$, and
3. $R$ is an $r \times k$ matrix with rank $(R) = r$.

**Example 5.2.8** 1. Let $A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$. Find an orthogonal matrix $Q$ and an upper triangular matrix $R$ such that $A = QR$.

**Solution:** From Example 5.2.2, we know that

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,0,1,0), \; \mathbf{v}_2 = \frac{1}{\sqrt{2}}(0,1,0,1), \; \mathbf{v}_3 = \frac{1}{\sqrt{2}}(0,-1,0,1). \tag{5.2.6}$$

We now compute $\mathbf{w}_4$. If we denote $\mathbf{u}_4 = (2,1,1,1)^t$ then by the Gram-Schmidt process,

$$\begin{aligned} \mathbf{w}_4 &= \mathbf{u}_4 - \langle \mathbf{u}_4, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_4, \mathbf{v}_2 \rangle \mathbf{v}_2 - \langle \mathbf{u}_4, \mathbf{v}_3 \rangle \mathbf{v}_3 \\ &= \frac{1}{2}(1,0,-1,0)^t. \end{aligned} \tag{5.2.7}$$

Thus, using (5.2.6) and (5.2.7), we get

$$Q = \big[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\big] = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

and

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

The readers are advised to check that $A = QR$ is indeed correct.

2. Let $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & -2 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 2 & 1 \end{bmatrix}$. Find a $4 \times 3$ matrix $Q$ satisfying $Q^t Q = I_3$ and an upper triangular matrix $R$ such that $A = QR$.

**Solution:** Let us apply the Gram Schmidt orthogonalisation to the columns of $A$. Or equivalently to the rows of $A^t$. So, we need to apply the process to the subset $\{(1,-1,1,1), (1,0,1,0), (1,-2,1,2), (0,1,0,1)\}$ of $\mathbb{R}^4$.

Let $\mathbf{u}_1 = (1, -1, 1, 1)$. Define $\mathbf{v}_1 = \dfrac{\mathbf{u}_1}{2}$. Let $\mathbf{u}_2 = (1, 0, 1, 0)$. Then

$$\mathbf{w}_2 = (1, 0, 1, 0) - \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1 = (1, 0, 1, 0) - \mathbf{v}_1 = \frac{1}{2}(1, 1, 1, -1).$$

Hence, $\mathbf{v}_2 = \dfrac{(1, 1, 1, -1)}{2}$. Let $\mathbf{u}_3 = (1, -2, 1, 2)$. Then

$$\mathbf{w}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2 = \mathbf{u}_3 - 3\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}.$$

So, we again take $\mathbf{u}_3 = (0, 1, 0, 1)$. Then

$$\mathbf{w}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2 = \mathbf{u}_3 - 0\mathbf{v}_1 - 0\mathbf{v}_2 = \mathbf{u}_3.$$

So, $\mathbf{v}_3 = \dfrac{(0, 1, 0, 1)}{\sqrt{2}}$. Hence,

$$Q = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{-1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{-1}{2} & \frac{1}{\sqrt{2}} \end{bmatrix}, \text{ and } R = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & \sqrt{2} \end{bmatrix}.$$

The readers are advised to check the following:

(a) rank $(A) = 3$,

(b) $A = QR$ with $Q^t Q = I_3$, and

(c) $R$ a $3 \times 4$ upper triangular matrix with rank $(R) = 3$.

**Exercise 5.2.9** 1. Determine an orthonormal basis of $\mathbb{R}^4$ containing the vectors $(1, -2, 1, 3)$ and $(2, 1, -3, 1)$.

2. Prove that the polynomials $1, x, \frac{3}{2}x^2 - \frac{1}{2}, \frac{5}{2}x^3 - \frac{3}{2}x$ form an orthogonal set of functions in the inner product space $C[-1, 1]$ with the inner product $\langle f, g \rangle = \int_{-1}^{1} f(t)\overline{g(t)}dt$. Find the corresponding functions, $f(x)$ with $\|f(x)\| = 1$.

3. Consider the vector space $C[-\pi, \pi]$ with the standard inner product defined in the above exercise. Find an orthonormal basis for the subspace spanned by $x$, $\sin x$ and $\sin(x + 1)$.

4. Let $M$ be a subspace of $\mathbb{R}^n$ and $\dim M = m$. A vector $x \in \mathbb{R}^n$ is said to be orthogonal to $M$ if $\langle x, y \rangle = 0$ for every $y \in M$.

   (a) How many linearly independent vectors can be orthogonal to $M$?

   (b) If $M = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$, determine a maximal set of linearly independent vectors orthogonal to $M$ in $\mathbb{R}^3$.

5. Determine an orthogonal basis of vector subspace spanned by $\{(1, 1, 0, 1), (-1, 1, 1, -1), (0, 2, 1, 0), (1, 0, 0, 0)\}$ in $\mathbb{R}^4$.

6. Let $S = \{(1, 1, 1, 1), (1, 2, 0, 1), (2, 2, 4, 0)\}$. Find an orthonormal basis of $L(S)$ in $\mathbb{R}^4$.

7. Let $\mathbb{R}^n$ be endowed with the standard inner product. Suppose we have a vector $\mathbf{x}^t = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, with $\|\mathbf{x}\| = 1$. Then prove the following:

   (a) the set $\{\mathbf{x}\}$ can always be extended to form an orthonormal basis of $\mathbb{R}^n$.

   (b) Let this basis be $\{\mathbf{x}, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. Suppose $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ is the standard basis of $\mathbb{R}^n$. Let $A = \begin{bmatrix} [\mathbf{x}]_{\mathcal{B}}, & [\mathbf{x}_2]_{\mathcal{B}}, & \ldots, & [\mathbf{x}_n]_{\mathcal{B}} \end{bmatrix}$. Then prove that $A$ is an orthogonal matrix.

8. Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n, n \geq 1$ with $\|\mathbf{u}\| = \|\mathbf{w}\| = 1$. Prove that there exists an orthogonal matrix $A$ such that $A\mathbf{v} = \mathbf{w}$. Prove also that $A$ can be chosen such that $\det(A) = 1$.

## 5.3 Orthogonal Projections and Applications

Recall that given a $k$-dimensional vector subspace of a vector space $V$ of dimension $n$, one can always find an $(n-k)$-dimensional vector subspace $W_0$ of $V$ (see Exercise 3.3.19.9) satisfying

$$W + W_0 = V \quad \text{and} \quad W \cap W_0 = \{\mathbf{0}\}.$$

The subspace $W_0$ is called the complementary subspace of $W$ in $V$. We now define an important class of linear transformations on an inner product space, called orthogonal projections.

**Definition 5.3.1 (Projection Operator)** Let $V$ be an $n$-dimensional vector space and let $W$ be a $k$-dimensional subspace of $V$. Let $W_0$ be a complement of $W$ in $V$. Then we define a map $P_W : V \longrightarrow V$ by

$$P_W(\mathbf{v}) = \mathbf{w}, \text{ whenever } \mathbf{v} = \mathbf{w} + \mathbf{w}_0, \; \mathbf{w} \in W, \; \mathbf{w}_0 \in W_0.$$

The map $P_W$ is called the projection of $V$ onto $W$ along $W_0$.

**Remark 5.3.2** *The map $P$ is well defined due to the following reasons:*

1. *$W + W_0 = V$ implies that for every $\mathbf{v} \in V$, we can find $w \in W$ and $w_0 \in W_0$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}_0$.*
2. *$W \cap W_0 = \{\mathbf{0}\}$ implies that the expression $\mathbf{v} = \mathbf{w} + \mathbf{w}_0$ is unique for every $\mathbf{v} \in V$.*

The next proposition states that the map defined above is a linear transformation from $V$ to $V$. We omit the proof, as it follows directly from the above remarks.

**Proposition 5.3.3** The map $P_W : V \longrightarrow V$ defined above is a linear transformation.

**Example 5.3.4** Let $V = \mathbb{R}^3$ and $W = \{(x, y, z) \in \mathbb{R}^3 : x + y - z = 0\}$.

1. Let $W_0 = L(\,(1, 2, 2)\,)$. Then $W \cap W_0 = \{\mathbf{0}\}$ and $W + W_0 = \mathbb{R}^3$. Also, for any vector $(x, y, z) \in \mathbb{R}^3$, note that $(x, y, z) = \mathbf{w} + \mathbf{w}_0$, where

   $$\mathbf{w} = (z - y, 2z - 2x - y, 3z - 2x - 2y), \text{ and } \mathbf{w}_0 = (x + y - z)(1, 2, 2).$$

   So, by definition,

   $$P_W((x, y, z)) = (z - y, 2z - 2x - y, 3z - 2x - 2y) = \begin{bmatrix} 0 & -1 & 1 \\ -2 & -1 & 2 \\ -2 & -2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

2. Let $W_0 = L(\,(1, 1, 1)\,)$. Then $W \cap W_0 = \{\mathbf{0}\}$ and $W + W_0 = \mathbb{R}^3$. Also, for any vector $(x, y, z) \in \mathbb{R}^3$, note that $(x, y, z) = \mathbf{w} + \mathbf{w}_0$, where

   $$\mathbf{w} = (z - y, z - x, 2z - x - y), \text{ and } \mathbf{w}_0 = (x + y - z)(1, 1, 1).$$

   So, by definition,

   $$P_W(\,(x, y, z)\,) = (z - y, z - x, 2z - x - y) = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

**Remark 5.3.5** 1. *The projection map $P_W$ depends on the complementary subspace $W_0$.*

2. *Observe that for a fixed subspace $W$, there are infinitely many choices for the complementary subspace $W_0$.*

3. *It will be shown later that if $V$ is an inner product space with inner product, $\langle\ ,\ \rangle$, then the subspace $W_0$ is unique if we put an additional condition that $W_0 = \{\mathbf{v} \in V\ : \langle \mathbf{v}, \mathbf{w}\rangle = 0 \ \text{for all}\ \ \mathbf{w} \in W\}$.*

We now prove some basic properties about projection maps.

**Theorem 5.3.6** Let $W$ and $W_0$ be complementary subspaces of a vector space $V$. Let $P_W : V \longrightarrow V$ be a projection operator of $V$ onto $W$ along $W_0$. Then

1. the null space of $P_W$, $\mathcal{N}(P_W) = \{\mathbf{v} \in V : P_W(\mathbf{v}) = \mathbf{0}\} = W_0$.
2. the range space of $P_W$, $\mathcal{R}(P_W) = \{P_W(\mathbf{v}) : \mathbf{v} \in V\} = W$.
3. $P_W^2 = P_W$. The condition $P_W^2 = P_W$ is equivalent to $P_W(I - P_W) = \mathbf{0} = (I - P_W)P_W$.

Proof. We only prove the first part of the theorem.
Let $\mathbf{w}_0 \in W_0$. Then $\mathbf{w}_0 = \mathbf{0} + \mathbf{w}_0$ for $\mathbf{0} \in W$. So, by definition, $P(\mathbf{w}_0) = \mathbf{0}$. Hence, $W_0 \subset \mathcal{N}(P_W)$.

Also, for any $\mathbf{v} \in V$, let $P_W(\mathbf{v}) = \mathbf{0}$ with $\mathbf{v} = \mathbf{w} + \mathbf{w}_0$ for some $\mathbf{w}_0 \in W_0$ and $\mathbf{w} \in W$. Then by definition $\mathbf{0} = P_W(\mathbf{v}) = \mathbf{w}$. That is, $\mathbf{w} = \mathbf{0}$ and $\mathbf{v} = \mathbf{w}_0$. Thus, $\mathbf{v} \in W_0$. Hence $\mathcal{N}(P_W) = W_0$. □

**Exercise 5.3.7** 1. Let $A$ be an $n \times n$ real matrix with $A^2 = A$. Consider the linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, defined by $T_A(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. Prove that

   (a) $T_A \circ T_A = T_A$ (use the condition $A^2 = A$).

   (b) $\mathcal{N}(T_A) \cap \mathcal{R}(T_A) = \{\mathbf{0}\}$.
   *Hint: Let $\mathbf{x} \in \mathcal{N}(T_A) \cap \mathcal{R}(T_A)$. This implies $T_A(\mathbf{x}) = \mathbf{0}$ and $\mathbf{x} = T_A(\mathbf{y})$ for some $\mathbf{y} \in \mathbb{R}^n$. So,*

   $$\mathbf{x} = T_A(\mathbf{y}) = (T_A \circ T_A)(\mathbf{y}) = T_A\big(T_A(\mathbf{y})\big) = T_A(\mathbf{x}) = \mathbf{0}.$$

   (c) $\mathbb{R}^n = \mathcal{N}(T_A) + \mathcal{R}(T_A)$.
   *Hint: Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $\mathcal{N}(T_A)$. Extend it to get a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$. Then by Rank-nullity Theorem 4.3.6, $\{T_A(\mathbf{v}_{k+1}), \ldots, T_A(\mathbf{v}_n)\}$ is a basis of $\mathcal{R}(T_A)$.*

   (d) Define $W = \mathcal{R}(T_A)$ and $W_0 = \mathcal{N}(T_A)$. Then $T_A$ is a projection operator of $\mathbb{R}^n$ onto $W$ along $W_0$.

   Recall that the first three parts of this exercise was also given in Exercise 4.3.10.7.

2. Find all $2 \times 2$ real matrices $A$ such that $A^2 = A$. Hence or otherwise, determine all projection operators of $\mathbb{R}^2$.

The next result uses the Gram-Schmidt orthogonalisation process to get the complementary subspace in such a way that the vectors in different subspaces are orthogonal.

**Definition 5.3.8 (Orthogonal Subspace of a Set)** Let $V$ be an inner product space. Let $S$ be a non-empty subset of $V$. We define

$$S^\perp = \{\mathbf{v} \in V\ : \langle \mathbf{v}, \mathbf{s}\rangle = 0 \text{ for all } \mathbf{s} \in S\}.$$

**Example 5.3.9** Let $V = \mathbb{R}$.

1. $S = \{0\}$. Then $S^\perp = \mathbb{R}$.
2. $S = \mathbb{R}$, Then $S^\perp = \{0\}$.
3. Let $S$ be any subset of $\mathbb{R}$ containing a non-zero real number. Then $S^\perp = \{0\}$.

**Theorem 5.3.10** Let $S$ be a subset of a finite dimensional inner product space $V$, with inner product $\langle\ ,\ \rangle$. Then

1. $S^{\perp}$ is a subspace of $V$.
2. Let $S$ be equal to a subspace $W$. Then the subspaces $W$ and $W^{\perp}$ are complementary. Moreover, if $\mathbf{w} \in W$ and $\mathbf{u} \in W^{\perp}$, then $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ and $V = W + W^{\perp}$.

Proof. We leave the prove of the first part for the reader. The prove of the second part is as follows: Let $\dim(V) = n$ and $\dim(W) = k$. Let $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ be a basis of $W$. By Gram-Schmidt orthogonalisation process, we get an orthonormal basis, say, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of $W$. Then, for any $\mathbf{v} \in V$,

$$\mathbf{v} - \sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{v}_i \rangle \mathbf{v}_i \in W^{\perp}.$$

So, $V \subset W + W^{\perp}$. Also, for any $\mathbf{v} \in W \cap W^{\perp}$, by definition of $W^{\perp}$, $0 = \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2$. So, $\mathbf{v} = \mathbf{0}$. That is, $W \cap W^{\perp} = \{\mathbf{0}\}$. □

**Definition 5.3.11 (Orthogonal Complement)** Let $W$ be a subspace of a vector space $V$. The subspace $W^{\perp}$ is called the orthogonal complement of $W$ in $V$.

**Exercise 5.3.12**

1. Let $W = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$. Find $W^{\perp}$ with respect to the standard inner product.
2. Let $W$ be a subspace of a finite dimensional inner product space $V$. Prove that $(W^{\perp})^{\perp} = W$.
3. Let $V$ be the vector space of all $n \times n$ real matrices. Then Exercise5.1.9.6 shows that $V$ is a real inner product space with the inner product given by $\langle A, B \rangle = \text{tr}(AB^t)$. If $W$ is the subspace given by $W = \{A \in V : \ A^t = A\}$, determine $W^{\perp}$.

**Definition 5.3.13 (Orthogonal Projection)** Let $W$ be a subspace of a finite dimensional inner product space $V$, with inner product $\langle\ ,\ \rangle$. Let $W^{\perp}$ be the orthogonal complement of $W$ in $V$. Define $P_W : V \longrightarrow V$ by

$$P_W(\mathbf{v}) = \mathbf{w} \ \text{ where } \ \mathbf{v} = \mathbf{w} + \mathbf{u}, \ \text{ with } \ \mathbf{w} \in W, \ \text{ and } \ \mathbf{u} \in W^{\perp}.$$

Then $P_W$ is called the orthogonal projection of $V$ onto $W$ along $W^{\perp}$.

**Definition 5.3.14 (Self-Adjoint Transformation/Operator)** Let $V$ be an inner product space with inner product $\langle\ ,\ \rangle$. A linear transformation $T : V \longrightarrow V$ is called a self-adjoint operator if $\langle T(\mathbf{v}), \mathbf{u} \rangle = \langle \mathbf{v}, T(\mathbf{u}) \rangle$ for every $\mathbf{u}, \mathbf{v} \in V$.

**Example 5.3.15**

1. Let $A$ be an $n \times n$ real symmetric matrix. That is, $A^t = A$. Then show that the linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by $T_A(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x}^t \in \mathbb{R}^n$ is a self-adjoint operator.
   **Solution:** By definition, for every $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^n$,
   $$\langle T_A(\mathbf{x}), \mathbf{y} \rangle = (\mathbf{y})^t A\mathbf{x} = (\mathbf{y})^t A^t \mathbf{x} = (A\mathbf{y})^t \mathbf{x} = \langle \mathbf{x}, T_A(\mathbf{y}) \rangle.$$
   Hence, the result follows.
2. Let $A$ be an $n \times n$ Hermitian matrix, that is, $A^* = A$. Then the linear transformation $T_A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ defined by $T_A(\mathbf{z}) = A\mathbf{z}$ for every $\mathbf{z}^t \in \mathbb{C}^n$ is a self-adjoint operator.

**Remark 5.3.16** *1. By Proposition 5.3.3, the map $P_W$ defined above is a linear transformation.*

2. $P_W^2 = P_W$, $(I - P_W)P_W = \mathbf{0} = P_W(I - P_W)$.

3. *Let* $\mathbf{u}, \mathbf{v} \in V$ *with* $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ *and* $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ *for some* $\mathbf{u}_1, \mathbf{v}_1 \in W$ *and* $\mathbf{u}_2, \mathbf{v}_2 \in W^{\perp}$. *Then we know that* $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = 0$ *whenever* $1 \leq i \neq j \leq 2$. *Therefore, for every* $\mathbf{u}, \mathbf{v} \in V$,

$$\begin{aligned} \langle P_W(\mathbf{u}), \mathbf{v} \rangle &= \langle \mathbf{u}_1, \mathbf{v} \rangle = \langle \mathbf{u}_1, \mathbf{v}_1 + \mathbf{v}_2 \rangle = \langle \mathbf{u}_1, \mathbf{v}_1 \rangle = \langle \mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 \rangle \\ &= \langle \mathbf{u}, P_W(\mathbf{v}) \rangle. \end{aligned}$$

*Thus, the orthogonal projection operator is a self-adjoint operator.*

4. *Let* $\mathbf{v} \in V$ *and* $\mathbf{w} \in W$. *Then* $P_W(\mathbf{w}) = \mathbf{w}$ *for all* $\mathbf{w} \in W$. *Therefore, using Remarks 5.3.16.2 and 5.3.16.3, we get*

$$\begin{aligned} \langle \mathbf{v} - P_W(\mathbf{v}), \mathbf{w} \rangle &= \langle \big(I - P_W\big)(\mathbf{v}), P_W(\mathbf{w}) \rangle = \langle P_W\big(I - P_W\big)(\mathbf{v}), \mathbf{w} \rangle \\ &= \langle \mathbf{0}(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{0}, \mathbf{w} \rangle = 0 \end{aligned}$$

*for every* $\mathbf{w} \in W$.

5. *In particular,* $\langle \mathbf{v} - P_W(\mathbf{v}), P_W(\mathbf{v}) - \mathbf{w} \rangle = 0$ *as* $P_W(\mathbf{v}) \in W$. *Thus,* $\langle \mathbf{v} - P_W(\mathbf{v}), P_W(\mathbf{v}) - \mathbf{w}' \rangle = 0$, *for every* $\mathbf{w}' \in W$. *Hence, for any* $\mathbf{v} \in V$ *and* $\mathbf{w} \in W$, *we have*

$$\begin{aligned} \|\mathbf{v} - \mathbf{w}\|^2 &= \|\mathbf{v} - P_W(\mathbf{v}) + P_W(\mathbf{v}) - \mathbf{w}\|^2 \\ &= \|\mathbf{v} - P_W(\mathbf{v})\|^2 + \|P_W(\mathbf{v}) - \mathbf{w}\|^2 \\ &\qquad + 2\langle \mathbf{v} - P_W(\mathbf{v}), P_W(\mathbf{v}) - \mathbf{w} \rangle \\ &= \|\mathbf{v} - P_W(\mathbf{v})\|^2 + \|P_W(\mathbf{v}) - \mathbf{w}\|^2. \end{aligned}$$

*Therefore,*

$$\|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{v} - P_W(\mathbf{v})\|$$

*and the equality holds if and only if* $\mathbf{w} = P_W(\mathbf{v})$. *Since* $P_W(\mathbf{v}) \in W$, *we see that*

$$d(\mathbf{v}, W) = \inf\{\|\mathbf{v} - \mathbf{w}\| \ : \mathbf{w} \in W\} = \|\mathbf{v} - P_W(\mathbf{v})\|.$$

*That is,* $P_W(\mathbf{v})$ *is the vector nearest to* $\mathbf{v} \in W$. *This can also be stated as: the vector* $P_W(\mathbf{v})$ *solves the following minimisation problem:*

$$\inf_{\mathbf{w} \in W} \|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v} - P_W(\mathbf{v})\|.$$

### 5.3.1 Matrix of the Orthogonal Projection

The minimization problem stated above arises in lot of applications. So, it will be very helpful if the matrix of the orthogonal projection can be obtained under a given basis.

To this end, let $W$ be a $k$-dimensional subspace of $\mathbb{R}^n$ with $W^{\perp}$ as its orthogonal complement. Let $P_W : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be the orthogonal projection of $\mathbb{R}^n$ onto $W$. Suppose, we are given an orthonormal basis $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ of $W$. Under the assumption that $\mathcal{B}$ is known, we explicitly give the matrix of $P_W$ with respect to an extended ordered basis of $\mathbb{R}^n$.

Let us extend the given ordered orthonormal basis $\mathcal{B}$ of $W$ to get an orthonormal ordered basis $\mathcal{B}_1 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1} \ldots, \mathbf{v}_n)$ of $\mathbb{R}^n$. Then by Theorem 5.1.11, for any $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} = \sum\limits_{i=1}^{n} \langle \mathbf{v}, \mathbf{v}_i \rangle \mathbf{v}_i$. Thus, by definition, $P_W(\mathbf{v}) = \sum\limits_{i=1}^{k} \langle \mathbf{v}, \mathbf{v}_i \rangle \mathbf{v}_i$. Let $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$. Consider the standard orthogonal ordered basis $\mathcal{B}_2 = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ of $\mathbb{R}^n$. Therefore, if $\mathbf{v}_i = \sum\limits_{j=1}^{n} a_{ji}\mathbf{e}_j$, for $1 \le i \le k$, then

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix}, \ [\mathbf{v}]_{\mathcal{B}_2} = \begin{bmatrix} \sum\limits_{i=1}^{n} a_{1i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \sum\limits_{i=1}^{n} a_{2i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \vdots \\ \sum\limits_{i=1}^{n} a_{ni}\langle \mathbf{v}, \mathbf{v}_i \rangle \end{bmatrix}$$

and

$$[P_W(\mathbf{v})]_{\mathcal{B}_2} = \begin{bmatrix} \sum\limits_{i=1}^{k} a_{1i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \sum\limits_{i=1}^{k} a_{2i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \vdots \\ \sum\limits_{i=1}^{k} a_{ni}\langle \mathbf{v}, \mathbf{v}_i \rangle \end{bmatrix}.$$

Then as observed in Remark 5.2.3.4, $A^tA = I_k$. That is, for $1 \le i, j \le k$,

$$\sum_{s=1}^{n} a_{si}a_{sj} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{5.3.1}$$

Thus, using the associativity of matrix product and (5.3.1), we get

$$\begin{aligned} (AA^t)(\mathbf{v}) &= A \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{bmatrix} \begin{bmatrix} \sum\limits_{i=1}^{n} a_{1i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \sum\limits_{i=1}^{n} a_{2i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \vdots \\ \sum\limits_{i=1}^{n} a_{ni}\langle \mathbf{v}, \mathbf{v}_i \rangle \end{bmatrix} \\ &= A \begin{bmatrix} \sum\limits_{s=1}^{n} a_{s1}\left(\sum\limits_{i=1}^{n} a_{si}\langle \mathbf{v}, \mathbf{v}_i \rangle\right) \\ \sum\limits_{s=1}^{n} a_{s2}\left(\sum\limits_{i=1}^{n} a_{si}\langle \mathbf{v}, \mathbf{v}_i \rangle\right) \\ \vdots \\ \sum\limits_{s=1}^{n} a_{sk}\left(\sum\limits_{i=1}^{n} a_{si}\langle \mathbf{v}, \mathbf{v}_i \rangle\right) \end{bmatrix} = A \begin{bmatrix} \sum\limits_{i=1}^{n}\left(\sum\limits_{s=1}^{n} a_{s1}a_{si}\right)\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \sum\limits_{i=1}^{n}\left(\sum\limits_{s=1}^{n} a_{s2}a_{si}\right)\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \vdots \\ \sum\limits_{i=1}^{n}\left(\sum\limits_{s=1}^{n} a_{sk}a_{si}\right)\langle \mathbf{v}, \mathbf{v}_i \rangle \end{bmatrix} \\ &= A \begin{bmatrix} \langle \mathbf{v}, \mathbf{v}_1 \rangle \\ \langle \mathbf{v}, \mathbf{v}_2 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{v}_k \rangle \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{k} a_{1i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \sum\limits_{i=1}^{k} a_{2i}\langle \mathbf{v}, \mathbf{v}_i \rangle \\ \vdots \\ \sum\limits_{i=1}^{k} a_{ni}\langle \mathbf{v}, \mathbf{v}_i \rangle \end{bmatrix} \\ &= [P_W(\mathbf{v})]_{\mathcal{B}_2}. \end{aligned}$$

Thus $P_W[\mathcal{B}_2, \mathcal{B}_2] = AA^t$. Thus, we have proved the following theorem.

**Theorem 5.3.17** Let $W$ be a $k$-dimensional subspace of $\mathbb{R}^n$ and let $P_W : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be the orthogonal projection of $\mathbb{R}^n$ onto $W$ along $W^{\perp}$. Suppose, $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is an orthonormal ordered basis of $W$. Define an $n \times k$ matrix $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$. Then the matrix of the linear transformation $P_W$ in the standard orthogonal ordered basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ is $AA^t$.

**Example 5.3.18** Let $W = \{(x,y,z,w) \in \mathbb{R}^4 : x = y, z = w\}$ be a subspace of $W$. Then an orthonormal ordered basis of $W$ is

$$\big(\frac{1}{\sqrt{2}}(1,1,0,0), \frac{1}{\sqrt{2}}(0,0,1,1)\big),$$

and that of $W^\perp$ is

$$\big(\frac{1}{\sqrt{2}}(1,-1,0,0), \frac{1}{\sqrt{2}}(0,0,1,-1)\big).$$

Therefore, if $P_W : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ is an orthogonal projection of $\mathbb{R}^4$ onto $W$ along $W^\perp$, then the corresponding matrix $A$ is given by

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

Hence, the matrix of the orthogonal projection $P_W$ in the ordered basis

$$\mathcal{B} = \big(\frac{1}{\sqrt{2}}(1,1,0,0), \frac{1}{\sqrt{2}}(0,0,1,1), \frac{1}{\sqrt{2}}(1,-1,0,0), \frac{1}{\sqrt{2}}(0,0,1,-1)\big)$$

is

$$P_W[\mathcal{B},\mathcal{B}] = AA^t = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

It is easy to see that

1. the matrix $P_W[\mathcal{B},\mathcal{B}]$ is symmetric,
2. $P_W[\mathcal{B},\mathcal{B}]^2 = P_W[\mathcal{B},\mathcal{B}]$, and
3. $\big(I_4 - P_W[\mathcal{B},\mathcal{B}]\big)P_W[\mathcal{B},\mathcal{B}] = \mathbf{0} = P_W[\mathcal{B},\mathcal{B}]\big(I_4 - P_W[\mathcal{B},\mathcal{B}]\big)$.

Also, for any $(x,y,z,w) \in \mathbb{R}^4$, we have

$$[(x,y,z,w)]_{\mathcal{B}} = \left(\frac{x+y}{\sqrt{2}}, \frac{z+w}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}, \frac{z-w}{\sqrt{2}}\right)^t.$$

Thus, $P_W\big((x,y,z,w)\big) = \frac{x+y}{2}(1,1,0,0) + \frac{z+w}{2}(0,0,1,1)$ is the closest vector to the subspace $W$ for any vector $(x,y,z,w) \in \mathbb{R}^4$.

**Exercise 5.3.19**

1. Show that for any non-zero vector $\mathbf{v}^t \in \mathbb{R}^n$, the rank of the matrix $\mathbf{v}\mathbf{v}^t$ is $1$.
2. Let $W$ be a subspace of a vector space $V$ and let $P : V \longrightarrow V$ be the orthogonal projection of $V$ onto $W$ along $W^\perp$. Let $\mathcal{B}$ be an orthonormal ordered basis of $V$. Then prove that corresponding matrix satisfies $P[\mathcal{B},\mathcal{B}]^t = P[\mathcal{B},\mathcal{B}]$.
3. Let $A$ be an $n \times n$ matrix with $A^2 = A$ and $A^t = A$. Consider the associated linear transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by $T_A(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v}^t \in \mathbb{R}^n$. Then prove that there exists a subspace $W$ of $\mathbb{R}^n$ such that $T_A$ is the orthogonal projection of $\mathbb{R}^n$ onto $W$ along $W^\perp$.
4. Let $W_1$ and $W_2$ be two distinct subspaces of a finite dimensional vector space $V$. Let $P_{W_1}$ and $P_{W_2}$ be the corresponding orthogonal projection operators of $V$ along $W_1^\perp$ and $W_2^\perp$, respectively. Then by constructing an example in $\mathbb{R}^2$, show that the map $P_{W_1} \circ P_{W_2}$ is a projection but not an orthogonal projection.

5. Let $W$ be an $(n-1)$-dimensional vector subspace of $\mathbb{R}^n$ and let $W^\perp$ be its orthogonal complement. Let $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n)$ be an orthogonal ordered basis of $\mathbb{R}^n$ with $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$ an ordered basis of $W$. Define a map
$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^n \quad \text{by } T(\mathbf{v}) = \mathbf{w}_0 - \mathbf{w}$$
whenever $\mathbf{v} = \mathbf{w} + \mathbf{w}_0$ for some $\mathbf{w} \in W$ and $\mathbf{w}_0 \in W^\perp$. Then

   (a) prove that $T$ is a linear transformation,

   (b) find the matrix, $T[\mathcal{B}, \mathcal{B}]$, and

   (c) prove that $T[\mathcal{B}, \mathcal{B}]$ is an orthogonal matrix.

   $T$ is called the reflection along $W^\perp$.

# Chapter 6

# Eigenvalues, Eigenvectors and Diagonalization

## 6.1 Introduction and Definitions

In this chapter, the linear transformations are from a given finite dimensional vector space $V$ to itself. Observe that in this case, the matrix of the linear transformation is a square matrix. So, in this chapter, all the matrices are square matrices and *a vector* $\mathbf{x}$ means $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ for some positive integer $n$.

**Example 6.1.1** Let $A$ be a real symmetric matrix. Consider the following problem:

$$\text{Maximize (Minimize) } \mathbf{x}^t A\mathbf{x} \text{ such that } \mathbf{x} \in \mathbb{R}^n \text{ and } \mathbf{x}^t\mathbf{x} = 1.$$

To solve this, consider the Lagrangian

$$L(\mathbf{x}, \lambda) = \mathbf{x}^t A\mathbf{x} - \lambda(\mathbf{x}^t\mathbf{x} - 1) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j - \lambda(\sum_{i=1}^{n} x_i^2 \quad - 1).$$

Partially differentiating $L(\mathbf{x}, \lambda)$ with respect to $x_i$ for $1 \le i \le n$, we get

$$\frac{\partial L}{\partial x_1} = 2a_{11}x_1 + 2a_{12}x_2 + \cdots + 2a_{1n}x_n - 2\lambda x_1,$$

$$\frac{\partial L}{\partial x_2} = 2a_{21}x_1 + 2a_{22}x_2 + \cdots + 2a_{2n}x_n - 2\lambda x_2,$$

and so on, till

$$\frac{\partial L}{\partial x_n} = 2a_{n1}x_1 + 2a_{n2}x_2 + \cdots + 2a_{nn}x_n - 2\lambda x_n.$$

Therefore, to get the points of extrema, we solve for

$$(0, 0, \ldots, 0)^t = (\frac{\partial L}{\partial x_1}, \frac{\partial L}{\partial x_2}, \ldots, \frac{\partial L}{\partial x_n})^t = \frac{\partial L}{\partial \mathbf{x}} = 2(A\mathbf{x} - \lambda\mathbf{x}).$$

We therefore need to find a $\lambda \in \mathbb{R}$ and $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for the extremal problem.

**Example 6.1.2** Consider a system of $n$ ordinary differential equations of the form

$$\frac{d\,\mathbf{y}(t)}{dt} = A\mathbf{y},\ t \geq 0; \tag{6.1.1}$$

where $A$ is a real $n \times n$ matrix and $\mathbf{y}$ is a column vector.
To get a solution, let us assume that

$$\mathbf{y}(t) = \mathbf{c}e^{\lambda t} \tag{6.1.2}$$

is a solution of (6.1.1) and look into what $\lambda$ and $\mathbf{c}$ has to satisfy, *i.e.,* we are investigating for a necessary condition on $\lambda$ and $\mathbf{c}$ so that (6.1.2) is a solution of (6.1.1). Note here that (6.1.1) has the zero solution, namely $y(t) \equiv 0$ and so we are looking for a non-zero $\mathbf{c}$. Differentiating (6.1.2) with respect to $t$ and substituting in (6.1.1), leads to

$$\lambda e^{\lambda t}\mathbf{c} = Ae^{\lambda t}\mathbf{c} \text{ or equivalently } (A - \lambda I)\mathbf{c} = \mathbf{0}. \tag{6.1.3}$$

So, (6.1.2) is a solution of the given system of differential equations if and only if $\lambda$ and $\mathbf{c}$ satisfy (6.1.3). That is, given an $n \times n$ matrix $A$, we are this lead to find a pair $(\lambda, \mathbf{c})$ such that $\mathbf{c} \neq \mathbf{0}$ and (6.1.3) is satisfied.

Let $A$ be a matrix of order $n$. In general, we ask the question:
For what values of $\lambda \in \mathbb{F}$, there exist a non-zero vector $\mathbf{x} \in \mathbb{F}^n$ such that

$$A\mathbf{x} = \lambda\mathbf{x}? \tag{6.1.4}$$

Here, $\mathbb{F}^n$ stands for either the vector space $\mathbb{R}^n$ over $\mathbb{R}$ or $\mathbb{C}^n$ over $\mathbb{C}$. Equation (6.1.4) is equivalent to the equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0}.$$

By Theorem 2.6.1, this system of linear equations has a non-zero solution, if

$$\text{rank } (A - \lambda I) < n, \quad \text{or equivalently} \quad \det(A - \lambda I) = 0.$$

So, to solve (6.1.4), we are forced to choose those values of $\lambda \in \mathbb{F}$ for which $\det(A - \lambda I) = 0$. Observe that $\det(A - \lambda I)$ is a polynomial in $\lambda$ of degree $n$. We are therefore lead to the following definition.

**Definition 6.1.3 (characteristic Polynomial)** Let $A$ be a matrix of order $n$. The polynomial $\det(A - \lambda I)$ is called the characteristic polynomial of $A$ and is denoted by $p(\lambda)$. The equation $p(\lambda) = 0$ is called the characteristic equation of $A$. If $\lambda \in \mathbb{F}$ is a solution of the characteristic equation $p(\lambda) = 0$, then $\lambda$ is called a characteristic value of $A$.

Some books use the term EIGENVALUE in place of characteristic value.

**Theorem 6.1.4** Let $A = [a_{ij}];\ a_{ij} \in \mathbb{F}, \ \text{for } 1 \leq i, j \leq n$. Suppose $\lambda = \lambda_0 \in \mathbb{F}$ is a root of the characteristic equation. Then there exists a non-zero $\mathbf{v} \in \mathbb{F}^n$ such that $A\mathbf{v} = \lambda_0\mathbf{v}$.

PROOF. Since $\lambda_0$ is a root of the characteristic equation, $\det(A - \lambda_0 I) = 0$. This shows that the matrix $A - \lambda_0 I$ is singular and therefore by Theorem 2.6.1 the linear system

$$(A - \lambda_0 I_n)\mathbf{x} = \mathbf{0}$$

has a non-zero solution. □

**Remark 6.1.5** *Observe that the linear system $A\mathbf{x} = \lambda\mathbf{x}$ has a solution $\mathbf{x} = \mathbf{0}$ for every $\lambda \in \mathbb{F}$. So, we consider only those $\mathbf{x} \in \mathbb{F}^n$ that are non-zero and are solutions of the linear system $A\mathbf{x} = \lambda\mathbf{x}$.*

**Definition 6.1.6 (Eigenvalue and Eigenvector)** If the linear system $A\mathbf{x} = \lambda\mathbf{x}$ has a non-zero solution $\mathbf{x} \in \mathbb{F}^n$ for some $\lambda \in \mathbb{F}$, then

1. $\lambda \in \mathbb{F}$ is called an eigenvalue of $A$,

2. $\mathbf{0} \neq \mathbf{x} \in \mathbb{F}^n$ is called an eigenvector corresponding to the eigenvalue $\lambda$ of $A$, and

3. the tuple $(\lambda, \mathbf{x})$ is called an eigenpair.

**Remark 6.1.7** *To understand the difference between a characteristic value and an eigenvalue, we give the following example.*

*Consider the matrix* $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. *Then the characteristic polynomial of $A$ is*

$$p(\lambda) = \lambda^2 + 1.$$

*Given the matrix $A$, recall the linear transformation $T_A : \mathbb{F}^2 \longrightarrow \mathbb{F}^2$ defined by*

$$T_A(\mathbf{x}) = A\mathbf{x} \quad \textit{for every} \quad \mathbf{x} \in \mathbb{F}^2.$$

1. *If $\mathbb{F} = \mathbb{C}$, that is, if $A$ is considered a* COMPLEX *matrix, then the roots of $p(\lambda) = 0$ in $\mathbb{C}$ are $\pm i$. So, $A$ has $(i, (1, i)^t)$ and $(-i, (i, 1)^t)$ as eigenpairs.*

2. *If $\mathbb{F} = \mathbb{R}$, that is, if $A$ is considered a* REAL *matrix, then $p(\lambda) = 0$ has no solution in $\mathbb{R}$. Therefore, if $\mathbb{F} = \mathbb{R}$, then $A$ has no eigenvalue but it has $\pm i$ as characteristic values.*

**Remark 6.1.8** *Note that if $(\lambda, \mathbf{x})$ is an eigenpair for an $n \times n$ matrix $A$ then for any non-zero $c \in \mathbb{F}$, $c \neq 0$, $(\lambda, c\mathbf{x})$ is also an eigenpair for $A$. Similarly, if $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ are eigenvectors of $A$ corresponding to the eigenvalue $\lambda$, then for any non-zero $(c_1, c_2, \ldots, c_r) \in \mathbb{F}^r$, it is easily seen that if $\sum\limits_{i=1}^{r} c_i\mathbf{x}_i \neq \mathbf{0}$, then $\sum\limits_{i=1}^{r} c_i\mathbf{x}_i$ is also an eigenvector of $A$ corresponding to the eigenvalue $\lambda$. Hence, when we talk of eigenvectors corresponding to an eigenvalue $\lambda$, we mean* LINEARLY INDEPENDENT EIGENVECTORS.

*Suppose $\lambda_0 \in \mathbb{F}$ is a root of the characteristic equation $\det(A - \lambda_0 I) = 0$. Then $A - \lambda_0 I$ is singular and rank $(A - \lambda_0 I) < n$. Suppose rank $(A - \lambda_0 I) = r < n$. Then by Corollary 4.3.9, the linear system $(A - \lambda_0 I)\mathbf{x} = \mathbf{0}$ has $n - r$ linearly independent solutions. That is, $A$ has $n - r$ linearly independent eigenvectors corresponding to the eigenvalue $\lambda_0$ whenever rank $(A - \lambda_0 I) = r < n$.*

**Example 6.1.9**

1. Let $A = \text{diag}(d_1, d_2, \ldots, d_n)$ with $d_i \in \mathbb{R}$ for $1 \leq i \leq n$. Then $p(\lambda) = \prod_{i=1}^{n}(\lambda - d_i)$ is the characteristic equation. So, the eigenpairs are

$$(d_1, (1, 0, \ldots, 0)^t), (d_2, (0, 1, 0, \ldots, 0)^t), \ldots, (d_n, (0, \ldots, 0, 1)^t).$$

2. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then $\det(A - \lambda I_2) = (1 - \lambda)^2$. Hence, the characteristic equation has roots $1, 1$. That is $1$ is a repeated eigenvalue. Now check that the equation $(A - I_2)\mathbf{x} = \mathbf{0}$ for $\mathbf{x} = (x_1, x_2)^t$ is equivalent to the equation $x_2 = 0$. And this has the solution $\mathbf{x} = (x_1, 0)^t$. Hence, from the above remark, $(1, 0)^t$ is a representative for the eigenvector. Therefore, HERE WE HAVE TWO EIGENVALUES $1, 1$ BUT ONLY ONE EIGENVECTOR.

3. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Then $\det(A - \lambda I_2) = (1 - \lambda)^2$. The characteristic equation has roots $1, 1$. Here, the matrix that we have is $I_2$ and we know that $I_2\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}^t \in \mathbb{R}^2$ and we can CHOOSE ANY TWO LINEARLY INDEPENDENT VECTORS $\mathbf{x}^t, \mathbf{y}^t$ from $\mathbb{R}^2$ to get $(1, \mathbf{x})$ and $(1, \mathbf{y})$ as the two eigenpairs.

   In general, if $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent vectors in $\mathbb{R}^n$, then $(1, \mathbf{x}_1)$, $(1, \mathbf{x}_2)$, $\ldots, (1, \mathbf{x}_n)$ are eigenpairs for the identity matrix, $I_n$.

4. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Then $\det(A - \lambda I_2) = (\lambda - 3)(\lambda + 1)$. The characteristic equation has roots $3, -1$. Now check that the eigenpairs are $(3, (1,1)^t)$, and $(-1, (1,-1)^t)$. In this case, we have TWO DISTINCT EIGENVALUES AND THE CORRESPONDING EIGENVECTORS ARE ALSO LINEARLY INDEPENDENT. The reader is required to prove the linear independence of the two eigenvectors.

5. Let $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Then $\det(A - \lambda I_2) = \lambda^2 - 2\lambda + 2$. The characteristic equation has roots $1+i, 1-i$. Hence, over $\mathbb{R}$, the matrix $A$ has no eigenvalue. Over $\mathbb{C}$, the reader is required to show that the eigenpairs are $(1+i, (i,1)^t)$ and $(1-i, (1,i)^t)$.

**Exercise 6.1.10** 1. Find the eigenvalues of a triangular matrix.

2. Find eigenpairs over $\mathbb{C}$, for each of the following matrices:
$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 1+i \\ 1-i & 1 \end{bmatrix}$, $\begin{bmatrix} i & 1+i \\ -1+i & i \end{bmatrix}$, $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, and $\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$.

3. Let $A$ and $B$ be similar matrices.

   (a) Then prove that $A$ and $B$ have the same set of eigenvalues.

   (b) Let $(\lambda, \mathbf{x})$ be an eigenpair for $A$ and $(\lambda, \mathbf{y})$ be an eigenpair for $B$. What is the relationship between the vectors $\mathbf{x}$ and $\mathbf{y}$?

   [*Hint: Recall that if the matrices $A$ and $B$ are similar, then there exists a non-singular matrix $P$ such that $B = PAP^{-1}$.*]

4. Let $A = (a_{ij})$ be an $n \times n$ matrix. Suppose that for all $i$, $1 \le i \le n$, $\sum\limits_{j=1}^{n} a_{ij} = a$. Then prove that $a$ is an eigenvalue of $A$. What is the corresponding eigenvector?

5. Prove that the matrices $A$ and $A^t$ have the same set of eigenvalues. Construct a $2 \times 2$ matrix $A$ such that the eigenvectors of $A$ and $A^t$ are different.

6. Let $A$ be a matrix such that $A^2 = A$ ($A$ is called an idempotent matrix). Then prove that its eigenvalues are either $0$ or $1$ or both.

7. Let $A$ be a matrix such that $A^k = \mathbf{0}$ ($A$ is called a nilpotent matrix) for some positive integer $k \ge 1$. Then prove that its eigenvalues are all $0$.

**Theorem 6.1.11** Let $A = [a_{ij}]$ be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, not necessarily distinct. Then $\det(A) = \prod\limits_{i=1}^{n} \lambda_i$ and $\text{tr}(A) = \sum\limits_{i=1}^{n} a_{ii} = \sum\limits_{i=1}^{n} \lambda_i$.

PROOF. Since $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the $n$ eigenvalues of $A$, by definition,

$$\det(A - \lambda I_n) = p(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n). \tag{6.1.5}$$

(6.1.5) is an identity in $\lambda$ as polynomials. Therefore, by substituting $\lambda = 0$ in (6.1.5), we get

$$\det(A) = (-1)^n(-1)^n \prod_{i=1}^{n} \lambda_i = \prod_{i=1}^{n} \lambda_i.$$

Also,

$$\det(A - \lambda I_n) = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix} \quad (6.1.6)$$

$$= a_0 - \lambda a_1 + \lambda^2 a_2 + \cdots + (-1)^{n-1}\lambda^{n-1} a_{n-1} + (-1)^n \lambda^n \quad (6.1.7)$$

for some $a_0, a_1, \ldots, a_{n-1} \in \mathbb{F}$. Note that $a_{n-1}$, the coefficient of $(-1)^{n-1}\lambda^{n-1}$, comes from the product

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

So, $a_{n-1} = \sum\limits_{i=1}^{n} a_{ii} = \operatorname{tr}(A)$ by definition of trace.

But , from (6.1.5) and (6.1.7), we get

$$\begin{aligned} & a_0 - \lambda a_1 + \lambda^2 a_2 + \cdots + (-1)^{n-1}\lambda^{n-1} a_{n-1} + (-1)^n \lambda^n \\ = \; & (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n). \end{aligned} \quad (6.1.8)$$

Therefore, comparing the coefficient of $(-1)^{n-1}\lambda^{n-1}$, we have

$$\operatorname{tr}(A) = a_{n-1} = (-1)\{(-1)\sum_{i=1}^{n} \lambda_i\} = \sum_{i=1}^{n} \lambda_i.$$

Hence, we get the required result. □

**Exercise 6.1.12**

1. Let $A$ be a skew symmetric matrix of order $2n+1$. Then prove that $0$ is an eigenvalue of $A$.
2. Let $A$ be a $3 \times 3$ orthogonal matrix $(AA^t = I)$.If $\det(A) = 1$, then prove that there exists a non-zero vector $\mathbf{v} \in \mathbb{R}^3$ such that $A\mathbf{v} = \mathbf{v}$.

Let $A$ be an $n \times n$ matrix. Then in the proof of the above theorem, we observed that the characteristic equation $\det(A - \lambda I) = 0$ is a polynomial equation of degree $n$ in $\lambda$. Also, for some numbers $a_0, a_1, \ldots, a_{n-1} \in \mathbb{F}$, it has the form

$$\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^2 + \cdots a_1\lambda + a_0 = 0.$$

Note that, in the expression $\det(A - \lambda I) = 0$, $\lambda$ is an element of $\mathbb{F}$. Thus, we can only substitute $\lambda$ by elements of $\mathbb{F}$.

It turns out that the expression

$$A^n + a_{n-1}A^{n-1} + a_{n-2}A^2 + \cdots a_1 A + a_0 I = \mathbf{0}$$

holds true as a matrix identity. This is a celebrated theorem called the Cayley Hamilton Theorem. We state this theorem without proof and give some implications.

**Theorem 6.1.13 (Cayley Hamilton Theorem)** Let $A$ be a square matrix of order $n$. Then $A$ satisfies its characteristic equation. That is,

$$A^n + a_{n-1}A^{n-1} + a_{n-2}A^2 + \cdots a_1 A + a_0 I = \mathbf{0}$$

holds true as a matrix identity.

Some of the implications of Cayley Hamilton Theorem are as follows.

**Remark 6.1.14** 1. *Let* $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. *Then its characteristic polynomial is* $p(\lambda) = \lambda^2$. *Also, for the function,* $f(x) = x, \ f(0) = 0$, *and* $f(A) = A \neq \mathbf{0}$. *This shows that the condition* $f(\lambda) = 0$ *for each eigenvalue* $\lambda$ *of* $A$ *does not imply that* $f(A) = \mathbf{0}$.

2. *Suppose we are given a square matrix* $A$ *of order* $n$ *and we are interested in calculating* $A^\ell$ *where* $\ell$ *is large compared to* $n$. *Then we can use the division algorithm to find numbers* $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ *and a polynomial* $f(\lambda)$ *such that*

$$\begin{aligned} \lambda^\ell &= f(\lambda)\big(\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^2 + \cdots a_1\lambda + a_0\big) \\ &\quad +\alpha_0 + \lambda\alpha_1 + \cdots + \lambda^{n-1}\alpha_{n-1}. \end{aligned}$$

*Hence, by the Cayley Hamilton Theorem,*

$$A^\ell = \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1}A^{n-1}.$$

*That is, we just need to compute the powers of* $A$ *till* $n-1$.

*In the language of graph theory, it says the following:*

*"Let* $G$ *be a graph on* $n$ *vertices. Suppose there is no path of length* $n-1$ *or less from a vertex* $v$ *to a vertex* $u$ *of* $G$. *Then there is no path from* $v$ *to* $u$ *of any length. That is, the graph* $G$ *is disconnected and* $v$ *and* $u$ *are in different components."*

3. *Let* $A$ *be a non-singular matrix of order* $n$. *Then note that* $a_n = \det(A) \neq 0$ *and*

$$A^{-1} = \frac{-1}{a_n}[A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_1 I].$$

*This matrix identity can be used to calculate the inverse.*

*Note that the vector* $A^{-1}$ *(as an element of the vector space of all* $n \times n$ *matrices) is a linear combination of the vectors* $I, A, \ldots, A^{n-1}$.

**Exercise 6.1.15** Find inverse of the following matrices by using the Cayley Hamilton Theorem

$$i) \begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \\ 1 & 1 & 2 \end{bmatrix} \quad ii) \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad iii) \begin{bmatrix} 1 & -2 & -1 \\ -2 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

**Theorem 6.1.16** If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of a matrix $A$ with corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, then the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is linearly independent.

Proof. The proof is by induction on the number $m$ of eigenvalues. The result is obviously true if $m = 1$ as the corresponding eigenvector is non-zero and we know that any set containing exactly one non-zero vector is linearly independent.

Let the result be true for $m, \ 1 \leq m < k$. We prove the result for $m+1$. We consider the equation

$$c_1 x_1 + c_2 x_2 + \cdots + c_{m+1}x_{m+1} = \mathbf{0} \tag{6.1.9}$$

for the unknowns $c_1, c_2, \ldots, c_{m+1}$. We have

$$\begin{aligned} \mathbf{0} = A\mathbf{0} &= A(c_1 x_1 + c_2 x_2 + \cdots + c_{m+1}x_{m+1}) \\ &= c_1 Ax_1 + c_2 Ax_2 + \cdots + c_{m+1}Ax_{m+1} \\ &= c_1\lambda_1 x_1 + c_2\lambda_2 x_2 + \cdots + c_{m+1}\lambda_{m+1}x_{m+1}. \end{aligned} \tag{6.1.10}$$

From equations (6.1.9) and (6.1.10), we get

$$c_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + c_3(\lambda_3 - \lambda_1)\mathbf{x}_3 + \cdots + c_{m+1}(\lambda_{m+1} - \lambda_1)\mathbf{x}_{m+1} = \mathbf{0}.$$

This is an equation in $m$ eigenvectors. So, by the induction hypothesis, we have

$$c_i(\lambda_i - \lambda_1) = 0 \quad \text{for} \quad 2 \le i \le m+1.$$

But the eigenvalues are distinct implies $\lambda_i - \lambda_1 \neq 0$ for $2 \le i \le m+1$. We therefore get $c_i = 0$ for $2 \le i \le m+1$. Also, $\mathbf{x}_1 \neq \mathbf{0}$ and therefore (6.1.9) gives $c_1 = 0$.

Thus, we have the required result. □

We are thus lead to the following important corollary.

**Corollary 6.1.17** The eigenvectors corresponding to distinct eigenvalues of an $n \times n$ matrix $A$ are linearly independent.

**Exercise 6.1.18**

1. For an $n \times n$ matrix $A$, prove the following.
   (a) $A$ and $A^t$ have the same set of eigenvalues.
   (b) If $\lambda$ is an eigenvalue of an invertible matrix $A$ then $\frac{1}{\lambda}$ is an eigenvalue of $A^{-1}$.
   (c) If $\lambda$ is an eigenvalue of $A$ then $\lambda^k$ is an eigenvalue of $A^k$ for any positive integer $k$.
   (d) If $A$ and $B$ are $n \times n$ matrices with $A$ nonsingular then $BA^{-1}$ and $A^{-1}B$ have the same set of eigenvalues.

   In each case, what can you say about the eigenvectors?

2. Let $A$ and $B$ be $2 \times 2$ matrices for which $\det(A) = \det(B)$ and $\text{tr}(A) = \text{tr}(B)$.
   (a) Do $A$ and $B$ have the same set of eigenvalues?
   (b) Give examples to show that the matrices $A$ and $B$ need not be similar.

3. Let $(\lambda_1, \mathbf{u})$ be an eigenpair for a matrix $A$ and let $(\lambda_2, \mathbf{u})$ be an eigenpair for another matrix $B$.
   (a) Then prove that $(\lambda_1 + \lambda_2, \mathbf{u})$ is an eigenpair for the matrix $A + B$.
   (b) Give an example to show that if $\lambda_1, \lambda_2$ are respectively the eigenvalues of $A$ and $B$, then $\lambda_1 + \lambda_2$ need not be an eigenvalue of $A + B$.

4. Let $\lambda_i, 1 \le i \le n$ be distinct non-zero eigenvalues of an $n \times n$ matrix $A$. Let $\mathbf{u}_i, 1 \le i \le n$ be the corresponding eigenvectors. Then show that $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ forms a basis of $\mathbb{F}^n(\mathbb{F})$. If $[\mathbf{b}]_{\mathcal{B}} = (c_1, c_2, \ldots, c_n)^t$ then show that $A\mathbf{x} = \mathbf{b}$ has the unique solution

$$\mathbf{x} = \frac{c_1}{\lambda_1}\mathbf{u}_1 + \frac{c_2}{\lambda_2}\mathbf{u}_2 + \cdots + \frac{c_n}{\lambda_n}\mathbf{u}_n.$$

## 6.2 diagonalization

Let $A$ be a square matrix of order $n$ and let $T_A : \mathbb{F}^n \longrightarrow \mathbb{F}^n$ be the corresponding linear transformation. In this section, we ask the question "does there exist a basis $\mathcal{B}$ of $\mathbb{F}^n$ such that $T_A[\mathcal{B}, \mathcal{B}]$, the matrix of the linear transformation $T_A$, is in the simplest possible form."

We know that, the simplest form for a matrix is the identity matrix and the diagonal matrix. In this section, we show that for a certain class of matrices $A$, we can find a basis $\mathcal{B}$ such that $T_A[\mathcal{B}, \mathcal{B}]$ is a diagonal matrix, consisting of the eigenvalues of $A$. This is equivalent to saying that $A$ is similar to a diagonal matrix. To show the above, we need the following definition.

**Definition 6.2.1 (Matrix Diagonalization)** A matrix $A$ is said to be diagonalizable if there exists a non-singular matrix $P$ such that $P^{-1}AP$ is a diagonal matrix.

**Remark 6.2.2** *Let $A$ be an $n \times n$ diagonalizable matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. By definition, $A$ is similar to a diagonal matrix $D$. Observe that $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ as similar matrices have the same set of eigenvalues and the eigenvalues of a diagonal matrix are its diagonal entries.*

**Example 6.2.3** Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then we have the following:

1. Let $V = \mathbb{R}^2$. Then $A$ has no real eigenvalue (see Example 6.1.8 and hence $A$ doesn't have eigenvectors that are vectors in $\mathbb{R}^2$. Hence, there does not exist any non-singular $2 \times 2$ real matrix $P$ such that $P^{-1}AP$ is a diagonal matrix.

2. In case, $V = \mathbb{C}^2(\mathbb{C})$, the two complex eigenvalues of $A$ are $-i, i$ and the corresponding eigenvectors are $(i,1)^t$ and $(-i,1)^t$, respectively. Also, $(i,1)^t$ and $(-i,1)^t$ can be taken as a basis of $\mathbb{C}^2(\mathbb{C})$. Define a $2 \times 2$ complex matrix by $U = \frac{1}{\sqrt{2}} \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$. Then
$$U^*AU = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}.$$

**Theorem 6.2.4** let $A$ be an $n \times n$ matrix. Then $A$ is diagonalizable if and only if $A$ has $n$ linearly independent eigenvectors.

PROOF. Let $A$ be diagonalizable. Then there exist matrices $P$ and $D$ such that
$$P^{-1}AP = D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$
Or equivalently, $AP = PD$. Let $P = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$. Then $AP = PD$ implies that
$$A\mathbf{u}_i = d_i \mathbf{u}_i \quad \text{for} \quad 1 \le i \le n.$$
Since $\mathbf{u}_i$'s are the columns of a non-singular matrix $P$, they are non-zero and so for $1 \le i \le n$, we get the eigenpairs $(d_i, \mathbf{u}_i)$ of $A$. Since, $\mathbf{u}_i$'s are columns of the non-singular matrix $P$, using Corollary 4.3.9, we get $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are linearly independent.

Thus we have shown that if $A$ is diagonalizable then $A$ has $n$ linearly independent eigenvectors.

Conversely, suppose $A$ has $n$ linearly independent eigenvectors $\mathbf{u}_i$, $1 \le i \le n$ with eigenvalues $\lambda_i$. Then $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$. Let $P = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$. Since $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are linearly independent, by Corollary 4.3.9, $P$ is non-singular. Also,
$$\begin{aligned} AP &= [A\mathbf{u}_1, A\mathbf{u}_2, \ldots, A\mathbf{u}_n] = [\lambda_1\mathbf{u}_1, \lambda_2\mathbf{u}_2, \ldots, \lambda_n\mathbf{u}_n] \\ &= [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix} = PD. \end{aligned}$$
Therefore the matrix $A$ is diagonalizable. □

**Corollary 6.2.5** let $A$ be an $n \times n$ matrix. Suppose that the eigenvalues of $A$ are distinct. Then $A$ is diagonalizable.

Proof. As $A$ is an $n \times n$ matrix, it has $n$ eigenvalues. Since all the eigenvalues of $A$ are distinct, by Corollary 6.1.17, the $n$ eigenvectors are linearly independent. Hence, by Theorem 6.2.4, $A$ is diagonalizable. □

**Corollary 6.2.6** Let $A$ be an $n \times n$ matrix with $\lambda_1, \lambda_2, \ldots, \lambda_k$ as its distinct eigenvalues and $p(\lambda)$ as its characteristic polynomial. Suppose that for each $i$, $1 \leq i \leq k$, $(x - \lambda_i)^{m_i}$ divides $p(\lambda)$ but $(x - \lambda_i)^{m_i+1}$ does not divides $p(\lambda)$ for some positive integers $m_i$. Then

$$A \text{ is diagonalizable if and only if } \dim\big(\ker(A - \lambda_i I)\big) = m_i \text{ for each } i,\ 1 \leq i \leq k.$$

Or equivalently $A$ is diagonalizable if and only if $\text{rank}(A - \lambda_i I) = n - m_i$ for each $i$, $1 \leq i \leq k$.

Proof. As $A$ is diagonalizable, by Theorem 6.2.4, $A$ has $n$ linearly independent eigenvalues. Also, $\sum\limits_{i=1}^{k} m_i = n$ as $\deg(p(\lambda)) = n$. Hence, for each eigenvalue $\lambda_i$, $1 \leq i \leq k$, $A$ has exactly $m_i$ linearly independent eigenvectors. Thus, for each $i$, $1 \leq i \leq k$, the homogeneous linear system $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$ has exactly $m_i$ linearly independent vectors in its solution set. Therefore, $\dim\big(\ker(A - \lambda_i I)\big) \geq m_i$. Indeed $\dim\big(\ker(A - \lambda_i I)\big) = m_i$ for $1 \leq i \leq k$ follows from a simple counting argument.

Now suppose that for each $i$, $1 \leq i \leq k$, $\dim\big(\ker(A - \lambda_i I)\big) = m_i$. Then for each $i$, $1 \leq i \leq k$, we can choose $m_i$ linearly independent eigenvectors. Also by Corollary 6.1.17, the eigenvectors corresponding to distinct eigenvalues are linearly independent. Hence $A$ has $n = \sum\limits_{i=1}^{k} m_i$ linearly independent eigenvectors. Hence by Theorem 6.2.4, $A$ is diagonalizable. □

**Example 6.2.7** 1. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. Then $\det(A - \lambda I) = (2 - \lambda)^2(1 - \lambda)$. Hence, $A$ has eigenvalues $1, 2, 2$. It is easily seen that $\big(1, (1, 0, -1)^t\big)$ and $(\big(2, (1, 1, -1)^t\big)$ are the only eigenpairs. That is, the matrix $A$ has exactly one eigenvector corresponding to the repeated eigenvalue 2. Hence, by Theorem 6.2.4, the matrix $A$ is not diagonalizable.

2. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Then $\det(A - \lambda I) = (4 - \lambda)(1 - \lambda)^2$. Hence, $A$ has eigenvalues $1, 1, 4$. It can be easily verified that $(1, -1, 0)^t$ and $(1, 0, -1)^t$ correspond to the eigenvalue 1 and $(1, 1, 1)^t$ corresponds to the eigenvalue 4. Note that the set $\{(1, -1, 0)^t, (1, 0, -1)^t\}$ consisting of eigenvectors corresponding to the eigenvalue 1 are not orthogonal. This set can be replaced by the orthogonal set $\{(1, 0, -1)^t, (1, -2, 1)^t\}$ which still consists of eigenvectors corresponding to the eigenvalue 1 as $(1, -2, 1) = 2(1, -1, 0) - (1, 0, -1)$. Also, the set $\{(1, 1, 1), (1, 0, -1), (1, -2, 1)\}$ forms a basis of $\mathbb{R}^3$. So, by Theorem 6.2.4, the matrix $A$ is diagonalizable. Also, if $U = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}$ is the corresponding unitary matrix then $U^*AU = \text{diag}(4, 1, 1)$.

Observe that the matrix $A$ is a symmetric matrix. In this case, the eigenvectors are mutually orthogonal. In general, for any $n \times n$ real symmetric matrix $A$, there always exist $n$ eigenvectors and they are mutually orthogonal. This result will be proved later.

**Exercise 6.2.8** 1. By finding the eigenvalues of the following matrices, justify whether or not $A = PDP^{-1}$ for some real non-singular matrix $P$ and a real diagonal matrix $D$.

$i)$ $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ $ii)$ $\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ for any $\theta$ with $0 \leq \theta \leq 2\pi$.

2. Let $A$ be an $n \times n$ matrix and $B$ an $m \times m$ matrix. Suppose $C = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix}$. Then show that $C$ is diagonalizable if and only if both $A$ and $B$ are diagonalizable.

3. Let $T : \mathbb{R}^5 \longrightarrow \mathbb{R}^5$ be a linear transformation with rank $(T - I) = 3$ and
$$\mathcal{N}(T) = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 \mid x_1 + x_4 + x_5 = 0, \ x_2 + x_3 = 0\}.$$
Then
   (a) determine the eigenvalues of $T$?
   (b) find the number of linearly independent eigenvectors corresponding to each eigenvalue?
   (c) is $T$ diagonalizable? Justify your answer.

4. Let $A$ be a non-zero square matrix such that $A^2 = \mathbf{0}$. Show that $A$ cannot be diagonalized. [*Hint: Use Remark 6.2.2.*]

5. Are the following matrices diagonalizable?
$$i) \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 2 & 3 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad ii) \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad iii) \begin{bmatrix} 1 & -3 & 3 \\ 0 & -5 & 6 \\ 0 & -3 & 4 \end{bmatrix}.$$

## 6.3 Diagonalizable matrices

In this section, we will look at some special classes of square matrices which are diagonalizable. We will also be dealing with matrices having complex entries and hence for a matrix $A = [a_{ij}]$, recall the following definitions.

**Definition 6.3.1 (Special Matrices)** 1. $A^* = (\overline{a_{ji}})$, is called the **conjugate transpose** of the matrix $A$.
Note that $A^* = \overline{A^t} = \overline{A}^t$.

2. A square matrix $A$ with complex entries is called
   (a) a Hermitian matrix if $A^* = A$.
   (b) a unitary matrix if $A\,A^* = A^*A = I_n$.
   (c) a skew-Hermitian matrix if $A^* = -A$.
   (d) a normal matrix if $A^*A = AA^*$.

3. A square matrix $A$ with real entries is called
   (a) a **symmetric** matrix if $A^t = A$.
   (b) an **orthogonal** matrix if $A\,A^t = A^tA = I_n$.
   (c) a **skew-symmetric** matrix if $A^t = -A$.

Note that a symmetric matrix is always Hermitian, a skew-symmetric matrix is always skew-Hermitian and an orthogonal matrix is always unitary. Each of these matrices are normal. If $A$ is a unitary matrix then $A^* = A^{-1}$.

**Example 6.3.2** 1. Let $B = \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix}$. Then $B$ is skew-Hermitian.

2. Let $A = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Then $A$ is a unitary matrix and $B$ is a normal matrix. Note that $\sqrt{2}A$ is also a normal matrix.

**Definition 6.3.3 (Unitary Equivalence)** Let $A$ and $B$ be two $n \times n$ matrices. They are called unitarily equivalent if there exists a unitary matrix $U$ such that $A = U^*BU$.

**Exercise 6.3.4**

1. Let $A$ be any matrix. Then $A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*)$ where $\frac{1}{2}(A + A^*)$ is the Hermitian part of $A$ and $\frac{1}{2}(A - A^*)$ is the skew-Hermitian part of $A$.

2. Every matrix can be uniquely expressed as $A = S + iT$ where both $S$ and $T$ are Hermitian matrices.

3. Show that $A - A^*$ is always skew-Hermitian.

4. Does there exist a unitary matrix $U$ such that $UAU^{-1} = B$ where $A = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -1 & 3\sqrt{2} \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 3 \end{bmatrix}$.

**Proposition 6.3.5** Let $A$ be an $n \times n$ Hermitian matrix. Then all the eigenvalues of $A$ are real.

PROOF. Let $(\lambda, \mathbf{x})$ be an eigenpair. Then $A\mathbf{x} = \lambda\mathbf{x}$ and $A = A^*$ implies

$$\mathbf{x}^*A = \mathbf{x}^*A^* = (A\mathbf{x})^* = (\lambda\mathbf{x})^* = \overline{\lambda}\mathbf{x}^*.$$

Hence

$$\lambda\mathbf{x}^*\mathbf{x} = \mathbf{x}^*(\lambda\mathbf{x}) = \mathbf{x}^*(A\mathbf{x}) = (\mathbf{x}^*A)\mathbf{x} = (\overline{\lambda}\mathbf{x}^*)\mathbf{x} = \overline{\lambda}\mathbf{x}^*\mathbf{x}.$$

But $\mathbf{x}$ is an eigenvector and hence $\mathbf{x} \neq \mathbf{0}$ and so the real number $\|\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x}$ is non-zero as well. Thus $\lambda = \overline{\lambda}$. That is, $\lambda$ is a real number. □

**Theorem 6.3.6** Let $A$ be an $n \times n$ Hermitian matrix. Then $A$ is unitarily diagonalizable. That is, there exists a unitary matrix $U$ such that $U^*AU = D$; where $D$ is a diagonal matrix with the eigenvalues of $A$ as the diagonal entries.

In other words, the eigenvectors of $A$ form an orthonormal basis of $\mathbb{C}^n$.

PROOF. We will prove the result by induction on the size of the matrix. The result is clearly true if $n = 1$. Let the result be true for $n = k - 1$. we will prove the result in case $n = k$. So, let $A$ be a $k \times k$ matrix and let $(\lambda_1, \mathbf{x})$ be an eigenpair of $A$ with $\|\mathbf{x}\| = 1$. We now extend the linearly independent set $\{\mathbf{x}\}$ to form an orthonormal basis $\{\mathbf{x}, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k\}$ (using *Gram-Schmidt Orthogonalisation*) of $\mathbb{C}^k$.

As $\{\mathbf{x}, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k\}$ is an orthonormal set,

$$\mathbf{u}_i^*\mathbf{x} = 0 \quad \text{for all} \quad i = 2, 3, \ldots, k.$$

Therefore, observe that for all $i$, $2 \leq i \leq k$,

$$(A\mathbf{u}_i)^*\mathbf{x} = (\mathbf{u}_i * A^*)\mathbf{x} = \mathbf{u}_i^*(A^*\mathbf{x}) = \mathbf{u}_i^*(A\mathbf{x}) = \mathbf{u}_i^*(\lambda_1\mathbf{x}) = \lambda_1(\mathbf{u}_i^*\mathbf{x}) = 0.$$

Hence, we also have $\mathbf{x}^*(A\mathbf{u}_i) = 0$ for $2 \le i \le k$. Now, define $U_1 = [\mathbf{x},\ \mathbf{u}_2,\ \cdots, \mathbf{u}_k]$ (with $\mathbf{x}, \mathbf{u}_2, \ldots, \mathbf{u}_k$ as columns of $U_1$). Then the matrix $U_1$ is a unitary matrix and

$$\begin{aligned}
U_1^{-1} A U_1 &= U_1^* A U_1 = U_1^*[A\mathbf{x}\ A\mathbf{u}_2\ \cdots A\mathbf{u}_k] \\
&= \begin{bmatrix} \mathbf{x}^* \\ \mathbf{u}_2^* \\ \vdots \\ \mathbf{u}_k^* \end{bmatrix} [\lambda_1\mathbf{x}\ A\mathbf{u}_2\ \cdots A\mathbf{u}_k] = \begin{bmatrix} \lambda_1 \mathbf{x}^*\mathbf{x} & \cdots & \mathbf{x}^* A\mathbf{u}_k \\ \mathbf{u}_2^*(\lambda_1\mathbf{x}) & \cdots & \mathbf{u}_2^*(A\mathbf{u}_k) \\ \vdots & \ddots & \vdots \\ \mathbf{u}_k^*(\lambda_1\mathbf{x}) & \cdots & \mathbf{u}_k^*(A\mathbf{u}_k) \end{bmatrix} \\
&= \left[\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0} & \\ \vdots & B \\ \mathbf{0} & \end{array}\right],
\end{aligned}$$

where $B$ is a $(k-1) \times (k-1)$ matrix. As the matrix $U_1$ is unitary, $U_1^* = U_1^{-1}$. So, $A^* = A$ gives $(U_1^{-1} A U_1)^* = U_1^{-1} A U_1$. This condition, together with the fact that $\lambda_1$ is a real number (use Proposition 6.3.5), implies that $B^* = B$. That is, $B$ is also a Hermitian matrix. Therefore, by induction hypothesis there exists a $(k-1) \times (k-1)$ unitary matrix $U_2$ such that

$$U_2^{-1} B U_2 = D_2 = \operatorname{diag}(\lambda_2, \ldots, \lambda_k).$$

Recall that , the entries $\lambda_i$, for $2 \le i \le k$ are the eigenvalues of the matrix $B$. We also know that two similar matrices have the same set of eigenvalues. Hence, the eigenvalues of $A$ are $\lambda_1, \lambda_2, \ldots, \lambda_k$. Define $U = U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix}$. Then $U$ is a unitary matrix and

$$\begin{aligned}
U^{-1} A U &= \left( U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \right)^{-1} A \left( U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \right) \\
&= \left( \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2^{-1} \end{bmatrix} U_1^{-1} \right) A \left( U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \right) \\
&= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2^{-1} \end{bmatrix} \left( U_1^{-1} A U_1 \right) \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & U_2^{-1} B U_2 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & D_2 \end{bmatrix}.
\end{aligned}$$

Thus, $U^{-1}AU$ is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_k$, the eigenvalues of $A$. Hence, the result follows. □

**Corollary 6.3.7** Let $A$ be an $n \times n$ real symmetric matrix. Then

1. the eigenvalues of $A$ are all real,

2. the corresponding eigenvectors can be chosen to have real entries, and

3. the eigenvectors also form an orthonormal basis of $\mathbb{R}^n$.

Proof. As $A$ is symmetric, $A$ is also an Hermitian matrix. Hence, by Proposition 6.3.5, the eigenvalues of $A$ are all real. Let $(\lambda,\ \mathbf{x})$ be an eigenpair of $A$. Suppose $\mathbf{x}^t \in \mathbb{C}^n$. Then there exist $\mathbf{y}^t, \mathbf{z}^t \in \mathbb{R}^n$ such that $\mathbf{x} = \mathbf{y} + i\mathbf{z}$. So,

$$A\mathbf{x} = \lambda\mathbf{x} \Longrightarrow A(\mathbf{y} + i\mathbf{z}) = \lambda(\mathbf{y} + i\mathbf{z}).$$

Comparing the real and imaginary parts, we get $A\mathbf{y} = \lambda\mathbf{y}$ and $A\mathbf{z} = \lambda\mathbf{z}$. Thus, we can choose the eigenvectors to have real entries.

To prove the orthonormality of the eigenvectors, we proceed on the lines of the proof of Theorem 6.3.6, Hence, the readers are advised to complete the proof. □

**Exercise 6.3.8** 1. Let $A$ be a skew-Hermitian matrix. Then all the eigenvalues of $A$ are either zero or purely imaginary. Also, the eigenvectors corresponding to distinct eigenvalues are mutually orthogonal. *[Hint: Carefully study the proof of Theorem 6.3.6.]*

2. Let $A$ be an $n \times n$ unitary matrix. Then

   (a) the rows of $A$ form an orthonormal basis of $\mathbb{C}^n$.

   (b) the columns of $A$ form an orthonormal basis of $\mathbb{C}^n$.

   (c) for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{n\times 1}$, $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$.

   (d) for any vector $\mathbf{x} \in \mathbb{C}^{n\times 1}$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$.

   (e) for any eigenvalue $\lambda$ $A$, $|\lambda| = 1$.

   (f) the eigenvectors $\mathbf{x}, \mathbf{y}$ corresponding to distinct eigenvalues $\lambda$ and $\mu$ satisfy $\langle \mathbf{x}, \mathbf{y}\rangle = 0$. That is, if $(\lambda, \mathbf{x})$ and $(\mu, \mathbf{y})$ are eigenpairs, with $\lambda \neq \mu$, then $\mathbf{x}$ and $\mathbf{y}$ are mutually orthogonal.

3. Let $A$ be a normal matrix. Then, show that if $(\lambda, \mathbf{x})$ is an eigenpair for $A$ then $(\overline{\lambda}, \mathbf{x})$ is an eigenpair for $A^*$.

4. Show that the matrices $A = \begin{bmatrix} 4 & 4 \\ 0 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 10 & 9 \\ -4 & -2 \end{bmatrix}$ are similar. Is it possible to find a unitary matrix $U$ such that $A = U^*BU$?

5. Let $A$ be a $2 \times 2$ orthogonal matrix. Then prove the following:

   (a) if $\det(A) = 1$, then $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ for some $\theta$, $0 \leq \theta < 2\pi$.

   (b) if $\det A = -1$, then there exists a basis of $\mathbb{R}^2$ in which the matrix of $A$ looks like $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

6. Describe all $2 \times 2$ orthogonal matrices.

7. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Determine $A^{301}$.

8. Let $A$ be a $3 \times 3$ orthogonal matrix. Then prove the following:

   (a) if $\det(A) = 1$, then $A$ is a rotation about a fixed axis, in the sense that $A$ has an eigenpair $(1, \mathbf{x})$ such that the restriction of $A$ to the plane $\mathbf{x}^\perp$ is a two dimensional rotation of $\mathbf{x}^\perp$.

   (b) if $\det A = -1$, then the action of $A$ corresponds to a reflection through a plane $P$, followed by a rotation about the line through the origin that is perpendicular to $P$.

**Remark 6.3.9** *In the previous exercise, we saw that the matrices* $A = \begin{bmatrix} 4 & 4 \\ 0 & 4 \end{bmatrix}$ *and* $B = \begin{bmatrix} 10 & 9 \\ -4 & -2 \end{bmatrix}$ *are similar but not unitarily equivalent, whereas unitary equivalence implies similarity equivalence as* $U^* = U^{-1}$. *But in numerical calculations, unitary transformations are preferred as compared to similarity transformations. The main reasons being:*

1. *Exercise 6.3.8.2 implies that an orthonormal change of basis leaves unchanged the sum of squares of the absolute values of the entries which need not be true under a non-orthonormal change of basis.*

2. *As $U^* = U^{-1}$ for a unitary matrix $U$, unitary equivalence is computationally simpler.*

3. *Also in doing "conjugate transpose", the loss of accuracy due to round-off errors doesn't occur.*

We next prove the Schur's Lemma and use it to show that normal matrices are unitarily diagonalizable.

**Lemma 6.3.10** (Schur's Lemma) Every $n \times n$ complex matrix is unitarily similar to an upper triangular matrix.

PROOF. We will prove the result by induction on the size of the matrix. The result is clearly true if $n = 1$. Let the result be true for $n = k - 1$. we will prove the result in case $n = k$. So, let $A$ be a $k \times k$ matrix and let $(\lambda_1, x)$ be an eigenpair for $A$ with $\|x\| = 1$. Now the linearly independent set $\{x\}$ is extended, using the *Gram-Schmidt Orthogonalisation,* to get an orthonormal basis $\{x, u_2, u_3, \ldots, u_k\}$. Then $U_1 = [x \; u_2 \; \cdots u_k]$ (with $x, u_2, \ldots, u_k$ as the columns of the matrix $U_1$ ) is a unitary matrix and

$$\begin{aligned} U_1^{-1} A U_1 &= U_1^* A U_1 = U_1^* [Ax \; Au_2 \; \cdots Au_k] \\ &= \begin{bmatrix} x^* \\ u_2^* \\ \vdots \\ u_k^* \end{bmatrix} [\lambda_1 x \; Au_2 \; \cdots Au_k] = \left[\begin{array}{c|c} \lambda_1 & * \\ \hline \mathbf{0} & \\ \vdots & B \\ \mathbf{0} & \end{array}\right] \end{aligned}$$

where $B$ is a $(k-1) \times (k-1)$ matrix. By induction hypothesis there exists a $(k-1) \times (k-1)$ unitary matrix $U_2$ such that $U_2^{-1} B U_2$ is an upper triangular matrix with diagonal entries $\lambda_2, \ldots, \lambda_k$, the eigen values of the matrix $B$. Observe that since the eigenvalues of $B$ are $\lambda_2, \ldots, \lambda_k$ the eigenvalues of $A$ are $\lambda_1, \lambda_2, \ldots, \lambda_k$. Define $U = U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix}$. Then check that $U$ is a unitary matrix and $U^{-1} A U$ is an upper triangular matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_k$, the eigenvalues of the matrix $A$. Hence, the result follows. □

**Exercise 6.3.11**

1. Let $A$ be an $n \times n$ real invertible matrix. Prove that there exists an orthogonal matrix $P$ and a diagonal matrix $D$ with positive diagonal entries such that $AA^t = PDP^{-1}$.

2. Show that matrices $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -1 & \sqrt{2} \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ are unitarily equivalent via the unitary matrix $U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$. Hence, conclude that the upper triangular matrix obtained in the "Schur's Lemma" need not be unique.

3. Show that the normal matrices are diagonalizable.
*[Hint: Show that the matrix $B$ in the proof of the above theorem is also a normal matrix and if $T$ is an upper triangular matrix with $T^*T = TT^*$ then $T$ has to be a diagonal matrix].*

**Remark 6.3.12 (The Spectral Theorem for Normal Matrices)** *Let $A$ be an $n \times n$ normal matrix. Then the above exercise shows that there exists an orthonormal basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ of $\mathbb{C}^n(\mathbb{C})$ such that $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for $1 \le i \le n$.*

4. Let $A$ be a normal matrix. Prove the following:

   (a) if all the eigenvalues of $A$ are $0$, then $A = \mathbf{0}$,

   (b) if all the eigenvalues of $A$ are $1$, then $A = I$.

We end this chapter with an application of the theory of diagonalization to the study of conic sections in analytic geometry and the study of maxima and minima in analysis.

## 6.4 Sylvester's Law of Inertia and Applications

**Definition 6.4.1 (Bilinear Form)** Let $A$ be a $n \times n$ matrix with real entries. A bilinear form in $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t$ is an expression of the type

$$Q(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t A \mathbf{y} = \sum_{i,j=1}^{n} a_{ij} x_i y_j.$$

Observe that if $A = I$ (the identity matrix) then the bilinear form reduces to the standard real inner product. Also, if we want it to be symmetric in $\mathbf{x}$ and $\mathbf{y}$ then it is necessary and sufficient that $a_{ij} = a_{ji}$ for all $i, j = 1, 2, \ldots, n$. Why? Hence, any symmetric bilinear form is naturally associated with a real symmetric matrix.

**Definition 6.4.2 (Sesquilinear Form)** Let $A$ be a $n \times n$ matrix with complex entries. A sesquilinear form in $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t$ is given by

$$H(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} a_{ij} x_i \overline{y_j}.$$

Note that if $A = I$ (the identity matrix) then the sesquilinear form reduces to the standard complex inner product. Also, it can be easily seen that this form is 'linear' in the first component and 'conjugate linear' in the second component. Also, if we want $H(\mathbf{x}, \mathbf{y}) = \overline{H(\mathbf{y}, \mathbf{x})}$ then the matrix $A$ need to be an Hermitian matrix. Note that if $a_{ij} \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then the sesquilinear form reduces to a bilinear form.

The expression $Q(\mathbf{x}, \mathbf{x})$ is called the quadratic form and $H(\mathbf{x}, \mathbf{x})$ the Hermitian form. We generally write $Q(\mathbf{x})$ and $H(\mathbf{x})$ in place of $Q(\mathbf{x}, \mathbf{x})$ and $H(\mathbf{x}, \mathbf{x})$, respectively. It can be easily shown that for any choice of $\mathbf{x}$, the Hermitian form $H(\mathbf{x})$ is a real number.

Therefore, in matrix notation, for a Hermitian matrix $A$, the Hermitian form can be rewritten as

$$H(\mathbf{x}) = \mathbf{x}^t A \mathbf{x}, \qquad \text{where } \mathbf{x} = (x_1, x_2, \ldots, x_n)^t, \text{ and } A = [a_{ij}].$$

**Example 6.4.3** Let $A = \begin{bmatrix} 1 & 2-i \\ 2+i & 2 \end{bmatrix}$. Then check that $A$ is an Hermitian matrix and for $\mathbf{x} = (x_1, x_2)^t$, the Hermitian form

$$\begin{aligned} H(\mathbf{x}) &= \mathbf{x}^* A \mathbf{x} = (\overline{x_1}, \overline{x_2}) \begin{bmatrix} 1 & 2-i \\ 2+i & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \overline{x_1} x_1 + 2\overline{x_2} x_2 + (2-i)\overline{x_1} x_2 + (2+i)\overline{x_2} x_1 \\ &= |x_1|^2 + 2|x_2|^2 + 2\text{Re}[(2-i)\overline{x_1} x_2] \end{aligned}$$

where 'Re' denotes the real part of a complex number. This shows that for every choice of $\mathbf{x}$ the Hermitian form is always real. Why?

The main idea is to express $H(\mathbf{x})$ as sum of squares and hence determine the possible values that it can take. Note that if we replace $\mathbf{x}$ by $c\mathbf{x}$, where $c$ is any complex number, then $H(\mathbf{x})$ simply gets multiplied by $|c|^2$ and hence one needs to study only those $\mathbf{x}$ for which $\|\mathbf{x}\| = 1$, i.e., $\mathbf{x}$ is a normalised vector.

From Exercise 6.3.11.3 one knows that if $A = A^*$ ($A$ is Hermitian) then there exists a unitary matrix $U$ such that $U^*AU = D$ ($D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ with $\lambda_i$'s the eigenvalues of the matrix $A$ which we know are real). So, taking $\mathbf{z} = U^*\mathbf{x}$ (i.e., choosing $z_i$'s as linear combination of $x_j$'s with coefficients coming from the entries of the matrix $U^*$), one gets

$$H(\mathbf{x}) = \mathbf{x}^*A\mathbf{x} = \mathbf{z}^*U^*AU\mathbf{z} = \mathbf{z}^*D\mathbf{z} = \sum_{i=1}^{n} \lambda_i |z_i|^2 = \sum_{i=1}^{n} \lambda_i \left| \sum_{j=1}^{n} u_{ji}{}^* x_j \right|^2. \tag{6.4.1}$$

Thus, one knows the possible values that $H(\mathbf{x})$ can take depending on the eigenvalues of the matrix $A$ in case $A$ is a Hermitian matrix. Also, for $1 \le i \le n$, $\sum\limits_{j=1}^{n} u_{ji}{}^* x_j$ represents the principal axes of the conic that they represent in the n-dimensional space.

Equation (6.4.1) gives one method of writing $H(\mathbf{x})$ as a sum of $n$ absolute squares of linearly independent linear forms. One can easily show that there are more than one way of writing $H(\mathbf{x})$ as sum of squares. The question arises, "what can we say about the coefficients when $H(\mathbf{x})$ has been written as sum of absolute squares".

This question is answered by 'Sylvester's law of inertia' which we state as the next lemma.

**Lemma 6.4.4** Every Hermitian form $H(\mathbf{x}) = \mathbf{x}^*A\mathbf{x}$ (with $A$ an Hermitian matrix) in $n$ variables can be written as

$$H(\mathbf{x}) = |y_1|^2 + |y_2|^2 + \cdots + |y_p|^2 - |y_{p+1}|^2 - \cdots - |y_r|^2$$

where $y_1, y_2, \ldots, y_r$ are linearly independent linear forms in $x_1, x_2, \ldots, x_n$, and the integers $p$ and $r$, $0 \le p \le r \le n$, depend only on $A$.

Proof. From Equation (6.4.1) it is easily seen that $H(\mathbf{x})$ has the required form. Need to show that $p$ and $r$ are uniquely given by $A$.

Hence, let us assume on the contrary that there exist positive integers $p, q, r, s$ with $p > q$ such that

$$\begin{aligned} H(\mathbf{x}) &= |y_1|^2 + |y_2|^2 + \cdots + |y_p|^2 - |y_{p+1}|^2 - \cdots - |y_r|^2 \\ &= |z_1|^2 + |z_2|^2 + \cdots + |z_q|^2 - |z_{q+1}|^2 - \cdots - |z_s|^2. \end{aligned}$$

Since, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t$ and $\mathbf{z} = (z_1, z_2, \ldots, z_n)^t$ are linear combinations of $x_1, x_2, \ldots, x_n$, we can find a matrix $B$ such that $\mathbf{z} = B\mathbf{y}$. Choose $y_{p+1} = y_{p+2} = \cdots = y_r = 0$. Since $p > q$, Theorem 2.6.1, gives the existence of finding nonzero values of $y_1, y_2, \ldots, y_p$ such that $z_1 = z_2 = \cdots = z_q = 0$. Hence, we get

$$|y_1|^2 + |y_2|^2 + \cdots + |y_p|^2 = -(|z_{q+1}|^2 + \cdots + |z_s|^2).$$

Now, this can hold only if $y_1 = y_2 = \cdots = y_p = 0$, which gives a contradiction. Hence $p = q$.

Similarly, the case $r > s$ can be resolved. □

**Note:** The integer $r$ is the rank of the matrix $A$ and the number $r - 2p$ is sometimes called the inertial degree of $A$.

We complete this chapter by understanding the graph of

$$ax^2 + 2hxy + by^2 + 2fx + 2gy + c = 0$$

for $a, b, c, f, g, h \in \mathbb{R}$. We first look at the following example.

**Example 6.4.5** Sketch the graph of $3x^2+4xy+3y^2=5$.

**Solution:** Note that

$$3x^2+4xy+3y^2=[x,\ \ y]\begin{bmatrix}3&2\\2&3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}.$$

The eigenpairs for $\begin{bmatrix}3&2\\2&3\end{bmatrix}$ are $(5,(1,1)^t),\ (1,(1,-1)^t)$. Thus,

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

Let

$$\begin{bmatrix}u\\v\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\frac{x+y}{\sqrt{2}}\\\frac{x-y}{\sqrt{2}}\end{bmatrix}.$$

Then

$$\begin{aligned}3x^2+4xy+3y^2 &= [x,\ \ y]\begin{bmatrix}3&2\\2&3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}\\ &= [x,\ \ y]\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}5&0\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}\\ &= [u,\ \ v]\begin{bmatrix}5&0\\0&1\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix}\\ &= 5u^2+v^2.\end{aligned}$$

Thus the given graph reduces to

$$5u^2+v^2=5 \quad \text{or equivalently} \quad u^2+\frac{v^2}{5}=1.$$

Therefore, the given graph represents an ellipse with the principal axes $u=0$ and $v=0$. That is, the principal axes are

$$y+x=0 \text{ and } \quad x-y=0.$$

The eccentricity of the ellipse is $e=\frac{2}{\sqrt{5}}$, the foci are at the points $S_1=(-\sqrt{2},\sqrt{2})$ and $S_2=(\sqrt{2},-\sqrt{2})$, and the equations of the directrices are $x-y=\pm\frac{5}{\sqrt{2}}$.

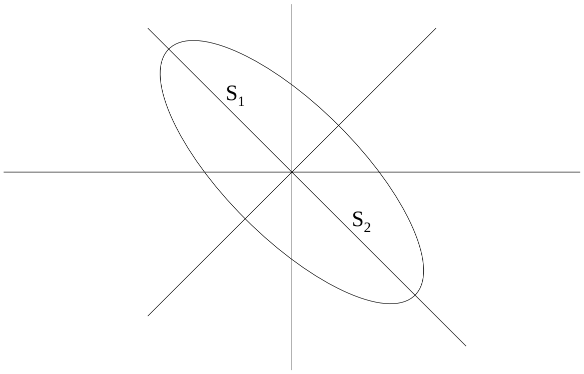


Figure 6.1: Ellipse

**Definition 6.4.6 (Associated Quadratic Form)** Let $ax^2+2hxy+by^2+2gx+2fy+c=0$ be the equation of a general conic. The quadratic expression

$$ax^2+2hxy+by^2 = \begin{bmatrix}x, & y\end{bmatrix}\begin{bmatrix}a & h\\ h & b\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}$$

is called the quadratic form associated with the given conic.

We now consider the general conic. We obtain conditions on the eigenvalues of the associated quadratic form to characterize the different conic sections in $\mathbb{R}^2$ (endowed with the standard inner product).

**Proposition 6.4.7** Consider the general conic

$$ax^2+2hxy+by^2+2gx+2fy+c=0.$$

Prove that this conic represents

1. an ellipse if $ab-h^2>0$,
2. a parabola if $ab-h^2=0$, and
3. a hyperbola if $ab-h^2<0$.

Proof. Let $A=\begin{bmatrix}a & h\\ h & b\end{bmatrix}$. Then the associated quadratic form

$$ax^2+2hxy+by^2 = \begin{bmatrix}x & y\end{bmatrix}A\begin{bmatrix}x\\ y\end{bmatrix}.$$

As $A$ is a symmetric matrix, by Corollary 6.3.7, the eigenvalues $\lambda_1,\lambda_2$ of $A$ are both real, the corresponding eigenvectors $\mathbf{u}_1,\mathbf{u}_2$ are orthonormal and $A$ is unitarily diagonalizable with

$$A=\begin{bmatrix}\mathbf{u}_1^t\\ \mathbf{u}_2^t\end{bmatrix}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}. \qquad (6.4.2)$$

Let $\begin{bmatrix}u\\ v\end{bmatrix}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}$. Then

$$ax^2+2hxy+by^2=\lambda_1u^2+\lambda_2v^2$$

and the equation of the conic section in the $(u,v)$-plane, reduces to

$$\lambda_1u^2+\lambda_2v^2+2g_1u+2f_1v+c=0.$$

Now, depending on the eigenvalues $\lambda_1,\lambda_2$, we consider different cases:

1. $\lambda_1=0=\lambda_2$.
   Substituting $\lambda_1=\lambda_2=0$ in (6.4.2) gives $A=\mathbf{0}$. Thus, the given conic reduces to a straight line $2g_1u+2f_1v+c=0$ in the $(u,v)$-plane.

2. $\lambda_1=0,\lambda_2\neq 0$.
   In this case, the equation of the conic reduces to

$$\lambda_2(v+d_1)^2=d_2u+d_3 \quad \text{for some} \quad d_1,d_2,d_3\in\mathbb{R}.$$

   (a) If $d_2=d_3=0$, then in the $(u,v)$-plane, we get the pair of coincident lines $v=-d_1$.

(b) If $d_2 = 0, \ d_3 \neq 0$.

i. If $\lambda_2 \cdot d_3 > 0$, then we get a pair of parallel lines $v = -d_1 \pm \sqrt{\frac{d_3}{\lambda_2}}$.

ii. If $\lambda_2 \cdot d_3 < 0$, the solution set corresponding to the given conic is an empty set.

(c) If $d_2 \neq 0$. Then the given equation is of the form $Y^2 = 4aX$ for some translates $X = x + \alpha$ and $Y = y + \beta$ and thus represents a parabola.

Also, observe that $\lambda_1 = 0$ implies that the $\det(A) = 0$. That is, $ab - h^2 = \det(A) = 0$.

3. $\lambda_1 > 0$ and $\lambda_2 < 0$.
Let $\lambda_2 = -\alpha_2$. Then the equation of the conic can be rewritten as

$$\lambda_1(u + d_1)^2 - \alpha_2(v + d_2)^2 = d_3 \quad \text{for some} \quad d_1, d_2, d_3 \in \mathbb{R}.$$

In this case, we have the following:

(a) suppose $d_3 = 0$. Then the equation of the conic reduces to

$$\lambda_1(u + d_1)^2 - \alpha_2(v + d_2)^2 = 0.$$

The terms on the left can be written as product of two factors as $\lambda_1, \alpha_2 > 0$. Thus, in this case, the given equation represents a pair of intersecting straight lines in the $(u, v)$-plane.

(b) suppose $d_3 \neq 0$. As $d_3 \neq 0$, we can assume $d_3 > 0$. So, the equation of the conic reduces to

$$\frac{\lambda_1(u + d_1)^2}{d_3} - \frac{\alpha_2(v + d_2)^2}{d_3} = 1.$$

This equation represents a hyperbola in the $(u, v)$-plane, with principal axes

$$u + d_1 = 0 \text{ and } \quad v + d_2 = 0.$$

As $\lambda_1\lambda_2 < 0$, we have

$$ab - h^2 = \det(A) = \lambda_1\lambda_2 < 0.$$

4. $\lambda_1, \lambda_2 > 0$.
In this case, the equation of the conic can be rewritten as

$$\lambda_1(u + d_1)^2 + \lambda_2(v + d_2)^2 = d_3, \quad \text{for some} \quad d_1, d_2, d_3 \in \mathbb{R}.$$

we now consider the following cases:

(a) suppose $d_3 = 0$. Then the equation of the ellipse reduces to a pair of perpendicular lines $u + d_1 = 0$ and $v + d_2 = 0$ in the $(u, v)$-plane.

(b) suppose $d_3 < 0$. Then there is no solution for the given equation. Hence, we do not get any real ellipse in the $(u, v)$-plane.

(c) suppose $d_3 > 0$. In this case, the equation of the conic reduces to

$$\frac{\lambda_1(u + d_1)^2}{d_3} + \frac{\alpha_2(v + d_2)^2}{d_3} = 1.$$

This equation represents an ellipse in the $(u, v)$-plane, with principal axes

$$u + d_1 = 0 \text{ and } \quad v + d_2 = 0.$$

Also, the condition $\lambda_1\lambda_2 > 0$ implies that

$$ab - h^2 = \det(A) = \lambda_1\lambda_2 > 0.$$

□

**Remark 6.4.8** *Observe that the condition*

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

*implies that the principal axes of the conic are functions of the eigenvectors* $\mathbf{u}_1$ *and* $\mathbf{u}_2$.

**Exercise 6.4.9** Sketch the graph of the following surfaces:

1. $x^2 + 2xy + y^2 - 6x - 10y = 3.$

2. $2x^2 + 6xy + 3y^2 - 12x - 6y = 5.$

3. $4x^2 - 4xy + 2y^2 + 12x - 8y = 10.$

4. $2x^2 - 6xy + 5y^2 - 10x + 4y = 7.$

As a last application, we consider the following problem that helps us in understanding the quadrics. Let

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz + 2lx + 2my + 2nz + q = 0 \qquad (6.4.3)$$

be a general quadric. Then we need to follow the steps given below to write the above quadric in the standard form and thereby get the picture of the quadric. The steps are:

1. Observe that this equation can be rewritten as

$$\mathbf{x}^t A\mathbf{x} + \mathbf{b}^t\mathbf{x} + q = 0,$$

   where

$$A = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2l \\ 2m \\ 2n \end{bmatrix}, \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

2. As the matrix $A$ is symmetric matrix, find an orthogonal matrix $P$ such that $P^tAP$ is a diagonal matrix.

3. Replace the vector $\mathbf{x}$ by $\mathbf{y} = P^t\mathbf{x}$. Then writing $\mathbf{y}^t = (y_1, y_2, y_3)$, the equation (6.4.3) reduces to

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 + 2l_1 y_1 + 2l_2 y_2 + 2l_3 y_3 + q' = 0 \qquad (6.4.4)$$

   where $\lambda_1, \ \lambda_2, \lambda_3$ are the eigenvalues of $A$.

4. Complete the squares, if necessary, to write the equation (6.4.4) in terms of the variables $z_1, z_2, z_3$ so that this equation is in the standard form.

5. Use the condition $\mathbf{y} = P^t\mathbf{x}$ to determine the centre and the planes of symmetry of the quadric in terms of the original system.

**Example 6.4.10** Determine the quadric $2x^2 + 2y^2 + 2z^2 + 2xy + 2xz + 2yz + 4x + 2y + 4z + 2 = 0.$

**Solution:** In this case, $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}$ and $q = 2$. Check that for the orthonormal matrix $P = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{bmatrix}$, $P^t A P = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. So, the equation of the quadric reduces to

$$4y_1^2 + y_2^2 + y_3^2 + \frac{10}{\sqrt{3}} y_1 + \frac{2}{\sqrt{2}} y_2 - \frac{2}{\sqrt{6}} y_3 + 2 = 0.$$

Or equivalently,

$$4(y_1 + \frac{5}{4\sqrt{3}})^2 + (y_2 + \frac{1}{\sqrt{2}})^2 + (y_3 - \frac{1}{\sqrt{6}})^2 = \frac{9}{12}.$$

So, the equation of the quadric in standard form is

$$4z_1^2 + z_2^2 + z_3^2 = \frac{9}{12},$$

where the point $(x, y, z)^t = P(\frac{-5}{4\sqrt{3}}, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{6}})^t = (\frac{-3}{4}, \frac{1}{4}, \frac{-3}{4})^t$ is the centre. The calculation of the planes of symmetry is left as an exercise to the reader.

# Part II

# Ordinary Differential Equation

# Chapter 7

# Differential Equations

## 7.1 Introduction and Preliminaries

There are many branches of science and engineering where differential equations naturally arise. Now a days there are applications to many areas in medicine, economics and social sciences. In this context, the study of differential equations assumes importance. In addition, in the elementary study of differential equations, we also see the applications of many results from analysis and linear algebra. Without spending more time on motivation, (which will be clear as we go along) let us start with the following notations. Suppose that $y$ is a dependent variable and $x$ is an independent variable. The derivatives of $y$ (with respect to $x$) are denoted by

$$y' = \frac{dy}{dx}, y'' = \frac{d^2y}{dx^2}, \ldots, y^{(k)} = \frac{d^{(k)}y}{dx^{(k)}} \quad \text{for} \quad k \geq 3.$$

The independent variable will be defined for an interval $I$; where $I$ is either $\mathbb{R}$ or an interval $a < x < b \subset \mathbb{R}$. With these notations, we ask the question: what is a differential equation?

A differential equation is a relationship between the independent variable and the unknown dependent functions along with its derivatives.

**Definition 7.1.1 (Ordinary Differential Equation, ODE)** An equation of the form

$$f\big(x, y, y', \ldots, y^{(n)}\big) = 0 \qquad \text{for} \quad x \in I \tag{7.1.1}$$

is called an ORDINARY DIFFERENTIAL EQUATION; where $f$ is a known function from $I \times \mathbb{R}^{n+1}$ to $\mathbb{R}$. Also, the unknown function $y$ is to be determined.

**Remark 7.1.2** *Usually, Equation (7.1.1) is written as $f\big(x, y, y', \ldots, y^{(n)}\big) = 0$, and the interval $I$ is not mentioned in most of the examples.*

Some examples of differential equations are

1. $y' = 6\sin x + 9$;

2. $y'' + 2y^2 = 0$;

3. $\sqrt{y'} = \sqrt{x} + \cos y$;

4. $(y')^2 + y = 0$.

5. $y' + y = 0$.

6. $y'' + y = 0$.

7. $y^{(3)} = 0$.

8. $y'' + m\sin(y) = 0$.

**Definition 7.1.3 (Order of a Differential Equation)** The ORDER of a differential equation is the order of the highest derivative occurring in the equation.

In Example 7.1, the order of Equations 1, 3, 4, 5 are one, that of Equations 2, 6 and 8 are two and the Equation 7 has order three.

**Definition 7.1.4 (Solution)** A function $y = f(x)$ is called a SOLUTION of a differential equation on $I$ if

1. $f$ is differentiable (as many times as the order of the equation) on $I$ and

2. $y$ satisfies the differential equation for all $x \in I$.

**Example 7.1.5**

1. Show that $y = ce^{-2x}$ is a solution of $y' + 2y = 0$ on $\mathbb{R}$ for a constant $c \in \mathbb{R}$.
   **Solution:** Let $x \in \mathbb{R}$. By direct differentiation we have $y' = -2ce^{-2x} = -2y$.

2. Show that for any constant $a \in \mathbb{R}$, $y = \dfrac{a}{1-x}$ is a solution of
   $$(1-x)y' - y = 0$$
   on $(-\infty, 1)$ or on $(1, \infty)$. Note that $y$ is not a solution on any interval containing 1.
   **Solution:** It can be easily checked.

**Remark 7.1.6** *Sometimes a solution $y$ is also called an* INTEGRAL. *A solution of the form $y = g(x)$ is called an* EXPLICIT SOLUTION. *If $y$ is given by an implicit relation $h(x, y) = 0$ and satisfies the differential equation, then $y$ is called an* IMPLICIT SOLUTION.

**Remark 7.1.7** *Since the solution is obtained by integration, we may expect a constant of integration (for each integration) to appear in a solution of a differential equation. If the order of the ODE is $n$, we expect $n (n \geq 1)$ arbitrary constants.*

To start with, let us try to understand the structure of a first order differential equation of the form

$$f(x, y, y') = 0 \tag{7.1.2}$$

and move to higher orders later. With this in mind let us look at:

**Definition 7.1.8 (General Solution)** A function $y(x, c)$ is called a general solution of Equation (7.1.2) on an interval $I \subset \mathbb{R}$, if $y(x, c)$ is a solution of Equation (7.1.2) for each $x \in I$, for a fixed $c \in \mathbb{R}$ but $c$ is arbitrary.

**Remark 7.1.9** *The family of functions $\{y(., c) : c \in \mathbb{R}\}$ is called a one parameter family of functions and $c$ is called a parameter. In other words, a general solution of Equation (7.1.2) is nothing but a one parameter family of solutions of the Equation (7.1.2).*

**Example 7.1.10**

1. Show that for each $k \in \mathbb{R}$, $y = ke^x$ is a solution of $y' = y$. This is a general solution as it is a one parameter family of solutions. Here the parameter is $k$.
   **Solution:** This can be easily verified.

2. Determine a differential equation for which a family of circles with center at $(1,0)$ and arbitrary radius, $a$ is an implicit solution.
   **Solution:** This family is represented by the implicit relation
   $$(x-1)^2 + y^2 = a^2, \tag{7.1.3}$$
   where $a$ is a real constant. Then $y$ is a solution of the differential equation
   $$(x-1) + y\frac{dy}{dx} = 0. \tag{7.1.4}$$
   The function $y$ satisfying Equation (7.1.3) is a one parameter family of solutions or a general solution of Equation (7.1.4).

3. Consider the one parameter family of circles with center at $(c,0)$ and unit radius. The family is represented by the implicit relation
   $$(x-c)^2 + y^2 = 1, \tag{7.1.5}$$
   where $c$ is a real constant. Show that $y$ satisfies $\left(yy'\right)^2 + y^2 = 1$.
   **Solution:** We note that, differentiation of the given equation, leads to
   $$(x-c) + yy' = 0.$$
   Now, eliminating $c$ from the two equations, we get
   $$(yy')^2 + y^2 = 1.$$

In Example 7.1.10.2, we see that $y$ is not defined explicitly as a function of $x$ but implicitly defined by Equation (7.1.3). On the other hand $y = \dfrac{1}{1-x}$ is an explicit solution in Example 7.1.5.2. Solving a differential equation means to find a solution.

Let us now look at some geometrical interpretations of the differential Equation (7.1.2). The Equation (7.1.2) is a relation between $x$, $y$ and the slope of the function $y$ at the point $x$. For instance, let us find the equation of the curve passing through $(0, \frac{1}{2})$ and whose slope at each point $(x,y)$ is $-\dfrac{x}{4y}$. If $y$ is the required curve, then $y$ satisfies
$$\frac{dy}{dx} = -\frac{x}{4y}, \; y(0) = \frac{1}{2}.$$
It is easy to verify that $y$ satisfies the equation $x^2 + 4y^2 = 1$.

**Exercise 7.1.11** 1. Find the order of the following differential equations:

   (a) $y^2 + \sin(y') = 1$.

   (b) $y + (y')^2 = 2x$.

   (c) $(y')^3 + y'' - 2y^4 = -1$.

2. Find a differential equation satisfied by the given family of curves:

   (a) $y = mx, \; m$ real (family of lines).

   (b) $y^2 = 4ax, \; a$ real (family of parabolas).

   (c) $x = r^2\cos\theta, \; y = r^2\sin\theta, \; \theta$ is a parameter of the curve and $r$ is a real number (family of circles in parametric representation).

3. Find the equation of the curve $C$ which passes through $(1,0)$ and whose slope at each point $(x,y)$ is $\dfrac{-x}{y}$.

## 7.2 Separable Equations

In general, it may not be possible to find solutions of

$$y' = f(x, y)$$

where $f$ is an arbitrary continuous function. But there are special cases of the function $f$ for which the above equation can be solved. One such set of equations is

$$y' = g(y)h(x). \tag{7.2.1}$$

Equation (7.2.1) is called a SEPARABLE EQUATION. The Equation (7.2.1) is equivalent to

$$\frac{1}{g(y)}\frac{dy}{dx} = h(x).$$

Integrating with respect to $x$, we get

$$H(x) = \int h(x)dx = \int \frac{1}{g(y)}\frac{dy}{dx}dy = \int \frac{dy}{g(y)} = G(y) + c,$$

where $c$ is a constant. Hence, its implicit solution is

$$G(y) + c = H(x).$$

**Example 7.2.1** 1. Solve: $y' = y(y-1)$.

**Solution:** Here, $g(y) = y\,(y-1)$ and $h(x) = 1$. Then

$$\int \frac{dy}{y\,(y-1)} = \int dx.$$

By using partial fractions and integrating, we get

$$y = \frac{1}{1 - e^{x+c}}\ ,$$

where $c$ is a constant of integration.

2. Solve $y' = y^2$.

**Solution:** It is easy to deduce that $y = -\dfrac{1}{x+c}$, where $c$ is a constant; is the required solution.

Observe that the solution is defined, only if $x + c \neq 0$ for any $x$. For example, if we let $y(0) = a$, then $y = -\dfrac{a}{ax-1}$ exists as long as $ax - 1 \neq 0$.

### 7.2.1 Equations Reducible to Separable Form

There are many equations which are not of the form 7.2.1, but by a suitable substitution, they can be reduced to the separable form. One such class of equation is

$$y' = \frac{g_1(x,y)}{g_2(x,y)} \quad \text{or equivalently} \quad y' = g(\frac{y}{x})$$

where $g_1$ and $g_2$ are homogeneous functions of the same degree in $x$ and $y$, and $g$ is a continuous function. In this case, we use the substitution, $y = xu(x)$ to get $y' = xu' + u$. Thus, the above equation after substitution becomes

$$xu' + u(x) = g(u),$$

which is a separable equation in $u$. For illustration, we consider some examples.

**Example 7.2.2** 1. Find the general solution of $2xyy' - y^2 + x^2 = 0$.

**Solution:** Let $I$ be any interval not containing $0$. Then

$$2\frac{y}{x}y' - (\frac{y}{x})^2 + 1 = 0.$$

Letting $y = xu(x)$, we have

$$2u(u'x + u) - u^2 + 1 = 0 \quad \text{or} \quad 2xuu' + u^2 + 1 = 0 \text{ or equivalently}$$

$$\frac{2u}{1+u^2}\frac{du}{dx} = -\frac{1}{x}.$$

On integration, we get

$$1 + u^2 = \frac{c}{x}$$

or

$$x^2 + y^2 - cx = 0.$$

The general solution can be re-written in the form

$$(x - \frac{c}{2})^2 + y^2 = \frac{c^2}{4}.$$

This represents a family of circles with center $(\frac{c}{2}, 0)$ and radius $\frac{c}{2}$.

2. Find the equation of the curve passing through $(0, 1)$ and whose slope at each point $(x, y)$ is $-\frac{x}{2y}$.

**Solution:** If $y$ is such a curve then we have

$$\frac{dy}{dx} = -\frac{x}{2y} \text{ and } y(0) = 1.$$

Notice that it is a separable equation and it is easy to verify that $y$ satisfies $x^2 + 2y^2 = 2$.

3. The equations of the type

$$\frac{dy}{dx} = \frac{a_1x + b_1y + c_1}{a_2x + b_2y + c_2}$$

can also be solved by the above method by replacing $x$ by $x + h$ and $y$ by $y + k$, where $h$ and $k$ are to be chosen such that

$$a_1h + b_1k + c_1 = 0 = a_2h + b_2k + c_2.$$

This condition changes the given differential equation into $\frac{dy}{dx} = \frac{a_1x + b_1y}{a_2x + b_2y}$. Thus, if $x \neq 0$ then the equation reduces to the form $y' = g(\frac{y}{x})$.

**Exercise 7.2.3** 1. Find the general solutions of the following:

(a) $\frac{dy}{dx} = -x(\ln x)(\ln y)$.

(b) $y^{-1}\cos^{-1} + (e^x + 1)\frac{dy}{dx} = 0$.

2. Find the solution of

(a) $(2a^2 + r^2) = r^2 \cos\frac{d\theta}{dr}, \; r(0) = a$.

(b) $xe^{x+y} = \frac{dy}{dx}, \;\; y(0) = 0$.

3. Obtain the general solutions of the following:

(a) $\{y - x\text{cosec}\,(\frac{y}{x})\} = x\frac{dy}{dx}$.

(b) $xy' = y + \sqrt{x^2 + y^2}$.

(c) $\frac{dy}{dx} = \frac{x - y + 2}{-x + y + 2}$.

4. Solve $y' = y - y^2$ and use it to determine $\lim\limits_{x \longrightarrow \infty} y$. This equation occurs in a model of population.

## 7.3 Exact Equations

As remarked, there are no general methods to find a solution of Equation (7.1.2). The EXACT EQUATIONS is yet another class of equations that can be easily solved. In this section, we introduce this concept.

Let $D$ be a region in $xy$-plane and let $M$ and $N$ be real valued functions defined on $D$. Consider an equation

$$M(x,y) + N(x,y)\frac{dy}{dx} = 0, \ (x,y) \in D. \tag{7.3.1}$$

In most of the books on Differential Equations, this equation is also written as

$$M(x,y)dx + N(x,y)dy = 0, \ (x,y) \in D. \tag{7.3.2}$$

**Definition 7.3.1 (Exact Equation)** Equation (7.3.1) is called Exact if there exists a real valued twice continuously differentiable function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ (or the domain is an open subset of $\mathbb{R}^2$) such that

$$\frac{\partial f}{\partial x} = M \ \text{ and } \ \frac{\partial f}{\partial y} = N. \tag{7.3.3}$$

**Remark 7.3.2** *If Equation (7.3.1) is exact, then*

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dx} = \frac{df(x,y)}{dx} = 0.$$

*This implies that $f(x,y) = c$ (where $c$ is a constant) is an implicit solution of Equation (7.3.1). In other words, the left side of Equation (7.3.1) is an exact differential.*

**Example 7.3.3** The equation $y + x\frac{dy}{dx} = 0$ is an exact equation. Observe that in this example, $f(x,y) = xy$.

The proof of the next theorem is given in Appendix 14.6.2.

**Theorem 7.3.4** Let $M$ and $N$ be twice continuously differentiable function in a region $D$. The Equation (7.3.1) is exact if and only if

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}. \tag{7.3.4}$$

**Note:** If the Equation (7.3.1) or Equation (7.3.2) is exact, then there is a function $f(x,y)$ satisfying $f(x,y) = c$ for some constant $c$, such that

$$d(f(x,y)) = M(x,y)dx + N(x,y)dy = 0.$$

Let us consider some examples, where Theorem 7.3.4 can be used to easily find the general solution.

**Example 7.3.5** 1. Solve

$$2xe^y + (x^2e^y + \cos y\,)\frac{dy}{dx} = 0.$$

**Solution:** With the above notations, we have

$$M = 2xe^y, \ N = x^2e^y + \cos y, \ \frac{\partial M}{\partial y} = 2xe^y \ \text{ and } \ \frac{\partial N}{\partial x} = 2xe^y.$$

Therefore, the given equation is exact. Hence, there exists a function $G(x,y)$ such that

$$\frac{\partial G}{\partial x} = 2xe^y \ \text{ and } \ \frac{\partial G}{\partial y} = x^2e^y + \cos y.$$

The first partial differentiation when integrated with respect to $x$ (assuming $y$ to be a constant) gives,

$$G(x,y) = x^2e^y + h(y).$$

But then

$$\frac{\partial G}{\partial y} = \frac{\partial(x^2e^y + h(y))}{\partial y} = N$$

implies $\frac{dh}{dy} = \cos y$ or $h(y) = \sin y + c$ where $c$ is an arbitrary constant. Thus, the general solution of the given equation is

$$x^2e^y + \sin y = c.$$

The solution in this case is in implicit form.

2. Find values of $\ell$ and $m$ such that the equation

$$\ell y^2 + mxy\frac{dy}{dx} = 0$$

is exact. Also, find its general solution.
**Solution:** In this example, we have

$$M = \ell y^2,\ N = mxy,\ \frac{\partial M}{\partial y} = 2\ell y \text{ and } \frac{\partial N}{\partial x} = my.$$

Hence for the given equation to be exact, $m = 2\ell$. With this condition on $\ell$ and $m$, the equation reduces to

$$\ell y^2 + 2\ell xy\frac{dy}{dx} = 0.$$

This equation is not meaningful if $\ell = 0$. Thus, the above equation reduces to

$$\frac{d}{dx}(xy^2) = 0$$

whose solution is

$$xy^2 = c$$

for some arbitrary constant $c$.

3. Solve the equation

$$(3x^2e^y - x^2)dx + (x^3e^y + y^2)dy = 0.$$

**Solution:** Here

$$M = 3x^2e^y - x^2 \text{ and } N = x^3e^y + y^2.$$

Hence, $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x} = 3x^2e^y$. Thus the given equation is exact. Therefore,

$$G(x,y) = \int(3x^2e^y - x^2)dx = x^3e^y - \frac{x^3}{3} + h(y)$$

(keeping $y$ as constant). To determine $h(y)$, we partially differentiate $G(x,y)$ with respect to $y$ and compare with $N$ to get $h(y) = \frac{y^3}{3}$. Hence

$$G(x,y) = x^3e^y - \frac{x^3}{3} + \frac{y^3}{3} = c$$

is the required implicit solution.

### 7.3.1 Integrating Factors

On may occasions,

$$M(x,y)+N(x,y)\frac{dy}{dx}=0, \text{ or equivalently } M(x,y)dx+N(x,y)dy=0$$

may not be exact. But the above equation may become exact, if we multiply it by a proper factor. For example, the equation

$$ydx-dy=0$$

is not exact. But, if we multiply it with $e^{-x}$, then the equation reduces to

$$e^{-x}ydx-e^{-x}dy=0, \text{ or equivalently } d\left(e^{-x}y\right)=0,$$

an exact equation. Such a factor (in this case, $e^{-x}$) is called an INTEGRATING FACTOR for the given equation. Formally

**Definition 7.3.6 (Integrating Factor)** A function $Q(x,y)$ is called an integrating factor for the Equation (7.3.1), if the equation

$$Q(x,y)M(x,y)dx+Q(x,y)N(x,y)dy=0$$

is exact.

**Example 7.3.7** 1. Solve the equation $ydx-xdy=0,\ \ x,y>0.$

**Solution:** It can be easily verified that the given equation is not exact. Multiplying by $\frac{1}{xy}$, the equation reduces to

$$\frac{1}{xy}ydx-\frac{1}{xy}xdy=0, \text{ or equivalently } d\left(\ln x-\ln y\right)=0.$$

Thus, by definition, $\dfrac{1}{xy}$ is an integrating factor. Hence, a general solution of the given equation is $G(x,y)=\dfrac{1}{xy}=c$, for some constant $c\in\mathbb{R}$. That is,

$$y=cx, \text{ for some constant } c\in\mathbb{R}.$$

2. Find a general solution of the differential equation

$$\big(4y^2+3xy\big)dx-\big(3xy+2x^2\big)dy=0.$$

**Solution:** It can be easily verified that the given equation is not exact.

METHOD 1: Here the terms $M=4y^2+3xy$ and $N=-(3xy+2x^2)$ are homogeneous functions of degree 2. It may be checked that an integrating factor for the given differential equation is

$$\frac{1}{Mx+Ny}=\frac{1}{xy\big(x+y\big)}.$$

Hence, we need to solve the partial differential equations

$$\frac{\partial G(x,y)}{\partial x} = \frac{y\big(4y+3x\big)}{xy\big(x+y\big)}=\frac{4}{x}-\frac{1}{x+y} \text{ and} \tag{7.3.5}$$

$$\frac{\partial G(x,y)}{\partial y} = \frac{-x(3y+2x)}{xy\big(x+y\big)}=-\frac{2}{y}-\frac{1}{x+y}. \tag{7.3.6}$$

Integrating (keeping $y$ constant) Equation (7.3.5), we have

$$G(x,y)=4\ln|x|-\ln|x+y|+h(y) \tag{7.3.7}$$

and integrating (keeping $x$ constant) Equation (7.3.6), we get

$$G(x,y) = -2\ln|y| - \ln|x+y| + g(x). \tag{7.3.8}$$

Comparing Equations (7.3.7) and (7.3.8), the required solution is

$$G(x,y) = 4\ln|x| - \ln|x+y| - 2\ln|y| = \ln c$$

for some real constant $c$. Or equivalently, the solution is

$$x^4 = c\big(x+y\big)y^2.$$

METHOD 2: Here the terms $M = 4y^2 + 3xy$ and $N = -(3xy + 2x^2)$ are polynomial in $x$ and $y$. Therefore, we suppose that $x^\alpha y^\beta$ is an integrating factor for some $\alpha, \beta \in \mathbb{R}$. We try to find this $\alpha$ and $\beta$.

Multiplying the terms $M(x,y)$ and $N(x,y)$ with $x^\alpha y^\beta$, we get

$$M(x,y) = x^\alpha y^\beta \big(4y^2 + 3xy\big), \text{ and } N(x,y) = -x^\alpha y^\beta (3xy + 2x^2).$$

For the new equation to be exact, we need $\dfrac{\partial M(x,y)}{\partial y} = \dfrac{\partial N(x,y)}{\partial x}$. That is, the terms

$$4(2+\beta)x^\alpha y^{1+\beta} + 3(1+\beta)x^{1+\alpha}y^\beta$$

and

$$-3(1+\alpha)x^\alpha y^{1+\beta} - 2(2+\alpha)x^{1+\alpha}y^\beta$$

must be equal. Solving for $\alpha$ and $\beta$, we get $\alpha = -5$ and $\beta = 1$. That is, the expression $\dfrac{y}{x^5}$ is also an integrating factor for the given differential equation. This integrating factor leads to

$$G(x,y) = -\frac{y^3}{x^4} - \frac{y^2}{x^3} + h(y)$$

and

$$G(x,y) = -\frac{y^3}{x^4} - \frac{y^2}{x^3} + g(x).$$

Thus, we need $h(y) = g(x) = c$, for some constant $c \in \mathbb{R}$. Hence, the required solution by this method is

$$y^2\big(y+x\big) = cx^4.$$

**Remark 7.3.8** 1. *If Equation (7.3.1) has a general solution, then it can be shown that Equation (7.3.1) admits an integrating factor.*

2. *If Equation (7.3.1) has an integrating factor, then it has many (in fact infinitely many) integrating factors.*

3. *Given Equation (7.3.1), whether or not it has an integrating factor, is a tough question to settle.*

4. *In some cases, we use the following rules to find the integrating factors.*

   (a) *Consider a homogeneous equation $M(x,y)dx + N(x,y)dy = 0$. If*

   $$Mx + Ny \neq 0, \quad \textit{then} \quad \frac{1}{Mx+Ny}$$

   *is an Integrating Factor.*

(b) *If the functions $M(x,y)$ and $N(x,y)$ are polynomial functions in $x,y$; then $x^\alpha y^\beta$ works as an integrating factor for some appropriate values of $\alpha$ and $\beta$.*

(c) *The equation $M(x,y)dx+N(x,y)dy = 0$ has $e^{\int f(x)dx}$ as an integrating factor, if $f(x) = \frac{1}{N}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right)$ is a function of $x$ alone.*

(d) *The equation $M(x,y)dx+N(x,y)dy = 0$ has $e^{-\int g(y)dy}$ as an integrating factor, if $g(y) = \frac{1}{M}\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right)$ is a function of $y$ alone.*

(e) *For the equation*

$$yM_1(xy)dx + xN_1(xy)dy = 0$$

*with $Mx - Ny \neq 0$, the function $\frac{1}{Mx - Ny}$ is an integrating factor.*

**Exercise 7.3.9** 1. Show that the following equations are exact and hence solve them.

(a) $(r+\sin\theta+\cos\theta)\frac{dr}{d\theta} + r(\cos\theta - \sin\theta) = 0.$

(b) $(e^{-x} - \ln y + \frac{y}{x}) + (-\frac{x}{y} + \ln x + \cos y)\frac{dy}{dx} = 0.$

2. Find conditions on the function $g(x,y)$ so that the equation

$$(x^2+xy^2) + \{ax^2y^2 + g(x,y)\}\frac{dy}{dx} = 0$$

is exact.

3. What are the conditions on $f(x)$, $g(y)$, $\phi(x)$, and $\psi(y)$ so that the equation

$$(\phi(x)+\psi(y)) + (f(x)+g(y))\frac{dy}{dx} = 0$$

is exact.

4. Verify that the following equations are not exact. Further find suitable integrating factors to solve them.

(a) $y + (x+x^3y^2)\frac{dy}{dx} = 0.$

(b) $y^2 + (3xy+y^2-1)\frac{dy}{dx} = 0.$

(c) $y + (x+x^3y^2)\frac{dy}{dx} = 0.$

(d) $y^2 + (3xy+y^2-1)\frac{dy}{dx} = 0.$

5. Find the solution of

(a) $(x^2y+2xy^2) + 2(x^3+3x^2y)\frac{dy}{dx} = 0$ with $y(1) = 0.$

(b) $y(xy+2x^2y^2) + x(xy - x^2y^2)\frac{dy}{dx} = 0$ with $y(1) = 1.$

## 7.4 Linear Equations

Some times we might think of a subset or subclass of differential equations which admit explicit solutions. This question is pertinent when we say that there are no means to find the explicit solution of $\frac{dy}{dx} = f(x,y)$ where $f$ is an arbitrary continuous function in $(x,y)$ in suitable domain of definition. In this context, we have a class of equations, called Linear Equations (to be defined shortly) which admit explicit solutions.

**Definition 7.4.1 (Linear/Nonlinear Equations)**

Let $p(x)$ and $q(x)$ be real-valued piecewise continuous functions defined on interval $I = [a,b]$. The equation

$$y' + p(x)y = q(x), \ x \in I \tag{7.4.1}$$

is called a linear equation, where $y'$ stands for $\frac{dy}{dx}$. Equation (7.4.1) is called Linear non-homogeneous if $q(x) \neq 0$ and is called Linear homogeneous if $q(x) = 0$ on $I$.

A first order equation is called a non-linear equation (in the independent variable) if it is neither a linear homogeneous nor a non-homogeneous linear equation.

**Example 7.4.2** 1. The equation $y' = \sin y$ is a non-linear equation.

2. The equation $y' + y = \sin x$ is a linear non-homogeneous equation.

3. The equation $y' + x^2 y = 0$ is a linear homogeneous equation.

Define the indefinite integral $P(x) = \int p(x)dx$ ( or $\int\limits_a^x p(s)ds)$. Multiplying Equation (7.4.1) by $e^{P(x)}$, we get

$$e^{P(x)}y' + e^{P(x)}p(x)y = e^{P(x)}q(x) \quad \text{or equivalently} \quad \frac{d}{dx}(e^{P(x)}y) = e^{P(x)}q(x).$$

On integration, we get

$$e^{P(x)}y = c + \int e^{P(x)}q(x)dx.$$

In other words,

$$y = ce^{-P(x)} + e^{-P(x)} \int e^{P(x)}q(x)dx \tag{7.4.2}$$

where $c$ is an arbitrary constant is the general solution of Equation (7.4.1).

**Remark 7.4.3** *If we let $P(x) = \int\limits_a^x p(s)ds$ in the above discussion, Equation (7.4.2) also represents*

$$y = y(a)e^{-P(x)} + e^{-P(x)} \int\limits_a^x e^{P(s)}q(s)ds. \tag{7.4.3}$$

As a simple consequence, we have the following proposition.

**Proposition 7.4.4** $y = ce^{-P(x)}$ (where $c$ is any constant) is the general solution of the linear homogeneous equation

$$y' + p(x)y = 0. \tag{7.4.4}$$

In particular, when $p(x) = k$, is a constant, the general solution is $y = ce^{-kx}$, with $c$ an arbitrary constant.

**Example 7.4.5** 1. Comparing the equation $y' = y$ with Equation (7.4.1), we have

$$p(x) = -1 \quad \text{and} \quad q(x) = 0.$$

Hence, $P(x) = \int(-1)dx = -x$. Substituting for $P(x)$ in Equation (7.4.2), we get $y = ce^x$ as the required general solution.

We can just use the second part of the above proposition to get the above result, as $k = -1$.

2. The general solution of $xy' = -y, \ x \in I \ (0 \notin I)$ is $y = cx^{-1}$, where $c$ is an arbitrary constant. Notice that no non-zero solution exists if we insist on the condition $\lim\limits_{x \to 0, x > 0} y = 0$.

A class of nonlinear Equations (7.4.1) (named after Bernoulli $(1654-1705)$) can be reduced to linear equation. These equations are of the type

$$y' + p(x)y = q(x)y^a. \tag{7.4.5}$$

If $a = 0$ or $a = 1$, then Equation (7.4.5) is a linear equation. Suppose that $a \neq 0, 1$. We then define $u(x) = y^{1-a}$ and therefore

$$u' = (1-a)y'y^{-a} = (1-a)(q(x) - p(x)u)$$

or equivalently

$$u' + (1-a)p(x)u = (1-a)q(x), \tag{7.4.6}$$

a linear equation. For illustration, consider the following example.

**Example 7.4.6** For $m, n$ constants and $m \neq 0$, solve $y' - my + ny^2 = 0$.
**Solution:** Let $u = y^{-1}$. Then $u(x)$ satisfies

$$u' + mu = n$$

and its solution is

$$u = Ae^{-mx} + e^{-mx} \int ne^{mx} dx = Ae^{-mx} + \frac{n}{m}.$$

Equivalently

$$y = \frac{1}{Ae^{-mx} + \frac{n}{m}}$$

with $m \neq 0$ and $A$ an arbitrary constant, is the general solution.

**Exercise 7.4.7** 1. In Example 7.4.6, show that $u' + mu = n$.

2. Find the genral solution of the following:

   (a) $y' + y = 4$.

   (b) $y' - 3y = 10$.

   (c) $y' - 2xy = 0$.

   (d) $y' - xy = 4x$.

   (e) $y' + y = e^{-x}$.

   (f) $\sinh xy' + y \cosh x = e^x$.

   (g) $(x^2 + 1)y' + 2xy = x^2$.

3. Solve the following IVP's:

   (a) $y' - 4y = 5, \ y(0) = 0$.

(b) $y' + (1+x^2)y = 3, \; y(0) = 0.$

(c) $y' + y = \cos x, \; y(\pi) = 0.$

(d) $y' - y^2 = 1, \; y(0) = 0.$

(e) $(1+x)y' + y = 2x^2, \; y(1) = 1.$

4. Let $y_1$ be a solution of $y' + a(x)y = b_1(x)$ and $y_2$ be a solution of $y' + a(x)y = b_2(x)$. Then show that $y_1 + y_2$ is a solution of
$$y' + a(x)y = b_1(x) + b_2(x).$$

5. Reduce the following to linear equations and hence solve:

(a) $y' + 2y = y^2.$

(b) $(xy + x^3 e^y)y' = y^2.$

(c) $y' \sin(y) + x\cos(y) = x.$

(d) $y' - y = xy^2.$

6. Find the solution of the IVP
$$y' + 4xy + xy^3 = 0, \;\; y(0) = \frac{1}{\sqrt{2}}.$$

## 7.5 Miscellaneous Remarks

In Section 7.4, we have learned to solve the linear equations. There are many other equations, though not linear, which are also amicable for solving. Below, we consider a few classes of equations which can be solved. In this section or in the sequel, $p$ denotes $\dfrac{dy}{dx}$ or $y'$. A word of caution is needed here. The method described below are more or less ad hoc methods.

1. EQUATIONS SOLVABLE FOR Y:
Consider an equation of the form
$$y = f(x, p). \tag{7.5.1}$$
Differentiating with respect to $x$, we get
$$\frac{dy}{dx} = p = \frac{\partial f(x,p)}{\partial x} + \frac{\partial f(x,p)}{\partial p} \cdot \frac{dp}{dx} \text{ of equivalently } \; p = g(x, p, \frac{dp}{dx}). \tag{7.5.2}$$
Equation (7.5.2) can be viewed as a differential equation in $p$ and $x$. We now assume that Equation (7.5.2) can be solved for $p$ and its solution is
$$h(x, p, c) = 0. \tag{7.5.3}$$
If we are able to eliminate $p$ between Equations (7.5.1) and (7.5.3), then we have an implicit solution of the Equation (7.5.1).

Solve $y = 2px - xp^2$.

**Solution:** Differentiating with respect to $x$ and replacing $\dfrac{dy}{dx}$ by $p$, we get
$$p = 2p - p^2 + 2x\frac{dp}{dx} - 2xp\frac{dp}{dx} \quad \text{or} \quad (p + 2x\frac{dp}{dx})(1-p) = 0.$$
So, either
$$p + 2x\frac{dp}{dx} = 0 \quad \text{or} \quad p = 1.$$

That is, either $p^2x = c$ or $p = 1$. Eliminating $p$ from the given equation leads to an explicit solution

$$y = 2x\sqrt{\frac{c}{x}} - c \quad \text{or} \quad y = x.$$

The first solution is a one-parameter family of solutions, giving us a general solution. The latter one is a solution but not a general solution since it is not a one parameter family of solutions.

2. Equations in which the independent variable $x$ is missing:
   These are equations of the type $f(y, p) = 0$. If possible we solve for $y$ and we proceed. Sometimes introducing an arbitrary parameter helps. We illustrate it below.

   Solve $y^2 + p^2 = a^2$ where $a$ is a constant.
   **Solution:** We equivalently rewrite the given equation, by (arbitrarily) introducing a new parameter $t$ by

   $$y = a\sin t, \quad p = a\cos t$$

   from which it follows

   $$\frac{dy}{dt} = a\cos t; \quad p = \frac{dy}{dx} = \frac{dy}{dt}\Big/\frac{dx}{dt}$$

   and so

   $$\frac{dx}{dt} = \frac{1}{p}\frac{dy}{dt} = 1 \quad \text{or} \quad x = t + c.$$

   Therefore, a general solution is

   $$y = a\sin(t + c).$$

3. Equations in which $y$ (dependent variable or the unknown) is missing:
   We illustrate this case by an example.

   Find the general solution of $x = p^3 - p - 1$.
   **Solution:** Recall that $p = \frac{dy}{dx}$. Now, from the given equation, we have

   $$\frac{dy}{dp} = \frac{dy}{dx} \cdot \frac{dx}{dp} = p(3p^2 - 1).$$

   Therefore,

   $$y = \frac{3}{4}p^4 - \frac{1}{2}p^2 + c$$

   (regarding $p$ as a parameter). The desired solution in this case is in the parametric form, given by

   $$x = t^3 - t - 1 \quad \text{and } y = \frac{3}{4}t^4 - \frac{1}{2}t^2 + c$$

   where $c$ is an arbitrary constant.

**Remark 7.5.1** *The readers are again informed that the methods discussed in* 1), 2), 3) *are more or less ad hoc methods. It may not work in all cases.*

**Exercise 7.5.2** 1. Find the general solution of $y = (1 + p)x + p^2$.
*Hint: Differentiate with respect to $x$ to get $\dfrac{dx}{dp} = -(x + 2p)$ ( a linear equation in $x$). Express the solution in the parametric form*

$$y(p) = (1 + p)x + p^2, \quad x(p) = 2(1 - p) + ce^{-p}.$$

2. Solve the following differential equations:

   (a) $8y = x^2 + p^2$.

(b) $y + xp = x^4 p^2$.

(c) $y^2 \log y - p^2 = 2xyp$.

(d) $2y + p^2 + 2p = 2x(p+1)$.

(e) $2y = 2x^2 + 4px + p^2$.

## 7.6 Initial Value Problems

As we had seen, there are no methods to solve a general equation of the form

$$y' = f(x, y) \tag{7.6.1}$$

and in this context two questions may be pertinent.

1. Does Equation (7.6.1) admit solutions at all (*i.e.,* the existence problem)?
2. Is there a method to find solutions of Equation (7.6.1) in case the answer to the above question is in the affirmative?

The answers to the above two questions are not simple. But there are partial answers if some additional restrictions on the function $f$ are imposed. The details are discussed in this section.

For $a, b \in \mathbb{R}$ with $a > 0, b > 0$, we define

$$S = \{(x, y) \in \mathbb{R}^2 : |x - x_0| \leq a,\ |y - y_0| \leq b\} \subset I \times \mathbb{R}.$$

**Definition 7.6.1 (Initial Value Problems)** Let $f : S \longrightarrow \mathbb{R}$ be a continuous function on a $S$. The problem of finding a solution $y$ of

$$y' = f(x, y),\ (x, y) \in S, x \in I \quad \text{with} \quad y(x_0) = y_0 \tag{7.6.2}$$

in a neighbourhood $I$ of $x_0$ (or an open interval $I$ containing $x_0$) is called an Initial Value Problem, henceforth denoted by IVP.

The condition $y(x_0) = y_0$ in Equation (7.6.2) is called the INITIAL CONDITION stated at $x = x_0$ and $y_0$ is called the INITIAL VALUE.

Further, we assume that $a$ and $b$ are finite. Let

$$M = \max\{|f(x, y)| : (x, y) \in S\}.$$

Such an $M$ exists since $S$ is a closed and bounded set and $f$ is a continuous function and let $h = \min(a, \frac{b}{M})$. The ensuing proposition is simple and hence the proof is omitted.

**Proposition 7.6.2** A function $y$ is a solution of IVP (7.6.2) if and only if $y$ satisfies

$$y = y_0 + \int_{x_0}^{x} f(s, y(s)) ds. \tag{7.6.3}$$

In the absence of any knowledge of a solution of IVP (7.6.2), we now try to find an approximate solution. Any solution of the IVP (7.6.2) must satisfy the initial condition $y(x_0) = y_0$. Hence, as a crude approximation to the solution of IVP (7.6.2), we define

$$y_0 = y_0 \text{ for all } x \in [x_0 - h, x_0 + h].$$

Now the Equation (7.6.3) appearing in Proposition 7.6.2, helps us to refine or improve the approximate solution $y_0$ with a hope of getting a better approximate solution. We define

$$y_1 = y_o + \int_{x_0}^{x} f(s, y_0) ds$$

and for $n = 2, 3, \ldots,$ we inductively define

$$y_n = y_0 + \int_{x_0}^{x} f(s, y_{n-1}(s)) ds \text{ for all } x \in [x_0 - h, x_0 + h].$$

As yet we have not checked a few things, like whether the point $(s, y_n(s)) \in S$ or not. We formalise the theory in the latter part of this section. To get ourselves motivated, let us apply the above method to the following IVP.

**Example 7.6.3** Solve the IVP

$$y' = -y, \; y(0) = 1, \; -1 \leq x \leq 1.$$

**Solution:** From Proposition 7.6.2, a function $y$ is a solution of the above IVP if and only if

$$y = 1 - \int_{x_0}^{x} y(s) ds.$$

We have $y_0 = y(0) \equiv 1$ and

$$y_1 = 1 - \int_{0}^{x} ds = 1 - x.$$

So,

$$y_2 = 1 - \int_{0}^{x} (1 - s) ds = 1 - x + \frac{x^2}{2!}.$$

By induction, one can easily verify that

$$y_n = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots + (-1)^n \frac{x^n}{n!}.$$

**Note:** The solution of the given IVP is

$$y = e^{-x} \quad \text{and that} \quad \lim_{n \longrightarrow \infty} y_n = e^{-x}.$$

This example justifies the use of the word approximate solution for the $y_n$'s.

We now formalise the above procedure.

**Definition 7.6.4 (Picard's Successive Approximations)** Consider the IVP (7.6.2). For $x \in I$ with $|x - x_0| \leq a$, define inductively

$$\begin{aligned} y_0(x) &= y_0 \quad \text{and for } n = 1, 2, \ldots, \\ y_n(x) &= y_0 + \int_{x_0}^{x} f(s, y_{n-1}(s)) ds. \end{aligned} \tag{7.6.4}$$

Then $y_0, y_1, \ldots, y_n, \ldots$ are called Picard's successive approximations to the IVP (7.6.2).

Whether Equation (7.6.4) is well defined or not is settled in the following proposition.

**Proposition 7.6.5** The Picard's approximates $y_n$'s, for the IVP (7.6.2) defined by Equation (7.6.4) is well defined on the interval $|x - x_0| \leq h = \min\{a, \frac{b}{M}\}$, *i.e.,* for $x \in [x_0 - h, x_0 + h]$.

PROOF. We have to verify that for each $n = 0, 1, 2, \ldots$, $(s, y_n)$ belongs to the domain of definition of $f$ for $|s - x_0| \leq h$. This is needed due to the reason that $f(s, y_n)$ appearing as integrand in Equation (7.6.4) may not be defined. For $n = 0$, it is obvious that $f(s, y_0) \in S$ as $|s - x_0| \leq a$ and $|y_0 - y_0| = 0 \leq b$. For $n = 1$, we notice that, if $|x - x_0| \leq h$ then

$$|y_1 - y_0| \leq M|x - x_0| \leq Mh \leq b.$$

So, $(x, y_1) \in S$ whenever $|x - x_0| \leq h$.

The rest of the proof is by the method of induction. We have established the result for $n = 1$, namely

$$(x, y_1) \in S \quad \text{if} \quad |x - x_0| \leq h.$$

Assume that for $k = 1, 2, \ldots, n-1$, $(x, y_k) \in S$ whenever $|x - x_0| \leq h$. Now, by definition of $y_n$, we have

$$y_n - y_0 = \int_{x_0}^{x} f(s, y_{n-1}) ds.$$

But then by induction hypotheses $(s, y_{n-1}) \in S$ and hence

$$|y_n - y_0| \leq M|x - x_0| \leq Mh \leq b.$$

This shows that $(x, y_n) \in S$ whenever $|x - x_0| \leq h$. Hence $(x, y_k) \in S$ for $k = n$ holds and therefore the proof of the proposition is complete. □

Let us again come back to Example 7.6.3 in the light of Proposition 7.6.2.

**Example 7.6.6** Compute the successive approximations to the IVP

$$y' = -y, \quad -1 \leq x \leq 1, \ |y - 1| \leq 1 \text{ and } y(0) = 1. \tag{7.6.5}$$

**Solution:** Note that $x_0 = 0, y_0 = 1, f(x, y) = -y$, and $a = b = 1$. The set $S$ on which we are studying the differential equation is

$$S = \{(x, y) : |x| \leq 1, |y - 1| \leq 1\}.$$

By Proposition 7.6.2, on this set

$$M = \max\{|y| : (x, y) \in S\} = 2 \quad \text{and} \quad h = \min\{1, 1/2\} = 1/2.$$

Therefore, the approximate solutions $y_n$'s are defined only for the interval $[-\frac{1}{2}, \frac{1}{2}]$, if we use Proposition 7.6.2.

Observe that the exact solution $y = e^{-x}$ and the approximate solutions $y_n$'s of Example 7.6.3 exist on $[-1, 1]$. But the approximate solutions as seen above are defined in the interval $[-\frac{1}{2}, \frac{1}{2}]$.

That is, for any IVP, the approximate solutions $y_n$'s may exist on a larger interval as compared to the interval obtained by the application of the Proposition 7.6.2.

We now consider another example.

**Example 7.6.7** Find the Picard's successive approximations for the IVP

$$y' = f(y), \quad 0 \leq x \leq 1, \ y \geq 0 \text{ and } y(0) = 0; \tag{7.6.6}$$

where

$$f(y) = \sqrt{y} \text{ for } y \geq 0.$$

**Solution:** By definition $y_0(x) = y_0 \equiv 0$ and

$$y_1(x) = y_0 + \int_0^x f(y_0)ds = 0 + \int_0^x \sqrt{0}ds = 0.$$

A similar argument implies that $y_n(x) \equiv 0$ for all $n = 2, 3, \ldots$ and $\lim\limits_{n \longrightarrow \infty} y_n(x) \equiv 0$. Also, it can be easily verified that $y(x) \equiv 0$ is a solution of the IVP (7.6.6).

Also $y(x) = \frac{x^2}{4},\ 0 \le x \le 1$ is a solution of Equation (7.6.6) and the $\{y_n\}$'s do not converge to $\frac{x^2}{4}$. Note here that the IVP (7.6.6) has at least two solutions.

The following result is about the existence of a unique solution to a class of IVPs. We state the theorem without proof.

**Theorem 7.6.8 (Picard's Theorem on Existence and Uniqueness)** Let $S = \{(x, y) : |x - x_0| \le a,\ |y - y_0| \le b\}$, and $a, b > 0$. Let $f : S \longrightarrow \mathbb{R}$ be such that $f$ as well as $\frac{\partial f}{\partial y}$ are continuous on $S$. Also, let $M, K \in \mathbb{R}$ be constants such that

$$|f| \le M,\ |\frac{\partial f}{\partial y}| \le K \text{ on } S.$$

Let $h = \min\{a, b/M\}$. Then the sequence of successive approximations $\{y_n\}$ (defined by Equation (7.6.4)) for the IVP (7.6.2) uniformly converges on $|x - x_0| \le h$ to a solution of IVP (7.6.2). Moreover the solution to IVP (7.6.2) is unique.

**Remark 7.6.9** *The theorem asserts the existence of a unique solution on a subinterval $|x - x_0| \le h$ of the given interval $|x - x_0| \le a$. In a way it is in a neighbourhood of $x_0$ and so this result is also called the local existence of a unique solution. A natural question is whether the solution exists on the whole of the interval $|x - x_0| \le a$. The answer to this question is beyond the scope of this book.*

*Whenever we talk of the Picard's theorem, we mean it in this local sense.*

**Exercise 7.6.10**

1. Compute the sequence $\{y_n\}$ of the successive approximations to the IVP

$$y' = y\ (y - 1),\ y(x_0) = 0, x_0 \ge 0.$$

2. Show that the solution of the IVP

$$y' = y\ (y - 1),\ y(x_0) = 1, x_0 \ge 0$$

is $y \equiv 1,\ x \ge x_0$.

3. The IVP

$$y' = \sqrt{y},\ y(0) = 0, x \ge 0$$

has solutions $y_1 \equiv 0$ as well as $y_2 = \frac{x^2}{4}, x \ge 0$. Why does the existence of the two solutions not contradict the Picard's theorem?

4. Consider the IVP

$$y' = y,\ y(0) = 1 \text{ in } \{(x, y) : |x| \le a, |y| \le b\}$$

for any $a, b > 0$.

(a) Compute the interval of existence of the solution of the IVP by using Theorem 7.6.8.

(b) Show that $y = e^x$ is the solution of the IVP which exists on whole of $\mathbb{R}$.

This again shows that the solution to an IVP may exist on a larger interval than what is being implied by Theorem 7.6.8.

### 7.6.1 Orthogonal Trajectories

One among the many applications of differential equations is to find curves that intersect a given family of curves at right angles. In other words, given a family $F$, of curves, we wish to find curve (or curves) $\Gamma$ which intersect orthogonally with any member of $F$ (whenever they intersect). It is important to note that we are not insisting that $\Gamma$ should intersect every member of $F$, but if they intersect, the angle between their tangents, at every point of intersection, is $90^\circ$. Such a family of curves $\Gamma$ is called "orthogonal trajectories" of the family $F$. That is, at the common point of intersection, the tangents are orthogonal. In case, the family $F_1$ and $F_2$ are identical, we say that the family is self-orthogonal.

Before procedding to an example, let us note that at the common point of intersection, the product of the slopes of the tangent is $-1$. In order to find the orthogonal trajectories of a family of curves $F$, parametrized by a constant $c$, we eliminate $c$ between $y$ and $y'$. This gives the slope at any point $(x, y)$ and is independent of the choice of the curve. Below, we illustrate, how to obtain the orthogonal trajectories.

**Example 7.6.11** Compute the orthogonal trajectories of the family $F$ of curves given by

$$F: \qquad y^2 = cx^3, \tag{7.6.7}$$

where $c$ is an arbitrary constant.

**Solution:** Differentiating Equation (7.6.7), we get

$$2yy' = 3cx^2. \tag{7.6.8}$$

Elimination of $c$ between Equations (7.6.7) and (7.6.8), leads to

$$y' = \frac{3cx^2}{2y} = \frac{3}{2x} \cdot \frac{cx^3}{y} = \frac{3y}{2x}. \tag{7.6.9}$$

At the point $(x, y)$, if any curve intersects orthogonally, then (if its slope is $y'$) we must have

$$y' = -\frac{2x}{3y}.$$

Solving this differential equation, we get

$$y^2 = -\frac{x^2}{3} + c.$$

Or equivalently, $y^2 + \frac{x^2}{3} = c$ is a family of curves which intersects the given family $F$ orthogonally.

Below, we summarize how to determine the orthogonal trajectories.

**Step 1:** Given the family $F(x, y, c) = 0$, determine the differential equation,

$$y' = f(x, y), \tag{7.6.10}$$

for which the given family $F$ are a general solution. Equation (7.6.10) is obtained by the elimination of the constant $c$ appearing in $F(x, y, c) = 0$ "using the equation obtained by differentiating this equation with respect to $x$".

**Step 2:** The differential equation for the orthogonal trajectories is then given by

$$y' = -\frac{1}{f(x, y)}. \tag{7.6.11}$$

**Final Step:** The general solution of Equation (7.6.11) is the orthogonal trajectories of the given family.

In the following, let us go through the steps.

**Example 7.6.12** Find the orthogonal trajectories of the family of stright lines

$$y = mx + 1, \tag{7.6.12}$$

where $m$ is a real parameter.

**Solution:** Differentiating Equation (7.6.12), we get $y' = m$. So, substituting $m$ in Equation (7.6.12), we have $y = y'x + 1$. Or equivalently,

$$y' = \frac{y-1}{x}.$$

So, by the final step, the orthogonal trajectories satisfy the differential equation

$$y' = \frac{x}{1-y}. \tag{7.6.13}$$

It can be easily verified that the general solution of Equation (7.6.13) is

$$x^2 + y^2 - 2y = c, \tag{7.6.14}$$

where $c$ is an arbitrary constant. In other words, the orthogonal trajectories of the family of straight lines (7.6.12) is the family of circles given by Equation (7.6.14).

**Exercise 7.6.13**

1. Find the orthogonal trajectories of the following family of curves (the constant $c$ appearing below is an arbitrary constant).
   (a) $y = x + c$.
   (b) $x^2 + y^2 = c$.
   (c) $y^2 = x + c$.
   (d) $y = cx^2$.
   (e) $x^2 - y^2 = c$.
2. Show that the one parameter family of curves $y^2 = 4k(k + x),\ k \in \mathbb{R}$ are self orthogonal.
3. Find the orthogonal trajectories of the family of circles passing through the points $(1, -2)$ and $(1, 2)$.

## 7.7 Numerical Methods

All said and done, the Picard's Successive approximations is not suitable for computations on computers. In the absence of methods for closed form solution (in the explicit form), we wish to explore "how computers can be used to find approximate solutions of IVP" of the form

$$y' = f(x, y), \qquad y(x_0) = y_0. \tag{7.7.1}$$

In this section, we study a simple method to find the "numerical solutions" of Equation (7.7.1). The study of differential equations has two important aspects (among other features) namely, the qualitative theory, the latter is called "Numerical methods" for solving Equation (7.7.1). What is presented here is at a very rudimentary level nevertheless it gives a flavour of the numerical method.

To proceed further, we assume that $f$ is a "good function" (there by meaning "sufficiently differentiable"). In such case, we have

$$y(x+h) = y + hy' + \frac{h^2}{2!}y'' + \cdots$$

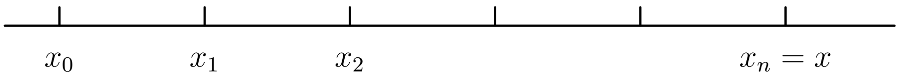


Figure 7.1: **Partitioning the interval**

which suggests a "crude" approximation $y(x+h) \simeq y + hf(x,y)$ (if $h$ is small enough), the symbol $\simeq$ means "approximately equal to". With this in mind, let us think of finding $y$, where $y$ is the solution of Equation (7.7.1) with $x > x_0$. Let $h = \dfrac{x - x_0}{n}$ and define

$$x_i = x_0 + ih, \;\; i = 0, 1, 2, \ldots, n.$$

That is, we have divided the interval $[x_0, x]$ into $n$ equal intervals with end points $x_0, x_1, \ldots, x = x_n$.

Our aim is to calculate $y$ : At the first step, we have $y(x+h) \simeq y_0 + hf\big(x_0, y_0\big)$. Define $y_1 = y_0 + hf(x_0, y_0)$. Error at first step is

$$|y(x_0 + h) - y_1| = E_1.$$

Similarly, we define $y_2 = y_1 + hf(x_1, y_1)$ and we approximate $y(x_0 + 2h) = y(x_2) \simeq y_1 + hf(x_1, y_1) = y_2$ and so on. In general, by letting $y_k = y_{k-1} + hf(x_{k-1}, y_{k-1})$, we define (inductively)

$$y(x_0 + (k+1)h) = y_{k+1} \simeq y_k + hf(x_k, y_k), \qquad\qquad k = 0, 1, 2, \ldots, n-1.$$

This method of calculation of $y_1, y_2, \ldots, y_n$ is called the Euler's method. The approximate solution of Equation (7.7.1) is obtained by "linear elements" joining $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$.

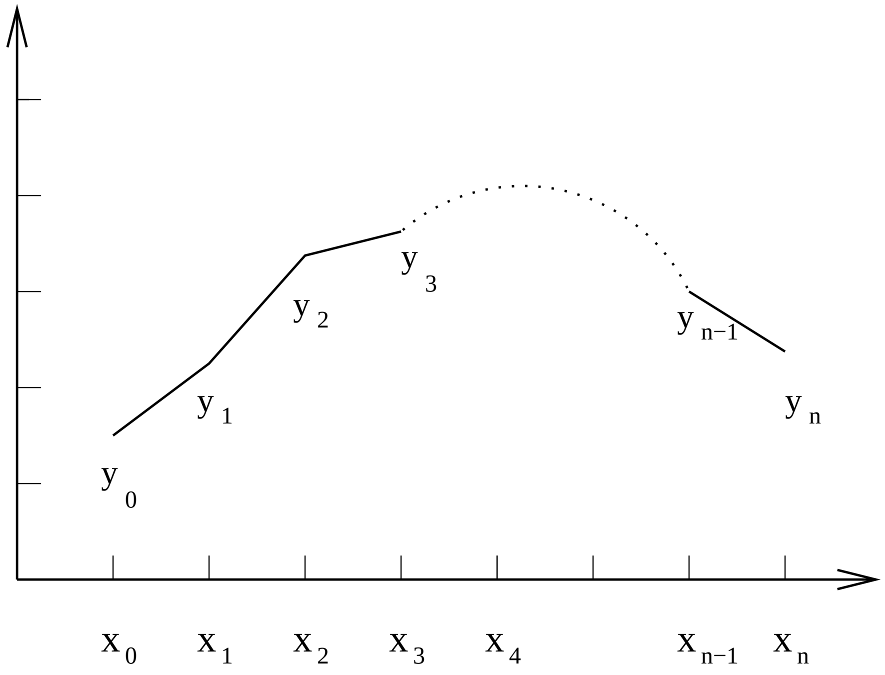


Figure 7.2: **Approximate Solution**

Chapter 8

# Second Order and Higher Order Equations

## 8.1 Introduction

Second order and higher order equations occur frequently in science and engineering (like pendulum problem etc.) and hence has its own importance. It has its own flavour also. We devote this section for an elementary introduction.

**Definition 8.1.1 (Second Order Linear Differential Equation)** The equation

$$p(x)y'' + q(x)y' + r(x)y = c(x), \quad x \in I \tag{8.1.1}$$

is called a SECOND ORDER LINEAR DIFFERENTIAL EQUATION.

Here $I$ is an interval contained in $\mathbb{R}$; and the functions $p(\cdot), q(\cdot), r(\cdot),$ and $c(\cdot)$ are real valued continuous functions defined on $\mathbb{R}$. The functions $p(\cdot), q(\cdot),$ and $r(\cdot)$ are called the coefficients of Equation (8.1.1) and $c(x)$ is called the non-homogeneous term or the force function.

Equation (8.1.1) is called linear homogeneous if $c(x) \equiv 0$ and non-homogeneous if $c(x) \neq 0$.

Recall that a second order equation is called nonlinear if it is not linear.

**Example 8.1.2** 1. The equation

$$y'' + \sqrt{\frac{9}{\ell}} \sin y = 0$$

is a second order equation which is nonlinear.

2. $y'' - y = 0$ is an example of a linear second order equation.

3. $y'' + y' + y = \sin x$ is a non-homogeneous linear second order equation.

4. $ax^2y'' + bxy' + cy = 0 \; c \neq 0$ is a homogeneous second order linear equation. This equation is called EULER EQUATION OF ORDER 2. Here $a, b,$ and $c$ are real constants.

**Definition 8.1.3** A function $y$ defined on $I$ is called a solution of Equation (8.1.1) if $y$ is twice differentiable and satisfies Equation (8.1.1).

**Example 8.1.4** 1. $e^x$ and $e^{-x}$ are solutions of $y'' - y = 0$.

2. $\sin x$ and $\cos x$ are solutions of $y'' + y = 0$.

We now state an important theorem whose proof is simple and is omitted.

**Theorem 8.1.5 (Superposition Principle)** Let $y_1$ and $y_2$ be two given solutions of

$$p(x)y'' + q(x)y' + r(x)y = 0, \ \ x \in I. \tag{8.1.2}$$

Then for any two real number $c_1, c_2$, the function $c_1y_1 + c_2y_2$ is also a solution of Equation (8.1.2).

It is to be noted here that Theorem 8.1.5 is not an existence theorem. That is, it does not assert the existence of a solution of Equation (8.1.2).

**Definition 8.1.6 (Solution Space)** The set of solutions of a differential equation is called the solution space.

For example, all the solutions of the Equation (8.1.2) form a solution space. Note that $y(x) \equiv 0$ is also a solution of Equation (8.1.2). Therefore, the solution set of a Equation (8.1.2) is non-empty. A moments reflection on Theorem 8.1.5 tells us that the solution space of Equation (8.1.2) forms a real vector space.

**Remark 8.1.7** *The above statements also hold for any homogeneous linear differential equation. That is, the solution space of a homogeneous linear differential equation is a real vector space.*

The natural question is to inquire about its dimension. This question will be answered in a sequence of results stated below.

We first recall the definition of Linear Dependence and Independence.

**Definition 8.1.8 (Linear Dependence and Linear Independence)** Let $I$ be an interval in $\mathbb{R}$ and let $f, g : I \longrightarrow \mathbb{R}$ be continuous functions. we say that $f, g$ are said to be linearly dependent if there are real numbers $a$ and $b$ (not both zero) such that

$$af(t) + bg(t) = 0 \quad \text{for all} \quad t \in I.$$

The functions $f(\cdot), g(\cdot)$ are said to be linearly independent if $f(\cdot), g(\cdot)$ are not linear dependent.

To proceed further and to simplify matters, we assume that $p(x) \equiv 1$ in Equation (8.1.2) and that the function $q(x)$ and $r(x)$ are continuous on $I$.

In other words, we consider a homogeneous linear equation

$$y'' + q(x)y' + r(x)y = 0, \quad x \in I, \tag{8.1.3}$$

where $q$ and $r$ are real valued continuous functions defined on $I$.

The next theorem, given without proof, deals with the existence and uniqueness of solutions of Equation (8.1.3) with initial conditions $y(x_0) = A, \ y'(x_0) = B$ for some $x_0 \in I$.

**Theorem 8.1.9 (Picard's Theorem on Existence and Uniqueness)** Consider the Equation (8.1.3) along with the conditions

$$y(x_0) = A, \ y'(x_0) = B, \ \ \text{for some} \ \ x_0 \in I \tag{8.1.4}$$

where $A$ and $B$ are prescribed real constants. Then Equation (8.1.3), with initial conditions given by Equation (8.1.4) has a unique solution on $I$.

**A word of Caution:** NOTE THAT THE COEFFICIENT OF $y''$ IN EQUATION (8.1.3) IS 1. BEFORE WE APPLY THEOREM 8.1.9, WE HAVE TO ENSURE THIS CONDITION.

An important application of Theorem 8.1.9 is that the equation (8.1.3) has exactly 2 linearly independent solutions. In other words, the set of all solutions over $\mathbb{R}$, forms a real vector space of dimension 2.

**Theorem 8.1.10** Let $q$ and $r$ be real valued continuous functions on $I$. Then Equation (8.1.3) has exactly two linearly independent solutions. Moreover, if $y_1$ and $y_2$ are two linearly independent solutions of Equation (8.1.3), then the solution space is a linear combination of $y_1$ and $y_2$.

PROOF. Let $y_1$ and $y_2$ be two unique solutions of Equation (8.1.3) with initial conditions

$$y_1(x_0) = 1,\ y_1'(x_0) = 0, \quad \text{and} \quad y_2(x_0) = 0,\ y_2'(x_0) = 1 \ \text{ for some } \ x_0 \in I. \tag{8.1.5}$$

The unique solutions $y_1$ and $y_2$ exist by virtue of Theorem 8.1.9. We now claim that $y_1$ and $y_2$ are linearly independent. Consider the system of linear equations

$$\alpha y_1(x) + \beta y_2(x) = 0, \tag{8.1.6}$$

where $\alpha$ and $\beta$ are unknowns. If we can show that the only solution for the system (8.1.6) is $\alpha = \beta = 0$, then the two solutions $y_1$ and $y_2$ will be linearly independent.

Use initial condition on $y_1$ and $y_2$ to show that the only solution is indeed $\alpha = \beta = 0$. Hence the result follows.

We now show that any solution of Equation (8.1.3) is a linear combination of $y_1$ and $y_2$. Let $\zeta$ be any solution of Equation (8.1.3) and let $d_1 = \zeta(x_0)$ and $d_2 = \zeta'(x_0)$. Consider the function $\phi$ defined by

$$\phi(x) = d_1 y_1(x) + d_2 y_2(x).$$

By Definition 8.1.3, $\phi$ is a solution of Equation (8.1.3). Also note that $\phi(x_0) = d_1$ and $\phi'(x_0) = d_2$. So, $\phi$ and $\zeta$ are two solution of Equation (8.1.3) with the same initial conditions. Hence by Picard's Theorem on Existence and Uniqueness (see Theorem 8.1.9), $\phi(x) \equiv \zeta(x)$ or

$$\zeta(x) = d_1 y_1(x) + d_2 y_2(x).$$

Thus, the equation (8.1.3) has two linearly independent solutions. □

**Remark 8.1.11**
1. *Observe that the solution space of Equation (8.1.3) forms a real vector space of dimension* 2.

2. *The solutions $y_1$ and $y_2$ corresponding to the initial conditions*

$$y_1(x_0) = 1,\ y_1'(x_0) = 0, \quad \textit{and} \quad y_2(x_0) = 0,\ y_2'(x_0) = 1 \ \textit{ for some } \ x_0 \in I,$$

*are called a* FUNDAMENTAL SYSTEM *of solutions for Equation (8.1.3).*

3. *Note that the fundamental system for Equation (8.1.3) is not unique.*
*Consider a $2 \times 2$ non-singular matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $a, b, c, d \in \mathbb{R}$. Let $\{y_1, y_2\}$ be a fundamental system for the differential Equation 8.1.3 and $\mathbf{y}^t = [y_1,\ y_2]$. Then the rows of the matrix $A\mathbf{y} = \begin{bmatrix} ay_1 + by_2 \\ cy_1 + dy_2 \end{bmatrix}$ also form a fundamental system for Equation 8.1.3. That is, if $\{y_1, y_2\}$ is a fundamental system for Equation 8.1.3 then $\{ay_1 + by_2, cy_1 + dy_2\}$ is also a fundamental system whenever $ad - bc = \det(A) \neq 0$.*

**Example 8.1.12** $\{1, \mathbf{x}\}$ is a fundamental system for $y'' = 0$.

Note that $\{1 - \mathbf{x}, 1 + \mathbf{x}\}$ is also a fundamental system. Here the matrix is $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

**Exercise 8.1.13**
1. State whether the following equations are SECOND-ORDER LINEAR or SECOND-ORDER NON-LINEAR equaitons.

(a) $y'' + y\sin x = 5.$

(b) $y'' + (y')^2 + y\sin x = 0.$

(c) $y'' + yy' = -2.$

(d) $(x^2+1)y'' + (x^2+1)^2 y' - 5y = \sin x.$

2. By showing that $y_1 = e^x$ and $y_2 = e^{-x}$ are solutions of

$$y'' - y = 0$$

conclude that $\sinh x$ and $\cosh x$ are also solutions of $y'' - y = 0$. Do $\sinh x$ and $\cosh x$ form a fundamental set of solutions?

3. Given that $\{\sin x, \cos x\}$ forms a basis for the solution space of $y'' + y = 0$, find another basis.

## 8.2 More on Second Order Equations

In this section, we wish to study some more properties of second order equations which have nice applications. They also have natural generalisations to higher order equations.

**Definition 8.2.1 (General Solution)** Let $y_1$ and $y_2$ be a fundamental system of solutions for

$$y'' + q(x)y' + r(x)y = 0, \; x \in I. \tag{8.2.1}$$

The general solution $y$ of Equation (8.2.1) is defined by

$$y = c_1 y_1 + c_2 y_2, \; x \in I$$

where $c_1$ and $c_2$ are arbitrary real constants. Note that $y$ is also a solution of Equation (8.2.1).

In other words, the general solution of Equation (8.2.1) is a 2-parameter family of solutions, the parameters being $c_1$ and $c_2$.

### 8.2.1 Wronskian

In this subsection, we discuss the linear independence or dependence of two solutions of Equation (8.2.1).

**Definition 8.2.2 (Wronskian)** Let $y_1$ and $y_2$ be two real valued continuously differentiable function on an interval $I \subset \mathbb{R}$. For $x \in I$, define

$$\begin{aligned} W(y_1, y_2) &:= \begin{vmatrix} y_1 & y_1' \\ y_2 & y_2' \end{vmatrix} \\ &= y_1 y_2' - y_1' y_2. \end{aligned}$$

$W$ is called the Wronskian of $y_1$ and $y_2$.

**Example 8.2.3** 1. Let $y_1 = \cos x$ and $y_2 = \sin x, \; x \in I \subset \mathbb{R}$. Then

$$W(y_1, y_2) = \begin{vmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{vmatrix} \equiv -1 \quad \text{for all} \quad x \in I. \tag{8.2.2}$$

Hence $\{\cos x, \sin x\}$ is a linearly independent set.

2. Let $y_1 = x^2|x|$, and $y_2 = x^3$ for $x \in (-1,1)$. Let us now compute $y_1'$ and $y_2'$. From analysis, we know that $y_1$ is differentiable at $x = 0$ and

$$y_1(x) = -3x^2 \text{ if } x < 0 \text{ and } y_1(x) = 3x^2 \text{ if } x \geq 0.$$

Therefore, for $x \geq 0$,

$$W(y_1, y_2) = \begin{vmatrix} y_1 & y_1' \\ y_2 & y_2' \end{vmatrix} = \begin{vmatrix} x^3 & 3x^2 \\ x^3 & 3x^2 \end{vmatrix} = 0$$

and for $x < 0$,

$$W(y_1, y_2) = \begin{vmatrix} y_1 & y_1' \\ y_2 & y_2' \end{vmatrix} = \begin{vmatrix} -x^3 & -3x^2 \\ x^3 & 3x^2 \end{vmatrix} = 0.$$

That is, for all $x \in (-1,1)$, $W(y_1, y_2) = 0$.

It is also easy to note that $y_1, y_2$ are linearly independent on $(-1,1)$. In fact,they are linearly independent on any interval $(a,b)$ containing $0$.

Given two solutions $y_1$ and $y_2$ of Equation (8.2.1), we have a characterisation for $y_1$ and $y_2$ to be linearly independent.

**Theorem 8.2.4** Let $I \subset \mathbb{R}$ be an interval. Let $y_1$ and $y_2$ be two solutions of Equation (8.2.1). Fix a point $x_0 \in I$. Then for any $x \in I$,

$$W(y_1, y_2) = W(y_1, y_2)(x_0) \exp(-\int_{x_0}^{x} q(s)ds). \tag{8.2.3}$$

Consequently,

$$W(y_1, y_2)(x_0) \neq 0 \Longleftrightarrow W(y_1, y_2) \neq 0 \quad \text{for all} \quad x \in I.$$

Proof. First note that, for any $x \in I$,

$$W(y_1, y_2) = y_1 y_2' - y_1' y_2.$$

So

$$\begin{aligned} \frac{d}{dx} W(y_1, y_2) &= y_1 y_2'' - y_1'' y_2 && (8.2.4) \\ &= y_1\left(-q(x)y_2' - r(x)y_2\right) - \left(-q(x)y_1' - r(x)y_1\right) y_2 && (8.2.5) \\ &= q(x)\big(y_1' y_2 - y_1 y_2'\big) && (8.2.6) \\ &= -q(x) W(y_1, y_2). && (8.2.7) \end{aligned}$$

So, we have

$$W(y_1, y_2) = W(y_1, y_2)(x_0) \ \exp\big(-\int_{x_0}^{x} q(s)ds\big).$$

This completes the proof of the first part.

The second part follows the moment we note that the exponential function does not vanish. Alternatively, $W(y_1, y_2)$ satisfies a first order linear homogeneous equation and therefore

$$W(y_1, y_2) \equiv 0 \quad \text{if and only if} \quad W(y_1, y_2)(x_0) = 0.$$

□

**Remark 8.2.5** *1. If the Wronskian $W(y_1, y_2)$ of two solutions $y_1, y_2$ of (8.2.1) vanish at a point $x_0 \in I$, then $W(y_1, y_2)$ is identically zero on $I$.*

2. *If any two solutions $y_1, y_2$ of Equation (8.2.1) are linearly dependent (on $I$), then $W(y_1, y_2) \equiv 0$ on $I$.*

**Theorem 8.2.6** Let $y_1$ and $y_2$ be any two solutions of Equation (8.2.1). Let $x_0 \in I$ be arbitrary. Then $y_1$ and $y_2$ are linearly independent on $I$ if and only if $W(y_1, y_2)(x_0) \neq 0$.

Proof. Let $y_1, y_2$ be linearly independent on $I$.
To show: $W(y_1, y_2)(x_0) \neq 0$.

Suppose not. Then $W(y_1, y_2)(x_0) = 0$. So, by Theorem 2.6.1 the equations

$$c_1 y_1(x_0) + c_2 y_2(x_0) = 0 \quad \text{and} \quad c_1 y_1'(x_0) + c_2 y_2'(x_0) = 0 \tag{8.2.8}$$

admits a non-zero solution $d_1, d_2$. (as $0 = W(y_1, y_2)(x_0) = y_1(x_0) y_2'(x_0) - y_1'(x_0) y_2(x_0)$.)

Let $y = d_1 y_1 + d_2 y_2$. Note that Equation (8.2.8) now implies

$$y(x_0) = 0 \ \text{ and } \ y'(x_0) = 0.$$

Therefore, by Picard's Theorem on existence and uniqueness of solutions (see Theorem 8.1.9), the solution $y \equiv 0$ on $I$. That is, $d_1 y_1 + d_2 y_2 \equiv 0$ for all $x \in I$ with $|d_1| + |d_2| \neq 0$. That is, $y_1, y_2$ is linearly dependent on $I$. A contradiction. Therefore, $W(y_1, y_2)(x_0) \neq 0$. This proves the first part.

Suppose that $W(y_1, y_2)(x_0) \neq 0$ for some $x_0 \in I$. Therefore, by Theorem 8.2.4, $W(y_1, y_2) \neq 0$ for all $x \in I$. Suppose that $c_1 y_1(x) + c_2 y_2(x) = 0$ for all $x \in I$. Therefore, $c_1 y_1'(x) + c_2 y_2'(x) = 0$ for all $x \in I$. Since $x_0 \in I$, in particular, we consider the linear system of equations

$$c_1 y_1(x_0) + c_2 y_2(x_0) = 0 \ \text{ and } \ c_1 y_1'(x_0) + c_2 y_2'(x_0) = 0. \tag{8.2.9}$$

But then by using Theorem 2.6.1 and the condition $W(y_1, y_2)(x_0) \neq 0$, the only solution of the linear system (8.2.9) is $c_1 = c_2 = 0$. So, by Definition 8.1.8, $y_1, y_2$ are linearly independent. □

**Remark 8.2.7** *Recall the following from Example 2:*

1. *The interval $I = (-1, 1)$.*

2. *$y_1 = x^2|x|$, $y_2 = x^3$ and $W(y_1, y_2) \equiv 0$ for all $x \in I$.*

3. *The functions $y_1$ and $y_2$ are linearly independent.*

*This example tells us that Theorem 8.2.6 may not hold if $y_1$ and $y_2$ are not solutions of Equation (8.2.1) but are just some arbitrary functions on $(-1, 1)$.*

The following corollary is a consequence of Theorem 8.2.6.

**Corollary 8.2.8** Let $y_1, y_2$ be two linearly independent solutions of Equation (8.2.1). Let $y$ be any solution of Equation (8.2.1). Then there exist unique real numbers $d_1, d_2$ such that

$$y = d_1 y_1 + d_2 y_2 \ \text{ on } \ I.$$

Proof. Let $x_0 \in I$. Let $y(x_0) = a$, $y'(x_0) = b$. Here $a$ and $b$ are known since the solution $y$ is given. Also for any $x_0 \in I$, by Theorem 8.2.6, $W(y_1, y_2)(x_0) \neq 0$ as $y_1, y_2$ are linearly independent solutions of Equation (8.2.1). Therefore by Theorem 2.6.1, the system of linear equations

$$c_1 y_1(x_0) + c_2 y_2(x_0) = a \ \text{ and } \ c_1 y_1'(x_0) + c_2 y_2'(x_0) = b \tag{8.2.10}$$

has a unique solution $d_1, d_2$.
Define $\zeta(x) = d_1 y_1 + d_2 y_2$ for $x \in I$. Note that $\zeta$ is a solution of Equation (8.2.1) with $\zeta(x_0) = a$ and $\zeta'(x_0) = b$. Hence, by Picard's Theorem on existence and uniqueness (see Theorem 8.1.9), $\zeta = y$ for all $x \in I$. That is, $y = d_1 y_1 + d_2 y_2$. □

**Exercise 8.2.9** 1. Let $y_1$ and $y_2$ be any two linearly independent solutions of $y'' + a(x)y = 0$. Find $W(y_1, y_2)$.

2. Let $y_1$ and $y_2$ be any two linearly independent solutions of

$$y'' + a(x)y' + b(x)y = 0, \; x \in I.$$

Show that $y_1$ and $y_2$ cannot vanish at any $x = x_0 \in I$.

3. Show that there is no equation of the type

$$y'' + a(x)y' + b(x)y = 0, \; x \in [0, 2\pi]$$

admiting $y_1 = \sin x$ and $y_2 = x - \pi$ as its solutions; where $a(x)$ and $b(x)$ are any continuous functions on $[0, 2\pi]$. *[Hint: Use Exercise 8.2.9.2.]*

### 8.2.2 Method of Reduction of Order

We are going to show that in order to find a fundamental system for Equation (8.2.1), it is sufficient to have the knowledge of a solution of Equation (8.2.1). In other words, if we know one (non-zero) solution $y_1$ of Equation (8.2.1), then we can determine a solution $y_2$ of Equation (8.2.1), so that $\{y_1, y_2\}$ forms a fundamental system for Equation (8.2.1). The method is described below and is usually called the method of reduction of order.

Let $y_1$ be an every where non-zero solution of Equation (8.2.1). Assume that $y_2 = u(x)y_1$ is a solution of Equation (8.2.1), where $u$ is to be determined. Substituting $y_2$ in Equation (8.2.1), we have (after a bit of simplification)

$$u''y_1 + u'(2y_1' + py_1) + u(y_1'' + py_1' + qy_1) = 0.$$

By letting $u' = v$, and observing that $y_1$ is a solution of Equation (8.2.1), we have

$$v'y_1 + v(2y_1' + py_1) = 0$$

which is same as

$$\frac{d}{dx}(vy_1^2) = -p(vy_1^2).$$

This is a linear equation of order one (hence the name, reduction of order) in $v$ whose solution is

$$vy_1^2 = exp(-\int_{x_0}^{x} p(s)ds), \quad x_0 \in I.$$

Substituting $v = u'$ and integrating we get

$$u = \int_{x_0}^{x} \frac{1}{y_1^2(s)} exp(-\int_{x_0}^{s} p(t)dt)ds, \; x_0 \in I$$

and hence a second solution of Equation (8.2.1) is

$$y_2 = y_1 \int_{x_0}^{x} \frac{1}{y_1^2(s)} exp(-\int_{x_0}^{s} p(t)dt)ds.$$

It is left as an exercise to show that $y_1, y_2$ are linearly independent. That is, $\{y_1, y_2\}$ form a fundamental system for Equation (8.2.1).

We illustrate the method by an example.

**Example 8.2.10** Given that e $y_1 = \frac{1}{x},\ x \geq 1$ is a solution of

$$x^2y'' + 4xy' + 2y = 0, \tag{8.2.11}$$

determine another solution $y_2$ of (8.2.11), such that the solutions $y_1, y_2,$ for $x \geq 1$ are linearly independent.
**Solution:** With the notations used above, note that $x_0 = 1,\ p(x) = \frac{4}{x},$ and $y_2(x) = u(x)y_1(x)$, where $u$ is given by

$$\begin{aligned} u &= \int_1^x \frac{1}{y_1^2(s)} \exp\Big(-\int_1^s p(t)dt\Big)\, ds \\ &= \int_1^x \frac{1}{y_1^2(s)} \exp\big(\ln(s^4)\big) ds \\ &= \int_1^x \frac{s^2}{s^4} ds = 1 - \frac{1}{x}; \end{aligned}$$

where $A$ and $B$ are constants. So,

$$y_2(x) = \frac{1}{x} - \frac{1}{x^2}.$$

Since the term $\frac{1}{x}$ already appears in $y_1$, we can take $y_2 = \frac{1}{x^2}$. So, $\frac{1}{x}$ and $\frac{1}{x^2}$ are the required two linearly independent solutions of (8.2.11).

**Exercise 8.2.11** In the following, use the given solution $y_1$, to find another solution $y_2$ so that the two solutions $y_1$ and $y_2$ are linearly independent.

1. $y'' = 0,\ y_1 = 1,\ x \geq 0.$
2. $y'' + 2y' + y = 0,\ y_1 = e^x,\ x \geq 0.$
3. $x^2y'' - xy' + y = 0,\ y_1 = x,\ x \geq 1.$
4. $xy'' + y' = 0,\ y_1 = 1,\ x \geq 1.$
5. $y'' + xy' - y = 0,\ y_1 = x,\ x \geq 1.$

## 8.3 Second Order equations with Constant Coefficients

**Definition 8.3.1** Let $a$ and $b$ be constant real numbers. An equation

$$y'' + ay' + by = 0 \tag{8.3.1}$$

is called a SECOND ORDER HOMOGENEOUS LINEAR EQUATION WITH CONSTANT COEFFICIENTS.

Let us assume that $y = e^{\lambda x}$ to be a solution of Equation (8.3.1) (where $\lambda$ is a constant, and is to be determined). To simplify the matter, we denote

$$L(y) = y'' + ay' + by$$

and

$$p(\lambda) = \lambda^2 + a\lambda + b.$$

It is easy to note that

$$L(e^{\lambda x}) = p(\lambda)e^{\lambda x}.$$

Now, it is clear that $e^{\lambda x}$ is a solution of Equation (8.3.1) if and only if

$$p(\lambda) = 0. \tag{8.3.2}$$

Equation (8.3.2) is called the CHARACTERISTIC EQUATION of Equation (8.3.1). Equation (8.3.2) is a quadratic equation and admits 2 roots (repeated roots being counted twice).

Case 1: Let $\lambda_1, \lambda_2$ be real roots of Equation (8.3.2) with $\lambda_1 \neq \lambda_2$.
Then $e^{\lambda_1 x}$ and $e^{\lambda_2 x}$ are two solutions of Equation (8.3.1) and moreover they are linearly independent (since $\lambda_1 \neq \lambda_2$). That is, $\{e^{\lambda_1 x}, e^{\lambda_2 x}\}$ forms a fundamental system of solutions of Equation (8.3.1).

Case 2: Let $\lambda_1 = \lambda_2$ be a repeated root of $p(\lambda) = 0$.
Then $p'(\lambda_1) = 0$. Now,

$$\frac{d}{dx}(L(e^{\lambda x})) = L(xe^{\lambda x}) = p'(\lambda)e^{\lambda x} + xp(\lambda)e^{\lambda x}.$$

But $p'(\lambda_1) = 0$ and therefore,

$$L(xe^{\lambda_1 x}) = 0.$$

Hence, $e^{\lambda_1 x}$ and $xe^{\lambda_1 x}$ are two linearly independent solutions of Equation (8.3.1). In this case, we have a fundamental system of solutions of Equation (8.3.1).

Case 3: Let $\lambda = \alpha + i\beta$ be a complex root of Equation (8.3.2).
So, $\alpha - i\beta$ is also a root of Equation (8.3.2). Before we proceed, we note:

**Lemma 8.3.2** Let $y = u + iv$ be a solution of Equation (8.3.1), where $u$ and $v$ are real valued functions. Then $u$ and $v$ are solutions of Equation (8.3.1). In other words, the real part and the imaginary part of a complex valued solution (of a real variable ODE Equation (8.3.1)) are themselves solution of Equation (8.3.1).

PROOF. exercise. □

Let $\lambda = \alpha + i\beta$ be a complex root of $p(\lambda) = 0$. Then

$$e^{\alpha x}(\cos(\beta x) + i\sin(\beta x))$$

is a complex solution of Equation (8.3.1). By Lemma 8.3.2, $y_1 = e^{\alpha x}\cos(\beta x)$ and $y_2 = \sin(\beta x)$ are solutions of Equation (8.3.1). It is easy to note that $y_1$ and $y_2$ are linearly independent. It is as good as saying $\{e^{\lambda x}\cos(\beta x), e^{\lambda x}\sin(\beta x)\}$ forms a fundamental system of solutions of Equation (8.3.1).

**Exercise 8.3.3** 1. Find the general solution of the follwoing equations.

(a) $y'' - 4y' + 3y = 0$.

(b) $2y'' + 5y = 0$.

(c) $y'' - 9y = 0$.

(d) $y'' + k^2 y = 0$, where $k$ is a real constant.

2. Solve the following IVP's.

(a) $y'' + y = 0, \ y(0) = 0, \ y'(0) = 1$.

(b) $y'' - y = 0, \ y(0) = 1, \ y'(0) = 1$.

(c) $y'' + 4y = 0, \ y(0) = -1, \ y'(0) = -3$.

(d) $y'' + 4y' + 4y = 0, \ y(0) = 1, \ y'(0) = 0$.

3. Find two linearly independent solutions $y_1$ and $y_2$ of the following equations.

(a) $y'' - 5y = 0$.

(b) $y'' + 6y' + 5y = 0$.

(c) $y'' + 5y = 0$.

(d) $y'' + 6y' + 9y = 0$. Also, in each case, find $W(y_1, y_2)$.

4. Show that the IVP

$$y'' + y = 0,\ y(0) = 0 \text{ and } y'(0) = B$$

has a unique solution for any real number $B$.

5. Consider the problem

$$y'' + y = 0,\ y(0) = 0 \text{ and } y'(\pi) = B. \tag{8.3.3}$$

Show that it has a solution if and only if $B = 0$. Compare this with Exercise 4. Also, show that if $B = 0$, then there are infinitely many solutions to (8.3.3).

## 8.4 Non Homogeneous Equations

Throughout this section, $I$ denotes an interval in $\mathbb{R}$. we assume that $q(\cdot), r(\cdot)$ and $f(\cdot)$ are real valued continuous function defined on $I$. Now, we focus the attention to the study of non-homogeneous equation of the form

$$y'' + q(x)y' + r(x)y = f(x). \tag{8.4.1}$$

We assume that the functions $q(\cdot), r(\cdot)$ and $f(\cdot)$ are known/given. The non-zero function $f(\cdot)$ in (8.4.1) is also called the non-homogeneous term or the forcing function. The equation

$$y'' + q(x)y' + r(x)y = 0. \tag{8.4.2}$$

is called the homogeneous equation corresponding to (8.4.1).

Consider the set of all twice differentiable functions defined on $I$. We define an operator $L$ on this set by

$$L(y) = y'' + q(x)y' + r(x)y.$$

Then (8.4.1) and (8.4.2) can be rewritten in the (compact) form

$$L(y) = f \tag{8.4.3}$$

$$L(y) = 0. \tag{8.4.4}$$

The ensuing result relates the solutions of (8.4.1) and (8.4.2).

**Theorem 8.4.1** 1. Let $y_1$ and $y_2$ be two solutions of (8.4.1) on $I$. Then $y = y_1 - y_2$ is a solution of (8.4.2).

2. Let $z$ be any solution of (8.4.1) on $I$ and let $z_1$ be any solution of (8.4.2). Then $y = z + z_1$ is a solution of (8.4.1) on $I$.

Proof. Observe that $L$ is a linear transformation on the set of twice differentiable function on $I$. We therefore have

$$L(y_1) = f \quad \text{and} \quad L(y_2) = f.$$

The linearity of $L$ implies that $L(y_1 - y_2) = 0$ or equivalently, $y = y_1 - y_2$ is a solution of (8.4.2).

For the proof of second part, note that

$$L(z) = f \quad \text{and} \quad L(z_1) = 0$$

implies that

$$L(z + z_1) = L(z) + L(z_1) = f.$$

Thus, $y = z + z_1$ is a solution of (8.4.1). □

The above result leads us to the following definition.

**Definition 8.4.2 (General Solution)** A general solution of (8.4.1) on $I$ is a solution of (8.4.1) of the form

$$y = y_h + y_p, \; x \in I$$

where $y_h = c_1 y_1 + c_2 y_2$ is a general solution of the corresponding homogeneous equation (8.4.2) and $y_p$ is any solution of (8.4.1) (preferably containing no arbitrary constants).

We now prove that the solution of (8.4.1) with initial conditions is unique.

**Theorem 8.4.3 (Uniqueness)** Suppose that $x_0 \in I$. Let $y_1$ and $y_2$ be two solutions of the IVP

$$y'' + qy' + ry = f, \;\; y(x_0) = a, \;\; y'(x_0) = b. \tag{8.4.5}$$

Then $y_1 = y_2$ for all $x \in I$.

PROOF. Let $z = y_1 - y_2$. Then $z$ satisfies

$$L(z) = 0, \;\; z(x_0) = 0, \;\; z'(x_0) = 0.$$

By the uniqueness theorem 8.1.9, we have $z \equiv 0$ on $I$. Or in other words, $y_1 \equiv y_2$ on $I$. □

**Remark 8.4.4** *The above results tell us that to solve (i.e., to find the general solution of (8.4.1)) or the IVP (8.4.5), we need to find the general solution of the homogeneous equation (8.4.2) and a particular solution $y_p$ of (8.4.1). To repeat, the two steps needed to solve (8.4.1), are:*

1. *compute the general solution of (8.4.2), and*
2. *compute a particular solution of (8.4.1).*

   *Then add the two solutions.*

*Step 1. has been dealt in the previous sections. The remainder of the section is devoted to step 2., i.e., we elaborate some methods for computing a particular solution $y_p$ of (8.4.1).*

**Exercise 8.4.5** 1. Find the general solution of the following equations:

   (a) $y'' + 5y' = -5$. (You may note here that $y = -x$ is a particular solution.)

   (b) $y'' - y = -2\sin x$. (First show that $y = \sin x$ is a particular solution.)

2. Solve the following IVPs:

   (a) $y'' + y = 2e^x, \; y(0) = 0 = y'(0)$. (It is given that $y = e^x$ is a particular solution.)

   (b) $y'' - y = -2\cos x, \;\; y(0) = 0, \; y'(0) = 1$. (First guess a particular solution using the idea given in Exercise 8.4.5.1b )

3. Let $f_1(x)$ and $f_2(x)$ be two continuous functions. Let $y_i$'s be particular solutions of

$$y'' + q(x)y' + r(x)y = f_i(x), \; i = 1, 2;$$

   where $q(x)$ and $r(x)$ are continuous functions. Show that $y_1 + y_2$ is a particular solution of $y'' + q(x)y' + r(x)y = f_1(x) + f_2(x)$.

## 8.5 Variation of Parameters

In the previous section, calculation of particular integrals/solutions for some special cases have been studied. Recall that the homogeneous part of the equation had constant coefficients. In this section, we deal with a useful technique of finding a particular solution when the coefficients of the homogeneous part are continuous functions and the forcing function $f(x)$ (or the non-homogeneous term) is piecewise continuous. Suppose $y_1$ and $y_2$ are two linearly independent solutions of

$$y'' + q(x)y' + r(x)y = 0 \tag{8.5.1}$$

on $I$, where $q(x)$ and $r(x)$ are arbitrary continuous functions defined on $I$. Then we know that

$$y = c_1 y_1 + c_2 y_2$$

is a solution of (8.5.1) for any constants $c_1$ and $c_2$. We now "vary" $c_1$ and $c_2$ to functions of $x$, so that

$$y = u(x)y_1 + v(x)y_2, \ \ x \in I \tag{8.5.2}$$

is a solution of the equation

$$y'' + q(x)y' + r(x)y = f(x), \quad \text{on} \ \ I, \tag{8.5.3}$$

where $f$ is a piecewise continuous function defined on $I$. The details are given in the following theorem.

**Theorem 8.5.1 (Method of Variation of Parameters)** Let $q(x)$ and $r(x)$ be continuous functions defined on $I$ and let $f$ be a piecewise continuous function on $I$. Let $y_1$ and $y_2$ be two linearly independent solutions of (8.5.1) on $I$. Then a particular solution $y_p$ of (8.5.3) is given by

$$y_p = -y_1 \int \frac{y_2 f(x)}{W} dx + y_2 \int \frac{y_1 f(x)}{W} dx, \tag{8.5.4}$$

where $W = W(y_1, y_2)$ is the Wronskian of $y_1$ and $y_2$. (Note that the integrals in (8.5.4) are the indefinite integrals of the respective arguments.)

Proof. Let $u(x)$ and $v(x)$ be continuously differentiable functions (to be determined) such that

$$y_p = uy_1 + vy_2, \ \ x \in I \tag{8.5.5}$$

is a particular solution of (8.5.3). Differentiation of (8.5.5) leads to

$$y_p' = uy_1' + vy_2' + u'y_1 + v'y_2. \tag{8.5.6}$$

We choose $u$ and $v$ so that

$$u'y_1 + v'y_2 = 0. \tag{8.5.7}$$

Substituting (8.5.7) in (8.5.6), we have

$$y_p' = uy_1' + vy_2', \ \ \text{and} \ \ y_p'' = uy_1'' + vy_2'' + u'y_1' + v'y_2'. \tag{8.5.8}$$

Since $y_p$ is a particular solution of (8.5.3), substitution of (8.5.5) and (8.5.8) in (8.5.3), we get

$$u\big(y_1'' + q(x)y_1' + r(x)y_1\big) + v\big(y_2'' + q(x)y_2' + r(x)y_2\big) + u'y_1' + v'y_2' = f(x).$$

As $y_1$ and $y_2$ are solutions of the homogeneous equation (8.5.1), we obtain the condition

$$u'y_1' + v'y_2' = f(x). \tag{8.5.9}$$

We now determine $u$ and $v$ from (8.5.7) and (8.5.9). By using the Cramer's rule for a linear system of equations, we get

$$u' = -\frac{y_2 f(x)}{W} \quad \text{and} \quad v' = \frac{y_1 f(x)}{W} \tag{8.5.10}$$

(note that $y_1$ and $y_2$ are linearly independent solutions of (8.5.1) and hence the Wronskian, $W \neq 0$ for any $x \in I$). Integration of (8.5.10) give us

$$u = -\int \frac{y_2 f(x)}{W} dx \quad \text{and} \quad v = \int \frac{y_1 f(x)}{W} dx \tag{8.5.11}$$

( without loss of generality, we set the values of integration constants to zero). Equations (8.5.11) and (8.5.5) yield the desired results. Thus the proof is complete. □

Before, we move onto some examples, the following comments are useful.

**Remark 8.5.2** 1. *The integrals in (8.5.11) exist, because $y_2$ and $W(\neq 0)$ are continuous functions and $f$ is a piecewise continuous function. Sometimes, it is useful to write (8.5.11) in the form*

$$u = -\int_{x_0}^{x} \frac{y_2(s) f(s)}{W(s)} ds \quad \text{and} \quad v = \int_{x_0}^{x} \frac{y_1(s) f(s)}{W(s)} ds$$

*where $x \in I$ and $x_0$ is a fixed point in $I$. In such a case, the particular solution $y_p$ as given by (8.5.4) assumes the form*

$$y_p = -y_1 \int_{x_0}^{x} \frac{y_2(s) f(s)}{W(s)} ds + y_2 \int )_{x_0}^{x} \frac{y_1(s) f(s)}{W(s)} ds \tag{8.5.12}$$

*for a fixed point $x_0 \in I$ and for any $x \in I$.*

2. *Again, we stress here that, $q$ and $r$ are assumed to be continuous. They need not be constants. Also, $f$ is a piecewise continuous function on $I$.*

3. *A word of caution. While using (8.5.4), one has to keep in mind that the coefficient of $y''$ in (8.5.3) is* 1.

**Example 8.5.3** 1. Find the general solution of

$$y'' + y = \frac{1}{2 + \sin x}, \quad x \geq 0.$$

**Solution:** The general solution of the corresponding homogeneous equation $y'' + y = 0$ is given by

$$y_h = c_1 \cos x + c_2 \sin x.$$

Here, the solutions $y_1 = \sin x$ and $y_2 = \cos x$ are linearly independent over $I = [0, \infty)$ and $W = W(\sin x, \cos x) = 1$. Therefore, a particular solution, $y_h$, by Theorem 8.5.1, is

$$\begin{aligned} y_p &= -y_1 \int \frac{y_2}{2 + \sin x} dx + y_2 \int \frac{y_1}{2 + \sin x} dx \\ &= \sin x \int \frac{\cos x}{2 + \sin x} dx + \cos x \int \frac{\sin x}{2 + \sin x} dx \\ &= -sinx \ \ln(2 + \sin x) + \cos x \ (x - 2\int \frac{1}{2 + \sin x} dx). \end{aligned} \tag{8.5.13}$$

So, the required general solution is

$$y = c_1 \cos x + c_2 \sin x + y_p$$

where $y_p$ is given by (8.5.13).

2. Find a particular solution of
$$x^2y'' - 2xy' + 2y = x^3, \ x > 0.$$

**Solution:** Verify that the given equation is
$$y'' - \frac{2}{x}y' + \frac{2}{x^2}y = x$$
and two linearly independent solutions of the corresponding homogeneous part are $y_1 = x$ and $y_2 = x^2$. Here
$$W = W(x, x^2) = \begin{vmatrix} x & x^2 \\ 1 & 2x \end{vmatrix} = x^2, \ \ x > 0.$$
By Theorem 8.5.1, a particular solution $y_p$ is given by
$$\begin{aligned} y_p &= -x \int \frac{x^2 \cdot x}{x^2} dx + x^2 \int \frac{x \cdot x}{x^2} dx \\ &= -\frac{x^3}{2} + x^3 = \frac{x^3}{2}. \end{aligned}$$
The readers should note that the methods of Section 8.7 are not applicable as the given equation is not an equation with constant coefficients.

**Exercise 8.5.4** 1. Find a particular solution for the following problems:

(a) $y'' + y = f(x), \ \ 0 \le x \le 1$ where $f(x) = \begin{cases} 0 & \text{if } \ 0 \le x < \frac{1}{2} \\ 1 & \text{if } \ \frac{1}{2} \le x \le 1. \end{cases}$

(b) $y'' + y = 2 \sec x$ for all $x \in (0, \frac{\pi}{2})$.

(c) $y'' - 3y' + 2y = -2\cos(e^{-x}), \ \ x > 0$.

(d) $x^2y'' + xy' - y = 2x, \ \ x > 0$.

2. Use the method of variation of parameters to find the general solution of

(a) $y'' - y = -e^x$ for all $x \in \mathbb{R}$.

(b) $y'' + y = \sin x$ for all $x \in \mathbb{R}$.

3. Solve the following IVPs:

(a) $y'' + y = f(x), \ \ x \ge 0$ where $f(x) = \begin{cases} 0 & \text{if } \ 0 \le x < 1 \\ 1 & \text{if } \ x \ge 1. \end{cases}$ with $y(0) = 0 = y'(0)$.

(b) $y'' - y = |x|$ for all $x \in [-1, \infty)$ with $y(-1) = 0$ and $y'(-1) = 1$.

## 8.6 Higher Order Equations with Constant Coefficients

This section is devoted to an introductory study of higher order linear equations with constant coefficients. This is an extension of the study of $2^{\text{nd}}$ order linear equations with constant coefficients (see, Section 8.3).

The standard form of a linear $n^{\text{th}}$ order differential equation with constant coefficients is given by
$$L_n(y) = f(x) \quad \text{on} \quad I, \tag{8.6.1}$$
where
$$L_n \equiv \frac{d^n}{dx^n} + a_1 \frac{d^{n-1}}{dx^{n-1}} + \cdots + a_{n-1}\frac{d}{dx} + a_n$$

is a linear differential operator of order $n$ with constant coefficients, $a_1, a_2, \dots, a_n$ being real constants (called the coefficients of the linear equation) and the function $f(x)$ is a piecewise continuous function defined on the interval $I$. We will be using the notation $y^{(n)}$ for the $n^{\text{th}}$ derivative of $y$. If $f(x) \equiv 0$, then (8.6.1) which reduces to

$$L_n(y) = 0 \quad \text{on} \quad I, \tag{8.6.2}$$

is called a homogeneous linear equation, otherwise (8.6.1) is called a non-homogeneous linear equation. The function $f$ is also known as the non-homogeneous term or a forcing term.

**Definition 8.6.1** A function $y$ defined on $I$ is called a **solution** of (8.6.1) if $y$ is $n$ times differentiable and $y$ along with its derivatives satisfy (8.6.1).

**Remark 8.6.2** *1. If $u$ and $v$ are any two solutions of (8.6.1), then $y = u - v$ is also a solution of (8.6.2). Hence, if $v$ is a solution of (8.6.2) and $y_p$ is a solution of (8.6.1), then $u = v + y_p$ is a solution of (8.6.1).*

*2. Let $y_1$ and $y_2$ be two solutions of (8.6.2). Then for any constants (need not be real) $c_1, c_2$,*

$$y = c_1 y_1 + c_2 y_2$$

*is also a solution of (8.6.2). The solution $y$ is called the superposition of $y_1$ and $y_2$.*

*3. Note that $y \equiv 0$ is a solution of (8.6.2). This, along with the super-position principle, ensures that the set of solutions of (8.6.2) forms a vector space over $\mathbb{R}$. This vector space is called the* SOLUTION SPACE *or space of solutions of (8.6.2).*

As in Section 8.3, we first take up the study of (8.6.2). It is easy to note (as in Section 8.3) that for a constant $\lambda$,

$$L_n(e^{\lambda x}) = p(\lambda) e^{\lambda x}$$

where,

$$p(\lambda) = \lambda^n + a_1 \lambda^{n-1} + \cdots + a_n \tag{8.6.3}$$

**Definition 8.6.3 (Characteristic Equation)** The equation $p(\lambda) = 0$, where $p(\lambda)$ is defined in (8.6.3), is called the CHARACTERISTIC EQUATION of (8.6.2).

Note that $p(\lambda)$ is of polynomial of degree $n$ with real coefficients. Thus, it has $n$ zeros (counting with multiplicities). Also, in case of complex roots, they will occur in conjugate pairs. In view of this, we have the following theorem. The proof of the theorem is omitted.

**Theorem 8.6.4** $e^{\lambda x}$ is a solution of (8.6.2) on any interval $I \subset \mathbb{R}$ if and only if $\lambda$ is a root of (8.6.3)

1. If $\lambda_1, \lambda_2, \dots, \lambda_n$ are distinct roots of $p(\lambda) = 0$, then

$$e^{\lambda_1 x}, e^{\lambda_2 x}, \dots, e^{\lambda_n x}$$

are the $n$ linearly independent solutions of (8.6.2).

2. If $\lambda_1$ is a repeated root of $p(\lambda) = 0$ of multiplicity $k$, *i.e.,* $\lambda_1$ is a zero of (8.6.3) repeated $k$ times, then

$$e^{\lambda_1 x}, x e^{\lambda_1 x}, \dots, x^{k-1} e^{\lambda_1 x}$$

are linearly independent solutions of (8.6.2), corresponding to the root $\lambda_1$ of $p(\lambda) = 0$.

3. If $\lambda_1 = \alpha + i\beta$ is a complex root of $p(\lambda) = 0$, then so is the complex conjugate $\overline{\lambda_1} = \alpha - i\beta$. Then the corresponding linearly independent solutions of (8.6.2) are

$$y_1 = e^{\alpha x}\big(\cos(\beta x) + i\sin(\beta x)\big) \quad \text{and} \quad y_2 = e^{\alpha x}\big(\cos(\beta x) - i\sin(\beta x)\big).$$

These are complex valued functions of $x$. However, using super-position principle, we note that

$$\frac{y_1 + y_2}{2} = e^{\alpha x}\cos(\beta x) \quad \text{and} \quad \frac{y_1 - y_2}{2i} = e^{\alpha x}\sin(\beta x)$$

are also solutions of (8.6.2). Thus, in the case of $\lambda_1 = \alpha + i\beta$ being a complex root of $p(\lambda) = 0$, we have the linearly independent solutions

$$e^{\alpha x}\cos(\beta x) \quad \text{and} \quad e^{\alpha x}\sin(\beta x).$$

**Example 8.6.5** 1. Find the solution space of the differential equation

$$y''' - 6y'' + 11y' - 6y = 0.$$

**Solution:** Its characteristic equation is

$$p(\lambda) = \lambda^3 - 6\lambda^2 + 11\lambda - 6 = 0.$$

By inspection, the roots of $p(\lambda) = 0$ are $\lambda = 1, 2, 3$. So, the linearly independent solutions are $e^x, e^{2x}, e^{3x}$ and the solution space is

$$\{c_1 e^x + c_2 e^{2x} + c_3 e^{3x} \;:\; c_1, c_2, c_3 \in \mathbb{R}\}.$$

2. Find the solution space of the differential equation

$$y''' - 2y'' + y' = 0.$$

**Solution:** Its characteristic equation is

$$p(\lambda) = \lambda^3 - 2\lambda^2 + \lambda = 0.$$

By inspection, the roots of $p(\lambda) = 0$ are $\lambda = 0, 1, 1$. So, the linearly independent solutions are $1, e^x, xe^x$ and the solution space is

$$\{c_1 + c_2 e^x + c_3 x e^x \;:\; c_1, c_2, c_3 \in \mathbb{R}\}.$$

3. Find the solution space of the differential equation

$$y^{(4)} + 2y'' + y = 0.$$

**Solution:** Its characteristic equation is

$$p(\lambda) = \lambda^4 + 2\lambda^2 + 1 = 0.$$

By inspection, the roots of $p(\lambda) = 0$ are $\lambda = i, i, -i, -i$. So, the linearly independent solutions are $\sin x, x\sin x, \cos x, x\cos x$ and the solution space is

$$\{c_1 \sin x + c_2 \cos x + c_3 x \sin x + c_4 x \cos x \;:\; c_1, c_2, c_3, c_4 \in \mathbb{R}\}.$$

From the above discussion, it is clear that the linear homogeneous equation (8.6.2), admits $n$ linearly independent solutions since the algebraic equation $p(\lambda) = 0$ has exactly $n$ roots (counting with multiplicity).

**Definition 8.6.6 (General Solution)** Let $y_1, y_2, \ldots, y_n$ be any set of $n$ linearly independent solution of (8.6.2). Then

$$y = c_1 y_1 + c_2 y_2 + \cdots + c_n y_n$$

is called a general solution of (8.6.2), where $c_1, c_2, \ldots, c_n$ are arbitrary real constants.

**Example 8.6.7** 1. Find the general solution of $y''' = 0$.

**Solution:** Note that $0$ is the repeated root of the characteristic equation $\lambda^3 = 0$. So, the general solution is

$$y = c_1 + c_2 x + c_3 x^2.$$

2. Find the general solution of

$$y''' + y'' + y' + y = 0.$$

**Solution:** Note that the roots of the characteristic equation $\lambda^3 + \lambda^2 + \lambda + 1 = 0$ are $-1, i, -i$. So, the general solution is

$$y = c_1 e^{-x} + c_2 \sin x + c_3 \cos x.$$

**Exercise 8.6.8** 1. Find the general solution of the following differential equations:

(a) $y''' + y' = 0$.

(b) $y''' + 5y' - 6y = 0$.

(c) $y^{iv} + 2y'' + y = 0$.

2. Find a linear differential equation with constant coefficients and of order $3$ which admits the following solutions:

(a) $\cos x, \sin x$ and $e^{-3x}$.

(b) $e^x, e^{2x}$ and $e^{3x}$.

(c) $1, e^x$ and $x$.

3. Solve the following IVPs:

(a) $y^{iv} - y = 0, \;\; y(0) = 0, y'(0) = 0, y''(0) = 0, y'''(0) = 1$.

(b) $2y''' + y'' + 2y' + y = 0, \;\; y(0) = 0, y'(0) = 1, y''(0) = 0$.

4. *Euler Cauchy Equations:*

Let $a_0, a_1, \ldots, a_{n-1} \in \mathbb{R}$ be given constants. The equation

$$x^n \frac{d^n y}{dx^n} + a_{n-1} x^{n-1} \frac{d^{n-1} y}{dx^{n-1}} + \cdots + a_0 y = 0, \;\; x \in I \tag{8.6.4}$$

is called the homogeneous Euler-Cauchy Equation (or just Euler's Equation) of degree $n$. (8.6.4) is also called the standard form of the Euler equation. We define

$$L(y) = x^n \frac{d^n y}{dx^n} + a_{n-1} x^{n-1} \frac{d^{n-1} y}{dx^{n-1}} + \cdots + a_0 y.$$

Then substituting $y = x^\lambda$, we get

$$L(x^\lambda) = \big(\lambda(\lambda-1)\cdots(\lambda-n+1) + a_{n-1}\lambda(\lambda-1)\cdots(\lambda-n+2) + \cdots + a_0\big)x^\lambda.$$

So, $x^\lambda$ is a solution of (8.6.4), if and only if

$$\lambda(\lambda-1)\cdots(\lambda-n+1) + a_{n-1}\lambda(\lambda-1)\cdots(\lambda-n+2) + \cdots + a_0 = 0. \tag{8.6.5}$$

Essentially, for finding the solutions of (8.6.4), we need to find the roots of (8.6.5), which is a polynomial in $\lambda$. With the above understanding, solve the following homogeneous Euler equations:

(a) $x^3y''' + 3x^2y'' + 2xy' = 0.$

(b) $x^3y''' - 6x^2y'' + 11xy' - 6y = 0.$

(c) $x^3y''' - x^2y'' + xy' - y = 0.$

For an alternative method of solving (8.6.4), see the next exercise.

5. Consider the Euler equation (8.6.4) with $x > 0$ and $x \in I$. Let $x = e^t$ or equivalently $t = \ln x$. Let $D = \frac{d}{dt}$ and $d = \frac{d}{dx}$. Then

   (a) show that $xd(y) = Dy(t)$, or equivalently $x\frac{dy}{dx} = \frac{dy)}{dt}$.

   (b) using mathematical induction, show that $x^n d^n y = \big(D(D-1)\cdots(D-n+1)\big)y(t)$.

   (c) with the new (independent) variable $t$, the Euler equation (8.6.4) reduces to an equation with constant coefficients. So, the questions in the above part can be solved by the method just explained.

We turn our attention toward the non-homogeneous equation (8.6.1). If $y_p$ is any solution of (8.6.1) and if $y_h$ is the general solution of the corresponding homogeneous equation (8.6.2), then

$$y = y_h + y_p$$

is a solution of (8.6.1). The solution $y$ involves $n$ arbitrary constants. Such a solution is called the GENERAL SOLUTION of (8.6.1).

Solving an equation of the form (8.6.1) usually means to find a general solution of (8.6.1). The solution $y_p$ is called a PARTICULAR SOLUTION which may not involve any arbitrary constants. Solving (8.6.1) essentially involves two steps (as we had seen in detail in Section 8.3).

Step 1: a) Calculation of the homogeneous solution $y_h$ and

b) Calculation of the particular solution $y_p$.

In the ensuing discussion, we describe the method of undetermined coefficients to determine $y_p$. Note that a particular solution is not unique. In fact, if $y_p$ is a solution of (8.6.1) and $u$ is any solution of (8.6.2), then $y_p + u$ is also a solution of (8.6.1). The undetermined coefficients method is applicable for equations (8.6.1).

## 8.7 Method of Undetermined Coefficients

In the previous section, we have seen than a general solution of

$$L_n(y) = f(x) \quad \text{on} \quad I \tag{8.7.6}$$

can be written in the form

$$y = y_h + y_p,$$

where $y_h$ is a general solution of $L_n(y) = 0$ and $y_p$ is a particular solution of (8.7.6). In view of this, in this section, we shall attempt to obtain $y_p$ for (8.7.6) using the method of undetermined coefficients in the following particular cases of $f(x)$;

1. $f(x) = ke^{\alpha x}$; $k \neq 0, \alpha$ a real constant
2. $f(x) = e^{\alpha x}\big(k_1 \cos(\beta x) + k_2 \sin(\beta x)\big)$; $k_1, k_2, \alpha, \beta \in \mathbb{R}$
3. $f(x) = x^m$.

**Case I.** $f(x) = ke^{\alpha x}$; $k \neq 0, \alpha$ a real constant.
We first assume that $\alpha$ is not a root of the characteristic equation, *i.e.,* $p(\alpha) \neq 0$. Note that $L_n(e^{\alpha x}) = p(\alpha)e^{\alpha x}$. Therefore, let us assume that a particular solution is of the form

$$y_p = Ae^{\alpha x},$$

where $A$, an unknown, is an undetermined coefficient. Thus

$$L_n(y_p) = Ap(\alpha)e^{\alpha x}.$$

Since $p(\alpha) \neq 0$, we can choose $A = \dfrac{k}{p(\alpha)}$ to obtain

$$L_n(y_p) = ke^{\alpha x}.$$

Thus, $y_p = \dfrac{k}{p(\alpha)}e^{\alpha x}$ is a particular solution of $L_n(y) = ke^{\alpha x}$.
**Modification Rule:** If $\alpha$ is a root of the characteristic equation, *i.e.,* $p(\alpha) = 0$, with multiplicity $r$, (*i.e.,* $p(\alpha) = p'(\alpha) = \cdots = p^{(r-1)}(\alpha) = 0$ and $p^{(r)}(\alpha) \neq 0$) then we take, $y_p$ of the form

$$y_p = Ax^r e^{\alpha x}$$

and obtain the value of $A$ by substituting $y_p$ in $L_n(y) = ke^{\alpha x}$.

**Example 8.7.1** 1. Find a particular solution of

$$y'' - 4y = 2e^x.$$

**Solution:** Here $f(x) = 2e^x$ with $k = 2$ and $\alpha = 1$. Also, the characteristic polynomial, $p(\lambda) = \lambda^2 - 4$. Note that $\alpha = 1$ is not a root of $p(\lambda) = 0$. Thus, we assume $y_p = Ae^x$. This on substitution gives

$$Ae^x - 4Ae^x = 2e^x \implies -3Ae^x = 2e^x.$$

So, we choose $A = \dfrac{-2}{3}$, which gives a particular solution as

$$y_p = \frac{-2e^x}{3}.$$

2. Find a particular solution of

$$y''' - 3y'' + 3y' - y = 2e^x.$$

**Solution:** The characteristic polynomial is $p(\lambda) = \lambda^3 - 3\lambda^2 + 3\lambda - 1 = (\lambda - 1)^3$ and $\alpha = 1$. Clearly, $p(1) = 0$ and $\lambda = \alpha = 1$ has multiplicity $r = 3$. Thus, we assume $y_p = Ax^3e^x$. Substituting it in the given equation,we have

$$\begin{aligned} Ae^x\left(x^3 + 9x^2 + 18x + 6\right) &- 3Ae^x\left(x^3 + 6x^2 + 6x\right) \\ &+ 3Ae^x\left(x^3 + 3x^2\right) - Ax^3e^x = 2e^x. \end{aligned}$$

Solving for $A$, we get $A = \dfrac{1}{3}$, and thus a particular solution is $y_p = \dfrac{x^3e^x}{3}$.

3. Find a particular solution of

$$y''' - y' = e^{2x}.$$

**Solution:** The characteristic polynomial is $p(\lambda) = \lambda^3 - \lambda$ and $\alpha = 2$. Thus, using $y_p = Ae^{2x}$, we get $A = \dfrac{1}{p(\alpha)} = \dfrac{1}{6}$, and hence a particular solution is $y_p = \dfrac{e^{2x}}{6}$.

4. Solve $y''' - 3y'' + 3y' - y = 2e^{2x}$.

**Exercise 8.7.2** Find a particular solution for the following differential equations:

1. $y'' - 3y' + 2y = e^x$.

2. $y'' - 9y = e^{3x}$.

3. $y''' - 3y'' + 6y' - 4y = e^{2x}$.

**Case II.** $f(x) = e^{\alpha x}\big(k_1 \cos(\beta x) + k_2 \sin(\beta x)\big);\;\; k_1, k_2, \alpha, \beta \in \mathbb{R}$
We first assume that $\alpha + i\beta$ is not a root of the characteristic equation, *i.e.,* $p(\alpha + i\beta) \neq 0$. Here, we assume that $y_p$ is of the form

$$y_p = e^{\alpha x}\big(A\cos(\beta x) + B\sin(\beta x)\big),$$

and then comparing the coefficients of $e^{\alpha x}\cos x$ and $e^{\alpha x}\sin x$ (why!) in $L_n(y) = f(x)$, obtain the values of $A$ and $B$.
**Modification Rule:** If $\alpha + i\beta$ is a root of the characteristic equation, *i.e.,* $p(\alpha + i\beta) = 0$, with multiplicity $r$, then we assume a particular solution as

$$y_p = x^r e^{\alpha x}\big(A\cos(\beta x) + B\sin(\beta x)\big),$$

and then comparing the coefficients in $L_n(y) = f(x)$, obtain the values of $A$ and $B$.

**Example 8.7.3** 1. Find a particular solution of

$$y'' + 2y' + 2y = 4e^x \sin x.$$

**Solution:** Here, $\alpha = 1$ and $\beta = 1$. Thus $\alpha + i\beta = 1 + i$, which is not a root of the characteristic equation $p(\lambda) = \lambda^2 + 2\lambda + 2 = 0$. Note that the roots of $p(\lambda) = 0$ are $-1 \pm i$.

Thus, let us assume $y_p = e^x\,(A\sin x + B\cos x)$. This gives us

$$(-4B + 4A)e^x \sin x + (4B + 4A)e^x \cos x = 4e^x \sin x.$$

Comparing the coefficients of $e^x \cos x$ and $e^x \sin x$ on both sides, we get $A - B = 1$ and $A + B = 0$. On solving for $A$ and $B$, we get $A = -B = \dfrac{1}{2}$. So, a particular solution is $y_p = \dfrac{e^x}{2}\,(\sin x - \cos x)$.

2. Find a particular solution of

$$y'' + y = \sin x.$$

**Solution:** Here, $\alpha = 0$ and $\beta = 1$. Thus $\alpha + i\beta = i$, which is a root with multiplicity $r = 1$, of the characteristic equation $p(\lambda) = \lambda^2 + 1 = 0$.

So, let $y_p = x\,(A\cos x + B\sin x)$. Substituting this in the given equation and comparing the coefficients of $\cos x$ and $\sin x$ on both sides, we get $B = 0$ and $A = -\dfrac{1}{2}$. Thus, a particular solution is $y_p = \dfrac{-1}{2} x\cos x$.

**Exercise 8.7.4** Find a particular solution for the following differential equations:

1. $y''' - y'' + y' - y = e^x \cos x$.
2. $y'''' + 2y'' + y = \sin x$.
3. $y'' - 2y' + 2y = e^x \cos x$.

**Case III.** $f(x) = x^m$.
Suppose $p(0) \neq 0$. Then we assume that

$$y_p = A_m x^m + A_{m-1} x^{m-1} + \cdots + A_0$$

and then compare the coefficient of $x^k$ in $L_n(y_p) = f(x)$ to obtain the values of $A_i$ for $0 \leq i \leq m$.
**Modification Rule:** If $\lambda = 0$ is a root of the characteristic equation, *i.e.,* $p(0) = 0$, with multiplicity $r$, then we assume a particular solution as

$$y_p = x^r \left( A_m x^m + A_{m-1} x^{m-1} + \cdots + A_0 \right)$$

and then compare the coefficient of $x^k$ in $L_n(y_p) = f(x)$ to obtain the values of $A_i$ for $0 \leq i \leq m$.

**Example 8.7.5** Find a particular solution of

$$y''' - y'' + y' - y = x^2.$$

**Solution:** As $p(0) \neq 0$, we assume

$$y_p = A_2 x^2 + A_1 x + A_0$$

which on substitution in the given differential equation gives

$$-2A_2 + (2A_2 x + A_1) - (A_2 x^2 + A_1 x + A_0) = x^2.$$

Comparing the coefficients of different powers of $x$ and solving, we get $A_2 = -1,\ A_1 = -2$ and $A_0 = 0$. Thus, a particular solution is

$$y_p = -(x^2 + 2x).$$

Finally, note that if $y_{p_1}$ is a particular solution of $L_n(y) = f_1(x)$ and $y_{p_2}$ is a particular solution of $L_n(y) = f_2(x)$, then a particular solution of

$$L_n(y) = k_1 f_1(x) + k_2 f_2(x)$$

is given by

$$y_p = k_1 y_{p_1} + k_2 y_{p_2}.$$

In view of this, one can use method of undetermined coefficients for the cases, where $f(x)$ is a linear combination of the functions described above.

**Example 8.7.6** Find a particular soltution of

$$y'' + y = 2\sin x + \sin 2x.$$

**Solution:** We can divide the problem into two problems:

1. $y'' + y = 2\sin x$.

2. $y'' + y = \sin 2x$.

For the first problem, a particular solution (Example 8.7.3.2) is $y_{p_1} = 2\,\frac{-1}{2} x \cos x = -x \cos x$.

For the second problem, one can check that $y_{p_2} = \frac{-1}{3} \sin(2x)$ is a particular solution.
Thus, a particular solution of the given problem is

$$y_{p_1} + y_{p_2} = -x \cos x - \frac{1}{3} \sin(2x).$$

**Exercise 8.7.7** Find a particular solution for the following differential equations:

1. $y''' - y'' + y' - y = 5e^x \cos x + 10e^{2x}$.
2. $y'' + 2y' + y = x + e^{-x}$.
3. $y'' + 3y' - 4y = 4e^x + e^{4x}$.
4. $y'' + 9y = \cos x + x^2 + x^3$.
5. $y''' - 3y'' + 4y' = x^2 + e^{2x} \sin x$.
6. $y'''' + 4y''' + 6y'' + 4y' + 5y = 2 \sin x + x^2$.

# Chapter 9

# Solutions Based on Power Series

## 9.1 Introduction

In the previous chapter, we had a discussion on the methods of solving

$$y'' + ay' + by = f(x);$$

where $a, b$ were real numbers and $f$ was a real valued continuous function. We also looked at Euler Equations which can be reduced to the above form. The natural question is:
what if $a$ and $b$ are functions of $x$?

In this chapter, we have a partial answer to the above question. In general, there are no methods of finding a solution of an equation of the form

$$y'' + q(x)y' + r(x)y = f(x), \quad x \in I$$

where $q(x)$ and $r(x)$ are real valued continuous functions defined on an interval $I \subset \mathbb{R}$. In such a situation, we look for a class of functions $q(x)$ and $r(x)$ for which we may be able to solve. One such class of functions is called the set of analytic functions.

**Definition 9.1.1 (Power Series)** Let $x_0 \in \mathbb{R}$ and $a_0, a_1, \ldots, a_n, \ldots \in \mathbb{R}$ be fixed. An expression of the type

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n \tag{9.1.1}$$

is called a power series in $x$ around $x_0$. The point $x_0$ is called the center, and $a_n$'s are called the coefficients.

In short, $a_0, a_1, \ldots, a_n, \ldots$ are called the coefficient of the power series and $x_0$ is called the center. Note here that $a_n \in \mathbb{R}$ is the coefficient of $(x - x_0)^n$ and that the power series converges for $x = x_0$. So, the set

$$S = \{x \in \mathbb{R} : \sum_{n=0}^{\infty} a_n (x - x_0)^n \text{ converges}\}$$

is a non-empty. It turns out that the set $S$ is an interval in $\mathbb{R}$. We are thus led to the following definition.

**Example 9.1.2** 1. Consider the power series

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots .$$

In this case, $x_0 = 0$ is the center, $a_0 = 0$ and $a_{2n} = 0$ for $n \geq 1$. Also, $a_{2n+1} = \dfrac{(-1)^n}{(2n+1)!}$, $n = 1, 2, \ldots$. Recall that the Taylor series expansion around $x_0 = 0$ of $\sin x$ is same as the above power series.

2. Any polynomial
$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$
is a power series with $x_0 = 0$ as the center, and the coefficients $a_m = 0$ for $m \geq n+1$.

**Definition 9.1.3 (Radius of Convergence)** A real number $R \geq 0$ is called the radius of convergence of the power series (9.1.1), if the expression (9.1.1) converges for all $x$ satisfying

$$|x - x_0| < R \quad \text{and } R \text{ is the largest such number.}$$

From what has been said earlier, it is clear that the set of points $x$ where the power series (9.1.1) is convergent is the interval $(-R + x_0,\ x_0 + R)$, whenever $R$ is the radius of convergence. If $R = 0$, the power series is convergent only at $x = x_0$.

Let $R > 0$ be the radius of convergence of the power series (9.1.1). Let $I = (-R + x_0,\ x_0 + R)$. In the interval $I$, the power series (9.1.1) converges. Hence, it defines a real valued function and we denote it by $f(x)$, *i.e.*,
$$f(x) = \sum_{n=1}^{\infty} a_n(x - x_0)^n,\ x \in I.$$
Such a function is well defined as long as $x \in I$. $f$ is called the function defined by the power series (9.1.1) on $I$. Sometimes, we also use the terminology that (9.1.1) induces a function $f$ on $I$.

It is a natural question to ask how to find the radius of convergence of a power series (9.1.1). We state one such result below but we do not intend to give a proof.

**Theorem 9.1.4** 1. Let $\sum\limits_{n=1}^{\infty} a_n(x-x_0)^n$ be a power series with center $x_0$. Then there exists a real number $R \geq 0$ such that

$$\sum_{n=1}^{\infty} a_n(x - x_0)^n \quad \text{converges for all } \ x \in (-R + x_0, x_0 + R).$$

In this case, the power series $\sum\limits_{n=1}^{\infty} a_n(x - x_0)^n$ converges absolutely and uniformly on

$$|x - x_0| \leq r \ \text{ for all } \ r < R$$

and diverges for all $x$ with
$$|x - x_0| > R.$$

2. Suppose $R$ is the radius of convergence of the power series (9.1.1). Suppose $\lim\limits_{n\longrightarrow\infty} \sqrt[n]{|a_n|}$ exists and equals $\ell$.

   (a) If $\ell \neq 0$, then $R = \dfrac{1}{\ell}$.

   (b) If $\ell = 0$, then the power series (9.1.1) converges for all $x \in \mathbb{R}$.

   Note that $\lim\limits_{n\longrightarrow\infty} \sqrt[n]{|a_n|}$ exists if $\lim\limits_{n\longrightarrow\infty} \left|\dfrac{a_{n+1}}{a_n}\right|$ and
$$\lim_{n\longrightarrow\infty} \sqrt[n]{|a_n|} = \lim_{n\longrightarrow\infty} \left|\frac{a_{n+1}}{a_n}\right|.$$

**Remark 9.1.5** *If the reader is familiar with the concept of* lim sup *of a sequence, then we have a modification of the above theorem.*

*In case, $\sqrt[n]{|a_n|}$ does not tend to a limit as $n \longrightarrow \infty$, then the above theorem holds if we replace* $\lim\limits_{n\longrightarrow\infty} \sqrt[n]{|a_n|}$ *by* $\limsup\limits_{n\longrightarrow\infty} \sqrt[n]{|a_n|}$.

**Example 9.1.6** 1. Consider the power series $\sum\limits_{n=0}^{\infty}(x+1)^n$. Here $x_0 = -1$ is the center and $a_n = 1$ for all $n \geq 0$. So, $\sqrt[n]{|a_n|} = \sqrt[n]{1} = 1$. Hence, by Theorem 9.1.4, the radius of convergence$R = 1$.

2. Consider the power series $\sum\limits_{n \geq 0} \dfrac{(-1)^n (x+1)^{2n+1}}{(2n+1)!}$. In this case, the center is

$$x_0 = -1, \ a_n = 0 \ \text{ for } n \text{ even and } a_{2n+1} = \frac{(-1)^n}{(2n+1)!}.$$

So,

$$\lim_{n \longrightarrow \infty} \sqrt[2n+1]{|a_{2n+1}|} = 0 \ \text{ and } \ \lim_{n \longrightarrow \infty} \sqrt[2n]{|a_{2n}|} = 0.$$

Thus, $\lim\limits_{n \longrightarrow \infty} \sqrt[n]{|a_n|}$ exists and equals $0$. Therefore, the power series converges for all $x \in \mathbb{R}$. Note that the series converges to $\sin(x+1)$.

3. Consider the power series $\sum\limits_{n=1}^{\infty} x^{2n}$. In this case, we have

$$a_{2n} = 1 \ \text{ and } a_{2n+1} = 0 \ \text{ for } n = 0, 1, 2, \ldots.$$

So,

$$\lim_{n \longrightarrow \infty} \sqrt[2n+1]{|a_{2n+1}|} = 0 \ \text{ and } \ \lim_{n \longrightarrow \infty} \sqrt[2n]{|a_{2n}|} = 1.$$

Thus, $\lim\limits_{n \longrightarrow \infty} \sqrt[n]{|a_n|}$ does not exist.

We let $u = x^2$. Then the power series $\sum\limits_{n=1}^{\infty} x^{2n}$ reduces to $\sum\limits_{n=1}^{\infty} u^n$. But then from Example 9.1.6.1, we learned that $\sum\limits_{n=1}^{\infty} u^n$ converges for all $u$ with $|u| < 1$. Therefore, the original power series converges whenever $|x^2| < 1$ or equivalently whenever $|x| < 1$. So, the radius of convergence is $R = 1$. Note that

$$\frac{1}{1 - x^2} = \sum_{n=1}^{\infty} x^{2n} \quad \text{for} \quad |x| < 1.$$

4. Consider the power series $\sum\limits_{n \geq 0} n^n x^n$. In this case, $\sqrt[n]{|a_n|} = \sqrt[n]{n^n} = n$. doesn't have any finite limit as $n \longrightarrow \infty$. Hence, the power series converges only for $x = 0$.

5. The power series $\sum\limits_{n \geq 0} \dfrac{x^n}{n!}$ has coefficients $a_n = \dfrac{1}{n!}$ and it is easily seen that $\lim\limits_{n \longrightarrow \infty} \left| \dfrac{1}{n!} \right|^{\frac{1}{n}} = 0$ and the power series converges for all $x \in \mathbb{R}$. Recall that it represents $e^x$.

**Definition 9.1.7** Let $f : I \longrightarrow \mathbb{R}$ be a function and $x_0 \in I$. $f$ is called analytic around $x_0$ if there exists a $\delta > 0$ such that

$$f(x) = \sum_{n \geq 0} a_n (x - x_0)^n \quad \text{for every } x \text{ with } |x - x_0| < \delta.$$

That is, $f$ has a power series representation in a neighbourhood of $x_0$.

### 9.1.1 Properties of Power Series

Now we quickly state some of the important properties of the power series. Consider two power series

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n \quad \text{and} \quad \sum_{n=0}^{\infty} b_n (x - x_0)^n$$

with radius of convergence $R_1 > 0$ and $R_2 > 0$, respectively. Let $F(x)$ and $G(x)$ be the functions defined by the two power series defined for all $x \in I$, where $I = (-R + x_0, x_0 + R)$ with $R = \min\{R_1, R_2\}$. Note that both the power series converge for all $x \in I$.

With $F(x)$, $G(x)$ and $I$ as defined above, we have the following properties of the power series.

1. Equality of Power Series

   The two power series defined by $F(x)$ and $G(x)$ are equal for all $x \in I$ if and only if
   $$a_n = b_n \text{ for all } n = 0, 1, 2, \ldots.$$
   In particular, if $\sum\limits_{n=0}^{\infty} a_n (x - x_0)^n = 0$ for all $x \in I$, then
   $$a_n = 0 \text{ for all } n = 0, 1, 2, \ldots.$$

2. Term by Term Addition

   For all $x \in I$, we have
   $$F(x) + G(x) = \sum_{n=0}^{\infty} (a_n + b_n)(x - x_0)^n$$
   Essentially, it says that in the common part of the regions of convergence, the two power series can be added term by term.

3. Multiplication of Power Series

   Let us define
   $$c_0 = a_0 b_0, \text{ and inductively } c_n = \sum_{j=1}^{n} a_{n-j} b_j.$$
   Then for all $x \in I$, the product of $F(x)$ and $G(x)$ is defined by
   $$H(x) = F(x)G(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^n.$$
   $H(x)$ is called the "Cauchy Product" of $F(x)$ and $G(x)$.

   Note that for any $n \geq o$, the coefficient of $x^n$ in
   $$\left( \sum_{j=0}^{\infty} a_j (x - x_0)^j \right) \cdot \left( \sum_{k=0}^{\infty} b_k (x - x_0)^k \right) \quad \text{is} \quad c_n = \sum_{j=1}^{n} a_{n-j} b_j.$$

4. Term by Term Differentiation

   The term by term differentiation of the power series function $F(x)$ is
   $$\sum_{n=1}^{\infty} n a_n (x - x_0)^n.$$
   Note that it also has $R_1$ as the radius of convergence as by Theorem 9.1.4 $\lim\limits_{n \longrightarrow \infty} \sqrt[n]{|a_n|} = \cdot \frac{1}{R_1}$ and
   $$\lim_{n \longrightarrow \infty} \sqrt[n]{|n a_n|} = \lim_{n \longrightarrow \infty} \sqrt[n]{|n|} \lim_{n \longrightarrow \infty} \sqrt[n]{|a_n|} = 1 \cdot \frac{1}{R_1}.$$
   Let $0 < r < R_1$. Then for all $x \in (-r + x_0, x_0 + r)$, we have
   $$\frac{d}{dx} F(x) = F'(x) = \sum_{n=1}^{\infty} n a_n (x - x_0)^n.$$
   In other words, inside the region of convergence, the power series can be differentiated term by term.

In the following, we shall consider power series with $x_0 = 0$ as the center. Note that by a transformation of $X = x - x_0$, the center of the power series can be shifted to the origin.

**Exercise 9.1.1** 1. which of the following represents a power series (with center $x_0$ indicated in the brackets) in $x$?

(a) $1 + x^2 + x^4 + \cdots + x^{2n} + \cdots \qquad (x_0 = 0).$

(b) $1 + \sin x + (\sin x)^2 + \cdots + (\sin x)^n + \cdots \qquad (x_0 = 0).$

(c) $1 + x|x| + x^2|x^2| + \cdots + x^n|x^n| + \cdots \qquad (x_0 = 0).$

2. Let $f(x)$ and $g(x)$ be two power series around $x_0 = 0$, defined by

$$\begin{aligned} f(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \cdots \\ \text{and} \quad g(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots . \end{aligned}$$

Find the radius of convergence of $f(x)$ and $g(x)$. Also, for each $x$ in the domain of convergence, show that

$$f'(x) = g(x) \quad \text{and} \quad g'(x) = -f(x).$$

*[Hint: Use Properties* $1, 2, 3$ *and* $4$ *mentioned above. Also, note that we usually call* $f(x)$ *by* $\sin x$ *and* $g(x)$ *by* $\cos x$.*]*

3. Find the radius of convergence of the following series centerd at $x_0 = -1$.

(a) $1 + (x+1) + \frac{(x+1)^2}{2!} + \cdots + \frac{(x+1)^n}{n!} + \cdots .$

(b) $1 + (x+1) + 2(x+1)^2 + \cdots + n(x+1)^n + \cdots .$

## 9.2 Solutions in terms of Power Series

Consider a linear second order equation of the type

$$y'' + a(x)y' + b(x)y = 0. \tag{9.2.1}$$

Let $a$ and $b$ be analytic around the point $x_0 = 0$. In such a case, we may hope to have a solution $y$ in terms of a power series, say

$$y = \sum_{k=0}^{\infty} c_k x^k. \tag{9.2.2}$$

In the absence of any information, let us assume that (9.2.1) has a solution $y$ represented by (9.2.2). We substitute (9.2.2) in Equation (9.2.1) and try to find the values of $c_k$'s. Let us take up an example for illustration.

**Example 9.2.1** Consider the differential equation

$$y'' + y = 0 \tag{9.2.3}$$

Here $a(x) \equiv 0,\ b(x) \equiv 1$, which are analytic around $x_0 = 0$.
**Solution:** Let

$$y = \sum_{n=0}^{\infty} c_n x^n. \tag{9.2.4}$$

Then $y' = \sum\limits_{n=0}^{\infty} nc_n x^{n-1}$ and $y'' = \sum\limits_{n=0}^{\infty} n(n-1)c_n x^{n-2}$. Substituting the expression for $y,\ y'$ and $y''$ in Equation (9.2.3), we get

$$\sum_{n=0}^{\infty} n(n-1)c_n x^{n-2} + \sum_{n=0}^{\infty} c_n x^n = 0$$

or, equivalently

$$0 = \sum_{n=0}^{\infty}(n+2)(n+1)c_{n+2}x^n + \sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty}\{(n+1)(n+2)c_{n+2} + c_n\}x^n.$$

Hence for all $n = 0, 1, 2, \ldots,$

$$(n+1)(n+2)c_{n+2} + c_n = 0 \text{ or } c_{n+2} = -\frac{c_n}{(n+1)(n+2)}.$$

Therefore, we have

$$\begin{array}{ll}
c_2 = -\frac{c_0}{2!}, & c_3 = -\frac{c_1}{3!}, \\
c_4 = (-1)^2\frac{c_0}{4!}, & c_5 = (-1)^2\frac{c_1}{5!}, \\
\vdots & \vdots \\
c_{2n} = (-1)^n\frac{c_0}{(2n)!}, & c_{2n+1} = (-1)^n\frac{c_1}{(2n+1)!}.
\end{array}$$

Here, $c_0$ and $c_1$ are arbitrary. So,

$$y = c_0\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{(2n)!} + c_1\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{(2n+1)!}$$

or $y = c_0\cos(x) + c_1\sin(x)$ where $c_0$ and $c_1$ can be chosen arbitrarily. For $c_0 = 1$ and $c_1 = 0$, we get $y = \cos(x)$. That is, $\cos(x)$ is a solution of the Equation (9.2.3). Similarly, $y = \sin(x)$ is also a solution of Equation (9.2.3).

**Exercise 9.2.2** Assuming that the solutions $y$ of the following differential equations admit power series representation, find $y$ in terms of a power series.

1. $y' = -y$, (center at $x_0 = 0$).
2. $y' = 1 + y^2$, (center at $x_0 = 0$).
3. Find two linearly independent solutions of
   (a) $y'' - y = 0$, (center at $x_0 = 0$).
   (b) $y'' + 4y = 0$, (center at $x_0 = 0$).

## 9.3 Statement of Frobenius Theorem for Regular (Ordinary) Point

Earlier, we saw a few properties of a power series and some uses also. Presently, we inquire the question, namely, whether an equation of the form

$$y'' + a(x)y' + b(x)y = f(x), \quad x \in I \tag{9.3.1}$$

admits a solution $y$ which has a power series representation around $x \in I$. In other words, we are interested in looking into an existence of a power series solution of (9.3.1) under certain conditions on $a(x), b(x)$ and $f(x)$. The following is one such result. We omit its proof.

**Theorem 9.3.1** Let $a(x), b(x)$ and $f(x)$ admit a power series representation around a point $x = x_0 \in I$, with non-zero radius of convergence $r_1, r_2$ and $r_3$, respectively. Let $R = \min\{r_1, r_2, r_3\}$. Then the Equation (9.3.1) has a solution $y$ which has a power series representation around $x_0$ with radius of convergence $R$.

**Remark 9.3.2** *We remind the readers that Theorem 9.3.1 is true for Equations (9.3.1), whenever the coefficient of $y''$ is* 1.

*Secondly, a point $x_0$ is called an* ORDINARY POINT *for (9.3.1) if $a(x), b(x)$ and $f(x)$ admit power series expansion (with non-zero radius of convergence) around $x = x_0$. $x_0$ is called a* SINGULAR POINT *for (9.3.1) if $x_0$ is not an ordinary point for (9.3.1).*

The following are some examples for illustration of the utility of Theorem 9.3.1.

**Exercise 9.3.3** 1. Examine whether the given point $x_0$ is an ordinary point or a singular point for the following differential equations.

(a) $(x-1)y'' + \sin xy = 0, \ x_0 = 0.$

(b) $y'' + \frac{\sin x}{x-1} y = 0, \ x_0 = 0.$

(c) Find two linearly independent solutions of

(d) $(1-x^2)y'' - 2xy' + n(n+1)y = 0, \ x_0 = 0, \ n$ is a real constant.

2. Show that the following equations admit power series solutions around a given $x_0$. Also, find the power series solutions if it exists.

(a) $y'' + y = 0, \ x_0 = 0.$

(b) $xy'' + y = 0, \ x_0 = 0.$

(c) $y'' + 9y = 0, \ x_0 = 0.$

## 9.4 Legendre Equations and Legendre Polynomials

### 9.4.1 Introduction

Legendre Equation plays a vital role in many problems of mathematical Physics and in the theory of quadratures (as applied to Numerical Integration).

**Definition 9.4.1** The equation

$$(1-x^2)y'' - 2xy' + p(p+1)y = 0, \ -1 < x < 1 \tag{9.4.1}$$

where $p \in \mathbb{R}$, is called a LEGENDRE EQUATION of order $p$.

Equation (9.4.1) was studied by Legendre and hence the name Legendre Equation.

Equation (9.4.1) may be rewritten as

$$y'' - \frac{2x}{(1-x^2)} y' + \frac{p(p+1)}{(1-x^2)} y = 0.$$

The functions $\frac{2x}{1-x^2}$ and $\frac{p(p+1)}{1-x^2}$ are analytic around $x_0 = 0$ (since they have power series expressions with center at $x_0 = 0$ and with $R = 1$ as the radius of convergence). By Theorem 9.3.1, a solution $y$ of (9.4.1) admits a power series solution (with center at $x_0 = 0$) with radius of convergence $R = 1$. Let us assume that $y = \sum\limits_{k=0}^{\infty} a_k x^k$ is a solution of (9.4.1). We have to find the value of $a_k$'s. Substituting the expression for

$$y' = \sum_{k=0}^{\infty} k a_k x^{k-1} \text{ and } y'' = \sum_{k=0}^{\infty} k(k-1) a_k x^{k-2}$$

in Equation (9.4.1), we get

$$\sum_{k=0}^{\infty} \left\{(k+1)(k+2)a_{k+2} + a_k(p-k)(p+k+1)\right\} x^k = 0.$$

Hence, for $k = 0, 1, 2, \ldots$

$$a_{k+2} = -\frac{(p-k)(p+k+1)}{(k+1)(k+2)} a_k.$$

It now follows that

$$\begin{aligned} a_2 &= -\tfrac{p(p+1)}{2!} a_0, & a_3 &= -\tfrac{(p-1)(p+2)}{3!} a_1, \\ a_4 &= -\tfrac{(p-2)(p+3)}{3 \cdot 4} a_2 & a_5 &= (-1)^2 \tfrac{(p-1)(p-3)(p+2)(p+4)}{5!} a_1 \\ &= (-1)^2 \tfrac{p(p-2)(p+1)(p+3)}{4!} a_0, \end{aligned}$$

etc. In general,

$$a_{2m} = (-1)^m \frac{p(p-2)\cdots(p-2m+2)(p+1)(p+3)\cdots(p+2m-1)}{(2m)!} a_0$$

and

$$a_{2m+1} = (-1)^m \frac{(p-1)(p-3)\cdots(p-2m+1)(p+2)(p+4)\cdots(p+2m)}{(2m+1)!} a_1.$$

It turns out that both $a_0$ and $a_1$ are arbitrary. So, by choosing $a_0 = 1, a_1 = 0$ and $a_0 = 0, a_1 = 1$ in the above expressions, we have the following two solutions of the Legendre Equation (9.4.1), namely,

$$y_1 = 1 - \frac{p(p+1)}{2!} x^2 + \cdots + (-1)^m \frac{(p-2m+2)\cdots(p+2m-1)}{(2m)!} x^{2m} + \cdots \tag{9.4.2}$$

and

$$y_2 = x - \frac{(p-1)(p+2)}{3!} x^3 + \cdots + (-1)^m \frac{(p-2m+1)\cdots(p+2m)}{(2m+1)!} x^{2m+1} + \cdots . \tag{9.4.3}$$

**Remark 9.4.2** *$y_1$ and $y_2$ are two linearly independent solutions of the Legendre Equation (9.4.1). It now follows that the general solution of (9.4.1) is*

$$y = c_1 y_1 + c_2 y_2 \tag{9.4.4}$$

*where $c_1$ and $c_2$ are arbitrary real numbers.*

### 9.4.2 Legendre Polynomials

In many problems, the real number $p$, appearing in the Legendre Equation (9.4.1), is a non-negative integer. Suppose $p = n$ is a non-negative integer. Recall

$$a_{k+2} = -\frac{(n-k)(n+k+1)}{(k+1)(k+2)} a_k, \; k = 0, 1, 2, \ldots . \tag{9.4.5}$$

Therefore, when $k = n$, we get

$$a_{n+2} = a_{n+4} = \cdots = a_{n+2m} = \cdots = 0 \text{ for all positive integer } m.$$

Case 1: Let $n$ be a positive even integer. Then $y_1$ in Equation (9.4.2) is a polynomial of degree $n$. In fact, $y_1$ is an even polynomial in the sense that the terms of $y_1$ are even powers of $x$ and hence $y_1(-x) = y_1(x)$.

Case 2: Now, let $n$ be a positive odd integer. Then $y_2(x)$ in Equation (9.4.3) is a polynomial of degree $n$. In this case, $y_2$ is an odd polynomial in the sense that the terms of $y_2$ are odd powers of $x$ and hence $y_2(-x) = -y_2(x)$.

In either case, we have a polynomial solution for Equation (9.4.1).

**Definition 9.4.3** A polynomial solution $P_n(x)$ of (9.4.1) is called a LEGENDRE POLYNOMIAL whenever $P_n(1) = 1$.

Fix a positive integer $n$ and consider $P_n(x) = a_0 + a_1x + \cdots + a_nx^n$. Then it can be checked that $P_n(1) = 1$ if we choose

$$a_n = \frac{(2n)!}{2^n(n!)^2} = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{n!}.$$

Using the recurrence relation, we have

$$a_{n-2} = -\frac{(n-1)n}{2(2n-1)}a_n = -\frac{(2n-2)!}{2^n(n-1)!(n-2)!}$$

by the choice of $a_n$. In general, if $n - 2m \geq 0$, then

$$a_{n-2m} = (-1)^m \frac{(2n-2m)!}{2^n m!(n-m)!(n-2m)!}.$$

Hence,

$$\sum_{m=0}^{M} (-1)^m \frac{(2n-2m)!}{2^n m!(n-m)!(n-2m)!} x^{n-2m}, \tag{9.4.6}$$

where $M = \frac{n}{2}$ when $n$ is even and $M = \frac{n-1}{2}$ when $n$ is odd.

**Proposition 9.4.4** Let $p = n$ be a non-negative even integer. Then any polynomial solution $y$ of (9.4.1) which has only even powers of $x$ is a multiple of $P_n(x)$.

Similarly, if $p = n$ is a non-negative odd integer, then any polynomial solution $y$ of (9.4.1) which has only odd powers of $x$ is a multiple of $P_n(x)$.

PROOF. Suppose that $n$ is a non-negative even integer. Let $y$ be a polynomial solution of (9.4.1). By (9.4.4)

$$y = c_1y_1 + c_2y_2,$$

where $y_1$ is a polynomial of degree $n$ (with even powers of $x$) and $y_2$ is a power series solution with odd powers only. Since $y$ is a polynomial, we have $c_2 = 0$ or $y = c_1y_1$ with $c_1 \neq 0$.
Similarly, $P_n(x) = c_1'y_1$ with $c_1' \neq 0$. which implies that $y$ is a multiple of $P_n(x)$. A similar proof holds when $n$ is an odd positive integer. □

We have an alternate way of evaluating $P_n(x)$. They are used later for the orthogonality properties of the Legendre polynomials, $P_n(x)$'s.

**Theorem 9.4.5 (Rodriguės Formula)** The Legendre polynomials $P_n(x)$ for $n = 1, 2, \ldots,$ are given by

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2-1)^n. \tag{9.4.7}$$

PROOF. Let $V(x) = (x^2-1)^n$. Then $\frac{d}{dx}V(x) = 2nx(x^2-1)^{n-1}$ or

$$(x^2-1)\frac{d}{dx}V(x) = 2nx(x^2-1)^n = 2nxV(x).$$

Now differentiating $(n+1)$ times (by the use of the Leibniz rule for differentiation), we get

$$\begin{aligned}(x^2-1)\frac{d^{n+2}}{dx^{n+2}}V(x) &+ 2(n+1)x\frac{d^{n+1}}{dx^{n+1}}V(x) + \frac{2n(n+1)}{1\cdot 2}\frac{d^n}{dx^n}V(x) \\ &- 2nx\frac{d^{n+1}}{dx^{n+1}}V(x) - 2n(n+1)\frac{d^n}{dx^n}V(x) = 0.\end{aligned}$$

By denoting, $U(x) = \frac{d^n}{dx^n}V(x)$, we have

$$\begin{aligned}(x^2-1)U'' + U'\{2(n+1)x - 2nx\} + U\{n(n+1) - 2n(n+1)\} &= 0 \\ \text{or} \quad (1-x^2)U'' - 2xU' + n(n+1)U &= 0.\end{aligned}$$

This tells us that $U(x)$ is a solution of the Legendre Equation (9.4.1). So, by Proposition 9.4.4, we have

$$P_n(x) = \alpha U(x) = \alpha\frac{d^n}{dx^n}(x^2-1)^n \quad \text{for some} \quad \alpha \in \mathbb{R}.$$

Also, let us note that

$$\begin{aligned}\frac{d^n}{dx^n}(x^2-1)^n &= \frac{d^n}{dx^n}\{(x-1)(x+1)\}^n \\ &= n!(x+1)^n + \text{ terms containing a factor of } (x-1).\end{aligned}$$

Therefore,

$$\left.\frac{d^n}{dx^n}(x^2-1)^n\right|_{x=1} = 2^n n! \quad \text{or, equivalently}$$

$$\left.\frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n\right|_{x=1} = 1$$

and thus

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n.$$

□

**Example 9.4.6** 1. When $n=0$, $P_0(x) = 1$.

2. When $n=1$, $P_1(x) = \frac{1}{2}\frac{d}{dx}(x^2-1) = x$.

3. When $n=2$, $P_2(x) = \frac{1}{2^2 2!}\frac{d^2}{dx^2}(x^2-1)^2 = \frac{1}{8}\{12x^2-4\} = \frac{3}{2}x^2 - \frac{1}{2}$.

One may observe that the Rodriguės formula is very useful in the computation of $P_n(x)$ for "small" values of $n$.

**Theorem 9.4.7** Let $P_n(x)$ denote, as usual, the Legendre Polynomial of degree $n$. Then

$$\int_{-1}^{1} P_n(x)P_m(x)\,dx = 0 \text{ if } m \neq n. \tag{9.4.8}$$

Proof. We know that the polynomials $P_n(x)$ and $P_m(x)$ satisfy

$$\big((1-x^2)P_n'(x)\big)' + n(n+1)P_n(x) = 0 \text{ and} \tag{9.4.9}$$

$$\big((1-x^2)P_m'(x)\big)' + m(m+1)P_m(x) = 0. \tag{9.4.10}$$

Multiplying Equation (9.4.9) by $P_m(x)$ and Equation (9.4.10) by $P_n(x)$ and subtracting, we get

$$\big(n(n+1) - m(m+1)\big)P_n(x)P_m(x) = \big((1-x^2)P_m'(x)\big)'P_n(x) - \big((1-x^2)P_n'(x)\big)'P_m(x).$$

Therefore,

$$\begin{aligned}\big(n(n+1) &- m(m+1)\big)\int_{-1}^{1} P_n(x)P_m(x)dx \\ &= \int_{-1}^{1}\Big(\big((1-x^2)P_m'(x)\big)'P_n(x) - \big((1-x^2)P_n'(x)\big)'P_m(x)\Big)dx \\ &= -\int_{-1}^{1}(1-x^2)P_m'(x)P_n'(x)dx + (1-x^2)P_m'(x)P_n(x)\Big|_{x=-1}^{x=1} \\ &\quad +\int_{-1}^{1}(1-x^2)P_n'(x)P_m'(x)dx + (1-x^2)P_n'(x)P_m(x)\Big|_{x=-1}^{x=1} \\ &= 0.\end{aligned}$$

Since $n \neq m,\ n(n+1) \neq m(m+1)$ and therefore, we have

$$\int_{-1}^{1} P_n(x)P_m(x)\ dx = 0 \ \text{ if } \ m \neq n.$$

□

**Theorem 9.4.8** For $n = 0, 1, 2, \ldots$

$$\int_{-1}^{1} P_n^2(x)\ dx = \frac{2}{2n+1}. \tag{9.4.11}$$

Proof. Let us write $V(x) = (x^2-1)^n$. By the Rodrigue's formula, we have

$$\int_{-1}^{1} P_n^2(x)\ dx = \int_{-1}^{1}\left(\frac{1}{n!2^n}\right)^2 \frac{d^n}{dx^n}V(x)\frac{d^n}{dx^n}V(x)dx.$$

Let us call $I = \int\limits_{-1}^{1} \frac{d^n}{dx^n}V(x)\frac{d^n}{dx^n}V(x)dx$. Note that for $0 \leq m < n$,

$$\frac{d^m}{dx^m}V(-1) = \frac{d^m}{dx^m}V(1) = 0. \tag{9.4.12}$$

Therefore, integrating $I$ by parts and using (9.4.12) at each step, we get

$$I = \int_{-1}^{1}\frac{d^{2n}}{dx^{2n}}V(x)\cdot(-1)^nV(x)dx = (2n)!\int_{-1}^{1}(1-x^2)^n dx = (2n)!\ 2\int_{0}^{1}(1-x^2)^n dx.$$

Now substitute $x = \cos\theta$ and use the value of the integral $\int\limits_{0}^{\frac{\pi}{2}} \sin^{2n}\theta\ \ d\theta$, to get the required result. □

We now state an important expansion theorem. The proof is beyond the scope of this book.

**Theorem 9.4.9** Let $f(x)$ be a real valued continuous function defined in $[-1, 1]$. Then

$$f(x) = \sum_{n=0}^{\infty} a_n P_n(x), \ \ x \in [-1, 1]$$

where $a_n = \frac{2n+1}{2}\int\limits_{-1}^{1} f(x)P_n(x)dx$.

Legendre polynomials can also be generated by a suitable function. To do that, we state the following result without proof.

**Theorem 9.4.10** Let $P_n(x)$ be the Legendre polynomial of degree $n$. Then

$$\frac{1}{\sqrt{1-2xt+t^2}} = \sum_{n=0}^{\infty} P_n(x)t^n, \quad t \neq 1. \tag{9.4.13}$$

The function $h(t) = \dfrac{1}{\sqrt{1-2xt+t^2}}$ admits a power series expansion in $t$ (for small $t$) and the coefficient of $t^n$ in $P_n(x)$. The function $h(t)$ is called the GENERATING FUNCTION for the Legendre polynomials.

**Exercise 9.4.11** 1. By using the Rodrigue's formula, find $P_0(x), P_1(x)$ and $P_2(x)$.

2. Use the generating function (9.4.13)

   (a) to find $P_0(x), P_1(x)$ and $P_2(x)$.

   (b) to show that $P_n(x)$ is an odd function whenever $n$ is odd and is an even function whenever$n$ is even.

Using the generating function (9.4.13), we can establish the following relations:

$$\begin{aligned}(n+1)P_{n+1}(x) &= (2n+1)\,x\,P_n(x) - n\,P_{n-1}(x) && (9.4.14)\\ nP_n(x) &= xP_n'(x) - P_{n-1}'(x) && (9.4.15)\\ P_{n+1}'(x) &= xP_n'(x) + (n+1)P_n(x). && (9.4.16)\end{aligned}$$

The relations (9.4.14), (9.4.15) and (9.4.16) are called recurrence relations for the Legendre polynomials, $P_n(x)$. The relation (9.4.14) is also known as Bonnet's recurrence relation. We will now give the proof of (9.4.14) using (9.4.13). The readers are required to proof the other two recurrence relations.

Differentiating the generating function (9.4.13) with respect to $t$ (keeping the variable $x$ fixed), we get

$$-\frac{1}{2}(1-2xt+t^2)^{-\frac{3}{2}}(-2x+2t) = \sum_{n=0}^{\infty} nP_n(x)t^{n-1}.$$

Or equivalently,

$$(x-t)(1-2xt+t^2)^{-\frac{1}{2}} = (1-2xt+t^2)\sum_{n=0}^{\infty} nP_n(x)t^{n-1}.$$

We now substitute $\sum\limits_{n=0}^{\infty} P_n(x)t^n$ in the left hand side for $(1-2xt+t^2)^{-\frac{1}{2}}$, to get

$$(x-t)\sum_{n=0}^{\infty} P_n(x)t^n = (1-2xt+t^2)\sum_{n=0}^{\infty} nP_n(x)t^{n-1}.$$

The two sides and power series in $t$ and therefore, comparing the coefficient of $t^n$, we get

$$xP_n(x) - P_{n-1}(x) = (n+1)P_n(x) + (n-1)P_{n-1}(x) - 2n\;x\;P_n(x).$$

This is clearly same as (9.4.14).

To prove (9.4.15), one needs to differentiate the generating function with respect to $x$ (keeping $t$ fixed) and doing a similar simplification. Now, use the relations (9.4.14) and (9.4.15) to get the relation (9.4.16). These relations will be helpful in solving the problems given below.

**Exercise 9.4.12** 1. Find a polynomial solution $y(x)$ of $(1-x^2)y'' - 2xy' + 20y = 0$ such that $y(1) = 10$.

2. Prove the following:

(a) $\int\limits_{-1}^{1} P_m(x)dx = 0$ for all positive integers $m \geq 1$.

(b) $\int\limits_{-1}^{1} x^{2n+1}P_{2m}(x)dx = 0$ whenever $m$ and $n$ are positive integers with $m \neq n$.

(c) $\int\limits_{-1}^{1} x^m P_n(x)dx = 0$ whenever $m$ and $n$ are positive integers with $m < n$.

3. Show that $P_n'(1) = \dfrac{n(n+1)}{2}$ and $P_n'(-1) = (-1)^{n-1}\dfrac{n(n+1)}{2}$.

4. Establish the following recurrence relations.

   (a) $(n+1)P_n(x) = P_{n+1}'(x) - xP_n'(x)$.

   (b) $(1-x^2)P_n'(x) = n\big[P_{n-1}(x) - xP_n(x)\big]$.

# Part III

# Laplace Transform

# Chapter 10

# Laplace Transform

## 10.1 Introduction

In many problems, a function $f(t),\ \ t \in [a,\ b]$ is transformed to another function $F(s)$ through a relation of the type:

$$F(s) = \int_a^b K(t,s) f(t) dt$$

where $K(t,s)$ is a known function. Here, $F(s)$ is called integral transform of $f(t)$. Thus, an integral transform sends a given function $f(t)$ into another function $F(s)$. This transformation of $f(t)$ into $F(s)$ provides a method to tackle a problem more readily. In some cases, it affords solutions to otherwise difficult problems. In view of this, the integral transforms find numerous applications in engineering problems. Laplace transform is a particular case of integral transform (where $f(t)$ is defined on $[0,\infty)$ and $K(s,t) = e^{-st}$). As we will see in the following, application of Laplace transform reduces a linear differential equation with constant coefficients to an algebraic equation, which can be solved by algebraic methods. Thus, it provides a powerful tool to solve differential equations.

It is important to note here that there is some sort of analogy with what we had learnt during the study of logarithms in school. That is, to multiply two numbers, we first calculate their logarithms, add them and then use the table of antilogarithm to get back the original product. In a similar way, we first transform the problem that was posed as a function of $f(t)$ to a problem in $F(s)$, make some calculations and then use the table of inverse Laplace transform to get the solution of the actual problem.

In this chapter, we shall see same properties of Laplace transform and its applications in solving differential equations.

## 10.2 Definitions and Examples

**Definition 10.2.1 (Piece-wise Continuous Function)** 1. A function $f(t)$ is said to be a piece-wise continuous function on a closed interval $[a,b] \subset \mathbb{R}$, if there exists finite number of points $a = t_0 < t_1 < t_2 < \cdots < t_N = b$ such that $f(t)$ is continuous in each of the intervals $(t_{i-1},\ t_i)$ for $1 \le i \le N$ and has finite limits as $t$ approaches the end points, see the Figure 10.1.

2. A function $f(t)$ is said to be a piece-wise continuous function for $t \ge 0$, if $f(t)$ is a piece-wise continuous function on every closed interval $[a,b] \subset [0,\infty)$. For example, see Figure 10.1.

**Definition 10.2.2 (Laplace Transform)** Let $f : [0,\infty) \longrightarrow \mathbb{R}$ and $s \in \mathbb{R}$. Then $F(s)$, for $s \in \mathbb{R}$ is called

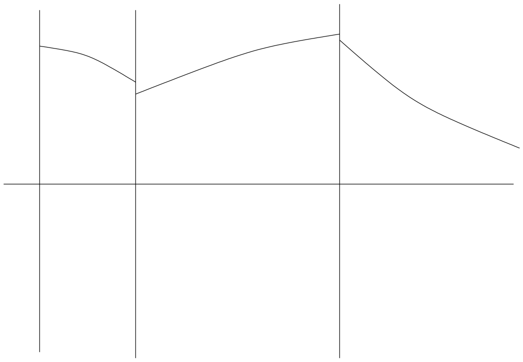


Figure 10.1: Piecewise Continuous Function

the LAPLACE TRANSFORM of $f(t)$, and is defined by

$$\mathcal{L}(f(t)) = F(s) = \int_0^\infty f(t)e^{-st}dt$$

whenever the integral exists.
(Recall that $\int\limits_0^\infty g(t)dt$ exists if $\lim\limits_{b\longrightarrow\infty} \int\limits_0^b g(t)d(t)$ exists and we define $\int\limits_0^\infty g(t)dt = \lim\limits_{b\longrightarrow\infty} \int\limits_0^b g(t)d(t)$.)

**Remark 10.2.3** *1. Let $f(t)$ be an* EXPONENTIALLY BOUNDED *function, i.e.,*

$$|f(t)| \le Me^{\alpha t} \quad \textit{for all} \quad t > 0 \;\; \textit{and for some real numbers} \;\; \alpha \;\; \textit{and} \;\; M \;\; \textit{with} \;\; M > 0.$$

*Then the Laplace transform of $f$ exists.*

*2. Suppose $F(s)$ exists for some function $f$. Then by definition, $\lim\limits_{b\longrightarrow\infty} \int_0^{\mathbf{b}} f(t)e^{-st}dt$ exists. Now, one can use the theory of improper integrals to conclude that*

$$\lim_{s\longrightarrow\infty} F(s) = 0.$$

*Hence, a function $F(s)$ satisfying*

$$\lim_{s\longrightarrow\infty} F(s) \;\; \textit{does not exist or} \;\; \lim_{s\longrightarrow\infty} F(s) \ne 0,$$

*cannot be a Laplace transform of a function $f$.*

**Definition 10.2.4 (Inverse Laplace Transform)** Let $\mathcal{L}(f(t)) = F(s)$. That is, $F(s)$ is the Laplace transform of the function $f(t)$. Then $f(t)$ is called the inverse Laplace transform of $F(s)$. In that case, we write $f(t) = \mathcal{L}^{-1}(F(s))$.

### 10.2.1 Examples

**Example 10.2.5** 1. Find $F(s) = \mathcal{L}(f(t))$, where $f(t) = 1, \;\; t \ge 0$.

**Solution:** $F(s) = \displaystyle\int_0^\infty e^{-st}dt = \lim_{b\longrightarrow\infty} \frac{e^{-st}}{-s}\Big|_0^b = \frac{1}{s} - \lim_{b\longrightarrow\infty} \frac{e^{-sb}}{s}.$
Note that if $s > 0$, then

$$\lim_{b\longrightarrow\infty} \frac{e^{-sb}}{s} = 0.$$

Thus,

$$F(s) = \frac{1}{s}, \quad \text{for} \quad s > 0.$$

In the remaining part of this chapter, whenever the improper integral is calculated, we will not explicitly write the limiting process. However, the students are advised to provide the details.

2. Find the Laplace transform $F(s)$ of $f(t)$, where $f(t) = t, \ \ t \geq 0$.
   **Solution:** Integration by parts gives

$$\begin{aligned} F(s) &= \int_0^\infty te^{-st}dt = \left.\frac{-te^{-st}}{s}\right|_0^\infty + \int_0^\infty \frac{e^{-st}}{s}dt \\ &= \frac{1}{s^2} \quad \text{for} \quad s > 0. \end{aligned}$$

3. Find the Laplace transform of $f(t) = t^n, \ \ n$ a positive integer.
   **Solution:** Substituting $st = \tau$, we get

$$\begin{aligned} F(s) &= \int_0^\infty e^{-st}t^n dt \\ &= \frac{1}{s^{n+1}} \int_0^\infty e^{-\tau}\tau^n \, d\tau \\ &= \frac{n!}{s^{n+1}} \quad \text{for} \quad s > 0. \end{aligned}$$

4. Find the Laplace transform of $f(t) = e^{at}, \ t \geq 0$.
   **Solution:** We have

$$\begin{aligned} \mathcal{L}(e^{at}) &= \int_0^\infty e^{at}e^{-st}dt = \int_0^\infty e^{-(s-a)t}dt \\ &= \frac{1}{s-a} \quad \text{for} \quad s > a. \end{aligned}$$

5. Compute the Laplace transform of $\cos(at), \ \ t \geq 0$.
   **Solution:**

$$\begin{aligned} \mathcal{L}(\cos(at)) &= \int_0^\infty \cos(at)e^{-st}dt \\ &= \cos(at)\left.\frac{e^{-st}}{-s}\right|_0^\infty - \int_0^\infty -a\sin(at)\cdot\frac{e^{-st}}{-s}dt \\ &= \frac{1}{s} - \left( \frac{a\sin(at)}{s}\left.\frac{e^{-st}}{-s}\right|_0^\infty - \int_0^\infty a^2\frac{\cos(at)}{s}\frac{e^{-st}}{-s}dt\right) \end{aligned}$$

   Note that the limits exist only when $s > 0$. Hence,

$$\frac{a^2+s^2}{s^2}\int_0^\infty \cos(at)e^{-st}dt = \frac{1}{s}. \quad \text{Thus} \quad \mathcal{L}(\cos(at)) = \frac{s}{a^2+s^2}; \qquad\qquad s > 0.$$

6. Similarly, one can show that

$$\mathcal{L}(\sin(at)) = \frac{a}{s^2+a^2}, \ \ s > 0.$$

7. Find the Laplace transform of $f(t) = \dfrac{1}{\sqrt{t}}, \ \ t > 0$.
   **Solution:** Note that $f(t)$ is not a bounded function near $t = 0$ (why!). We will still show that the Laplace transform of $f(t)$ exists.

$$\begin{aligned} \mathcal{L}(\frac{1}{\sqrt{t}}) &= \int_0^\infty \frac{1}{\sqrt{t}}e^{-st}dt = \int_0^\infty \frac{\sqrt{s}}{\sqrt{\tau}}e^{-\tau}\frac{d\tau}{s} \quad (\text{ substitute } \tau = st) \\ &= \frac{1}{\sqrt{s}}\int_0^\infty \tau^{-\frac{1}{2}}e^{-\tau}d\tau = \frac{1}{\sqrt{s}}\int_0^\infty \tau^{\frac{1}{2}-1}e^{-\tau}d\tau. \end{aligned}$$

Recall that for calculating the integral $\int_0^\infty \tau^{\frac{1}{2}-1}e^{-\tau}d\tau$, one needs to consider the double integral

$$\int_0^\infty\int_0^\infty e^{-(x^2+y^2)}dxdy = \Big(\int_0^\infty e^{-x^2}dx\Big)^2 = \Big(\frac{1}{2}\int_0^\infty \tau^{\frac{1}{2}-1}e^{-\tau}d\tau\Big)^2.$$

It turns out that

$$\int_0^\infty \tau^{\frac{1}{2}-1}e^{-\tau}d\tau = \sqrt{\pi}.$$

Thus, $\mathcal{L}(\frac{1}{\sqrt{t}}) = \frac{\sqrt{\pi}}{\sqrt{s}}$ for $s > 0$.

We now put the above discussed examples in tabular form as they constantly appear in applications of Laplace transform to differential equations.

| $f(t)$ | $\mathcal{L}(f(t))$ | $f(t)$ | $\mathcal{L}(f(t))$ |
|---|---|---|---|
| $1$ | $\frac{1}{s}, \quad s > 0$ | $t$ | $\frac{1}{s^2}, \quad s > 0$ |
| $t^n$ | $\frac{n!}{s^{n+1}}, \quad s > 0$ | $e^{at}$ | $\frac{1}{s-a}, \quad s > a$ |
| $\sin(at)$ | $\frac{a}{s^2+a^2}, \quad s > 0$ | $\cos(at)$ | $\frac{s}{s^2+a^2}, \quad s > 0$ |
| $\sinh(at)$ | $\frac{a}{s^2-a^2}, \quad s > a$ | $\cosh(at)$ | $\frac{s}{s^2-a^2}, \quad s > a$ |

Table 10.1: Laplace transform of some Elementary Functions

## 10.3 Properties of Laplace Transform

**Lemma 10.3.1 (Linearity of Laplace Transform)** 1. Let $a, b \in \mathbb{R}$. Then

$$\begin{aligned}\mathcal{L}\big(af(t)+bg(t)\big) &= \int_0^\infty \big(af(t)+bg(t)\big)e^{-st}dt \\ &= a\mathcal{L}(f(t)) + b\mathcal{L}(g(t)).\end{aligned}$$

2. If $F(s) = \mathcal{L}(f(t)),$ and $G(s) = \mathcal{L}(g(t))$, then

$$\mathcal{L}^{-1}\big(aF(s)+bG(s)\big) = af(t)+bg(t).$$

The above lemma is immediate from the definition of Laplace transform and the linearity of the definite integral.

**Example 10.3.2** 1. Find the Laplace transform of $\cosh(at)$.

**Solution:** $\cosh(at) = \frac{e^{at}+e^{-at}}{2}$. Thus

$$\mathcal{L}(\cosh(at)) = \frac{1}{2}\left(\frac{1}{s-a}+\frac{1}{s+a}\right) = \frac{s}{s^2-a^2}, \qquad s > |a|.$$

2. Similarly,

$$\mathcal{L}(\sinh(at)) = \frac{1}{2}\left(\frac{1}{s-a}-\frac{1}{s+a}\right) = \frac{a}{s^2-a^2}, \qquad s > |a|.$$

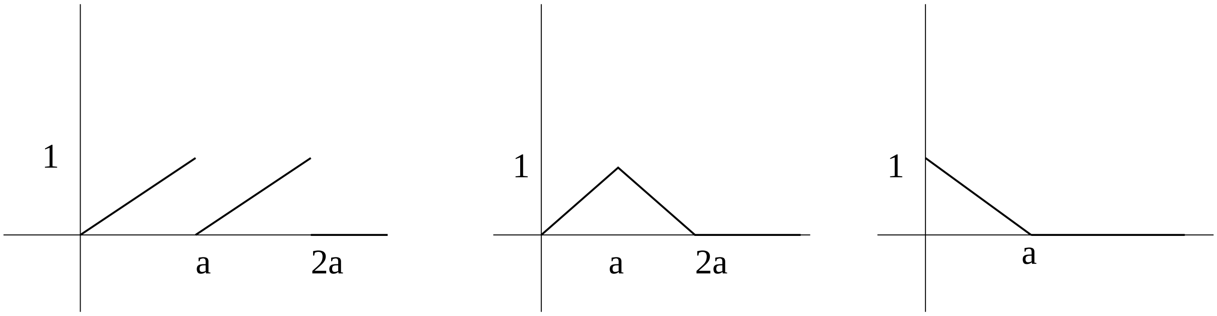


Figure 10.2: $f(t)$

3. Find the inverse Laplace transform of $\dfrac{1}{s(s+1)}$.

   **Solution:**

$$\begin{aligned}\mathcal{L}^{-1}\left(\frac{1}{s(s+1)}\right) &= \mathcal{L}^{-1}\left(\frac{1}{s} - \frac{1}{s+1}\right) \\ &= \mathcal{L}^{-1}\left(\frac{1}{s}\right) - \mathcal{L}^{-1}\left(\frac{1}{s+1}\right) = 1 - e^{-t}.\end{aligned}$$

   Thus, the inverse Laplace transform of $\dfrac{1}{s(s+1)}$ is $f(t) = 1 - e^{-t}$.

**Theorem 10.3.3 (Scaling by $a$)** Let $f(t)$ be a piecewise continuous function with Laplace transform $F(s)$. Then for $a > 0$, $\mathcal{L}(f(at)) = \dfrac{1}{a}F(\dfrac{s}{a})$.

Proof. By definition and the substitution $z = at$, we get

$$\begin{aligned}\mathcal{L}(f(at)) &= \int_0^\infty e^{-st} f(at)dt = \frac{1}{a}\int_0^\infty e^{-s\frac{z}{a}} f(z)dz \\ &= \frac{1}{a}\int_0^\infty e^{-\frac{s}{a}z} f(z)dz = \frac{1}{a}F(\frac{s}{a}).\end{aligned}$$

□

**Exercise 10.3.4**

1. Find the Laplace transform of

$$t^2 + at + b, \quad \cos(wt + \theta), \quad \cos^2 t, \quad \sinh^2 t;$$

   where $a, b, w$ and $\theta$ are arbitrary constants.

2. Find the Laplace transform of the function $f(\cdot)$ given by the graphs in Figure 10.2.

3. If $\mathcal{L}(f(t)) = \dfrac{1}{s^2+1} + \dfrac{1}{2s+1}$, find $f(t)$.

The next theorem relates the Laplace transform of the function $f'(t)$ with that of $f(t)$.

**Theorem 10.3.5 (Laplace Transform of Differentiable Functions)** Let $f(t)$, for $t > 0$, be a differentiable function with the derivative, $f'(t)$, being continuous. Suppose that there exist constants $M$ and $T$ such that $|f(t)| \le Me^{\alpha t}$ for all $t \ge T$. If $\mathcal{L}(f(t)) = F(s)$ then

$$\mathcal{L}\left(f'(t)\right) = sF(s) - f(0) \quad \text{for} \quad s > \alpha. \tag{10.3.1}$$

Proof. Note that the condition $|f(t)| \leq Me^{\alpha t}$ for all $t \geq T$ implies that

$$\lim_{b \longrightarrow \infty} f(b)e^{-sb} = 0 \quad \text{for} \quad s > \alpha.$$

So, by definition,

$$\begin{aligned} \mathcal{L}\big(f'(t)\big) &= \int_0^\infty e^{-st} f'(t)dt = \lim_{b \longrightarrow \infty} \int_0^b e^{-st} f'(t)dt \\ &= \lim_{b \longrightarrow \infty} f(t)e^{-st}\Big|_0^b - \lim_{b \longrightarrow \infty} \int_0^b f(t)(-s)e^{-st}dt \\ &= -f(0) + sF(s). \end{aligned}$$

□

We can extend the above result for $n^{\text{th}}$ derivative of a function $f(t)$, if $f'(t), \ldots, f^{(n-1)}(t), f^{(n)}(t)$ exist and $f^{(n)}(t)$ is continuous for $t \geq 0$. In this case, a repeated use of Theorem 10.3.5, gives the following corollary.

**Corollary 10.3.6** Let $f(t)$ be a function with $\mathcal{L}(f(t)) = F(s)$. If $f'(t), \ldots, f^{(n-1)}(t), f^{(n)}(t)$ exist and $f^{(n)}(t)$ is continuous for $t \geq 0$, then

$$\mathcal{L}\big(f^{(n)}(t)\big) = s^n F(s) - s^{n-1} f(0) - s^{n-2} f'(0) - \cdots - f^{(n-1)}(0). \tag{10.3.2}$$

In particular, for $n = 2$, we have

$$\mathcal{L}\big(f''(t)\big) = s^2 F(s) - sf(0) - f'(0). \tag{10.3.3}$$

**Corollary 10.3.7** Let $f'(t)$ be a piecewise continuous function for $t \geq 0$. Also, let $f(0) = 0$. Then

$$\mathcal{L}(f'(t)) = sF(s) \quad \text{or equivalently} \quad \mathcal{L}^{-1}(sF(s)) = f'(t).$$

**Example 10.3.8**

1. Find the inverse Laplace transform of $\dfrac{s}{s^2+1}$.

   **Solution:** We know that $\mathcal{L}^{-1}(\dfrac{1}{s^2+1}) = \sin t$. Then $\sin(0) = 0$ and therefore, $\mathcal{L}^{-1}(\dfrac{s}{s^2+1}) = \cos t$.

2. Find the Laplace transform of $f(t) = \cos^2(t)$.

   **Solution:** Note that $f(0) = 1$ and $f'(t) = -2\cos t \, \sin t = -\sin(2t)$. Also,

   $$\mathcal{L}(-\sin(2t)) = \frac{-2}{s^2+4}.$$

   Now, using Theorem 10.3.5, we get

   $$\mathcal{L}(f(t)) = \frac{1}{s}\left(-\frac{2}{s^2+4} + 1\right) = \frac{s^2+2}{s(s^2+4)}.$$

**Lemma 10.3.9 (Laplace Transform of $tf(t)$)** Let $f(t)$ be a piecewise continuous function with $\mathcal{L}(f(t)) = F(s)$. If the function $F(s)$ is differentiable, then

$$\mathcal{L}(tf(t)) = -\frac{d}{ds}F(s).$$

$$\text{Equivalently,} \quad \mathcal{L}^{-1}(-\frac{d}{ds}F(s)) = tf(t).$$

PROOF. By definition, $F(s) = \int_0^\infty e^{-st} f(t)dt$. The result is obtained by differentiating both sides with respect to $s$. □

Suppose we know the Laplace transform of a $f(t)$ and we wish to find the Laplace transform of the function $g(t) = \frac{f(t)}{t}$. Suppose that $G(s) = \mathcal{L}(g(t))$ exists. Then writing $f(t) = tg(t)$ gives

$$F(s) = \mathcal{L}(f(t)) = \mathcal{L}(tg(t)) = -\frac{d}{ds}G(s).$$

Thus, $G(s) = -\int\limits_a^s F(p)dp$ for some real number $a$. As $\lim\limits_{s\longrightarrow\infty} G(s) = 0$, we get $G(s) = \int\limits_s^\infty F(p)dp$.

Hence,we have the following corollary.

**Corollary 10.3.10** Let $\mathcal{L}(f(t)) = F(s)$ and $g(t) = \frac{f(t)}{t}$. Then

$$\mathcal{L}(g(t)) = G(s) = \int\limits_s^\infty F(p)dp.$$

**Example 10.3.11** 1. Find $\mathcal{L}(t\sin(at))$.

**Solution:** We know $\mathcal{L}(\sin(at)) = \frac{a}{s^2+a^2}$. Hence $\mathcal{L}(t\sin(at)) = \frac{2as}{(s^2+a^2)^2}$.

2. Find the function $f(t)$ such that $F(s) = \frac{4}{(s-1)^3}$.

**Solution:** We know $\mathcal{L}(e^t) = \frac{1}{s-1}$ and

$\frac{4}{(s-1)^3} = 2\frac{d}{ds}\left(-\frac{1}{(s-1)^2}\right) = 2\frac{d^2}{ds^2}\left(\frac{1}{s-1}\right)$.

By lemma 10.3.9, we know that $\mathcal{L}(tf(t)) = -\frac{d}{ds}F(s)$. Suppose $\frac{d}{ds}F(s) = G(s)$. Then $g(t) = \mathcal{L}^{-1}G(s) = \mathcal{L}^{-1}\frac{d}{ds}F(s) = -tf(t)$. Therefore,

$$\mathcal{L}^{-1}\left(\frac{d^2}{ds^2}F(s)\right) = \mathcal{L}^{-1}\left(\frac{d}{ds}G(s)\right) = -tg(t) = t^2 f(t).$$

Thus we get $f(t) = 2t^2e^t$.

**Lemma 10.3.12 (Laplace Transform of an Integral)** If $F(s) = \mathcal{L}(f(t))$ then

$$\mathcal{L}\left[\int_0^t f(\tau)d\tau\right] = \frac{F(s)}{s}.$$

Equivalently, $\mathcal{L}^{-1}\left(\frac{F(s)}{s}\right) = \int_0^t f(\tau)d\tau$.

PROOF. By definition,

$$\mathcal{L}(\int_0^t f(\tau)\,d\tau) = \int_0^\infty e^{-st}\left(\int_0^t f(\tau)\,d\tau\right)dt = \int_0^\infty\int_0^t e^{-st}f(\tau)\,d\tau dt.$$

We don't go into the details of the proof of the change in the order of integration. We assume that the order of the integrations can be changed and therefore

$$\int_0^\infty\int_0^t e^{-st}f(\tau)\,d\tau dt = \int_0^\infty\int_\tau^\infty e^{-st}f(\tau)\,dt\,d\tau.$$

Thus,

$$\begin{aligned}
\mathcal{L}\big(\int_0^t f(\tau)\, d\tau\big) &= \int_0^\infty \int_0^t e^{-st} f(\tau)\, d\tau dt \\
&= \int_0^\infty \int_\tau^\infty e^{-st} f(\tau)\, dt\, d\tau = \int_0^\infty \int_\tau^\infty e^{-s(t-\tau)-s\tau} f(\tau)\, dt\, d\tau \\
&= \int_0^\infty e^{-s\tau} f(\tau) d\tau \left( \int_\tau^\infty e^{-s(t-\tau)} dt \right) \\
&= \int_0^\infty e^{-s\tau} f(\tau) d\tau \left( \int_0^\infty e^{-sz} dz \right) = F(s)\frac{1}{s}.
\end{aligned}$$

□

**Example 10.3.13** 1. Find $\mathcal{L}(\int_0^t \sin(az)dz)$.
**Solution:** We know $\mathcal{L}(\sin(at)) = \dfrac{a}{s^2+a^2}$. Hence

$$\mathcal{L}(\int_0^t \sin(az)dz) = \frac{1}{s}\cdot\frac{a}{(s^2+a^2)} = \frac{a}{s(s^2+a^2)}.$$

2. Find $\mathcal{L}\left(\int_0^t \tau^2 d\tau\right)$.
**Solution:** By Lemma 10.3.12

$$\mathcal{L}\left(\int_0^t \tau^2 d\tau\right) = \frac{\mathcal{L}\left(t^2\right)}{s} = \frac{1}{s}\cdot\frac{2!}{s^3} = \frac{2}{s^4}.$$

3. Find the function $f(t)$ such that $F(s) = \dfrac{4}{s(s-1)}$.
**Solution:** We know $\mathcal{L}(e^t) = \dfrac{1}{s-1}$. So,

$$\mathcal{L}^{-1}\left(\frac{4}{s(s-1)}\right) = 4\mathcal{L}^{-1}\left(\frac{1}{s}\frac{1}{s-1}\right) = 4\int_0^t e^\tau d\tau = 4(e^t-1).$$

**Lemma 10.3.14 ($s$-Shifting)** Let $\mathcal{L}(f(t)) = F(s)$. Then $\mathcal{L}(e^{at}f(t)) = F(s-a)$ for $s > a$.

PROOF.

$$\begin{aligned}
\mathcal{L}(e^{at}f(t)) &= \int_0^\infty e^{at} f(t) e^{-st} dt = \int_0^\infty f(t) e^{-(s-a)t} dt \\
&= F(s-a) \qquad s > a.
\end{aligned}$$

□

**Example 10.3.15** 1. Find $\mathcal{L}(e^{at}\sin(bt))$.
**Solution:** We know $\mathcal{L}(\sin(bt)) = \dfrac{b}{s^2+b^2}$. Hence $\mathcal{L}(e^{at}\sin(bt)) = \dfrac{b}{(s-a)^2+b^2}$.

2. Find $\mathcal{L}^{-1}\left(\frac{s-5}{(s-5)^2+36}\right)$.
**Solution:** By $s$-Shifting, if $\mathcal{L}(f(t)) = F(s)$ then $\mathcal{L}(e^{at}f(t)) = F(s-a)$. Here, $a = 5$ and

$$\mathcal{L}^{-1}\left(\frac{s}{s^2+36}\right) = \mathcal{L}^{-1}\left(\frac{s}{s^2+6^2}\right) = \cos(6t).$$

Hence, $f(t) = e^{5t}\cos(6t)$.

### 10.3.1 Inverse Transforms of Rational Functions

Let $F(s)$ be a rational function of $s$. We give a few examples to explain the methods for calculating the inverse Laplace transform of $F(s)$.

**Example 10.3.16** 1. Denominator of $F$ has Distinct Real Roots:

$$\text{If} \quad F(s) = \frac{(s+1)(s+3)}{s(s+2)(s+8)} \quad \text{find } f(t).$$

**Solution:** $F(s) = \dfrac{3}{16s} + \dfrac{1}{12(s+2)} + \dfrac{35}{48(s+8)}$. Thus,
$f(t) = \dfrac{3}{16} + \dfrac{1}{12}e^{-2t} + \dfrac{35}{48}e^{-8t}$.

2. Denominator of $F$ has Distinct Complex Roots:

$$\text{If} \quad F(s) = \frac{4s+3}{s^2+2s+5} \quad \text{find} \quad f(t).$$

**Solution:** $F(s) = 4\dfrac{s+1}{(s+1)^2+2^2} - \dfrac{1}{2}\cdot\dfrac{2}{(s+1)^2+2^2}$. Thus,
$f(t) = 4e^{-t}\cos(2t) - \dfrac{1}{2}e^{-t}\sin(2t)$.

3. Denominator of $F$ has Repeated Real Roots:

$$\text{If} \quad F(s) = \frac{3s+4}{(s+1)(s^2+4s+4)} \quad \text{find} \quad f(t).$$

**Solution:** Here,

$$F(s) = \frac{3s+4}{(s+1)(s^2+4s+4)} = \frac{3s+4}{(s+1)(s+2)^2} = \frac{a}{s+1} + \frac{b}{s+2} + \frac{c}{(s+2)^2}.$$

Solving for $a, b$ and $c$, we get $F(s) = \frac{1}{s+1} - \frac{1}{s+2} + \frac{2}{(s+2)^2} = \frac{1}{s+1} - \frac{1}{s+2} + 2\frac{d}{ds}\left(-\frac{1}{(s+2)}\right)$. Thus, $f(t) = e^{-t} - e^{-2t} + 2te^{-2t}$.

### 10.3.2 Transform of Unit Step Function

**Definition 10.3.17 (Unit Step Function)** The Unit-Step function is defined by

$$U_a(t) = \begin{cases} 0 & \text{if } 0 \le t < a \\ 1 & \text{if } t \ge a \end{cases}.$$

**Example 10.3.18** $\mathcal{L}\big(U_a(t)\big) = \displaystyle\int_a^\infty e^{-st}dt = \frac{e^{-sa}}{s}, \ s > 0.$

**Lemma 10.3.19 ($t$-Shifting)** Let $\mathcal{L}(f(t)) = F(s)$. Define $g(t)$ by

$$g(t) = \begin{cases} 0 & \text{if } 0 \le t < a \\ f(t-a) & \text{if } t \ge a \end{cases}.$$

Then $g(t) = U_a(t)f(t-a)$ and

$$\mathcal{L}\big(g(t)\big) = e^{-as}F(s).$$

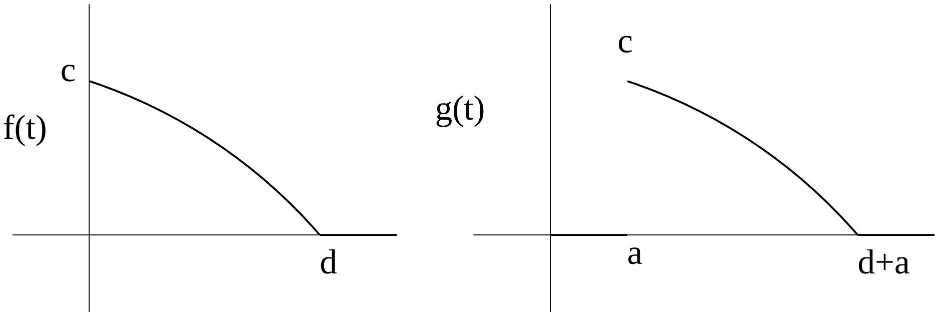


Figure 10.3: Graphs of $f(t)$ and $U_a(t)f(t-a)$

PROOF. Let $0 \le t < a$. Then $U_a(t) = 0$ and so, $U_a(t)f(t-a) = 0 = g(t)$.
If $t \ge a$, then $U_a(t) = 1$ and $U_a(t)f(t-a) = f(t-a) = g(t)$. Since the functions $g(t)$ and $U_a(t)f(t-a)$ take the same value for all $t \ge 0$, we have $g(t) = U_a(t)f(t-a)$. Thus,

$$\begin{aligned}\mathcal{L}(g(t)) &= \int_0^\infty e^{-st}g(t)dt = \int_a^\infty e^{-st}f(t-a)dt \\ &= \int_0^\infty e^{-s(t+a)}f(t)dt = e^{-as}\int_0^\infty e^{-st}f(t)dt \\ &= e^{-as}F(s).\end{aligned}$$

□

**Example 10.3.20** Find $\mathcal{L}^{-1}\left(\frac{e^{-5s}}{s^2-4s-5}\right)$.
**Solution:** Let $G(s) = \frac{e^{-5s}}{s^2-4s-5} = e^{-5s}F(s)$, with $F(s) = \frac{1}{s^2-4s-5}$. Since $s^2-4s-5 = (s-2)^2 - 3^2$

$$\mathcal{L}^{-1}(F(s)) = \mathcal{L}^{-1}\left(\frac{1}{3}\cdot\frac{3}{(s-2)^2-3^2}\right) = \frac{1}{3}\sinh(3t)e^{2t}.$$

Hence, by Lemma 10.3.19

$$\mathcal{L}^{-1}(G(s)) = \frac{1}{3}\,U_5(t)\sinh\big(3(t-5)\big)e^{2(t-5)}.$$

**Example 10.3.21** Find $\mathcal{L}(f(t))$, where $f(t) = \begin{cases} 0 & t < 2\pi \\ t\cos t & t > 2\pi. \end{cases}$

**Solution:** Note that

$$f(t) = \begin{cases} 0 & t < 2\pi \\ (t-2\pi)\cos(t-2\pi) + 2\pi\cos(t-2\pi) & t > 2\pi. \end{cases}$$

Thus, $\mathcal{L}(f(t)) = e^{-2\pi s}\left(\frac{s^2-1}{(s^2+1)^2} + 2\pi\frac{s}{s^2+1}\right)$

**Note: To be filled by a graph**

## 10.4 Some Useful Results

### 10.4.1 Limiting Theorems

The following two theorems give us the behaviour of the function $f(t)$ when $t \longrightarrow 0^+$ and when $t \longrightarrow \infty$.

**Theorem 10.4.1 (First Limit Theorem)** Suppose $\mathcal{L}(f(t))$ exists. Then

$$\lim_{t\longrightarrow 0^+} f(t) = \lim_{s\longrightarrow\infty} sF(s).$$

Proof. We know $sF(s) - f(0) = \mathcal{L}(f'(t))$. Therefore

$$\begin{aligned}\lim_{s\longrightarrow\infty} sF(s) &= f(0) + \lim_{s\longrightarrow\infty}\int_0^\infty e^{-st}f'(t)dt\\ &= f(0) + \int_0^\infty \lim_{s\longrightarrow\infty} e^{-st}f'(t)dt = f(0).\end{aligned}$$

as $\lim\limits_{s\longrightarrow\infty} e^{-st} = 0$. □

**Example 10.4.2** 1. For $t \geq 0$, let $Y(s) = \mathcal{L}(y(t)) = a(1+s^2)^{-1/2}$. Determine $a$ such that $y(0) = 1$.
**Solution:** Theorem 10.4.1 implies
$1 = \lim\limits_{s\longrightarrow\infty} sY(s) = \lim\limits_{s\longrightarrow\infty} \dfrac{as}{(1+s^2)^{1/2}} = \lim\limits_{s\longrightarrow\infty} \dfrac{a}{(\frac{1}{s^2}+1)^{1/2}}$. Thus, $a = 1$.

2. If $F(s) = \dfrac{(s+1)(s+3)}{s(s+2)(s+8)}$ find $f(0^+)$.
**Solution:** Theorem 10.4.1 implies

$$f(0^+) = \lim_{s\longrightarrow\infty} sF(s) = \lim_{s\longrightarrow\infty} s\cdot\frac{(s+1)(s+3)}{s(s+2)(s+8)} = 1.$$

On similar lines, one has the following theorem. But this theorem is valid only when $f(t)$ is bounded as $t$ approaches infinity.

**Theorem 10.4.3 (Second Limit Theorem)** Suppose $\mathcal{L}(f(t))$ exists. Then

$$\lim_{t\longrightarrow\infty} f(t) = \lim_{s\longrightarrow 0} sF(s)$$

provided that $sF(s)$ converges to a finite limit as $s$ tends to $0$.

Proof.

$$\begin{aligned}\lim_{s\longrightarrow 0} sF(s) &= f(0) + \lim_{s\longrightarrow 0}\int_0^\infty e^{-st}f'(t)dt\\ &= f(0) + \lim_{s\longrightarrow 0}\lim_{t\longrightarrow\infty}\int_0^t e^{-s\tau}f'(\tau)d\tau\\ &= f(0) + \lim_{t\longrightarrow\infty}\int_0^t \lim_{s\longrightarrow 0} e^{-s\tau}f'(\tau)d\tau = \lim_{t\longrightarrow\infty} f(t).\end{aligned}$$

□

**Example 10.4.4** If $F(s) = \dfrac{2(s+3)}{s(s+2)(s+8)}$ find $\lim\limits_{t\longrightarrow\infty} f(t)$.
**Solution:** From Theorem 10.4.3, we have

$$\lim_{t\longrightarrow\infty} f(t) = \lim_{s\longrightarrow 0} sF(s) = \lim_{s\longrightarrow 0} s\cdot\frac{2(s+3)}{s(s+2)(s+8)} = \frac{6}{16} = \frac{3}{8}.$$

We now generalise the lemma on Laplace transform of an integral as convolution theorem.

**Definition 10.4.5 (Convolution of Functions)** Let $f(t)$ and $g(t)$ be two smooth functions. The convolution, $f\star g$, is a function defined by

$$(f\star g)(t) = \int_0^t f(\tau)g(t-\tau)d\tau.$$

Check that

1. $(f \star g)(t) = g \star f(t)$.
2. If $f(t) = \cos(t)$ then $(f \star f)(t) = \dfrac{t\cos(t) + \sin(t)}{2}$.

**Theorem 10.4.6 (Convolution Theorem)** If $F(s) = \mathcal{L}(f(t))$ and $G(s) = \mathcal{L}(g(t))$ then

$$\mathcal{L}\left[\int_0^t f(\tau)g(t-\tau)d\tau\right] = F(s) \cdot G(s).$$

**Remark 10.4.7** *Let $g(t) = 1$ for all $t \geq 0$. Then we know that $\mathcal{L}(g(t)) = G(s) = \dfrac{1}{s}$. Thus, the Convolution Theorem 10.4.6 reduces to the Integral Lemma 10.3.12.*

## 10.5 Application to Differential Equations

Consider the following example.

**Example 10.5.1** Solve the following Initial Value Problem:

$$af''(t) + bf'(t) + cf(t) = g(t) \text{ with } f(0) = f_0, \; f'(0) = f_1.$$

**Solution:** Let $\mathcal{L}(g(t)) = G(s)$. Then

$$G(s) = a(s^2F(s) - sf(0) - f'(0)) + b(sF(s) - f(0)) + cF(s)$$

and the initial conditions imply

$$G(s) = (as^2 + bs + c)F(s) - (as + b)f_0 - af_1.$$

Hence,

$$F(s) = \underbrace{\frac{G(s)}{as^2+bs+c}}_{non-homogeneous\ \ part} + \underbrace{\frac{(as+b)f_0}{as^2+bs+c} + \frac{af_1}{as^2+bs+c}}_{initial\ \ conditions}. \tag{10.5.1}$$

Now, if we know that $G(s)$ is a rational function of $s$ then we can compute $f(t)$ from $F(s)$ by using the method of PARTIAL FRACTIONS (see Subsection 10.3.1 ).

**Example 10.5.2** 1. Solve the IVP

$$y'' - 4y' - 5y = f(t) = \begin{cases} t & \text{if } 0 \leq t < 5 \\ t+5 & \text{if } t \geq 5 \end{cases}.$$

with $y(0) = 1$ and $y'(0) = 4$.
**Solution:** Note that $f(t) = t + U_5(t)$. Thus,

$$\mathcal{L}(f(t)) = \frac{1}{s^2} + \frac{e^{-5s}}{s}.$$

Taking Laplace transform of the above equation, we get

$$\left(s^2Y(s) - sy(0) - y'(0)\right) - 4\left(sY(s) - y(0)\right) - 5Y(s) = \mathcal{L}(f(t)) = \frac{1}{s^2} + \frac{e^{-5s}}{s}.$$

Which gives

$$\begin{aligned} Y(s) &= \frac{s}{(s+1)(s-5)} + \frac{e^{-5s}}{s(s+1)(s-5)} + \frac{1}{s^2(s+1)(s-5)} \\ &= \frac{1}{6}\left[\frac{5}{s-5} + \frac{1}{s+1}\right] + \frac{e^{-5s}}{30}\left[-\frac{6}{s} + \frac{5}{s+1} + \frac{1}{s-5}\right] \\ & \quad + \frac{1}{150}\left[-\frac{30}{s^2} + \frac{24}{s} - \frac{25}{s+1} + \frac{1}{s-5}\right]. \end{aligned}$$

Hence,

$$\begin{aligned} y(t) &= \frac{5e^{5t}}{6} + \frac{e^{-t}}{6} + U_5(t)\left[-\frac{1}{5} + \frac{e^{-(t-5)}}{6} + \frac{e^{5(t-5)}}{30}\right] \\ & \quad + \frac{1}{150}\left[-30t + 24 - 25e^{-t} + e^{5t}\right]. \end{aligned}$$

**Remark 10.5.3** *Even though $f(t)$ is a* DISCONTINUOUS *function at $t = 5$, the solution $y(t)$ and $y'(t)$ are continuous functions of $t$, as $y''$ exists. In general, the following is always true:*
*Let $y(t)$ be a solution of $ay'' + by' + cy = f(t)$. Then both $y(t)$ and $y'(t)$ are continuous functions of time.*

**Example 10.5.4** 1. Consider the IVP $ty''(t) + y'(t) + ty(t) = 0$, with $y(0) = 1$ and $y'(0) = 0$. Find $\mathcal{L}(y(t))$.

**Solution:** Applying Laplace transform, we have

$$-\frac{d}{ds}\left[s^2Y(s) - sy(0) - y'(0)\right] + (sY(s) - y(0)) - \frac{d}{ds}Y(s) = 0.$$

Using initial conditions, the above equation reduces to

$$\frac{d}{ds}\left[(s^2+1)Y(s) - s\right] - sY(s) + 1 = 0.$$

This equation after simplification can be rewritten as

$$\frac{Y'(s)}{Y(s)} = -\frac{s}{s^2+1}.$$

Therefore, $Y(s) = a(1+s^2)^{-\frac{1}{2}}$. From Example 10.4.2.1, we see that $a = 1$ and hence

$$Y(s) = (1+s^2)^{-\frac{1}{2}}.$$

2. Show that $y(t) = \int_0^t f(\tau)g(t-\tau)d\tau$ is a solution of

$$y''(t) + ay'(t) + by(t) = f(t), \quad \text{with} \quad y(0) = y'(0) = 0;$$

where $\mathcal{L}[g(t)] = \dfrac{1}{s^2 + as + b}$.

**Solution:** Here, $Y(s) = \dfrac{F(s)}{s^2 + as + b} = F(s) \cdot \dfrac{1}{s^2 + as + b}$. Hence,

$$y(t) = (f \star g)(t) = \int_0^t f(\tau)g(t-\tau)d\tau.$$

3. Show that $y(t) = \dfrac{1}{a}\int_0^t f(\tau)\sin(a(t-\tau))d\tau$ is a solution of

$$y''(t) + a^2y(t) = f(t), \quad \text{with} \quad y(0) = y'(0) = 0.$$

**Solution:** Here, $Y(s) = \frac{F(s)}{s^2+a^2} = \frac{1}{a}\left(F(s)\cdot\frac{a}{s^2+a^2}\right)$. Hence,

$$y(t) = \frac{1}{a}f(t)\star\sin(at) = \frac{1}{a}\int_0^t f(\tau)\sin(a(t-\tau))d\tau.$$

4. Solve the following IVP.

$$y'(t) = \int_0^t y(\tau)d\tau + t - 4\sin t, \quad \text{with} \quad y(0) = 1.$$

**Solution:** Taking Laplace transform of both sides and using Theorem 10.3.5, we get

$$sY(s) - 1 = \frac{Y(s)}{s} + \frac{1}{s^2} - 4\frac{1}{s^2+1}.$$

Solving for $Y(s)$, we get

$$Y(s) = \frac{s^2-1}{s(s^2+1)} = \frac{1}{s} - 2\frac{1}{s^2+1}.$$

So,

$$y(t) = 1 - 2\int_0^t \sin(\tau)d\tau = 1 + 2(\cos t - 1) = 2\cos t - 1.$$

## 10.6 Transform of the Unit-Impulse Function

Consider the following example.

**Example 10.6.1** Find the Laplace transform, $D_h(s)$, of

$$\delta_h(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{h} & 0 \le t < h \\ 0 & t > h. \end{cases}$$

**Solution:** Note that $\delta_h(t) = \frac{1}{h}\big(U_0(t) - U_h(t)\big)$. By linearity of the Laplace transform, we get

$$D_h(s) = \frac{1}{h}\left(\frac{1-e^{-hs}}{s}\right).$$

**Remark 10.6.2** *1. Observe that in Example 10.6.1, if we allow $h$ to approach $0$, we obtain a new function, say $\delta(t)$. That is, let*

$$\delta(t) = \lim_{h\longrightarrow 0}\delta_h(t).$$

*This new function is zero everywhere except at the origin. At origin, this function tends to infinity. In other words, the graph of the function appears as a line of infinite height at the origin. This new function, $\delta(t)$, is called the* UNIT-IMPULSE FUNCTION *(or Dirac's delta function).*

*2. We can also write*

$$\delta(t) = \lim_{h\longrightarrow 0}\delta_h(t) = \lim_{h\longrightarrow 0}\frac{1}{h}\big(U_0(t) - U_h(t)\big).$$

*3. In the strict mathematical sense $\lim_{h\longrightarrow 0}\delta_h(t)$ does not exist. Hence, mathematically speaking, $\delta(t)$ does not represent a function.*

*4. However, note that*

$$\int_0^\infty \delta_h(t)dt = 1, \quad \textit{for all} \quad h.$$

5. *Also, observe that* $\mathcal{L}(\delta_h(t)) = \dfrac{1-e^{-hs}}{hs}$*. Now, if we take the limit of both sides, as* $h$ *approaches zero (apply L'Hospital's rule), we get*

$$\mathcal{L}(\delta(t)) = \lim_{h\longrightarrow 0} \frac{1-e^{-hs}}{hs} = \lim_{h\longrightarrow 0} \frac{se^{-hs}}{s} = 1.$$

# Part IV

# Numerical Applications

# Chapter 11

# Newton's Interpolation Formulae

## 11.1 Introduction

In many practical situations, for a function $y = f(x)$, which either may not be explicitly specified or may be difficult to handle, we often have a tabulated data $(x_i, y_i)$, where $y_i = f(x_i)$, and $x_i < x_{i+1}$ for $i = 0, 1, 2, \ldots, N$. In such cases, it may be required to represent or replace the given function by a simpler function, which coincides with the values of $f$ at the $N + 1$ tabular points $x_i$. This process is known as INTERPOLATION. Interpolation is also used to estimate the value of the function at the non tabular points. Here, we shall consider only those functions which are sufficiently smooth, *i.e.,* they are differentiable sufficient number of times. Many of the interpolation methods, where the tabular points are equally spaced, use difference operators. Hence, in the following we introduce various difference operators and study their properties before looking at the interpolation methods.

We shall assume here that the TABULAR POINTS $x_0, x_1, x_2, \ldots, x_N$ are equally spaced, *i.e.,* $x_k - x_{k-1} = h$ for each $k = 1, 2, \ldots, N$. The real number $h$ is called the STEP LENGTH. This gives us $x_k = x_0 + kh$. Further, $y_k = f(x_k)$ gives the value of the function $y = f(x)$ at the $k^{\text{th}}$ tabular point. The points $y_1, y_2, \ldots, y_N$ are known as NODES or NODAL VALUES.

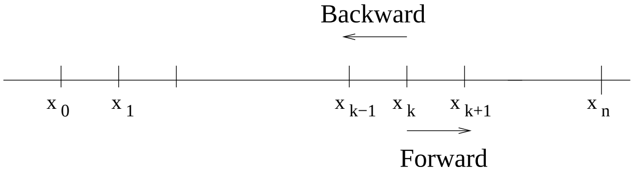
## 11.2 Difference Operator

### 11.2.1 Forward Difference Operator

**Definition 11.2.1 (First Forward Difference Operator)** We define the FORWARD DIFFERENCE OPERATOR, denoted by $\Delta$, as

$$\Delta f(x) = f(x + h) - f(x).$$

The expression $f(x + h) - f(x)$ gives the FIRST FORWARD DIFFERENCE of $f(x)$ and the operator $\Delta$ is called the FIRST FORWARD DIFFERENCE OPERATOR. Given the step size $h$, this formula uses the values at $x$ and $x + h$, the point at the next step. As it is moving in the forward direction, it is called the forward difference operator.



**Definition 11.2.2 (Second Forward Difference Operator)** The second forward difference operator, $\Delta^2$, is defined as

$$\Delta^2 f(x) = \Delta\big(\Delta f(x)\big) = \Delta f(x+h) - \Delta f(x).$$

We note that

$$\begin{aligned}\Delta^2 f(x) &= \Delta f(x+h) - \Delta f(x)\\ &= \big(f(x+2h) - f(x+h)\big) - \big(f(x+h) - f(x)\big)\\ &= f(x+2h) - 2f(x+h) + f(x).\end{aligned}$$

In particular, for $x = x_k$, we get,

$$\Delta y_k = y_{k+1} - y_k$$

and

$$\Delta^2 y_k = \Delta y_{k+1} - \Delta y_k = y_{k+2} - 2y_{k+1} + y_k.$$

**Definition 11.2.3 ($r^{\text{th}}$ Forward Difference Operator)** The $r^{\text{th}}$ forward difference operator, $\Delta^r$, is defined as

$$\begin{aligned}\Delta^r f(x) &= \Delta^{r-1} f(x+h) - \Delta^{r-1} f(x), \qquad\qquad r = 1, 2, \ldots,\\ \text{with} \quad & \Delta^0 f(x) = f(x).\end{aligned}$$

**Exercise 11.2.4** Show that $\Delta^3 y_k = \Delta^2(\Delta y_k) = \Delta(\Delta^2 y_k)$. In general, show that for any positive integers $r$ and $m$ with $r > m$,

$$\Delta^r y_k = \Delta^{r-m}(\Delta^m y_k) = \Delta^m(\Delta^{r-m} y_k).$$

**Example 11.2.5** For the tabulated values of $y = f(x)$ find $\Delta y_3$ and $\Delta^3 y_2$

| $i$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $x_i$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $y_i$ | 0.05 | 0.11 | 0.26 | 0.35 | 0.49 | 0.67 |

.

**Solution:** Here,

$$\Delta y_3 = y_4 - y_3 = 0.49 - 0.35 = 0.14, \quad \text{and}$$

$$\begin{aligned}\Delta^3 y_2 &= \Delta(\Delta^2 y_2) = \Delta(y_4 - 2y_3 + y_2)\\ &= (y_5 - y_4) - 2(y_4 - y_3) + (y_3 - y_2)\\ &= y_5 - 3y_4 + 3y_3 - y_2\\ &= 0.67 - 3 \times 0.49 + 3 \times 0.35 - 0.26 = -0.01.\end{aligned}$$

**Remark 11.2.6** *Using mathematical induction, it can be shown that*

$$\Delta^r y_k = \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} y_{k+j}.$$

*Thus the $r^{th}$ forward difference at $y_k$ uses the values at $y_k, y_{k+1}, \ldots, y_{k+r}$.*

**Example 11.2.7** If $f(x) = x^2 + ax + b$, where $a$ and $b$ are real constants, calculate $\Delta^r f(x)$.

**Solution:** We first calculate $\Delta f(x)$ as follows:

$$\begin{aligned}\Delta f(x) &= f(x+h) - f(x) = \big[(x+h)^2 + a(x+h) + b\big] - \big[x^2 + ax + b\big] \\ &= 2xh + h^2 + ah.\end{aligned}$$

Now,

$$\begin{aligned}\Delta^2 f(x) &= \Delta f(x+h) - \Delta f(x) = [2(x+h)h + h^2 + ah] - [2xh + h^2 + ah] = 2h^2, \\ \text{and} \qquad \Delta^3 f(x) &= \Delta^2 f(x) - \Delta^2 f(x) = 2h^2 - 2h^2 = 0.\end{aligned}$$

Thus, $\Delta^r f(x) = 0 \qquad \text{for all} \;\; r \geq 3.$

**Remark 11.2.8** *In general, if $f(x) = x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1}x + a_n$ is a polynomial of degree $n$, then it can be shown that*

$$\Delta^n f(x) = n!\, h^n \quad \textit{and} \quad \Delta^{n+r} f(x) = 0 \qquad \textit{for} \;\; r = 1, 2, \ldots.$$

The reader is advised to prove the above statement.

**Remark 11.2.9** 1. *For a set of tabular values, the horizontal forward difference table is written as:*

$$\begin{array}{cccccc}
x_0 & y_0 & \Delta y_0 = y_1 - y_0 & \Delta^2 y_0 = \Delta y_1 - \Delta y_0 & \cdots & \Delta^n y_0 = \Delta^{n-1} y_1 - \Delta^{n-1} y_0 \\
x_1 & y_1 & \Delta y_1 = y_2 - y_1 & \Delta^2 y_1 = \Delta y_2 - \Delta y_1 & \cdots & \\
x_2 & y_2 & \Delta y_2 = y_3 - y_2 & \Delta^2 y_2 = \Delta y_3 - \Delta y_2 & & \\
\vdots & & & & & \\
x_{n-1} & y_{n-1} & \Delta y_{n-1} = y_n - y_{n-1} & & & \\
x_n & y_n & & & &
\end{array}$$

2. *In many books, a diagonal form of the difference table is also used. This is written as:*

$$\begin{array}{cccccc}
x_0 & y_0 & & & & \\
 & & \Delta y_0 & & & \\
x_1 & y_1 & & \Delta^2 y_0 & & \\
 & & \Delta y_1 & & \Delta^3 y_0 & \\
x_2 & y_2 & & \Delta^2 y_1 & & \\
\vdots & & & & & \Delta y_{n-1} \\
x_{n-2} & y_{n-2} & & \Delta^2 y_{n-3} & & \\
 & & \Delta y_{n-2} & & \Delta^3 y_{n-3} & \\
x_{n-1} & y_{n-1} & & \Delta^2 y_{n-2} & & \\
 & & \Delta y_{n-1} & & & \\
x_n & y_n & & & &
\end{array}$$

*However, in the following, we shall mostly adhere to horizontal form only.*

### 11.2.2 Backward Difference Operator

**Definition 11.2.10 (First Backward Difference Operator)** The FIRST BACKWARD DIFFERENCE OPERATOR, denoted by $\nabla$, is defined as

$$\nabla f(x) = f(x) - f(x-h).$$

Given the step size $h$, note that this formula uses the values at $x$ and $x-h$, the point at the previous step. As it moves in the backward direction, it is called the backward difference operator.

**Definition 11.2.11 ($r^{\text{th}}$ Backward Difference Operator)** The $r^{\text{th}}$ backward difference operator, $\nabla^r$, is defined as

$$\begin{aligned} \nabla^r f(x) &= \nabla^{r-1} f(x) - \nabla^{r-1} f(x-h), \qquad\qquad r = 1, 2, \ldots, \\ \text{with} \quad & \nabla^0 f(x) = f(x). \end{aligned}$$

In particular, for $x = x_k$, we get

$$\nabla y_k = y_k - y_{k-1} \text{ and } \nabla^2 y_k = y_k - 2y_{k-1} + y_{k-2}.$$

Note that $\nabla^2 y_k = \Delta^2 y_{k-2}$.

**Example 11.2.12** Using the tabulated values in Example 11.2.5, find $\nabla y_4$ and $\nabla^3 y_3$.

**Solution:** We have $\nabla y_4 = y_4 - y_3 = 0.49 - 0.35 = 0.14$, and

$$\begin{aligned} \nabla^3 y_3 &= \nabla^2 y_3 - \nabla^2 y_2 = (y_3 - 2y_2 + y_1) - (y_2 - 2y_1 + y_0) \\ &= y_3 - 3y_2 + 3y_1 - y_0 \\ &= 0.35 - 3 \times 0.26 + 3 \times 0.11 - 0.05 = -0.15. \end{aligned}$$

**Example 11.2.13** If $f(x) = x^2 + ax + b$, where $a$ and $b$ are real constants, calculate $\nabla^r f(x)$.

**Solution:** We first calculate $\nabla f(x)$ as follows:

$$\begin{aligned} \nabla f(x) &= f(x) - f(x-h) = \big[x^2 + ax + b\big] - \big[(x-h)^2 + a(x-h) + b\big] \\ &= 2xh - h^2 + ah. \end{aligned}$$

Now,

$$\begin{aligned} \nabla^2 f(x) &= \nabla f(x) - \Delta f(x-h) = [2xh - h^2 + ah] - [2(x-h)h - h^2 + ah] = 2h^2, \\ \text{and} \quad \nabla^3 f(x) &= \nabla^2 f(x) - \nabla^2 f(x) = 2h^2 - 2h^2 = 0. \end{aligned}$$

Thus, $\nabla^r f(x) = 0 \qquad \text{for all} \;\; r \geq 3.$

**Remark 11.2.14** *For a set of tabular values, backward difference table in the horizontal form is written as:*

| | | | | | |
|---|---|---|---|---|---|
| $x_0$ | $y_0$ | | | | |
| $x_1$ | $y_1$ | $\nabla y_1 = y_1 - y_0$ | | | |
| $x_2$ | $y_2$ | $\nabla y_2 = y_2 - y_1$ | $\nabla^2 y_2 = \nabla y_2 - \nabla y_1$ | | |
| $\vdots$ | | | | | |
| $x_{n-2}$ | $y_{n-2}$ | $\cdots$ | $\cdots$ | | |
| $x_{n-1}$ | $y_{n-1}$ | $\nabla y_{n-1} = y_{n-1} - y_{n-2}$ | $\cdots$ | $\cdots$ | |
| $x_n$ | $y_n$ | $\nabla y_n = y_n - y_{n-1}$ | $\nabla^2 y_n = \nabla y_n - \nabla y_{n-1}$ | $\cdots$ | $\nabla^n y_n = \nabla^{n-1} y_n - \nabla^{n-1} y_{n-1}$ |

**Example 11.2.15** For the following set of tabular values $(x_i, y_i)$, write the forward and backward difference tables.

| $x_i$ | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|
| $y_i$ | 5.0 | 5.4 | 6.0 | 6.8 | 7.5 | 8.7 |

**Solution:** The forward difference table is written as

| $x$ | $y$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ | $\Delta^4 y$ | $\Delta^5 y$ |
|---|---|---|---|---|---|---|
| 9 | 5 | 0.4 = 5.4 - 5 | 0.2 = 0.6 - 0.4 | 0= 0.2-0.2 | -.3 = -0.3 - 0.0 | 0.6 = 0.3 - (-0.3) |
| 10 | 5.4 | 0.6 | 0.2 | -0.3 | 0.3 | |
| 11 | 6.0 | 0.8 | -0.1 | 0.0 | | |
| 12 | 6.8 | 0.7 | -0.1 | | | |
| 13 | 7.5 | 0.6 | | | | |
| 14 | 8.1 | | | | | |

In the similar manner, the backward difference table is written as follows:

| $x$ | $y$ | $\nabla y$ | $\nabla^2 y$ | $\nabla^3 y$ | $\nabla^4 y$ | $\nabla^5 y$ |
|---|---|---|---|---|---|---|
| 9 | 5 | | | | | |
| 10 | 5.4 | 0.4 | | | | |
| 11 | 6 | 0.6 | 0.2 | | | |
| 12 | 6.8 | 0.8 | 0.2 | 0.0 | | |
| 13 | 7.5 | 0.7 | -0.1 | - 0.3 | -0.3 | |
| 14 | 8.1 | 0.6 | -0.1 | 0.0 | 0.3 | 0.6 |

Observe from the above two tables that $\Delta^3 y_1 = \nabla^3 y_4$, $\Delta^2 y_3 = \nabla^2 y_5$ , $\Delta^4 y_1 = \nabla^4 y_5$ etc.

**Exercise 11.2.16** 1. Show that $\Delta^3 y_4 = \nabla^3 y_7$.

2. Prove that $\Delta(\nabla y_k) = \Delta^2 y_{k+1} = \nabla^2 y_{k-1}$.

3. Obtain $\nabla^k y_k$ in terms of $y_0, y_1, y_2, \ldots, y_k$. Hence show that $\nabla^k y_k = \Delta^k y_0$.

**Remark 11.2.17** *In general it can be shown that* $\Delta^k f(x) = \nabla^k f(x + kh)$ *or* $\Delta^k y_m = \nabla^k y_{k+m}$

**Remark 11.2.18** *In view of the remarks (11.2.8) and (11.2.17) it is obvious that, if* $y = f(x)$ *is a polynomial function of degree* $n$*, then* $\nabla^n f(x)$ *is constant and* $\nabla^{n+r} f(x) = 0$ *for* $r > 0$.

### 11.2.3 Central Difference Operator

**Definition 11.2.19 (Central Difference Operator)** The FIRST CENTRAL DIFFERENCE OPERATOR, denoted by $\delta$, is defined by

$$\delta f(x) = f(x + \frac{h}{2}) - f(x - \frac{h}{2})$$

and the $r^{\text{TH}}$ CENTRAL DIFFERENCE OPERATOR is defined as

$$\begin{aligned} \delta^r f(x) &= \delta^{r-1} f(x + \frac{h}{2}) - \delta^{r-1} f(x - \frac{h}{2}) \\ \text{with} \quad & \delta^0 f(x) = f(x). \end{aligned}$$

Thus, $\delta^2 f(x) = f(x+h) - 2f(x) + f(x-h)$.

In particular, for $x = x_k$, define $y_{k+\frac{1}{2}} = f(x_k + \frac{h}{2})$, and $y_{k-\frac{1}{2}} = f(x_k - \frac{h}{2})$, then

$$\delta y_k = y_{k+\frac{1}{2}} - y_{k-\frac{1}{2}} \quad \text{and} \quad \delta^2 y_k = y_{k+1} - 2y_k + y_{k-1}.$$

Thus, $\delta^2$ uses the table of $(x_k, y_k)$. It is easy to see that only the even central differences use the tabular point values $(x_k, y_k)$.

### 11.2.4 Shift Operator

**Definition 11.2.20 (Shift Operator)** A SHIFT OPERATOR, denoted by $E$, is the operator which shifts the value at the next point with step $h$, *i.e.,*

$$Ef(x) = f(x+h).$$

Thus,

$$Ey_i = y_{i+1}, \quad E^2 y_i = y_{i+2}, \quad \text{and} \quad E^k y_i = y_{i+k}.$$

### 11.2.5 Averaging Operator

**Definition 11.2.21 (Averaging Operator)** The AVERAGING OPERATOR, denoted by $\mu$, gives the average value between two central points, *i.e.,*

$$\mu f(x) = \frac{1}{2}\big[f(x+\frac{h}{2}) + f(x-\frac{h}{2})\big].$$

Thus $\mu\, y_i = \frac{1}{2}(y_{i+\frac{1}{2}} + y_{i-\frac{1}{2}})$ and

$$\mu^2\, y_i = \frac{1}{2}\left[\mu\, y_{i+\frac{1}{2}} + \mu\, y_{i-\frac{1}{2}}\right] = \frac{1}{4}\left[y_{i+1} + 2y_i + y_{i-1}\right].$$

## 11.3 Relations between Difference operators

1. We note that

$$Ef(x) = f(x+h) \quad = [f(x+h) - f(x)] + f(x) \quad = \Delta f(x) + f(x) = (\Delta + 1)f(x).$$

Thus,

$$\boxed{E \equiv 1 + \Delta} \quad \text{or} \quad \Delta \equiv E - 1.$$

2. Further, $\nabla(E(f(x)) = \nabla(f(x+h)) = f(x+h) - f(x)$. Thus,

$$(1-\nabla)Ef(x) = E(f(x)) - \nabla(E(f(x)) = f(x+h) - [f(x+h) - f(x)] = f(x).$$

Thus $E \equiv 1 + \Delta$, gives us

$$(1-\nabla)(1+\Delta)f(x) = f(x) \text{ for all } x.$$

So we write,

$$(1+\Delta)^{-1} = 1 - \nabla \quad \text{or} \quad \boxed{\nabla = 1 - (1+\Delta)^{-1},} \quad \text{and}$$
$$(1-\nabla)^{-1} = 1 + \Delta = E.$$

Similarly,

$$\Delta = (1-\nabla)^{-1} - 1.$$

3. Let us denote by $E^{\frac{1}{2}}f(x) = f(x+\frac{h}{2})$. Then, we see that

$$\delta f(x) = f(x+\frac{h}{2}) - f(x-\frac{h}{2}) = E^{\frac{1}{2}}f(x) - E^{-\frac{1}{2}}f(x).$$

Thus,

$$\boxed{\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}.}$$

Recall,

$$\delta^2 f(x) = f(x+h) - 2f(x) + f(x-h) = [f(x+h) + 2f(x) + f(x-h)] - 4f(x) = 4(\mu^2 - 1)f(x).$$

So, we have,

$$\boxed{\mu^2 \equiv \frac{\delta^2}{4}+1} \quad \text{or} \quad \boxed{\mu \equiv \sqrt{1+\frac{\delta^2}{4}}}.$$

That is, the action of $\sqrt{1+\dfrac{\delta^2}{4}}$ is same as that of $\mu$.

4. We further note that,

$$\begin{aligned}\Delta f(x) &= f(x+h)-f(x) = \frac{1}{2}\big[f(x+h)-2f(x)+f(x-h)\big]+\frac{1}{2}\big[f(x+h)-f(x-h)\big]\\ &= \frac{1}{2}\delta^2(f(x))+\frac{1}{2}\big[f(x+h)-f(x-h)\big]\end{aligned}$$

and

$$\begin{aligned}\delta\mu f(x) &= \delta\left[\frac{1}{2}\left\{f(x+\frac{h}{2})+f(x-\frac{h}{2})\right\}\right] = \frac{1}{2}\big[\{f(x+h)-f(x)\}+\{f(x)-f(x-h)\}\big]\\ &= \frac{1}{2}\left[f(x+h)-f(x-h)\right].\end{aligned}$$

Thus,

$$\Delta f(x) = \left[\frac{1}{2}\delta^2+\delta\mu\right]f(x),$$

*i.e.*,

$$\Delta \equiv \frac{1}{2}\delta^2+\delta\mu \equiv \frac{1}{2}\delta^2+\delta\sqrt{1+\frac{\delta^2}{4}}.$$

In view of the above discussion, we have the following table showing the relations between various difference operators:

| | E | $\Delta$ | $\nabla$ | $\delta$ |
|---|---|---|---|---|
| E | E | $\Delta+1$ | $(1-\nabla)^{-1}$ | $\frac{1}{2}\delta^2+\delta\sqrt{1+\frac{\delta^2}{4}}+1$ |
| $\Delta$ | $E-1$ | $\Delta$ | $(1-\nabla)^{-1}-1$ | $\frac{1}{2}\delta^2+\delta\sqrt{1+\frac{1}{4}\delta^2}$ |
| $\nabla$ | $1-E^{-1}$ | $1-(1+\nabla)^{-1}$ | $\nabla$ | $-\frac{1}{2}\delta^2+\delta\sqrt{1+\frac{1}{4}\delta^2}$ |
| $\delta$ | $E^{1/2}-E^{-1/2}$ | $\Delta(1+\Delta)^{-1/2}$ | $\nabla(1-\nabla)^{-1/2}$ | $\delta$ |

**Exercise 11.3.1** 1. Verify the validity of the above table.

2. Obtain the relations between the averaging operator and other difference operators.

3. Find $\Delta^2 y_2$, $\nabla^2 y_2$, $\delta^2 y_2$ and $\mu^2 y_2$ for the following tabular values:

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $x_i$ | 93.0 | 96.5 | 100.0 | 103.5 | 107.0 |
| $y_i$ | 11.3 | 12.5 | 14.0 | 15.2 | 16.0 |

## 11.4 Newton's Interpolation Formulae

As stated earlier, interpolation is the process of approximating a given function, whose values are known at $N+1$ tabular points, by a suitable polynomial, $P_N(x)$, of degree $N$ which takes the values $y_i$ at $x=x_i$ for $i=0,1,\ldots,N$. Note that if the given data has errors, it will also be reflected in the polynomial so obtained.

In the following, we shall use forward and backward differences to obtain polynomial function approximating $y=f(x)$, when the tabular points $x_i$'s are equally spaced. Let

$$f(x) \approx P_N(x),$$

where the polynomial $P_N(x)$ is given in the following form:

$$\begin{aligned} P_N(x) &= a_0 + a_1(x-x_0) + a_2(x-x_0)(x-x_1) + \cdots + a_k(x-x_0)(x-x_1)\cdots(x-x_{k-1}) \\ &\quad + a_N(x-x_0)(x-x_1)\cdots(x-x_{N-1}). \end{aligned} \tag{11.4.1}$$

for some constants $a_0, a_1, ...a_N$, to be determined using the fact that $P_N(x_i) = y_i$ for $i = 0, 1, \ldots, N$.

So, for $i = 0$, substitute $x = x_0$ in (11.4.1) to get $P_N(x_0) = y_0$. This gives us $a_0 = y_0$. Next,

$$P_N(x_1) = y_1 \Rightarrow y_1 = a_0 + (x_1 - x_0)a_1.$$

So, $a_1 = \frac{y_1 - y_0}{h} = \dfrac{\Delta y_0}{h}$. For $i = 2$, $\quad y_2 = a_0 + (x_2 - x_0)a_1 + (x_2 - x_1)(x_2 - x_0)a_2$, or equivalently

$$2h^2 a_2 = y_2 - y_0 - 2h(\frac{\Delta y_0}{h}) = y_2 - 2y_1 + y_0 = \Delta^2 y_0.$$

Thus, $a_2 = \dfrac{\Delta^2 y_0}{2h^2}$. Now, using mathematical induction, we get

$$a_k = \frac{\Delta^k y_0}{k!\, h^k} \quad \text{for} \quad k = 0, 1, 2, \ldots, N.$$

Thus,

$$\begin{aligned} P_N(x) &= y_0 + \frac{\Delta y_0}{h}(x-x_0) + \frac{\Delta^2 y_0}{2!\, h^2}(x-x_0)(x-x_1) + \cdots + \frac{\Delta^k y_0}{k!\, h^k}(x-x_0)\cdots(x-x_{k-1}) \\ &\quad + \frac{\Delta^N y_0}{N!\, h^N}(x-x_0)...(x-x_{N-1}). \end{aligned}$$

As this uses the forward differences, it is called NEWTON'S FORWARD DIFFERENCE FORMULA for interpolation, or simply, forward interpolation formula.

**Exercise 11.4.1** Show that

$$a_3 = \frac{\Delta^3 y_0}{3!\, h^3} \quad \text{and} \quad a_4 = \frac{\Delta^4 y_0}{4!\, h^2}$$

and in general,

$$a_k = \frac{\Delta^k y_0}{k! h^k}, \quad \text{for} \quad k = 0, 1, 2, \ldots, N.$$

For the sake of numerical calculations, we give below a convenient form of the forward interpolation formula.

Let $u = \dfrac{x - x_0}{h}$, then

$$x - x_1 = hu + x_0 - (x_0 + h) = h(u-1), x - x_2 = h(u-2), \ldots, x - x_k = h(u-k), \quad \text{etc..}$$

With this transformation the above forward interpolation formula is simplified to the following form:

$$\begin{aligned} P_N(u) &= y_0 + \frac{\Delta y_0}{h}(hu) + \frac{\Delta^2 y_0}{2!\, h^2}\{(hu)(h(u-1))\} + \cdots + \frac{\Delta^k y_0 h^k}{k!\, h^k}\big[u(u-1)\cdots(u-k+1)\big] \\ &\quad + \cdots + \frac{\Delta^N y_0}{N!\, h^N}\Big[(hu)\big(h(u-1)\big)\cdots\big(h(u-N+1)\big)\Big]. \\ &= y_0 + \Delta y_0(u) + \frac{\Delta^2 y_0}{2!}(u(u-1)) + \cdots + \frac{\Delta^k y_0}{k!}\Big[u(u-1)\cdots(u-k+1)\Big] \\ &\quad + \cdots + \frac{\Delta^N y_0}{N!}\Big[u(u-1)...(u-N+1)\Big]. \end{aligned} \tag{11.4.2}$$

If $N$=1, we have a linear interpolation given by

$$f(u) \approx y_0 + \Delta y_0(u). \tag{11.4.3}$$

For $N = 2$, we get a quadratic interpolating polynomial:

$$f(u) \approx y_0 + \Delta y_0(u) + \frac{\Delta^2 y_0}{2!}[u(u-1)] \tag{11.4.4}$$

and so on.

It may be pointed out here that if $f(x)$ is a polynomial function of degree $N$ then $P_N(x)$ coincides with $f(x)$ on the given interval. Otherwise, this gives only an approximation to the true values of $f(x)$.

If we are given additional point $x_{N+1}$ also, then the error, denoted by $R_N(x) = |P_N(x) - f(x)|$, is estimated by

$$R_N(x) \simeq \left| \frac{\Delta^{N+1} y_0}{h^{N+1}(N+1)!}(x - x_0) \cdots (x - x_N) \right|.$$

Similarly, if we assume, $P_N(x)$ is of the form

$$P_N(x) = b_0 + b_1(x - x_N) + b_1(x - x_N)(x - x_{N-1}) + \cdots + b_N(x - x_N)(x - x_{N-1}) \cdots (x - x_1),$$

then using the fact that $P_N(x_i) = y_i$, we have

$$\begin{aligned}
b_0 &= y_N \\
b_1 &= \frac{1}{h}(y_N - y_{N-1}) = \frac{1}{h}\nabla y_N \\
b_2 &= \frac{y_N - 2y_{N-1} + y_{N-2}}{2h^2} = \frac{1}{2h^2}(\nabla^2 y_N) \\
&\vdots \\
b_k &= \frac{1}{k!\, h^k}\nabla^k y_N.
\end{aligned}$$

Thus, using backward differences and the transformation $x = x_N + hu$, we obtain the Newton's backward interpolation formula as follows:

$$P_N(u) = y_N + u\nabla y_N + \frac{u(u+1)}{2!}\nabla^2 y_N + \cdots + \frac{u(u+1)\cdots(u+N-1)}{N!}\nabla^N y_N. \tag{11.4.5}$$

**Exercise 11.4.2** Derive the Newton's backward interpolation formula (11.4.5) for $N = 3$.

**Remark 11.4.3** *If the interpolating point lies closer to the beginning of the interval then one uses the Newton's forward formula and if it lies towards the end of the interval then Newton's backward formula is used.*

**Remark 11.4.4** *For a given set of n tabular points, in general, all the n points need not be used for interpolating polynomial. In fact N is so chosen that $N^{th}$ forward/backward difference almost remains constant. Thus N is less than or equal to n.*

**Example 11.4.5** 1. Obtain the Newton's forward interpolating polynomial, $P_5(x)$ for the following tabular data and interpolate the value of the function at $x = 0.0045$.

| | | | | | | |
|---|---|---|---|---|---|---|
| x | 0 | 0.001 | 0.002 | 0.003 | 0.004 | 0.005 |
| y | 1.121 | 1.123 | 1.1255 | 1.127 | 1.128 | 1.1285 |

**Solution:** For this data, we have the Forward difference difference table

| $x_i$ | $y_i$ | $\Delta y_i$ | $\Delta^2 y_3$ | $\Delta^3 y_i$ | $\Delta^4 y_i$ | $\Delta^5 y_i$ |
|---|---|---|---|---|---|---|
| 0 | 1.121 | 0.002 | 0.0005 | -0.0015 | 0.002 | -.0025 |
| .001 | 1.123 | 0.0025 | -0.0010 | 0.0005 | -0.0005 | |
| .002 | 1.1255 | 0.0015 | -0.0005 | 0.0 | | |
| .003 | 1.127 | 0.001 | -0.0005 | | | |
| .004 | 1.128 | 0.0005 | | | | |
| .005 | 1.1285 | | | | | |

Thus, for $x = x_0 + hu$, where $x_0 = 0,\ h = 0.001$ and $u = \dfrac{x - x_0}{h}$, we get

$$\begin{aligned} P_5(x) &= 1.121 + u \times .002 + \frac{u(u-1)}{2}(.0005) + \frac{u(u-1)(u-2)}{3!} \times (-.0015) \\ &\quad + \frac{u(u-1)(u-2)(u-3)}{4!}(.002) + \frac{u(u-1)(u-2)(u-3)(u-4)}{5!} \times (-.0025). \end{aligned}$$

Thus,

$$\begin{aligned} P_5(0.0045) &= P_5(0 + 0.001 \times 4.5) \\ &= 1.121 + 0.002 \times 4.5 + \frac{0.0005}{2} \times 4.5 \times 3.5 - \frac{0.0015}{6} \times 4.5 \times 3.5 \times 2.5 \\ &\quad + \frac{0.002}{24} \times 4.5 \times 3.5 \times 2.5 \times 1.5 - \frac{0.0025}{120} \times 4.5 \times 3.5 \times 2.5 \times 1.5 \times 0.5 \\ &= 1.12840045. \end{aligned}$$

2. Using the following table for $\tan x$, approximate its value at $0.71$. Also, find an error estimate (Note $\tan(0.71) = 0.85953$).

| $x_i$ | 0.70 | 72 | 0.74 | 0.76 | 0.78 |
|---|---|---|---|---|---|
| $\tan x_i$ | 0.84229 | 0.87707 | 0.91309 | 0.95045 | 0.98926 |

**Solution:** As the point $x = 0.71$ lies towards the initial tabular values, we shall use Newton's Forward formula. The forward difference table is:

| $x_i$ | $y_i$ | $\Delta y_i$ | $\Delta^2 y_i$ | $\Delta^3 y_i$ | $\Delta^4 y_i$ |
|---|---|---|---|---|---|
| 0.70 | 0.84229 | 0.03478 | 0.00124 | 0.0001 | 0.00001 |
| 0.72 | 0.87707 | 0.03602 | 0.00134 | 0.00011 | |
| 0.74 | 0.91309 | 0.03736 | 0.00145 | | |
| 0.76 | 0.95045 | 0.03881 | | | |
| 0.78 | 0.98926 | | | | |

In the above table, we note that $\Delta^3 y$ is almost constant, so we shall attempt $3^{\text{rd}}$ degree polynomial interpolation.

Note that $x_0 = 0.70,\ h = 0.02$ gives $u = \dfrac{0.71 - 0.70}{0.02} = 0.5$. Thus, using forward interpolating polynomial of degree $3$, we get

$$P_3(u) = 0.84229 + 0.03478u + \frac{0.00124}{2!}u(u-1) + \frac{0.0001}{3!}u(u-1)(u-2).$$

$$\text{Thus,} \quad \begin{aligned} \tan(0.71) &\approx 0.84229 + 0.03478(0.5) + \frac{0.00124}{2!} \times 0.5 \times (-0.5) \\ &\quad + \frac{0.0001}{3!} \times 0.5 \times (-0.5) \times (-1.5) \\ &= 0.859535. \end{aligned}$$

An error estimate for the approximate value is

$$\left. \frac{\Delta^4 y_0}{4!} u(u-1)(u-2)(u-3) \right|_{u=0.5} = 0.00000039.$$

Note that exact value of $\tan(0.71)$ (upto $5$ decimal place) is $0.85953$. and the approximate value, obtained using the Newton's interpolating polynomial is very close to this value. This is also reflected by the error estimate given above.

3. Apply $3^{\text{rd}}$ degree interpolation polynomial for the set of values given in Example 11.2.15, to estimate the value of $f(10.3)$ by taking

$$(i)\ x_0 = 9.0, \qquad (ii)\ x_0 = 10.0.$$

Also, find approximate value of $f(13.5)$.

**Solution:** Note that $x = 10.3$ is closer to the values lying in the beginning of tabular values, while $x = 13.5$ is towards the end of tabular values. Therefore, we shall use forward difference formula for $x = 10.3$ and the backward difference formula for $x = 13.5$. Recall that the interpolating polynomial of degree 3 is given by

$$f(x_0 + hu) = y_0 + \Delta y_0 u + \frac{\Delta^2 y_0}{2!}u(u-1) + \frac{\Delta^3 y_0}{3!}u(u-1)(u-2).$$

Therefore,

(a) for $x_0 = 9.0,\ h = 1.0$ and $x = 10.3$, we have $u = \dfrac{10.3 - 9.0}{1} = 1.3$. This gives,

$$\begin{aligned} f(10.3) &\approx 5 + .4 \times 1.3 + \frac{.2}{2!}(1.3) \times .3 + \frac{.0}{3!}(1.3) \times .3 \times (-0.7) \\ &= 5.559. \end{aligned}$$

(b) for $x_0 = 10.0,\ h = 1.0$ and $x = 10.3$, we have $u = \dfrac{10.3 - 10.0}{1} = .3$. This gives,

$$\begin{aligned} f(10.3) &\approx 5.4 + .6 \times .3 + \frac{.2}{2!}(.3) \times (-0.7) + \frac{-0.3}{3!}(.3) \times (-0.7) \times (-1.7) \\ &= 5.54115. \end{aligned}$$

**Note:** as $x = 10.3$ is closer to $x = 10.0$, we may expect estimate calculated using $x_0 = 10.0$ to be a better approximation.

(c) for $x_0 = 13.5$, we use the backward interpolating polynomial, which gives,

$$f(x_N + hu) \approx y_0 + \nabla y_N u + \frac{\nabla^2 y_N}{2!}u(u+1) + \frac{\Delta^3 y_N}{3!}u(u+1)(u+2).$$

Therefore, taking $x_N = 14,\ h = 1.0$ and $x = 13.5$, we have $u = \dfrac{13.5 - 14}{1} = -0.5$. This gives,

$$\begin{aligned} f(13.5) &\approx 8.1 + .6 \times (-0.5) + \frac{-0.1}{2!}(-0.5) \times 0.5 + \frac{0.0}{3!}(-0.5) \times 0.5 \times (1.5) \\ &= 7.8125. \end{aligned}$$

**Exercise 11.4.6** 1. Following data is available for a function $y = f(x)$

| x | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
|---|---|---|---|---|---|---|
| y | 1.0 | 0.808 | 0.664 | 0.616 | 0.712 | 1.0 |

Compute the value of the function at $x = 0.3$ and $x = 1.1$

2. The speed of a train, running between two station is measured at different distances from the starting station. If $x$ is the distance in $km.$ from the starting station, then $v(x)$, the speed (in $km/hr$) of the train at the distance $x$ is given by the following table:

| x | 0 | 50 | 100 | 150 | 200 | 250 |
|---|---|---|---|---|---|---|
| v(x) | 0 | 60 | 80 | 110 | 90 | 0 |

Find the approximate speed of the train at the mid point between the two stations.

3. Following table gives the values of the function $S(x) = \int\limits_0^x sin(\frac{\pi}{2}t^2)dt$ at the different values of the tabular points $x$,

| x | 0 | 0.04 | 0.08 | 0.12 | 0.16 | 0.20 |
|---|---|---|---|---|---|---|
| S(x) | 0 | 0.00003 | 0.00026 | 0.00090 | 0.00214 | 0.00419 |

Obtain a fifth degree interpolating polynomial for $S(x)$. Compute $S(0.02)$ and also find an error estimate for it.

4. Following data gives the temperatures (in $^oC$) between 8.00 am to 8.00 pm. on May 10, 2005 in Kanpur:

| Time | 8 am | 12 noon | 4 pm | 8pm |
|---|---|---|---|---|
| Temperature | 30 | 37 | 43 | 38 |

Obtain Newton's backward interpolating polynomial of degree $3$ to compute the temperature in Kanpur on that day at 5.00 pm.

# Chapter 12

# Lagrange's Interpolation Formula

## 12.1 Introduction

In the previous chapter, we derived the interpolation formula when the values of the function are given at equidistant tabular points $x_0, x_1, \ldots, x_N$. However, it is not always possible to obtain values of the function, $y = f(x)$ at equidistant interval points, $x_i$. In view of this, it is desirable to derive an interpolating formula, which is applicable even for unequally distant points. Lagrange's Formula is one such interpolating formula. Unlike the previous interpolating formulas, it does not use the notion of differences, however we shall introduce the concept of divided differences before coming to it.

## 12.2 Divided Differences

**Definition 12.2.1 (First Divided Difference)** The ratio

$$\frac{f(x_i) - f(x_j)}{x_i - x_j}$$

for any two points $x_i$ and $x_j$ is called the FIRST DIVIDED DIFFERENCE of $f(x)$ relative to $x_i$ and $x_j$. It is denoted by $\delta[x_i, x_j]$.

Let us assume that the function $y = f(x)$ is linear. Then $\delta[x_i, x_j]$ is constant for any two tabular points $x_i$ and $x_j$, *i.e.,* it is independent of $x_i$ and $x_j$. Hence,

$$\delta[x_i, x_j] = \frac{f(x_i) - f(x_j)}{x_i - x_j} = \delta[x_j, x_i].$$

Thus, for a linear function $f(x)$, if we take the points $x, x_0$ and $x_1$ then, $\delta[x_0, x] = \delta[x_0, x_1]$, *i.e.,*

$$\frac{f(x) - f(x_0)}{x - x_0} = \delta[x_0, x_1].$$

Thus, $f(x) = f(x_0) + (x - x_0)\delta[x_0, x_1]$.

So, if $f(x)$ is approximated with a linear polynomial, then the value of the function at any point $x$ can be calculated by using $f(x) \approx P_1(x) = f(x_0) + (x - x_0)\delta[x_0, x_1]$, where $\delta[x_0, x_1]$ is the first divided difference of $f$ relative to $x_0$ and $x_1$.

**Definition 12.2.2 (Second Divided Difference)** The ratio

$$\delta[x_i, x_j, x_k] = \frac{\delta[x_j, x_k] - \delta[x_i, x_j]}{x_k - x_i}$$

is defined as SECOND DIVIDED DIFFERENCE of $f(x)$ relative to $x_i, x_j$ and $x_k$.

If $f(x)$ is a second degree polynomial then $\delta[x_0, x]$ is a linear function of $x$. Hence,

$$\delta[x_i, x_j, x_k] = \frac{\delta[x_j, x_k] - \delta[x_i, x_j]}{x_k - x_i} \quad \text{is constant.}$$

In view of the above, for a polynomial function of degree 2, we have $\delta[x, x_0, x_1] = \delta[x_0, x_1, x_2]$. Thus,

$$\frac{\delta[x, x_0] - \delta[x_0, x_1]}{x - x_1} = \delta[x_0, x_1, x_2].$$

This gives,

$$\delta[x, x_0] = \delta[x_0, x_1] + (x - x_1)\delta[x_0, x_1, x_2].$$

From this we obtain,

$$f(x) = f(x_0) + (x - x_0)\delta[x_0, x_1] + (x - x_0)(x - x_1)\delta[x_0, x_1, x_2].$$

So, whenever $f(x)$ is approximated with a second degree polynomial, the value of $f(x)$ at any point $x$ can be computed using the above polynomial, which uses the values at three points $x_0, x_1$ and $x_2$.

**Example 12.2.3** Using the following tabular values for a function $y = f(x)$, obtain its second degree polynomial approximation.

| $i$ | 0 | 1 | 2 |
|---|---|---|---|
| $x_i$ | 0.1 | 0.16 | 0.2 |
| $f(x_i)$ | 1.12 | 1.24 | 1.40 |

Also, find the approximate value of the function at $x = 0.13$.

**Solution:** We shall first calculate the desired divided differences.

$$\begin{aligned}
\delta[x_0, x_1] &= (1.24 - 1.12)/(0.16 - 0.1) = 2, \\
\delta[x_1, x_2] &= (1.40 - 1.24)/(0.2 - 0.16) = 4, \quad \text{and} \\
\delta[x_0, x_1, x_2] &= \frac{\delta[x_1, x_2] - \delta[x_0, x_1]}{x_2 - x_0} = (4 - 2)/(0.2 - 0.1) = 20.
\end{aligned}$$

Thus,

$$f(x) \approx P_2(x) = 1.12 + 2(x - 0.1) + 20(x - 0.1)(x - 0.16).$$

Therefore

$$f(0.13) \approx 1.12 + 2(0.13 - 0.1) + 20(0.13 - 0.1)(0.13 - 0.16) = 1.162.$$

**Exercise 12.2.4**

1. Using the following table, which gives values of $\log(x)$ corresponding to certain values of $x$, find approximate value of $\log(323.5)$ with the help of a second degree polynomial.

| $x$ | 322.8 | 324.2 | 325 |
|---|---|---|---|
| $\log(x)$ | 2.50893 | 2.51081 | 2.5118 |

2. Show that

$$\delta[x_0, x_1, x_2] = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)}.$$

So, $\delta[x_0, x_1, x_2] = \delta[x_0, x_2, x_1] = \delta[x_1, x_0, x_2] = \delta[x_1, x_2, x_0] = \delta[x_2, x_0, x_1] = \delta[x_2, x_1, x_0]$. That is, the second divided difference remains unchanged regardless of how its arguments are interchanged.

3. Show that for equidistant points $x_0$, $x_1$ and $x_2$, $\delta[x_0, x_1, x_2] = \dfrac{\Delta^2 y_0}{2h^2} = \dfrac{\nabla^2 y_2}{2h^2}$, where $y_k = f(x_k)$, and $h = x_1 - x_0 = x_2 - x_1$.

4. Show that for a linear function, the second divided difference with respect to any three points, $x_i, x_j$ and $x_k$, is always zero.

Now, we define the $k^{\text{th}}$ divided difference.

**Definition 12.2.5 ($k^{\text{th}}$ Divided Difference)** The $k^{\text{TH}}$ DIVIDED DIFFERENCE of $f(x)$ relative to the tabular points $x_0, x_1, \ldots, x_k$, is defined recursively as

$$\delta[x_0, x_1, \ldots, x_k] = \frac{\delta[x_1, x_2, \ldots, x_k] - \delta[x_0, x_1, \ldots, x_{k-1}]}{x_k - x_0}.$$

It can be shown by mathematical induction that for equidistant points,

$$\delta[x_0, x_1, \ldots, x_k] = \frac{\Delta^k y_0}{k! h^k} = \frac{\nabla^k y_k}{k! h^k} \tag{12.2.1}$$

where, $y_0 = f(x_0)$, and $h = x_1 - x_0 = x_2 - x_1 = \cdots = x_k - x_{k-1}$.

In general,

$$\delta[x_i, x_{i+1}, \ldots, x_{i+n}] = \frac{\Delta^n y_i}{n! h^n},$$

where $y_i = f(x_i)$ and $h$ is the length of the interval for $i = 0, 1, 2, \ldots$.

**Remark 12.2.6** *In view of the remark (11.2.18) and (12.2.1), it is easily seen that for a polynomial function of degree $n$, the $n^{th}$ divided difference is constant and the $(n+1)^{th}$ divided difference is zero.*

**Example 12.2.7** Show that $f(x)$ can be written as

$$f(x) = f(x_0) + \delta[x_0, x_1](x - x_0) + \delta[x, x_0, x_1](x - x_0)(x - x_1).$$

**Solution:**By definition, we have

$$\delta[x, x_0, x_1] = \frac{\delta[x, x_0] - \delta[x_0, x_1]}{(x - x_1)},$$

so, $\delta[x, x_0] = \delta[x_0, x_1] + (x - x_0)\delta[x, x_0, x_1]$. Now since,

$$\delta[x, x_0] = \frac{f(x) - f(x_0)}{(x - x_0)},$$

we get the desired result.

**Exercise 12.2.8** Show that $f(x)$ can be written in the following form:

$$f(x) = P_2(x) + R_3(x),$$

where, $P_2(x) = f(x_0) + \delta[x_0, x_1](x - x_0) + \delta[x_0, x_1, x_2](x - x_0)(x - x_1)$
and $R_3(x) = \delta[x, x_0, x_1, x_2](x - x_0)(x - x_1)(x - x_2)$.

Further show that $P_2(x_i) = f(x_i)$ for $i = 0, 1$.

**Remark 12.2.9** *In general it can be shown that $f(x) = P_n(x) + R_{n+1}(x)$, where,*

$$\begin{aligned} P_n(x) &= f(x_0) + \delta[x_0, x_1](x - x_0) + \delta[x_0, x_1, x_2](x - x_0)(x - x_1) + \cdots \\ &\quad + \delta[x_0, x_1, x_2, \ldots, x_n](x - x_0)(x - x_1)(x - x_2) \cdots (x - x_{n-1}), \end{aligned}$$

*and $R_{n+1}(x) = (x - x_0)(x - x_1)(x - x_2) \cdots (x - x_n)\delta[x, x_0, x_1, x_2, \ldots, x_n]$.*

*Here, $R_{n+1}(x)$ is called the remainder term.*

*It may be observed here that the expression $P_n(x)$ is a polynomial of degree $'n'$ and $P_n(x_i) = f(x_i)$ for $i = 0, 1, \cdots, (n-1)$.*

*Further, if $f(x)$ is a polynomial of degree $n$, then in view of the Remark 12.2.6, the remainder term, $R_{n+1}(x) = 0$, as it is a multiple of the $(n+1)^{th}$ divided difference, which is $0$.*

## 12.3 Lagrange's Interpolation formula

In this section, we shall obtain an interpolating polynomial when the given data has unequal tabular points. However, before going to that, we see below an important result.

**Theorem 12.3.1** The $k^{\text{th}}$ divided difference $\delta[x_0, x_1, \ldots, x_k]$ can be written as:

$$\begin{aligned}\delta[x_0, x_1, \ldots, x_k] &= \frac{f(x_0)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_k)} + \frac{f(x_1)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_k)} \\ &\quad + \cdots + \frac{f(x_k)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})} \\ &= \frac{f(x_0)}{\prod\limits_{j=1}^{k}(x_0-x_j)} + \cdots + \frac{f(x_l)}{\prod\limits_{j=0,\, j\neq l}^{k}(x_l-x_j)} + \cdots + \frac{f(x_k)}{\prod\limits_{j=0,\, j\neq k}^{k}(x_k-x_j)}\end{aligned}$$

Proof. We will prove the result by induction on $k$. The result is trivially true for $k = 0$. For $k = 1$,

$$\delta[x_0, x_1] = \frac{f(x_1)-f(x_0)}{x_1-x_0} = \frac{f(x_0)}{x_0-x_1} + \frac{f(x_1)}{x_1-x_0}.$$

Let us assume that the result is true for $k = n$, *i.e.,*

$$\begin{aligned}\delta[x_0, x_1, \ldots, x_n] &= \frac{f(x_0)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)} + \frac{f(x_1)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)} \\ &\quad + \cdots + \frac{f(x_n)}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})}.\end{aligned}$$

Consider $k = n+1$, then the $(n+1)^{\text{th}}$ divided difference is

$$\begin{aligned}\delta[x_0, x_1, \ldots, x_{n+1}] &= \frac{\delta[x_1, x_2, \ldots, x_{n+1}] - \delta[x_0, x_1, \ldots, x_n]}{x_{n+1}-x_0} \\ &= \frac{1}{x_{n+1}-x_0}\left[\frac{f(x_1)}{(x_1-x_2)\cdots(x_1-x_{n+1})} + \frac{f(x_2)}{(x_2-x_1)(x_2-x_3)\cdots(x_2-x_{n+1})} + \right. \\ &\quad \left. \cdots + \frac{f(x_{n+1})}{(x_{n+1}-x_1)\cdots(x_{n+1}-x_n)}\right] - \frac{1}{x_{n+1}-x_0}\left[\frac{f(x_0)}{(x_0-x_1)\cdots(x_0-x_n)} + \right. \\ &\quad \left. \frac{f(x_1)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)} + \cdots + \frac{f(x_n)}{(x_n-x_0)\cdots(x_n-x_{n-1})}\right]\end{aligned}$$

which on rearranging the terms gives the desired result. Therefore, by mathematical induction, the proof of the theorem is complete. □

**Remark 12.3.2** *In view of the theorem 12.3.1 the $k^{\text{th}}$ divided difference of a function $f(x)$, remains unchanged regardless of how its arguments are interchanged, i.e., it is independent of the order of its arguments.*

Now, if a function is approximated by a polynomial of degree $n$, then , its $(n+1)^{\text{th}}$ divided difference relative to $x, x_0, x_1, \ldots, x_n$ will be zero,(Remark 12.2.6) *i.e.,*

$$\delta[x, x_0, x_1, \ldots, x_n] = 0$$

Using this result, Theorem 12.3.1 gives

$$\begin{aligned}&\frac{f(x)}{(x-x_0)(x-x_1)\cdots(x-x_n)} + \frac{f(x_0)}{(x_0-x)(x_0-x_1)\cdots(x_0-x_n)} + \\ &\qquad \frac{f(x_1)}{(x_1-x)(x_1-x_2)\cdots(x_1-x_n)} + \cdots + \frac{f(x_n)}{(x_n-x)(x_n-x_0)\cdots(x_n-x_{n-1})} = 0,\end{aligned}$$

or,

$$\begin{aligned}\frac{f(x)}{(x-x_0)(x-x_1)\cdots(x-x_n)} &= -\left[\frac{f(x_0)}{(x_0-x)(x_0-x_1)\cdots(x_0-x_n)} + \frac{f(x_1)}{(x_1-x)(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)}\right.\\ &\left. + \cdots + \frac{f(x_n)}{(x_n-x)(x_n-x_0)\cdots(x_n-x_{n-1})}\right],\end{aligned}$$

which gives ,

$$\begin{aligned} f(x) &= \frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)\cdots(x_0-x_n)}f(x_0) + \frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)}f(x_1) \\ &+ \cdots + \frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})}f(x_n) \\ &= \sum_{i=0}^{n}\left(\prod_{j=0,\, j\neq i}^{n}\frac{x-x_j}{x_i-x_j}\right)f(x_i) = \sum_{i=0}^{n}\frac{\prod_{j=0}^{n}(x-x_j)}{(x-x_i)\prod_{j=0,\, j\neq i}^{n}(x_i-x_j)}f(x_i) \\ &= \prod_{j=0}^{n}(x-x_j)\sum_{i=0}^{n}\frac{f(x_i)}{(x-x_i)\prod_{j=0,\, j\neq i}^{n}(x_i-x_j)}.\end{aligned}$$

Note that the expression on the right is a polynomial of degree $n$ and takes the value $f(x_i)$ at $x = x_i$ for $i = 0, 1, \cdots, (n-1)$.

This polynomial approximation is called LAGRANGE'S INTERPOLATION FORMULA.

**Remark 12.3.3** *In view of the Remark (12.2.9), we can observe that $P_n(x)$ is another form of Lagrange's Interpolation polynomial formula as obtained above. Also the remainder term $R_{n+1}$ gives an estimate of error between the true value and the interpolated value of the function.*

**Remark 12.3.4** *We have seen earlier that the divided differences are independent of the order of its arguments. As the Lagrange's formula has been derived using the divided differences, it is not necessary here to have the tabular points in the increasing order. Thus one can use Lagrange's formula even when the points $x_0, x_1, \cdots, x_k, \cdots, x_n$ are in any order, which was not possible in the case of Newton's Difference formulae.*

**Remark 12.3.5** *One can also use the Lagrange's Interpolating Formula to compute the value of $x$ for a given value of $y = f(x)$. This is done by interchanging the roles of $x$ and $y$, i.e. while using the table of values, we take tabular points as $y_k$ and nodal points are taken as $x_k$.*

**Example 12.3.6** Using the following data, find by Lagrange's formula, the value of $f(x)$ at $x = 10$ :

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $x_i$ | 9.3 | 9.6 | 10.2 | 10.4 | 10.8 |
| $y_i = f(x_i)$ | 11.40 | 12.80 | 14.70 | 17.00 | 19.80 |

Also find the value of $x$ where $f(x) = 16.00$.

**Solution:** To compute $f(10)$, we first calculate the following products:

$$\begin{aligned}\prod_{j=0}^{4}(x-x_j) &= \prod_{j=0}^{4}(10-x_j) \\ &= (10-9.3)(10-9.6)(10-10.2)(10-10.4)(10-10.8) = -0.01792, \\ \prod_{j=1}^{4}(x_0-x_j) &= 0.4455, \quad \prod_{j=0,\, j\neq 1}^{n}(x_1-x_j) = -0.1728, \quad \prod_{j=0,\, j\neq 2}^{n}(x_2-x_j) = 0.0648, \\ \prod_{j=0,\, j\neq 3}^{n}(x_3-x_j) &= -0.0704, \text{ and } \quad \prod_{j=0,\, j\neq 4}^{n}(x_4-x_j) = 0.4320.\end{aligned}$$

Thus,

$$\begin{aligned} f(10) &\approx -0.01792 \times \left[\frac{11.40}{0.7 \times 0.4455} + \frac{12.80}{0.4 \times (-0.1728)} + \frac{14.70}{(-0.2) \times 0.0648}\right. \\ &\qquad \left. + \frac{17.00}{(-0.4) \times (-0.0704)} + \frac{19.80}{(-0.8) \times 0.4320}\right] \\ &= 13.197845. \end{aligned}$$

Now to find the value of $x$ such that $f(x) = 16$, we interchange the roles of $x$ and $y$ and calculate the following products:

$$\begin{aligned} \prod_{j=0}^{4}(y - y_j) &= \prod_{j=0}^{4}(16 - y_j) \\ &= (16 - 11.4)(16 - 12.8)(16 - 14.7)(16 - 17.0)(16 - 19.8) = 72.7168, \\ \prod_{j=1}^{4}(y_0 - y_j) &= 217.3248, \quad \prod_{j=0,\ j\neq 1}^{n} (y_1 - y_j) = -78.204, \quad \prod_{j=0,\ j\neq 2}^{n} (y_2 - y_j) = 73.5471, \\ \prod_{j=0,\ j\neq 3}^{n} (y_3 - y_j) &= -151.4688, \text{ and } \quad \prod_{j=0,\ j\neq 4}^{n} (y_4 - y_j) = 839.664. \end{aligned}$$

Thus,the required value of $x$ is obtained as:

$$\begin{aligned} x &\approx 217.3248 \times \left[\frac{9.3}{4.6 \times 217.3248} + \frac{9.6}{3.2 \times (-78.204)} + \frac{10.2}{1.3 \times 73.5471}\right. \\ &\qquad \left. + \frac{10.40}{(-1.0) \times (-151.4688)} + \frac{10.80}{(-3.8) \times 839.664}\right] \\ &\approx 10.39123. \end{aligned}$$

**Exercise 12.3.7** The following table gives the data for steam pressure $P$ *vs* temperature $T$:

| $T$ | 360 | 365 | 373 | 383 | 390 |
|---|---|---|---|---|---|
| $P = f(T)$ | 154.0 | 165.0 | 190.0 | 210.0 | 240.0 |

Compute the pressure at $T = 375$.

**Exercise 12.3.8** Compute from following table the value of $y$ for $x = 6.20$ :

| $x$ | 5.60 | 5.90 | 6.50 | 6.90 | 7.20 |
|---|---|---|---|---|---|
| $y$ | 2.30 | 1.80 | 1.35 | 1.95 | 2.00 |

Also find the value of $x$ where $y = 1.00$

## 12.4 Gauss's and Stirling's Formulas

In case of equidistant tabular points a convenient form for interpolating polynomial can be derived from Lagrange's interpolating polynomial. The process involves renaming or re-designating the tabular points. We illustrate it by deriving the interpolating formula for 6 tabular points. This can be generalized for more number of points. Let the given tabular points be $x_0, x_1 = x_0 + h, x_2 = x_0 - h, x_3 = x_0 + 2h, x_4 = x_0 - 2h, x_5 = x_0 + 3h$. These six points in the given order are not equidistant. We re-designate them for the sake of convenience as follows: $x^*_{-2} = x_4, x^*_{-1} = x_2, x^*_0 = x_0, x^*_1 = x_1, x^*_2 = x_3, x^*_3 = x_5$. These re-designated tabular points in their given order are equidistant. Now recall from remark (12.3.3) that Lagrange's interpolating polynomial can also be written as :

$$\begin{aligned} f(x) &\approx f(x_0) + \delta[x_0, x_1](x - x_0) + \delta[x_0, x_1, x_2](x - x_0)(x - x_1) \\ &+\delta[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) \\ &+\delta[x_0, x_1, x_2, x_3, x_4](x - x_0)(x - x_1)(x - x_2)(x - x_3) \\ &+\delta[x_0, x_1, x_2, x_3, x_4, x_5](x - x_0)(x - x_1)(x - x_2)(x - x_3)(x - x_4), \end{aligned}$$

which on using the re-designated points give:

$$\begin{aligned} f(x) &\approx f(x_0^*) + \delta[x_0^*, x_1^*](x - x_0^*) + \delta[x_0^*, x_1^*, x_{-1}^*](x - x_0^*)(x - x_1^*) \\ &+\delta[x_0^*, x_1^*, x_{-1}^*, x_2^*](x - x_0^*)(x - x_1^*)(x - x_{-1}^*) \\ &+\delta[x_0^*, x_1^*, x_{-1}^*, x_2^*, x_{-2}^*](x - x_0^*)(x - x_1^*)(x - x_{-1}^*)(x - x_2^*) \\ &+\delta[x_0^*, x_1^*, x_{-1}^*, x_2^*, x_{-2}^*, x_3^*](x - x_0^*)(x - x_1^*)(x - x_{-1}^*)(x - x_2^*)(x - x_{-2}^*). \end{aligned}$$

Now note that the points $x_{-2}^*, x_{-1}^*, x_0^*, x_1^*, x_2^*$ and $x_3^*$ are equidistant and the divided difference are independent of the order of their arguments. Thus, we have

$$\begin{aligned} &\delta[x_0^*, x_1^*] = \frac{\Delta y_0^*}{h}, \qquad \delta[x_0^*, x_1^*, x_{-1}^*] = \delta[x_{-1}^*, x_0^*, x_1^*] = \frac{\Delta^2 y_{-1}^*}{2h^2}, \\ &\delta[x_0^*, x_1^*, x_{-1}^*, x_2^*] = \delta[x_{-1}^*, x_0^*, x_1^*, x_2^*] = \frac{\Delta^3 y_{-1}^*}{3!h^3}, \\ &\delta[x_0^*, x_1^*, x_{-1}^*, x_2^*, x_{-2}^*] = \delta[x_{-2}^*, x_{-1}^*, x_0^*, x_1^*, x_2^*] = \frac{\Delta^4 y_{-2}^*}{4!h^4}, \\ &\delta[x_0^*, x_1^*, x_{-1}^*, x_2^*, x_{-2}^*, x_3^*] = \delta[x_{-2}^*, x_{-1}^*, x_0^*, x_1^*, x_2^*, x_3^*] = \frac{\Delta^5 y_{-2}^*}{5!h^5}, \end{aligned}$$

where $y_i^* = f(x_i^*)$ for $i = -2, -1, 0, 1, 2$. Now using the above relations and the transformation $x = x_0^* + hu$, we get

$$\begin{aligned} f(x_0^* + hu) &\approx y_0^* + \frac{\Delta y_0^*}{h}(hu) + \frac{\Delta^2 y_{-1}^*}{2h^2}(hu)(hu - h) + \frac{\Delta^3 y_{-1}^*}{3!h^3}(hu)(hu - h)(hu + h) \\ &+\frac{\Delta^4 y_{-2}^*}{4!h^4}(hu)(hu - h)(hu + h)(hu - 2h) \\ &+\frac{\Delta^5 y_{-2}^*}{5!h^5}(hu)(hu - h)(hu + h)(hu - 2h)(hu + 2h). \end{aligned}$$

Thus we get the following form of interpolating polynomial

$$\begin{aligned} f(x_0^* + hu) &\approx y_0^* + u\Delta y_0^* + u(u - 1)\frac{\Delta^2 y_{-1}^*}{2!} + u(u^2 - 1)\frac{\Delta^3 y_{-1}^*}{3!} \\ &+u(u^2 - 1)(u - 2)\frac{\Delta^4 y_{-2}^*}{4!} + u(u^2 - 1)(u^2 - 2^2)\frac{\Delta^5 y_{-2}^*}{5!}. \end{aligned} \tag{12.4.1}$$

Similarly using the tabular points $x_0, x_1 = x_0 - h, x_2 = x_0 + h, x_3 = x_0 - 2h, x_4 = x_0 + 2h, x_5 = x_0 - 3h$, and the re-designating them, as $x_{-3}^*, x_{-2}^*, x_{-1}^*, x_0^*, x_1^*$ and $x_2^*$, we get another form of interpolating polynomial as follows:

$$\begin{aligned} f(x_0^* + hu) &\approx y_0^* + u\Delta y_{-1}^* + u(u + 1)\frac{\Delta^2 y_{-1}^*}{2!} + u(u^2 - 1)\frac{\Delta^3 y_{-2}^*}{3!} \\ &+u(u^2 - 1)(u + 2)\frac{\Delta^4 y_{-2}^*}{4!} + u(u^2 - 1)(u^2 - 2^2)\frac{\Delta^5 y_{-3}^*}{5!}. \end{aligned} \tag{12.4.2}$$

Now taking the average of the two interpoating polynomials (12.4.1) and (12.4.2) (called GAUSS'S FIRST AND SECOND INTERPOLATING FORMULAS, respectively), we obtain Sterling's Formula of interpolation:

$$\begin{aligned} f(x_0^* + hu) &\approx y_0^* + u\frac{\Delta y_{-1}^* + \Delta y_0^*}{2} + u^2\frac{\Delta^2 y_{-1}^*}{2!} + \frac{u(u^2-1)}{2}\left[\frac{\Delta^3 y_{-2}^* + \Delta^3 y_{-1}^*}{3!}\right] \\ &\quad + u^2(u^2-1)\frac{\Delta^4 y_{-2}^*}{4!} + \frac{u(u^2-1)(u^2-2^2)}{2}\left[\frac{\Delta^5 y_{-3}^* + \Delta^5 y_{-2}^*}{5!}\right] + \cdots . \qquad (12.4.3) \end{aligned}$$

These are very helpful when, the point of interpolation lies near the middle of the interpolating interval. In this case one usually writes the diagonal form of the difference table.

**Example 12.4.1** Using the following data, find by Sterling's formula, the value of $f(x) = cot(\pi x)$ at $x = 0.225$ :

| $x$ | 0.20 | 0.21 | 0.22 | 0.23 | 0.24 |
|---|---|---|---|---|---|
| $f(x)$ | 1.37638 | 1.28919 | 1.20879 | 1.13427 | 1.06489 |

Here the point $x = 0.225$ lies near the central tabular point $x = 0.22$. Thus , we define $x_{-2} = 0.20, x_{-1} = 0.21, x_0 = 0.22, x_1 = 0.23, x_2 = 0.24$, to get the difference table in diagonal form as:

| | | | | | |
|---|---|---|---|---|---|
| $x_{-2} = 0.20$ | $y_{-2} = 1.37638$ | | | | |
| | | $\Delta y_{-2} = -.08719$ | | | |
| $x_{-1} = .021$ | $y_{-1} = 1.28919$ | | $\Delta^2 y_{-2} = .00679$ | | |
| | | $\Delta y_{-1} = -.08040$ | | $\Delta^3 y_{-2} = -.00091$ | |
| $x_0 = 0.22$ | $y_0 = 1.20879$ | | $\Delta^2 y_{-1} = .00588$ | | $\Delta^4 y_{-2} = .00017$ |
| | | $\Delta y_0 = -.07452$ | | $\Delta^3 y_{-1} = -.00074$ | |
| $x_1 = 0.23$ | $y_1 = 1.13427$ | | $\Delta^2 y_0 = .00514$ | | |
| | | $\Delta y_1 = -.06938$ | | | |
| $x_2 = 0.24$ | $y_2 = 1.06489$ | | | | |

(here, $\Delta y_0 = y_1 - y_0 = 1.13427 - 1.20879 = -.07452; \Delta y_{-1} = 1.20879 - 1.28919 = -0.08040$; and $\Delta^2 y_{-1} = \Delta y_0 - \Delta y_{-1} = .00588$, etc.).

Using the Sterling's formula with $u = \dfrac{0.225 - 0.22}{0.01} = 0.5$, we get $f(0.225)$ as follows:

$$\begin{aligned} f(0.225) &= 1.20879 + 0.5\frac{-.08040 - .07452}{2} + (-0.5)^2\frac{.00588}{2} \\ &+ \frac{-0.5(0.5^2-1)}{2}\frac{(-.00091 - .00074)}{3!}0.5^2(0.5^2-1)\frac{.00017}{4!} \\ &= 1.1708457 \end{aligned}$$

Note that tabulated value of $cot(\pi x)$ at $x = 0.225$ is 1.1708496.

**Exercise 12.4.2** Compute from the following table the value of $y$ for $x = 0.05$ :

| $x$ | 0.00 | 0.02 | 0.04 | 0.06 | 0.08 |
|---|---|---|---|---|---|
| $y$ | 0.00000 | 0.02256 | 0.04511 | 0.06762 | 0.09007 |

# Chapter 13

# Numerical Differentiation and Integration

## 13.1 Introduction

Numerical differentiation/ integration is the process of computing the value of the derivative of a function, whose analytical expression is not available, but is specified through a set of values at certain tabular points $x_0, x_1, \cdots, x_n$ In such cases, we first determine an interpolating polynomial approximating the function (either on the whole interval or in sub-intervals) and then differentiate/integrate this polynomial to approximately compute the value of the derivative at the given point.

## 13.2 Numerical Differentiation

In the case of differentiation, we first write the interpolating formula on the interval $(x_0, x_n)$. and the differentiate the polynomial term by term to get an approximated polynomial to the derivative of the function. When the tabular points are equidistant, one uses either the Newton's Forward/ Backward Formula or Sterling's Formula; otherwise Lagrange's formula is used. Newton's Forward/ Backward formula is used depending upon the location of the point at which the derivative is to be computed. In case the given point is near the mid point of the interval, Sterling's formula can be used. We illustrate the process by taking (i) Newton's Forward formula, and (ii) Sterling's formula.

Recall, that the Newton's forward interpolating polynomial is given by

$$\begin{aligned} f(x) = f(x_0 + hu) \quad &\approx \quad y_0 + \Delta y_0 u + \frac{\Delta^2 y_0}{2!}(u(u-1)) + \cdots + \frac{\Delta^k y_0}{k!}\{u(u-1)\cdots(u-k+1)\} \\ &\quad + \cdots + \frac{\Delta^n y_0}{n!}\{u(u-1)...(u-n+1)\}. \end{aligned} \tag{13.2.1}$$

Differentiating (13.2.1), we get the approximate value of the first derivative at $x$ as

$$\begin{aligned} \frac{df}{dx} = \frac{1}{h}\frac{df}{du} \quad &\approx \quad \frac{1}{h}\left[\Delta y_0 + \frac{\Delta^2 y_0}{2!}(2u-1) + \frac{\Delta^3 y_0}{3!}(3u^2 - 6u + 2) + \cdots \right. \\ &\quad \left. + \frac{\Delta^n y_0}{n!}\left(nu^{n-1} - \frac{n(n-1)^2}{2}u^{n-2} + \cdots + (-1)^{(n-1)}(n-1)!\right)\right]. \end{aligned} \tag{13.2.2}$$

where, $u = \dfrac{x - x_0}{h}$.

Thus, an approximation to the value of first derivative at $x = x_0$ *i.e.* $u = 0$ is obtained as :

$$\left.\frac{df}{dx}\right|_{x=x_0} = \frac{1}{h}\left[\Delta y_0 - \frac{\Delta^2 y_0}{2} + \frac{\Delta^3 y_0}{3} - \cdots + (-1)^{(n-1)}\frac{\Delta^n y_0}{n}\right]. \tag{13.2.3}$$

Similarly, using Stirling's formula:

$$\begin{aligned} f(x_0^* + hu) &\approx y_0^* + u\frac{\Delta y_{-1}^* + \Delta y_0^*}{2} + u^2\frac{\Delta^2 y_{-1}^*}{2!} + \frac{u(u^2-1)}{2}\frac{\Delta^3 y_{-2}^* + \Delta^3 y_{-1}^*}{3!} \\ &\quad + u^2(u^2-1)\frac{\Delta^4 y_{-2}^*}{4!} + \frac{u(u^2-1)(u^2-2^2)}{2}\frac{\Delta^5 y_{-3}^* + \Delta^5 y_{-2}^*}{5!} + \cdots \end{aligned} \tag{13.2.4}$$

Therefore,

$$\begin{aligned} \frac{df}{dx} &= \frac{1}{h}\frac{df}{du} \approx \frac{1}{h}\left[\frac{\Delta y_{-1}^* + \Delta y_0^*}{2} + u\Delta^2 y_{-1}^* + \frac{(3u^2-1)}{2} \times \frac{(\Delta^3 y_{-2}^* + \Delta^3 y_{-1}^*)}{3!}\right. \\ &\quad \left. + 2u(2u^2-1)\frac{\Delta^4 y_{-2}^*}{4!} + \frac{(5u^4 - 15u^2 + 4)(\Delta^5 y_{-3}^* + \Delta^5 y_{-2}^*)}{2 \times 5!} + \cdots\right] \end{aligned} \tag{13.2.5}$$

Thus, the derivative at $x = x_0^*$ is obtained as:

$$\left.\frac{df}{dx}\right|_{x=x_0^*} = \frac{1}{h}\left[\frac{\Delta y_{-1}^* + \Delta y_0^*}{2} - \frac{(1)}{2} \times \frac{(\Delta^3 y_{-2}^* + \Delta^3 y_{-1}^*)}{3!} + \frac{4 \times (\Delta^5 y_{-3}^* + \Delta^5 y_{-2}^*)}{2 \times 5!} + \cdots\right]. \tag{13.2.6}$$

**Remark 13.2.1** *Numerical differentiation using Stirling's formula is found to be more accurate than that with the Newton's difference formulae. Also it is more convenient to use.*

Now higher derivatives can be found by successively differentiating the interpolating polynomials. Thus *e.g.* using (13.2.2), we get the second derivative at $x = x_0$ as

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x_0} = \frac{1}{h^2}\left[\Delta^2 y_0 - \Delta^3 y_0 + \frac{2 \times 11 \times \Delta^4 y_0}{4!} - \cdots\right].$$

**Example 13.2.2** Compute from following table the value of the derivative of $y = f(x)$ at $x = 1.7489$ :

| $x$ | 1.73 | 1.74 | 1.75 | 1.76 | 1.77 |
|---|---|---|---|---|---|
| $y$ | 1.772844100 | 1.155204006 | 1.737739435 | 1.720448638 | 1.703329888 |

Solution: We note here that $x_0 = 1.75, h = 0.01$, so $u = (1.7489 - 1.75)/0.01 = -0.11$, and $\Delta y_0 = -.0017290797, \Delta^2 y_0 = .0000172047, \Delta^3 y_0 = -.0000001712,$
$\Delta y_{-1} = -.0017464571, \Delta^2 y_{-1} = .0000173774, \Delta^3 y_{-1} = -.0000001727,$
$\Delta^3 y_{-2} = -.0000001749, \Delta^4 y_{-2} = -.0000000022$
Thus, $f'(1.7489)$ is obtained as:
(i) Using Newton's Forward difference formula,

$$\begin{aligned} f'(1.4978) &\approx \frac{1}{0.01}\left[-0.0017290797 + (2 \times -0.11 - 1) \times \frac{0.0000172047}{2}\right. \\ &\quad \left. + (3 \times (-0.11)^2 - 6 \times -0.11 + 2) \times \frac{-0.0000001712}{3!}\right] = -0.173965150143. \end{aligned}$$

(ii) Using Stirling's formula, we get:

$$\begin{aligned} f'(1.4978) &\approx \frac{1}{.01}\left[\frac{(-.0017464571) + (-.0017290797)}{2} + (-0.11) \times .0000173774\right. \\ &\quad + \frac{(3 \times (-0.11)^2 - 1)}{2}\frac{((-.0000001749) + (-.0000001727))}{3!} \\ &\quad \left. + 2 \times (-0.11) \times (2(-0.11)^2 - 1) \times \frac{(-.0000000022)}{4!}\right] \\ &= -0.17396520185 \end{aligned}$$

It may be pointed out here that the above table is for $f(x) = e^{-x}$, whose derivative has the value -0.1739652000 at $x = 1.7489$.

**Example 13.2.3** Using only the first term in the formula (13.2.6) show that

$$f'(x_0^*) \approx \frac{y_1^* - y_{-1}^*}{2h}.$$

Hence compute from following table the value of the derivative of $y = e^x$ at $x = 1.15$ :

| $x$ | 1.05 | 1.15 | 1.25 |
|---|---|---|---|
| $e^x$ | 2.8577 | 3.1582 | 3.4903 |

Solution: Truncating the formula (13.2.6)after the first term, we get:

$$\begin{aligned} f'(x_0^*) &\approx \frac{1}{h}\left[\frac{\Delta y_{-1}^* + \Delta y_0^*}{2}\right] \\ &= \frac{(y_0^* - y_{-1}^*) + (y_1^* - y_0^*)}{2h} \\ &= \frac{y_1^* - y_{-1}^*}{2h}. \end{aligned}$$

Now from the given table, taking $x_0^* = 1.15$, we have

$$f'(1.15) \approx \frac{3.4903 - 2.8577}{2 \times 0.1} = 3.1630.$$

Note the error between the computed value and the true value is $3.1630 - 3.1582 = 0.0048$.

**Exercise 13.2.4** Retaining only the first two terms in the formula (13.2.3), show that

$$f'(x_0) \approx \frac{-3y_0 + 4y_1 - y_2}{2h}.$$

Hence compute the derivative of $y = e^x$ at $x = 1.15$ from the following table:

| $x$ | 1.15 | 1.20 | 1.25 |
|---|---|---|---|
| $e^x$ | 3.1582 | 3.3201 | 3.4903 |

Also compare your result with the computed value in the example (13.2.3).

**Exercise 13.2.5** Retaining only the first two terms in the formula (13.2.6), show that

$$f'(x_0^*) \approx \frac{y_{-2}^* - 8y_{-1}^* + 8y_1^* - y_2^*}{12h}.$$

Hence compute from following table the value of the derivative of $y = e^x$ at $x = 1.15$ :

| $x$ | 1.05 | 1.10 | 1.15 | 1.20 | 1.25 |
|---|---|---|---|---|---|
| $e^x$ | 2.8577 | 3.0042 | 3.1582 | 3.3201 | 3.4903 |

**Exercise 13.2.6** Following table gives the values of $y = f(x)$ at the tabular points $x = 0 + 0.05 \times k$, $k = 0, 1, 2, 3, 4, 5$.

| $x$ | 0.00 | 0.05 | 0.10 | 0.15 | 0.20 | 0.25 |
|---|---|---|---|---|---|---|
| $y$ | 0.00000 | 0.10017 | 0.20134 | 0.30452 | 0.41075 | 0.52110 |

Compute (i)the derivatives $y\prime$ and $y\prime\prime$ at $x = 0.0$ by using the formula (13.2.2). (ii)the second derivative $y\prime\prime$ at $x = 0.1$ by using the formula (13.2.6).

Similarly, if we have tabular points which are not equidistant, one can use Lagrange's interpolating polynomial, which is differentiated to get an estimate of first derivative. We shall see the result for four tabular points and then give the general formula. Let $x_0, x_1, x_2, x_3$ be the tabular points, then the corresponding Lagrange's formula gives us:

$$
\begin{aligned}
f(x) &\approx \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f(x_0) + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f(x_1) \\
&\quad + \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f(x_2) + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}f(x_3)
\end{aligned}
$$

Differentiation of the above interpolating polynomial gives:

$$
\begin{aligned}
\frac{df(x)}{dx} &\approx \frac{(x-x_2)(x-x_3)+(x-x_1)(x-x_2)+(x-x_1)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f(x_0) \\
&\quad + \frac{(x-x_2)(x-x_3)+(x-x_0)(x-x_2)+(x-x_0)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f(x_1) \\
&\quad + \frac{(x-x_1)(x-x_2)+(x-x_0)(x-x_1)+(x-x_0)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f(x_2) \\
&\quad + \frac{(x-x_1)(x-x_2)+(x-x_0)(x-x_2)+(x-x_0)(x-x_1)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}f(x_3) \\
&= \left(\prod_{r=0}^{3}(x-x_r)\right)\left[\sum_{i=0}^{3}\frac{f(x_i)}{(x-x_i)\prod\limits_{j=0,\ j\neq i}^{3}(x_i-x_j)}\left(\sum_{k=0,\ k\neq i}^{3}\frac{1}{(x-x_k)}\right)\right]. \qquad (13.2.7)
\end{aligned}
$$

In particular, the value of the derivative at $x = x_0$ is given by

$$
\begin{aligned}
\left.\frac{df}{dx}\right|_{x=x_0} &\approx \left[\frac{1}{(x_0-x_1)}+\frac{1}{(x_0-x_2)}+\frac{1}{(x_0-x_3)}\right]f(x_0) + \frac{(x_0-x_2)(x_0-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f(x_1) \\
&\quad + \frac{(x_0-x_1)(x_0-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f(x_2) + \frac{(x_0-x_1)(x_0-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}f(x_3).
\end{aligned}
$$

Now, generalizing Equation (13.2.7) for $n+1$ tabular points $x_0, x_1, \cdots, x_n$ we get:

$$
\frac{df}{dx} = \prod_{r=0}^{n}(x-x_r)\left[\sum_{i=0}^{n}\frac{f(x_i)}{(x-x_i)\prod\limits_{j=0,\ j\neq i}^{n}(x_i-x_j)}\left(\sum_{k=0,\ k\neq i}^{n}\frac{1}{(x-x_k)}\right)\right].
$$

**Example 13.2.7** Compute from following table the value of the derivative of $y = f(x)$ at $x = 0.6$ :

| | | | |
|---|---|---|---|
| $x$ | 0.4 | 0.6 | 0.7 |
| $y$ | 3.3836494 | 4.2442376 | 4.7275054 |

Solution: The given tabular points are not equidistant, so we use Lagrange's interpolating polynomial with three points: $x_0 = 0.4, x_1 = 0.6, x_2 = 0.7$ . Now differentiating this polynomial the derivative of the function at $x = x_1$ is obtained in the following form:

$$
\left.\frac{df}{dx}\right|_{x=x_1} \approx \frac{(x_1-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \left[\frac{1}{(x_1-x_2)}+\frac{1}{(x_1-x_0)}\right]f(x_1) + \frac{(x_1-x_0)}{(x_2-x_0)(x_2-x_1)}f(x_2).
$$

<u>Note</u>: The reader is advised to derive the above expression.

Now, using the values from the table, we get:

$$
\begin{aligned}
\left.\frac{df}{dx}\right|_{x=0.6} &\approx \frac{(0.6-0.7)}{(0.4-0.6)(0.4-0.7)}\times 3.3836494 + \left[\frac{1}{(0.6-0.7)}+\frac{1}{(0.6-0.4)}\right]\times 4.2442376 \\
&\quad + \frac{(0.6-0.4)}{(0.7-0.4)(0.7-0.6)}\times 4.7225054 \\
&= -5.63941567 - 21.221188 + 31.48336933 = 4.6227656.
\end{aligned}
$$

For the sake of comparison, it may be pointed out here that the above table is for the function $f(x) = 2e^x + x$, and the value of its derivative at $x = 0.6$ is $4.6442376$.

**Exercise 13.2.8** For the function, whose tabular values are given in the above example(13.2.8), compute the value of its derivative at $x = 0.5$.

**Remark 13.2.9** *It may be remarked here that the numerical differentiation for higher derivatives does not give very accurate results and so is not much preferred.*

## 13.3 Numerical Integration

Numerical Integration is the process of computing the value of a definite integral, $\int\limits_a^b f(x)dx$, when the values of the integrand function, $y = f(x)$ are given at some tabular points. As in the case of Numerical differentiation, here also the integrand is first replaced with an interpolating polynomial, and then the integrating polynomial is integrated to compute the value of the definite integral. This gives us 'quadrature formula' for numerical integration. In the case of equidistant tabular points, either the Newton's formulae or Stirling's formula are used. Otherwise, one uses Lagrange's formula for the interpolating polynomial. We shall consider below the case of equidistant points: $x_0, x_1, \cdots, x_n$.

### 13.3.1 A General Quadrature Formula

Let $f(x_k) = y_k$ be the nodal value at the tabular point $x_k$ for $k = 0, 1, \cdots, x_n$, where $x_0 = a$ and $x_n = x_0 + nh = b$. Now, a general quadrature formula is obtained by replacing the integrand by Newton's forward difference interpolating polynomial. Thus, we get,

$$\begin{aligned}\int\limits_a^b f(x)dx &= \int\limits_a^b \Big[y_0 + \frac{\Delta y_0}{h}(x-x_0) + \frac{\Delta^2 y_0}{2!h^2}(x-x_0)(x-x_1) + \frac{\Delta^3 y_0}{3!h^3}(x-x_0)(x-x_1)(x-x_2) \\ &\quad + \frac{\Delta^4 y_0}{4!h^4}(x-x_0)(x-x_1)(x-x_2)(x-x_3) + \cdots\Big] dx\end{aligned}$$

This on using the transformation $x = x_0 + hu$ gives:

$$\begin{aligned}\int\limits_a^b f(x)dx &= h\int\limits_0^n \Big[y_0 + u\Delta y_0 + \frac{\Delta^2 y_0}{2!}u(u-1) + \frac{\Delta^3 y_0}{3!}u(u-1)(u-2) \\ &\quad + \frac{\Delta^4 y_0}{4!}u(u-1)(u-2)(u-3) + \cdots\Big] du\end{aligned}$$

which on term by term integration gives,

$$\begin{aligned}\int\limits_a^b f(x)dx &= h\Big[ny_0 + \frac{n^2}{2}\Delta y_0 + \frac{\Delta^2 y_0}{2!}\left(\frac{n^3}{3} - \frac{n^2}{2}\right) + \frac{\Delta^3 y_0}{3!}\left(\frac{n^4}{4} - n^3 + n^2\right) \\ &\quad + \frac{\Delta^4 y_0}{4!}\left(\frac{n^5}{5} - \frac{3n^4}{2} + \frac{11n^3}{3} - 3n^2\right) + \cdots\Big] \qquad (13.3.1)\end{aligned}$$

For $n = 1$, *i.e.,* when linear interpolating polynomial is used then, we have

$$\int\limits_a^b f(x)dx = h\left[y_0 + \frac{\Delta y_0}{2}\right] = \frac{h}{2}\left[y_0 + y_1\right]. \qquad (13.3.2)$$

Similarly, using interpolating polynomial of degree 2 (*i.e.* $n = 2$), we obtain,

$$\begin{aligned}\int\limits_a^b f(x)dx &= h\left[2y_0 + 2\Delta y_0 + \left(\frac{8}{3} - \frac{4}{2}\right)\frac{\Delta^2 y_0}{2}\right] \\ &= 2h\left[y_0 + (y_1 - y_0) + \frac{1}{3} \times \frac{y_2 - 2y_1 + y_0}{2}\right] = \frac{h}{3}\left[y_0 + 4y_1 + y_2\right]. \qquad (13.3.3)\end{aligned}$$

In the above we have replaced the integrand by an interpolating polynomial over the whole interval $[a, b]$ and then integrated it term by term. However, this process is not very useful. More useful Numerical integral formulae are obtained by dividing the interval $[a, b]$ in $n$ sub-intervals $[x_k, x_{k+1}]$, where, $x_k = x_0 + kh$ for $k = 0, 1, \cdots, n$ with $x_0 = a, x_n = x_0 + nh = b$.

### 13.3.2 Trapezoidal Rule

Here, the integral is computed on each of the sub-intervals by using linear interpolating formula, *i.e.* for $n = 1$ and then summing them up to obtain the desired integral.

Note that

$$\int\limits_a^b f(x)dx = \int\limits_{x_0}^{x_1} f(x)dx + \int\limits_{x_1}^{x_2} f(x)dx + \cdots + \int\limits_{x_{k+1}}^{x_k} f(x)dx + \cdots + \int\limits_{x_n}^{x_{n-1}} f(x)dx$$

Now using the formula ( 13.3.2) for $n = 1$ on the interval $[x_k, x_{k+1}]$, we get,

$$\int\limits_{x_k}^{x_{k+1}} f(x)dx = \frac{h}{2}\left[y_k + y_{k+1}\right].$$

Thus, we have,

$$\int\limits_a^b f(x)dx = \frac{h}{2}\left[y_0 + y_1\right] + \frac{h}{2}\left[y_1 + y_2\right] + \cdots + \frac{h}{2}\left[y_k + y_{k+1}\right] + \cdots + \frac{h}{2}\left[y_{n-2} + y_{n-1}\right] + \frac{h}{2}\left[y_{n-1} + y_n\right]$$

*i.e.*

$$\begin{aligned}\int\limits_a^b f(x)dx &= \frac{h}{2}\left[y_0 + 2y_1 + 2y_2 + \cdots + 2y_k + \cdots + 2y_{n-1} + y_n\right] \\ &= h\left[\frac{y_0 + y_n}{2} + \sum_{i=1}^{n-1} y_i\right]. \qquad (13.3.4)\end{aligned}$$

This is called TRAPEZOIDAL RULE. It is a simple quadrature formula, but is not very accurate.

**Remark 13.3.1** *An estimate for the error $E_1$ in numerical integration using the Trapezoidal rule is given by*

$$E_1 = -\frac{b-a}{12}\overline{\Delta^2 y},$$

*where $\overline{\Delta^2 y}$ is the average value of the second forward differences.*

*Recall that in the case of linear function, the second forward differences is zero, hence, the Trapezoidal rule gives exact value of the integral if the integrand is a linear function.*

**Example 13.3.2** Using Trapezoidal rule compute the integral $\int\limits_0^1 e^{x^2}dx$, where the table for the values of $y = e^{x^2}$ is given below:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| $y$ | 1.00000 | 1.01005 | 1.04081 | 1.09417 | 1.17351 | 1.28402 | 1.43332 | 1.63231 | 1.89648 | 2.2479 | 2.71828 |

**Solution:** Here, $h = 0.1,\ n = 10$,

$$\frac{y_0 + y_{10}}{2} = \frac{1.0 + 2.71828}{2} = 1.85914,$$

and

$$\sum_{i=1}^{9} y_i = 12.81257.$$

Thus,

$$\int\limits_0^1 e^{x^2}dx = 0.1 \times [1.85914 + 12.81257] = 1.467171$$

### 13.3.3 Simpson's Rule

If we are given odd number of tabular points,*i.e.* $n$ is even, then we can divide the given integral of integration in even number of sub-intervals $[x_{2k}, x_{2k+2}]$. Note that for each of these sub-intervals, we have the three tabular points $x_{2k}, x_{2k+1}, x_{2k+2}$ and so the integrand is replaced with a quadratic interpolating polynomial. Thus using the formula (13.3.3), we get,

$$\int\limits_{x_{2k}}^{x_{2k+2}} f(x)dx = \frac{h}{3}\left[y_{2k} + 4y_{2k+1} + y_{2k+2}\right].$$

In view of this, we have

$$\begin{aligned}\int\limits_a^b f(x)dx &= \int\limits_{x_0}^{x_2} f(x)dx + \int\limits_{x_2}^{x_4} f(x)dx + \cdots + \int\limits_{x_{2k}}^{x_{2k+2}} f(x)dx + \cdots + \int\limits_{x_{n-2}}^{x_n} f(x)dx \\ &= \frac{h}{3}\left[(y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + \cdots + (y_{n-2} + 4y_{n-1} + y_n)\right] \\ &= \frac{h}{3}\left[y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n\right],\end{aligned}$$

which gives the second quadrature formula as follows:

$$\begin{aligned}\int\limits_a^b f(x)dx &= \frac{h}{3}[(y_0 + y_n) + 4 \times (y_1 + y_3 + \cdots + y_{2k+1} + \cdots + y_{n-1}) \\ &\quad + 2 \times (y_2 + y_4 + \cdots + y_{2k} + \cdots + y_{n-2})] \\ &= \frac{h}{3}\left[(y_0 + y_n) + 4 \times \left(\sum_{i=1,\ i-odd}^{n-1} y_i\right) + 2 \times \left(\sum_{i=2,\ i-even}^{n-2} y_i\right)\right]. \qquad (13.3.5)\end{aligned}$$

This is known as SIMPSON'S RULE.

**Remark 13.3.3** *An estimate for the error $E_2$ in numerical integration using the Simpson's rule is given by*

$$E_2 = -\frac{b-a}{180}\overline{\Delta^4 y}, \qquad (13.3.6)$$

*where $\overline{\Delta^4 y}$ is the average value of the forth forward differences.*

**Example 13.3.4** Using the table for the values of $y = e^{x^2}$ as is given in Example 13.3.2, compute the integral $\int\limits_0^1 e^{x^2}dx$, by Simpson's rule. Also estimate the error in its calculation and compare it with the error using Trapezoidal rule.

**Solution:** Here, $h = 0.1,\ n = 10$, thus we have odd number of nodal points. Further,

$$y_0 + y_{10} = 1.0 + 2.71828 = 3.71828, \qquad \sum_{i=1,\ i-odd}^{9} y_i = y_1 + y_3 + y_5 + y_7 + y_9 = 7.26845,$$

and

$$\sum_{i=2,\ i-even}^{8} y_i = y_2 + y_4 + y_6 + y_8 = 5.54412.$$

Thus,

$$\int\limits_0^1 e^{x^2}dx = \frac{0.1}{3} \times [3.71828 + 4 \times 7.268361 + 2 \times 5.54412] = 1.46267733$$

To find the error estimates, we consider the forward difference table, which is given below:

| $x_i$ | $y_i$ | $\Delta y_i$ | $\Delta^2 y_i$ | $\Delta^3 y_i$ | $\Delta^4 y_i$ |
|---|---|---|---|---|---|
| 0.0 | 1.00000 | 0.01005 | 0.02071 | 0.00189 | 0.00149 |
| 0.1 | 1.01005 | 0.03076 | 0.02260 | 0.00338 | 0.00171 |
| 0.2 | 1.04081 | 0.05336 | 0.02598 | 0.00519 | 0.00243 |
| 0.3 | 1.09417 | 0.07934 | 0.03117 | 0.00762 | 0.00320 |
| 0.4 | 1.17351 | 0.11051 | 0.3879 | 0.01090 | 0.00459 |
| 0.5 | 1.28402 | 0.14930 | 0.04969 | 0.01549 | 0.00658 |
| 0.6 | 1.43332 | 0.19899 | 0.06518 | 0.02207 | 0.00964 |
| 0.7 | 1.63231 | 0.26417 | 0.08725 | 0.03171 | |
| 0.8 | 1.89648 | 0.35142 | 0.11896 | | |
| 0.9 | 2.24790 | 0.47038 | | | |
| 1.0 | 2.71828 | | | | |

Thus, error due to Trapezoidal rule is,

$$\begin{aligned} E_1 &= -\frac{1-0}{12}\overline{\Delta^2 y} \\ &= -\frac{1}{12} \times \frac{0.02071 + 0.02260 + 0.02598 + 0.03117 + 0.03879 + 0.04969 + 0.06518 + 0.08725 + 0.11896}{9} \\ &= -0.004260463. \end{aligned}$$

Similarly, error due to Simpson's rule is,

$$\begin{aligned} E_2 &= -\frac{1-0}{180}\overline{\Delta^4 y} \\ &= -\frac{1}{180} \times \frac{0.00149 + 0.00171 + 0.00243 + 0.00328 + 0.00459 + 0.00658 + 0.00964}{7} \\ &= -2.35873 \times 10^{-5}. \end{aligned}$$

It shows that the error in numerical integration is much less by using Simpson's rule.

**Example 13.3.5** Compute the integral $\int\limits_{0.05}^{1} f(x)dx$, where the table for the values of $y = f(x)$ is given below:

| $x$ | 0.05 | 0.1 | 0.15 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 0.0785 | 0.1564 | 0.2334 | 0.3090 | 0.4540 | 0.5878 | 0.7071 | 0.8090 | 0.8910 | 0.9511 | 0.9877 | 1.0000 |

Solution: Note that here the points are not given to be equidistant, so as such we can not use any of the above two formulae. However, we notice that the tabular points $0.05, 0.10, 0, 15$ and $0.20$ are equidistant and so are the tabular points $0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ and $1.0$. Now we can divide the interval in two subinterval: $[0.05, 0.2]$ and $[0.2, 1.0]$; thus,

$$\int_{0.05}^{1} f(x)dx = \int_{0.05}^{0.2} f(x)dx + \int_{0.2}^{1} f(x)dx$$

. The integrals then can be evaluated in each interval. We observe that the second set has odd number of points. Thus, the first integral is evaluated by using Trapezoidal rule and the second one by Simpson's rule (of course, one could have used Trapezoidal rule in both the subintervals).
For the first integral $h = 0.05$ and for the second one $h = 0.1$. Thus,

$$\int_{0.05}^{0.2} f(x)dx = 0.05 \times \left[\frac{0.0785 + 0.3090}{2} + 0.1564 + 0.2334\right] = 0.0291775,$$

$$\begin{aligned} \text{and} \quad \int_{0.2}^{1.0} f(x)dx &= \frac{0.1}{3} \times \Big[(0.3090 + 1.0000) + 4 \times (0.4540 + 0.7071 + 0.8910 + 0.9877) \\ &\qquad\qquad +2 \times (0.5878 + 0.8090 + 0.9511)\Big] \\ &= 0.6054667, \end{aligned}$$

which gives,

$$\int_{0.05}^{1} f(x)dx = 0.0291775 + 0.6054667 = 0.6346442$$

It may be mentioned here that in the above integral, $f(x) = sin(\pi x/2)$ and that the value of the integral is $0.6346526$. It will be interesting for the reader to compute the two integrals using Trapezoidal rule and compare the values.

**Exercise 13.3.6**

1. Using Trapezoidal rule, compute the integral $\int_{a}^{b} f(x)dx$, where the table for the values of $y = f(x)$ is given below. Also find an error estimate for the computed value.

(a)

| $x$ | a=1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | b=10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 0.09531 | 0.18232 | 0.26236 | 0.33647 | 0.40546 | 0.47000 | 0.53063 | 0.58779 | 0.64185 | 0.69314 |

(b)

| $x$ | a=1.50 | 1.55 | 1.60 | 1.65 | 1.70 | 1.75 | b=1.80 |
|---|---|---|---|---|---|---|---|
| $y$ | 0.40546 | 0.43825 | 0.47000 | 0.5077 | 0.53063 | 0.55962 | 0.58779 |

(c)

| $x$ | a = 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | b = 3.5 |
|---|---|---|---|---|---|---|
| $y$ | 1.1752 | 2.1293 | 3.6269 | 6.0502 | 10.0179 | 16.5426 |

2. Using Simpson's rule, compute the integral $\int_{a}^{b} f(x)dx$. Also get an error estimate of the computed integral.

(a) Use the table given in Exercise 13.3.6.1b.

(b)

| $x$ | a = 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | b = 3.5 |
|---|---|---|---|---|---|---|---|
| $y$ | 0.493 | 0.946 | 1.325 | 1.605 | 1.778 | 1.849 | 1.833 |

3. Compute the integral $\int_{0}^{1.5} f(x)dx$, where the table for the values of $y = f(x)$ is given below:

| $x$ | 0.0 | 0.5 | 0.7 | 0.9 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 |
|---|---|---|---|---|---|---|---|---|---|
| $y$ | 0.00 | 0.39 | 0.77 | 1.27 | 1.90 | 2.26 | 2.65 | 3.07 | 3.53 |

# Chapter 14

# Appendix

## 14.1 System of Linear Equations

**Theorem 14.1.1 (Existence and Non-existence)** Consider a linear system $A\mathbf{x} = \mathbf{b}$, where $A$ is a $m \times n$ matrix, and $\mathbf{x}$, $\mathbf{b}$ are vectors with orders $n \times 1$, and $m \times 1$, respectively. Suppose rank $(A) = r$ and rank$([A \;\; \mathbf{b}]) = r_a$. Then exactly one of the following statement holds:

1. if $r_a = r < n$, the set of solutions of the linear system is an infinite set and has the form

$$\{\mathbf{u}_0 + k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_{n-r}\mathbf{u}_{n-r} \;:\; k_i \in \mathbb{R},\; 1 \le i \le n-r\},$$

   where $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-r}$ are $n \times 1$ vectors satisfying $A\mathbf{u}_0 = \mathbf{b}$ and $A\mathbf{u}_i = \mathbf{0}$ for $1 \le i \le n-r$.

2. if $r_a = r = n$, the solution set of the linear system has a unique $n \times 1$ vector $\mathbf{x}_0$ satisfying $A\mathbf{x}_0 = \mathbf{0}$.

3. If $r < r_a$, the linear system has no solution.

Proof. Suppose $[C \;\; \mathbf{d}]$ is the row reduced echelon form of the augmented matrix $[A \;\; \mathbf{b}]$. Then by Theorem 2.3.4, the solution set of the linear system $[C \;\; \mathbf{d}]$ is same as the solution set of the linear system $[A \;\; \mathbf{b}]$. So, the proof consists of understanding the solution set of the linear system $C\mathbf{x} = \mathbf{d}$.

1. Let $r = r_a < n$.

   Then $[C \;\; \mathbf{d}]$ has its first $r$ rows as the non-zero rows. So, by Remark 2.4.5, the matrix $C = [c_{ij}]$ has $r$ leading columns. Let the leading columns be $1 \le i_1 < i_2 < \cdots < i_r \le n$. Then we observe the following:

   (a) the entries $c_{li_l}$ for $1 \le l \le r$ are leading terms. That is, for $1 \le l \le r$, all entries in the $i_l^{\text{th}}$ column of $C$ is zero, except the entry $c_{li_l}$. The entry $c_{li_l} = 1$;

   (b) corresponding is each leading column, we have $r$ basic variables, $x_{i_1}, x_{i_2}, \ldots, x_{i_r}$;

   (c) the remaining $n-r$ columns correspond to the $n-r$ free variables (see Remark 2.4.5), $x_{j_1}, x_{j_2}, \ldots, x_{j_{n-r}}$. So, the free variables correspond to the columns $1 \le j_1 < j_2 < \cdots < j_{n-r} \le n$.

   For $1 \le l \le r$, consider the $l^{\text{th}}$ row of $[C \;\; \mathbf{d}]$. The entry $c_{li_l} = 1$ and is the leading term. Also, the first $r$ rows of the augmented matrix $[C \;\; \mathbf{d}]$ give rise to the linear equations

$$x_{i_l} + \sum_{k=1}^{n-r} c_{lj_k} x_{j_k} = d_l, \quad \text{for} \quad 1 \le l \le r.$$

These equations can be rewritten as

$$x_{i_l} = d_l - \sum_{k=1}^{n-r} c_{lj_k} x_{j_k} = d_l, \quad \text{for} \quad 1 \le l \le r.$$

Let $\mathbf{y}^t = (x_{i_1}, \ldots, x_{i_r}, x_{j_1}, \ldots, x_{j_{n-r}})$. Then the set of solutions consists of

$$\mathbf{y} = \begin{bmatrix} x_{i_1} \\ \vdots \\ x_{i_r} \\ x_{j_1} \\ \vdots \\ x_{j_{n-r}} \end{bmatrix} = \begin{bmatrix} d_1 - \sum\limits_{k=1}^{n-r} c_{1j_k} x_{j_k} \\ \vdots \\ d_r - \sum\limits_{k=1}^{n-r} c_{rj_k} x_{j_k} \\ x_{j_1} \\ \vdots \\ x_{j_{n-r}} \end{bmatrix}. \tag{14.1.1}$$

As $x_{j_s}$ for $1 \le s \le n-r$ are free variables, let us assign arbitrary constants $k_s \in \mathbb{R}$ to $x_{j_s}$. That is, for $1 \le s \le n-r$, $x_{j_s} = k_s$. Then the set of solutions is given by

$$\begin{aligned} \mathbf{y} \quad &= \begin{bmatrix} d_1 - \sum\limits_{s=1}^{n-r} c_{1j_s} x_{j_s} \\ \vdots \\ d_r - \sum\limits_{s=1}^{n-r} c_{rj_s} x_{j_s} \\ x_{j_1} \\ \vdots \\ x_{j_{n-r}} \end{bmatrix} = \begin{bmatrix} d_1 - \sum\limits_{s=1}^{n-r} c_{1j_s} k_s \\ \vdots \\ d_r - \sum\limits_{s=1}^{n-r} c_{rj_s} k_s \\ k_1 \\ \vdots \\ k_{n-r} \end{bmatrix} \\ &= \begin{bmatrix} d_1 \\ \vdots \\ d_r \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} - k_1 \begin{bmatrix} c_{1j_1} \\ \vdots \\ c_{rj_1} \\ -1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} - k_2 \begin{bmatrix} c_{1j_2} \\ \vdots \\ c_{rj_2} \\ 0 \\ -1 \\ \vdots \\ 0 \\ 0 \end{bmatrix} - \cdots - k_{n-r} \begin{bmatrix} c_{1j_{n-r}} \\ \vdots \\ c_{rj_{n-r}} \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{bmatrix}. \end{aligned}$$

Let us write ${\mathbf{v}_0}^t = (d_1, d_2, \ldots, d_r, 0, \ldots, 0)^t$. Also, for $1 \le i \le n-r$, let $\mathbf{v}_i$ be the vector associated with $k_i$ in the above representation of the solution $\mathbf{y}$. Observe the following:

(a) if we assign $k_s = 0$, for $1 \le s \le n-r$, we get

$$C\mathbf{v}_0 = C\mathbf{y} = \mathbf{d}. \tag{14.1.2}$$

(b) if we assign $k_1 = 1$ and $k_s = 0$, for $2 \le s \le n-r$, we get

$$\mathbf{d} = C\mathbf{y} = C(\mathbf{v}_0 + \mathbf{v}_1). \tag{14.1.3}$$

So, using (14.1.2), we get $C\mathbf{v}_1 = \mathbf{0}$.

(c) in general, if we assign $k_t = 1$ and $k_s = 0$, for $1 \le s \ne t \le n-r$, we get

$$\mathbf{d} = C\mathbf{y} = C(\mathbf{v}_0 + \mathbf{v}_t). \tag{14.1.4}$$

So, using (14.1.2), we get $C\mathbf{v}_t = \mathbf{0}$.

Note that a rearrangement of the entries of $\mathbf{y}$ will give us the solution vector $\mathbf{x}^t = (x_1, x_2, \ldots, x_n)^t$. Suppose that for $0 \le i \le n-r$, the vectors $\mathbf{u}_i$'s are obtained by applying the same rearrangement to the entries of $\mathbf{v}_i$'s which when applied to $\mathbf{y}$ gave $\mathbf{x}$. Therefore, we have $C\mathbf{u}_0 = \mathbf{d}$ and for $1 \le i \le n-r$, $C\mathbf{u}_i = \mathbf{0}$. Now, using equivalence of the linear system $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ gives

$$A\mathbf{u}_0 = \mathbf{b} \quad \text{and for} \quad 1 \le i \le n-r, \; A\mathbf{u}_i = \mathbf{0}.$$

Thus, we have obtained the desired result for the case $r = r_1 < n$.

2. $r = r_a = n, \; m \ge n.$

   Here the first $n$ rows of the row reduced echelon matrix $[C \;\; \mathbf{d}]$ are the non-zero rows. Also, the number of columns in $C$ equals $n =$ rank $(A) =$ rank $(C)$. So, by Remark 2.4.5, all the columns of $C$ are leading columns and all the variables $x_1, x_2, \ldots, x_n$ are basic variables. Thus, the row reduced echelon form $[C \;\; \mathbf{d}]$ of $[A \;\; \mathbf{b}]$ is given by

$$[C \;\; \mathbf{d}] = \begin{bmatrix} I_n & \tilde{\mathbf{d}} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

   Therefore, the solution set of the linear system $C\mathbf{x} = \mathbf{d}$ is obtained using the equation $I_n\mathbf{x} = \tilde{\mathbf{d}}$. This gives us, a solution as $\mathbf{x}_0 = \tilde{\mathbf{d}}$. Also, by Theorem 2.4.11, the row reduced form of a given matrix is unique, the solution obtained above is the only solution. That is, the solution set consists of a single vector $\tilde{\mathbf{d}}$.

3. $r < r_a.$

   As $C$ has $n$ columns, the row reduced echelon matrix $[C \;\; \mathbf{d}]$ has $n+1$ columns. The condition, $r < r_a$ implies that $r_a = r+1$. We now observe the following:

   (a) as rank$(C) = r$, the $(r+1)$th row of $C$ consists of only zeros.

   (b) Whereas the condition $r_a = r+1$ implies that the $(r+1)^{\text{th}}$ row of the matrix $[C \;\; \mathbf{d}]$ is non-zero.

   Thus, the $(r+1)^{\text{th}}$ row of $[C \;\; \mathbf{d}]$ is of the form $(0, \ldots, 0, 1)$. Or in other words, $\mathbf{d}_{r+1} = 1$.

   Thus, for the equivalent linear system $C\mathbf{x} = \mathbf{d}$, the $(r+1)^{\text{th}}$ equation is

$$0\,x_1 + 0\,x_2 + \cdots + 0\,x_n = 1.$$

   This linear equation has no solution. Hence, in this case, the linear system $C\mathbf{x} = \mathbf{d}$ has no solution. Therefore, by Theorem 2.3.4, the linear system $A\mathbf{x} = \mathbf{b}$ has no solution.

□

We now state a corollary whose proof is immediate from previous results.

**Corollary 14.1.2** Consider the linear system $A\mathbf{x} = \mathbf{b}$. Then the two statements given below cannot hold together.

1. The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.

2. The system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.

## 14.2 Determinant

In this section, $S$ denotes the set $\{1, 2, \ldots, n\}$.

**Definition 14.2.1** 1. A function $\sigma : S \longrightarrow S$ is called a permutation on $n$ elements if $\sigma$ is both one to one and onto.

2. The set of all functions $\sigma : S \longrightarrow S$ that are both one to one and onto will be denoted by $\mathcal{S}_n$. That is, $\mathcal{S}_n$ is the set of all permutations of the set $\{1, 2, \ldots, n\}$.

**Example 14.2.2** 1. In general, we represent a permutation $\sigma$ by $\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$. This representation of a permutation is called a TWO ROW NOTATION for $\sigma$.

2. For each positive integer $n$, $\mathcal{S}_n$ has a special permutation called the identity permutation, denoted $Id_n$, such that $Id_n(i) = i$ for $1 \le i \le n$. That is, $Id_n = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}$.

3. Let $n = 3$. Then

$$\begin{aligned} \mathcal{S}_3 \;=\; & \left\{ \tau_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \tau_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \tau_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \right. \\ & \left. \tau_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \tau_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \tau_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \right\} \end{aligned} \quad (14.2.5)$$

**Remark 14.2.3** *1. Let $\sigma \in \mathcal{S}_n$. Then $\sigma$ is determined if $\sigma(i)$ is known for $i = 1, 2, \ldots, n$. As $\sigma$ is both one to one and onto, $\{\sigma(1), \sigma(2), \ldots, \sigma(n)\} = S$. So, there are $n$ choices for $\sigma(1)$ (any element of $S$), $n-1$ choices for $\sigma(2)$ (any element of $S$ different from $\sigma(1)$), and so on. Hence, there are $n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1 = n!$ possible permutations. Thus, the number of elements in $\mathcal{S}_n$ is $n!$. That is, $|\mathcal{S}_n| = n!$.*

*2. Suppose that $\sigma, \tau \in \mathcal{S}_n$. Then both $\sigma$ and $\tau$ are one to one and onto. So, their composition map $\sigma \circ \tau$, defined by $(\sigma \circ \tau)(i) = \sigma\big(\tau(i)\big)$, is also both one to one and onto. Hence, $\sigma \circ \tau$ is also a permutation. That is, $\sigma \circ \tau \in \mathcal{S}_n$.*

*3. Suppose $\sigma \in \mathcal{S}_n$. Then $\sigma$ is both one to one and onto. Hence, the function $\sigma^{-1} : S \longrightarrow S$ defined by $\sigma^{-1}(m) = \ell$ if and only if $\sigma(\ell) = m$ for $1 \le m \le n$, is well defined and indeed $\sigma^{-1}$ is also both one to one and onto. Hence, for every element $\sigma \in \mathcal{S}_n$, $\sigma^{-1} \in \mathcal{S}_n$ and is the inverse of $\sigma$.*

*4. Observe that for any $\sigma \in \mathcal{S}_n$, the compositions $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = Id_n$.*

**Proposition 14.2.4** Consider the set of all permutations $\mathcal{S}_n$. Then the following holds:

1. Fix an element $\tau \in \mathcal{S}_n$. Then the sets $\{\sigma \circ \tau : \sigma \in \mathcal{S}_n\}$ and $\{\tau \circ \sigma : \sigma \in \mathcal{S}_n\}$ have exactly $n!$ elements. Or equivalently,

$$\mathcal{S}_n = \{\tau \circ \sigma : \sigma \in \mathcal{S}_n\} = \{\sigma \circ \tau : \sigma \in \mathcal{S}_n\}.$$

2. $\mathcal{S}_n = \{\sigma^{-1} : \sigma \in \mathcal{S}_n\}$.

PROOF. For the first part, we need to show that given any element $\alpha \in \mathcal{S}_n$, there exists elements $\beta, \gamma \in \mathcal{S}_n$ such that $\alpha = \tau \circ \beta = \gamma \circ \tau$. It can easily be verified that $\beta = \tau^{-1} \circ \alpha$ and $\gamma = \alpha \circ \tau^{-1}$.

For the second part, note that for any $\sigma \in \mathcal{S}_n$, $(\sigma^{-1})^{-1} = \sigma$. Hence the result holds. □

**Definition 14.2.5** Let $\sigma \in \mathcal{S}_n$. Then the number of inversions of $\sigma$, denoted $n(\sigma)$, equals

$$|\{(i,j) : \; i < j, \; \sigma(i) > \sigma(j) \; \}|.$$

Note that, for any $\sigma \in \mathcal{S}_n$, $n(\sigma)$ also equals

$$\sum_{i=1}^{n} |\{\sigma(j) < \sigma(i), \;\; \text{for} \;\; j = i+1, i+2, \ldots, n\}|.$$

**Definition 14.2.6** A permutation $\sigma \in \mathcal{S}_n$ is called a transposition if there exists two positive integers $m, r \in \{1, 2, \ldots, n\}$ such that $\sigma(m) = r, \; \sigma(r) = m$ and $\sigma(i) = i$ for $1 \leq i \neq m, r \leq n$.

For the sake of convenience, a transposition $\sigma$ for which $\sigma(m) = r, \; \sigma(r) = m$ and $\sigma(i) = i$ for $1 \leq i \neq m, r \leq n$ will be denoted simply by $\sigma = (m \; r)$ or $(r \; m)$. Also, note that for any transposition $\sigma \in \mathcal{S}_n$, $\sigma^{-1} = \sigma$. That is, $\sigma \circ \sigma = Id_n$.

**Example 14.2.7** 1. The permutation $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$ is a transposition as $\tau(1) = 3, \tau(3) = 1, \; \tau(2) = 2$ and $\tau(4) = 4$. Here note that $\tau = (1 \; 3) = (3 \; 1)$. Also, check that

$$n(\tau) = |\{(1,2), (1,3), (2,3)\}| = 3.$$

2. Let $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 2 & 3 & 5 & 1 & 9 & 8 & 7 & 6 \end{pmatrix}$. Then check that

$$n(\tau) = 3 + 1 + 1 + 1 + 0 + 3 + 2 + 1 = 12.$$

3. Let $\ell, m$ and $r$ be distinct element from $\{1, 2, \ldots, n\}$. Suppose $\tau = (m \; r)$ and $\sigma = (m \; \ell)$. Then

$$\begin{aligned} (\tau \circ \sigma)(\ell) &= \tau\big(\sigma(\ell)\big) = \tau(m) = r, \quad (\tau \circ \sigma)(m) = \tau\big(\sigma(m)\big) = \tau(\ell) = \ell \\ (\tau \circ \sigma)(r) &= \tau\big(\sigma(r)\big) = \tau(r) = m, \quad \text{and} \;\; (\tau \circ \sigma)(i) = \tau\big(\sigma(i)\big) = \tau(i) = i \;\; \text{if} \; i \neq \ell, m, r. \end{aligned}$$

Therefore, $\tau \circ \sigma = (m \; r) \circ (m \; \ell) = \begin{pmatrix} 1 & 2 & \cdots & \ell & \cdots & m & \cdots & r & \cdots & n \\ 1 & 2 & \cdots & r & \cdots & \ell & \cdots & m & \cdots & n \end{pmatrix} = (r \; l) \circ (r \; m).$

Similarly check that $\sigma \circ \tau = \begin{pmatrix} 1 & 2 & \cdots & \ell & \cdots & m & \cdots & r & \cdots & n \\ 1 & 2 & \cdots & m & \cdots & r & \cdots & \ell & \cdots & n \end{pmatrix}$.

With the above definitions, we state and prove two important results.

**Theorem 14.2.8** For any $\sigma \in \mathcal{S}_n$, $\sigma$ can be written as composition (product) of transpositions.

Proof. We will prove the result by induction on $n(\sigma)$, the number of inversions of $\sigma$. If $n(\sigma) = 0$, then $\sigma = Id_n = (1 \; 2) \circ (1 \; 2)$. So, let the result be true for all $\sigma \in \mathcal{S}_n$ with $n(\sigma) \leq k$.

For the next step of the induction, suppose that $\tau \in \mathcal{S}_n$ with $n(\tau) = k+1$. Choose the smallest positive number, say $\ell$, such that

$$\tau(i) = i, \;\; \text{for} \; i = 1, 2, \ldots, \ell - 1 \;\; \text{and} \; \tau(\ell) \neq \ell.$$

As $\tau$ is a permutation, there exists a positive number, say $m$, such that $\tau(\ell) = m$. Also, note that $m > \ell$. Define a transposition $\sigma$ by $\sigma = (\ell \; m)$. Then note that

$$(\sigma \circ \tau)(i) = i, \;\; \text{for} \; i = 1, 2, \ldots, \ell.$$

So, the definition of "number of inversions" and $m > \ell$ implies that

$$\begin{aligned}
n(\sigma \circ \tau) &= \sum_{i=1}^{n} |\{(\sigma \circ \tau)(j) < (\sigma \circ \tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&= \sum_{i=1}^{\ell} |\{(\sigma \circ \tau)(j) < (\sigma \circ \tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&\qquad\qquad + \sum_{i=\ell+1}^{n} |\{(\sigma \circ \tau)(j) < (\sigma \circ \tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&= \sum_{i=\ell+1}^{n} |\{(\sigma \circ \tau)(j) < (\sigma \circ \tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&\leq \sum_{i=\ell+1}^{n} |\{\tau(j) < \tau(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \ \text{ as } \ m > \ell, \\
&< (m - \ell) + \sum_{i=\ell+1}^{n} |\{\tau(j) < \tau(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&= n(\tau).
\end{aligned}$$

Thus, $n(\sigma \circ \tau) < k+1$. Hence, by the induction hypothesis, the permutation $\sigma \circ \tau$ is a composition of transpositions. That is, there exist transpositions, say $\alpha_i, \ 1 \leq i \leq t$ such that

$$\sigma \circ \tau = \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_t.$$

Hence, $\tau = \sigma \circ \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_t$ as $\sigma \circ \sigma = Id_n$ for any transposition $\sigma \in \mathcal{S}_n$. Therefore, by mathematical induction, the proof of the theorem is complete. □

Before coming to our next important result, we state and prove the following lemma.

**Lemma 14.2.9** Suppose there exist transpositions $\alpha_i, \ 1 \leq i \leq t$ such that

$$Id_n = \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_t,$$

then $t$ is even.

Proof. Observe that $t \neq 1$ as the identity permutation is not a transposition. Hence, $t \geq 2$. If $t = 2$, we are done. So, let us assume that $t \geq 3$. We will prove the result by the method of mathematical induction. The result clearly holds for $t = 2$. Let the result be true for all expressions in which the number of transpositions $t \leq k$. Now, let $t = k+1$.

Suppose $\alpha_1 = (m \ r)$. Note that the possible choices for the composition $\alpha_1 \circ \alpha_2$ are

$$(m\ r) \circ (m\ r) = Id_n, \ (m\ r) \circ (m\ \ell) = (r\ \ell) \circ (r\ m), \ (m\ r) \circ (r\ \ell) = (\ell\ r) \circ (\ell\ m) \text{ and } (m\ r) \circ (\ell\ s) = (\ell\ s) \circ (m\ r),$$

where $\ell$ and $s$ are distinct elements of $\{1, 2, \ldots, n\}$ and are different from $m, \ r$. In the first case, we can remove $\alpha_1 \circ \alpha_2$ and obtain $Id_n = \alpha_3 \circ \alpha_4 \circ \cdots \circ \alpha_t$. In this expression for identity, the number of transpositions is $t-2 = k-1 < k$. So, by mathematical induction, $t-2$ is even and hence $t$ is also even.

In the other three cases, we replace the original expression for $\alpha_1 \circ \alpha_2$ by their counterparts on the right to obtain another expression for identity in terms of $t = k+1$ transpositions. But note that in the new expression for identity, the positive integer $m$ doesn't appear in the first transposition, but appears in the second transposition. We can continue the above process with the second and third transpositions. At this step, either the number of transpositions will reduce by 2 (giving us the result by mathematical induction) or the positive number $m$ will get shifted to the third transposition. The continuation of this process will at some stage lead to an expression for identity in which the number of transpositions is

$t-2=k-1$ (which will give us the desired result by mathematical induction), or else we will have an expression in which the positive number $m$ will get shifted to the right most transposition. In the later case, the positive integer $m$ appears exactly once in the expression for identity and hence this expression does not fix $m$ whereas for the identity permutation $Id_n(m)=m$. So the later case leads us to a contradiction.

Hence, the process will surely lead to an expression in which the number of transpositions at some stage is $t-2=k-1$. Therefore, by mathematical induction, the proof of the lemma is complete.

□

**Theorem 14.2.10** Let $\alpha \in \mathcal{S}_n$. Suppose there exist transpositions $\tau_1, \tau_2, \ldots, \tau_k$ and $\sigma_1, \sigma_2, \ldots, \sigma_\ell$ such that

$$\alpha = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_\ell$$

then either $k$ and $\ell$ are both even or both odd.

Proof. Observe that the condition $\tau_1 \circ \tau_2 \circ \cdots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_\ell$ and $\sigma \circ \sigma = Id_n$ for any transposition $\sigma \in \mathcal{S}_n$, implies that

$$Id_n = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k \circ \sigma_\ell \circ \sigma_{\ell-1} \circ \cdots \circ \sigma_1.$$

Hence by Lemma 14.2.9, $k+\ell$ is even. Hence, either $k$ and $\ell$ are both even or both odd. Thus the result follows. □

**Definition 14.2.11** A permutation $\sigma \in \mathcal{S}_n$ is called an even permutation if $\sigma$ can be written as a composition (product) of an even number of transpositions. A permutation $\sigma \in \mathcal{S}_n$ is called an odd permutation if $\sigma$ can be written as a composition (product) of an odd number of transpositions.

**Remark 14.2.12** *Observe that if $\sigma$ and $\tau$ are both even or both odd permutations, then the permutations $\sigma \circ \tau$ and $\tau \circ \sigma$ are both even. Whereas if one of them is odd and the other even then the permutations $\sigma \circ \tau$ and $\tau \circ \sigma$ are both odd. We use this to define a function on $\mathcal{S}_n$, called the sign of a permutation, as follows:*

**Definition 14.2.13** Let sgn : $\mathcal{S}_n \longrightarrow \{1,-1\}$ be a function defined by

$$\text{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is an even permutation} \\ -1 & \text{if } \sigma \text{ is an odd permutation} \end{cases}.$$

**Example 14.2.14**

1. The identity permutation, $Id_n$ is an even permutation whereas every transposition is an odd permutation. Thus, $\text{sgn}(Id_n)=1$ and for any transposition $\sigma \in \mathcal{S}_n$, $\text{sgn}(\sigma)=-1$.

2. Using Remark 14.2.12, $\text{sgn}(\sigma \circ \tau) = \text{sgn}(\sigma) \cdot \text{sgn}(\tau)$ for any two permutations $\sigma, \tau \in \mathcal{S}_n$.

We are now ready to define determinant of a square matrix $A$.

**Definition 14.2.15** Let $A=[a_{ij}]$ be an $n \times n$ matrix with entries from $\mathbb{F}$. The determinant of $A$, denoted $\det(A)$, is defined as

$$\det(A) = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)} = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}.$$

**Remark 14.2.16**

1. *Observe that $\det(A)$ is a scalar quantity. The expression for $\det(A)$ seems complicated at the first glance. But this expression is very helpful in proving the results related with "properties of determinant".*

2. *If $A = [a_{ij}]$ is a $3 \times 3$ matrix, then using (14.2.5),*

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in \mathcal{S}_n} sgn(\sigma) \prod_{i=1}^{3} a_{i\sigma(i)} \\
&= sgn(\tau_1) \prod_{i=1}^{3} a_{i\tau_1(i)} + sgn(\tau_2) \prod_{i=1}^{3} a_{i\tau_2(i)} + sgn(\tau_3) \prod_{i=1}^{3} a_{i\tau_3(i)} + \\
&\qquad sgn(\tau_4) \prod_{i=1}^{3} a_{i\tau_4(i)} + sgn(\tau_5) \prod_{i=1}^{3} a_{i\tau_5(i)} + sgn(\tau_6) \prod_{i=1}^{3} a_{i\tau_6(i)} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.
\end{aligned}$$

*Observe that this expression for* $\det(A)$ *for a $3 \times 3$ matrix $A$ is same as that given in (2.8.1).*

## 14.3 Properties of Determinant

**Theorem 14.3.1 (Properties of Determinant)** Let $A = [a_{ij}]$ be an $n \times n$ matrix. Then

1. if $B$ is obtained from $A$ by interchanging two rows, then $\det(B) = -\det(A)$.

2. if $B$ is obtained from $A$ by multiplying a row by $c$ then $\det(B) = c\det(A)$.

3. if all the elements of one row is $0$ then $\det(A) = 0$.

4. if $A$ is a square matrix having two rows equal then $\det(A) = 0$.

5. Let $B = [b_{ij}]$ and $C = [c_{ij}]$ be two matrices which differ from the matrix $A = [a_{ij}]$ only in the $m^{\text{th}}$ row for some $m$. If $c_{mj} = a_{mj} + b_{mj}$ for $1 \leq j \leq n$ then $\det(C) = \det(A) + \det(B)$.

6. if $B$ is obtained from $A$ by replacing the $\ell$th row by itself plus $k$ times the $m$th row, for $\ell \neq m$ then $\det(B) = \det(A)$.

7. if $A$ is a triangular matrix then $\det(A) = a_{11}a_{22}\cdots a_{nn}$, the product of the diagonal elements.

8. If $E$ is an elementary matrix of order $n$ then $\det(EA) = \det(E)\det(A)$.

9. $A$ is invertible if and only if $\det(A) \neq 0$.

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

11. $\det(A) = \det(A^t)$, where recall that $A^t$ is the transpose of the matrix $A$.

PROOF. **Proof of Part 1.** Suppose $B = [b_{ij}]$ is obtained from $A = [a_{ij}]$ by the interchange of the $\ell^{\text{th}}$ and $m^{\text{th}}$ row. Then $b_{\ell j} = a_{mj},\ b_{mj} = a_{\ell j}$ for $1 \leq j \leq n$ and $b_{ij} = a_{ij}$ for $1 \leq i \neq \ell, m \leq n,\ 1 \leq j \leq n$.

Let $\tau = (\ell\ m)$ be a transposition. Then by Proposition 14.2.4, $\mathcal{S}_n = \{\sigma \circ \tau : \sigma \in \mathcal{S}_n\}$. Hence by the definition of determinant and Example 14.2.14.2, we have

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} = \sum_{\sigma\circ\tau \in \mathcal{S}_n} \operatorname{sgn}(\sigma \circ \tau) \prod_{i=1}^{n} b_{i(\sigma\circ\tau)(i)} \\
&= \sum_{\sigma\circ\tau \in \mathcal{S}_n} \operatorname{sgn}(\tau)\cdot\operatorname{sgn}(\sigma)\ b_{1(\sigma\circ\tau)(1)} b_{2(\sigma\circ\tau)(2)} \cdots b_{\ell(\sigma\circ\tau)(\ell)} \cdots b_{m(\sigma\circ\tau)(m)} \cdots b_{n(\sigma\circ\tau)(n)} \\
&= \operatorname{sgn}(\tau) \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma)\ b_{1\sigma(1)} \cdot b_{2\sigma(2)} \cdots b_{\ell\sigma(m)} \cdots b_{m\sigma(\ell)} \cdots b_{n\sigma(n)} \\
&= -\left( \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma)\ a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{\ell\sigma(\ell)} \cdots a_{n\sigma(n)} \right) \quad \text{as } \operatorname{sgn}(\tau) = -1 \\
&= -\det(A).
\end{aligned}$$

**Proof of Part 2.** Suppose that $B = [b_{ij}]$ is obtained by multiplying the $m^{\text{th}}$ row of $A$ by $c \neq 0$. Then $b_{mj} = c\, a_{mj}$ and $b_{ij} = a_{ij}$ for $1 \le i \neq m \le n,\ 1 \le j \le n$. Then

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{m\sigma(m)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots c a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= c \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= c \det(A).
\end{aligned}$$

**Proof of Part 3.** Note that $\det(A) = \sum\limits_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}$. So, each term in the expression for determinant, contains one entry from each row. Hence, from the condition that $A$ has a row consisting of all zeros, the value of each term is 0. Thus, $\det(A) = 0$.

**Proof of Part 4.** Suppose that the $\ell^{\text{th}}$ and $m^{\text{th}}$ row of $A$ are equal. Let $B$ be the matrix obtained from $A$ by interchanging the $\ell^{\text{th}}$ and $m^{\text{th}}$ rows. Then by the first part, $\det(B) = -\det(A)$. But the assumption implies that $B = A$. Hence, $\det(B) = \det(A)$. So, we have $\det(B) = -\det(A) = \det(A)$. Hence, $\det(A) = 0$.

**Proof of Part 5.** By definition and the given assumption, we have

$$\begin{aligned}
\det(C) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots c_{m\sigma(m)} \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots (b_{m\sigma(m)} + a_{m\sigma(m)}) \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{m\sigma(m)} \cdots b_{n\sigma(n)} \\
&\qquad\qquad + \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= \det(B) + \det(A).
\end{aligned}$$

**Proof of Part 6.** Suppose that $B = [b_{ij}]$ is obtained from $A$ by replacing the $\ell$th row by itself plus $k$ times the $m$th row, for $\ell \neq m$. Then $b_{\ell j} = a_{\ell j} + k\, a_{mj}$ and $b_{ij} = a_{ij}$ for $1 \le i \neq m \le n,\ 1 \le j \le n$.

Then

$$\begin{aligned}
\det(B) &= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{\ell\sigma(\ell)} \cdots b_{m\sigma(m)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots (a_{\ell\sigma(\ell)} + k a_{m\sigma(m)}) \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{\ell\sigma(\ell)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&\qquad\qquad + k \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{\ell\sigma(\ell)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \qquad \text{use Part } 4 \\
&= \det(A).
\end{aligned}$$

**Proof of Part 7.** First let us assume that $A$ is an upper triangular matrix. Observe that if $\sigma \in \mathcal{S}_n$ is different from the identity permutation then $n(\sigma) \geq 1$. So, for every $\sigma \neq Id_n \in \mathcal{S}_n$, there exists a positive integer $m$, $1 \leq m \leq n-1$ (depending on $\sigma$) such that $m > \sigma(m)$. As $A$ is an upper triangular matrix, $a_{m\sigma(m)} = 0$ for each $\sigma(\neq Id_n) \in \mathcal{S}_n$. Hence the result follows.

A similar reasoning holds true, in case $A$ is a lower triangular matrix.

**Proof of Part 8.** Let $I_n$ be the identity matrix of order $n$. Then using Part 7, $\det(I_n) = 1$. Also, recalling the notations for the elementary matrices given in Remark 2.4.14, we have $\det(E_{ij}) = -1$, (using Part 1) $\det(E_i(c)) = c$ (using Part 2) and $\det(E_{ij}(k) = 1$ (using Part 6). Again using Parts 1, 2 and 6, we get $\det(EA) = \det(E)\det(A)$.

**Proof of Part 9.** Suppose $A$ is invertible. Then by Theorem 2.7.7, $A$ is a product of elementary matrices. That is, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $A = E_1 E_2 \cdots E_k$. Now a repeated application of Part 8 implies that $\det(A) = \det(E_1)\det(E_2)\cdots\det(E_k)$. But $\det(E_i) \neq 0$ for $1 \leq i \leq k$. Hence, $\det(A) \neq 0$.

Now assume that $\det(A) \neq 0$. We show that $A$ is invertible. On the contrary, assume that $A$ is not invertible. Then by Theorem 2.7.7, the matrix $A$ is not of full rank. That is there exists a positive integer $r < n$ such that $\operatorname{rank}(A) = r$. So, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $E_1 E_2 \cdots E_k A = \begin{bmatrix} B \\ \mathbf{0} \end{bmatrix}$. Therefore, by Part 3 and a repeated application of Part 8,

$$\det(E_1)\det(E_2)\cdots\det(E_k)\det(A) = \det(E_1 E_2 \cdots E_k A) = \det\left(\begin{bmatrix} B \\ \mathbf{0} \end{bmatrix}\right) = 0.$$

But $\det(E_i) \neq 0$ for $1 \leq i \leq k$. Hence, $\det(A) = 0$. This contradicts our assumption that $\det(A) \neq 0$. Hence our assumption is false and therefore $A$ is invertible.

**Proof of Part 10.** Suppose $A$ is not invertible. Then by Part 9, $\det(A) = 0$. Also, the product matrix $AB$ is also not invertible. So, again by Part 9, $\det(AB) = 0$. Thus, $\det(AB) = \det(A)\det(B)$.

Now suppose that $A$ is invertible. Then by Theorem 2.7.7, $A$ is a product of elementary matrices. That is, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $A = E_1 E_2 \cdots E_k$. Now a repeated application of Part 8 implies that

$$\begin{aligned}
\det(AB) &= \det(E_1 E_2 \cdots E_k B) = \det(E_1)\det(E_2)\cdots\det(E_k)\det(B) \\
&= \det(E_1 E_2 \cdots E_k)\det(B) = \det(A)\det(B).
\end{aligned}$$

**Proof of Part 11.** Let $B = [b_{ij}] = A^t$. Then $b_{ij} = a_{ji}$ for $1 \le i, j \le n$. By Proposition 14.2.4, we know that $\mathcal{S}_n = \{\sigma^{-1} : \ \sigma \in \mathcal{S}_n\}$. Also $\text{sgn}(\sigma) = \text{sgn}(\sigma^{-1})$. Hence,

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma^{-1}) b_{\sigma^{-1}(1)\ 1}\, b_{\sigma^{-1}(2)\ 2} \cdots b_{\sigma^{-1}(n)\ n} \\
&= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma^{-1}) a_{1\sigma^{-1}(1)} b_{2\sigma^{-1}(2)} \cdots b_{n\sigma^{-1}(n)} \\
&= \det(A).
\end{aligned}$$

□

**Remark 14.3.2** 1. *The result that $\det(A) = \det(A^t)$ implies that in the statements made in Theorem 14.3.1, where ever the word "row" appears it can be replaced by "column".*

2. *Let $A = [a_{ij}]$ be a matrix satisfying $a_{11} = 1$ and $a_{1j} = 0$ for $2 \le j \le n$. Let $B$ be the submatrix of $A$ obtained by removing the first row and the first column. Then it can be easily shown that $\det(A) = \det(B)$. The reason being is as follows:*
*for every $\sigma \in \mathcal{S}_n$ with $\sigma(1) = 1$ is equivalent to saying that $\sigma$ is a permutation of the elements $\{2, 3, \ldots, n\}$. That is, $\sigma \in \mathcal{S}_{n-1}$. Hence,*

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in \mathcal{S}_n} sgn(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = \sum_{\sigma \in \mathcal{S}_n, \sigma(1)=1} sgn(\sigma) a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_{n-1}} sgn(\sigma) b_{1\sigma(1)} \cdots b_{n\sigma(n)} = \det(B).
\end{aligned}$$

We are now ready to relate this definition of determinant with the one given in Definition 2.8.2.

**Theorem 14.3.3** Let $A$ be an $n \times n$ matrix. Then $\det(A) = \sum\limits_{j=1}^{n} (-1)^{1+j} a_{1j} \det\big(A(1|j)\big)$, where recall that $A(1|j)$ is the submatrix of $A$ obtained by removing the $1^{\text{st}}$ row and the $j^{\text{th}}$ column.

Proof. For $1 \le j \le n$, define two matrices

$$B_j = \begin{bmatrix} 0 & 0 & \cdots & a_{1j} & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}_{n \times n} \quad \text{and} \quad C_j = \begin{bmatrix} a_{1j} & 0 & 0 & \cdots & 0 \\ a_{2j} & a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{nj} & a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}_{n \times n}.$$

Then by Theorem 14.3.1.5,

$$\det(A) = \sum_{j=1}^{n} \det(B_j). \tag{14.3.6}$$

We now compute $\det(B_j)$ for $1 \le j \le n$. Note that the matrix $B_j$ can be transformed into $C_j$ by $j-1$ interchanges of columns done in the following manner:
first interchange the $1^{\text{st}}$ and $2^{\text{nd}}$ column, then interchange the $2^{\text{nd}}$ and $3^{\text{rd}}$ column and so on (the last process consists of interchanging the $(j-1)^{\text{th}}$ column with the $j^{\text{th}}$ column. Then by Remark 14.3.2 and Parts 1 and 2 of Theorem 14.3.1, we have $\det(B_j) = a_{1j}(-1)^{j-1} \det(C_j)$. Therefore by (14.3.6),

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det\big(A(1|j)\big) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det\big(A(1|j)\big).$$

□

## 14.4 Dimension of $M+N$

**Theorem 14.4.1** Let $V(\mathbb{F})$ be a finite dimensional vector space and let $M$ and $N$ be two subspaces of $V$. Then

$$\dim(M)+\dim(N)=\dim(M+N)+\dim(M\cap N). \tag{14.4.7}$$

PROOF. Since $M\cap N$ is a vector subspace of $V$, consider a basis $\mathcal{B}_1=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k\}$ of $M\cap N$. As, $M\cap N$ is a subspace of the vector spaces $M$ and $N$, we extend the basis $\mathcal{B}_1$ to form a basis $\mathcal{B}_M=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k,\mathbf{v}_1,\ldots,\mathbf{v}_r\}$ of $M$ and also a basis $\mathcal{B}_N=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k,\mathbf{w}_1,\ldots,\mathbf{w}_s\}$ of $N$.

We now proceed to prove that that the set $\mathcal{B}_2=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k,\mathbf{w}_1,\ldots,\mathbf{w}_s,\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_r\}$ is a basis of $M+N$.

To do this, we show that

1. the set $\mathcal{B}_2$ is linearly independent subset of $V$, and

2. $L(\mathcal{B}_2)=M+N$.

The second part can be easily verified. To prove the first part, we consider the linear system of equations

$$\alpha_1\mathbf{u}_1+\cdots+\alpha_k\mathbf{u}_k+\beta_1\mathbf{w}_1+\cdots+\beta_s\mathbf{w}_s+\gamma_1\mathbf{v}_1+\cdots+\gamma_r\mathbf{v}_r=\mathbf{0}. \tag{14.4.8}$$

This system can be rewritten as

$$\alpha_1\mathbf{u}_1+\cdots+\alpha_k\mathbf{u}_k+\beta_1\mathbf{w}_1+\cdots+\beta_s\mathbf{w}_s=-(\gamma_1\mathbf{v}_1+\cdots+\gamma_r\mathbf{v}_r).$$

The vector $\mathbf{v}=-(\gamma_1\mathbf{v}_1+\cdots+\gamma_r\mathbf{v}_r)\in M$, as $\mathbf{v}_1,\ldots,\mathbf{v}_r\in\mathcal{B}_M$. But we also have $\mathbf{v}=\alpha_1\mathbf{u}_1+\cdots+\alpha_k\mathbf{u}_k+\beta_1\mathbf{w}_1+\cdots+\beta_s\mathbf{w}_s\in N$ as the vectors $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k,\mathbf{w}_1,\ldots,\mathbf{w}_s\in\mathcal{B}_N$. Hence, $\mathbf{v}\in M\cap N$ and therefore, there exists scalars $\delta_1,\ldots,\delta_k$ such that $\mathbf{v}=\delta_1\mathbf{u}_1+\delta_2\mathbf{u}_2+\cdots+\delta_k\mathbf{u}_k$.

Substituting this representation of $\mathbf{v}$ in Equation (14.4.8), we get

$$(\alpha_1-\delta_1)\mathbf{u}_1+\cdots+(\alpha_k-\delta_k)\mathbf{u}_k+\beta_1\mathbf{w}_1+\cdots+\beta_s\mathbf{w}_s=\mathbf{0}.$$

But then, the vectors $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k,\mathbf{w}_1,\ldots,\mathbf{w}_s$ are linearly independent as they form a basis. Therefore, by the definition of linear independence, we get

$$\alpha_i-\delta_i=0,\ \text{ for }\ 1\le i\le k\ \text{ and }\ \beta_j=0\ \text{ for }\ 1\le j\le s.$$

Thus the linear system of Equations (14.4.8) reduces to

$$\alpha_1\mathbf{u}_1+\cdots+\alpha_k\mathbf{u}_k+\gamma_1\mathbf{v}_1+\cdots+\gamma_r\mathbf{v}_r=\mathbf{0}.$$

The only solution for this linear system is

$$\alpha_i=0,\ \text{ for }\ 1\le i\le k\ \text{ and }\ \gamma_j=0\ \text{ for }\ 1\le j\le r.$$

Thus we see that the linear system of Equations (14.4.8) has no non-zero solution. And therefore, the vectors are linearly independent.

Hence, the set $\mathcal{B}_2$ is a basis of $M+N$. We now count the vectors in the sets $\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_M$ and $\mathcal{B}_N$ to get the required result. □

## 14.5 Proof of Rank-Nullity Theorem

**Theorem 14.5.1** Let $T : V \longrightarrow W$ be a linear transformation and $\{u_1, u_2, \ldots, u_n\}$ be a basis of $V$ . Then

1. $\mathcal{R}(T) = L(T(u_1), T(u_2), \ldots, T(u_n))$.
2. $T$ is one-one $\Longleftrightarrow$ $\mathcal{N}(T) = \{\mathbf{0}\}$ is the zero subspace of $V$ $\Longleftrightarrow$ $\{T(u_i) : 1 \leq i \leq n\}$ is a basis of $\mathcal{R}(T)$.
3. If $V$ is finite dimensional vector space then $\dim(\mathcal{R}(T)) \leq \dim(V)$. The equality holds if and only if $\mathcal{N}(T) = \{\mathbf{0}\}$.

Proof. Part 1) can be easily proved. For 2), let $T$ be one-one. Suppose $u \in \mathcal{N}(T)$. This means that $T(u) = \mathbf{0} = T(\mathbf{0})$. But then $T$ is one-one implies that $u = \mathbf{0}$. If $\mathcal{N}(T) = \{\mathbf{0}\}$ then $T(u) = T(v) \Longleftrightarrow T(u - v) = \mathbf{0}$ implies that $u = v$. Hence, $T$ is one-one.

The other parts can be similarly proved. Part 3) follows from the previous two parts. □

The proof of the next theorem is immediate from the fact that $T(\mathbf{0}) = \mathbf{0}$ and the definition of linear independence/dependence.

**Theorem 14.5.2** Let $T : V \longrightarrow W$ be a linear transformation. If $\{T(u_1), T(u_2), \ldots, T(u_n)\}$ is linearly independent in $\mathcal{R}(T)$ then $\{u_1, u_2, \ldots, u_n\} \subset V$ is linearly independent.

**Theorem 14.5.3 (Rank Nullity Theorem)** Let $T : V \longrightarrow W$ be a linear transformation and $V$ be a finite dimensional vector space. Then

$$\dim(\text{ Range}(T)) + \dim(\mathcal{N}(T)) = \dim(V),$$

or $\rho(T) + \nu(T) = n$.

Proof. Let $\dim(V) = n$ and $\dim(\mathcal{N}(T)) = r$. Suppose $\{u_1, u_2, \ldots, u_r\}$ is a basis of $\mathcal{N}(T)$. Since $\{u_1, u_2, \ldots, u_r\}$ is a linearly independent set in $V$, we can extend it to form a basis of $V$. Now there exists vectors $\{u_{r+1}, u_{r+2}, \ldots, u_n\}$ such that the set $\{u_1, \ldots, u_r, u_{r+1}, \ldots, u_n\}$ is a basis of $V$. Therefore,

$$\begin{aligned}\text{Range } (T) &= L(T(u_1), T(u_2), \ldots, T(u_n)) \\ &= L(\mathbf{0}, \ldots, \mathbf{0}, T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)) \\ &= L(T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n))\end{aligned}$$

which is equivalent to showing that Range $(T)$ is the span of $\{T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)\}$.

We now prove that the set $\{T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)\}$ is a linearly independent set. Suppose the set is linearly dependent. Then, there exists scalars, $\alpha_{r+1}, \alpha_{r+2}, \ldots, \alpha_n$, not all zero such that

$$\alpha_{r+1}T(u_{r+1}) + \alpha_{r+2}T(u_{r+2}) + \cdots + \alpha_n T(u_n) = \mathbf{0}.$$

Or $T(\alpha_{r+1}u_{r+1} + \alpha_{r+2}u_{r+2} + \cdots + \alpha_n u_n) = \mathbf{0}$ which in turn implies $\alpha_{r+1}u_{r+1} + \alpha_{r+2}u_{r+2} + \cdots + \alpha_n u_n \in \mathcal{N}(T) = L(u_1, \ldots, u_r)$. So, there exists scalars $\alpha_i,\ 1 \leq i \leq r$ such that

$$\alpha_{r+1}u_{r+1} + \alpha_{r+2}u_{r+2} + \cdots + \alpha_n u_n = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r.$$

That is,

$$\alpha_1 u_1 + + \cdots + \alpha_r u_r - \alpha_{r+1}u_{r+1} - \cdots - \alpha_n u_n = \mathbf{0}.$$

Thus $\alpha_i = 0$ for $1 \leq i \leq n$ as $\{u_1, u_2, \ldots, u_n\}$ is a basis of $V$. In other words, we have shown that the set $\{T(u_{r+1}), T(u_{r+2}), \ldots, T(u_n)\}$ is a basis of Range $(T)$. Now, the required result follows. □

we now state another important implication of the Rank-nullity theorem.

**Corollary 14.5.4** Let $T : V \longrightarrow V$ be a linear transformation on a finite dimensional vector space $V$. Then

$$T \text{ is one-one } \Longleftrightarrow T \text{ is onto} \Longleftrightarrow T \text{ has an inverse.}$$

PROOF. Let $\dim(V) = n$ and let $T$ be one-one. Then $\dim(\mathcal{N}(T)) = 0$. Hence, by the rank-nullity Theorem 14.5.3 $\dim(\text{ Range } (T)) = n = \dim(V)$. Also, Range$(T)$ is a subspace of $V$. Hence, Range$(T) = V$. That is, $T$ is onto.

Suppose $T$ is onto. Then Range$(T) = V$. Hence, $\dim(\text{ Range } (T)) = n$. But then by the rank-nullity Theorem 14.5.3, $\dim(\mathcal{N}(T)) = 0$. That is, $T$ is one-one.

Now we can assume that $T$ is one-one and onto. Hence, for every vector $\mathbf{u}$ in the range, there is a unique vectors $\mathbf{v}$ in the domain such that $T(\mathbf{v}) = \mathbf{u}$. Therefore, for every $\mathbf{u}$ in the range, we define

$$T^{-1}(\mathbf{u}) = \mathbf{v}.$$

That is, $T$ has an inverse.

Let us now assume that $T$ has an inverse. Then it is clear that $T$ is one-one and onto. □

## 14.6 Condition for Exactness

Let $D$ be a region in $xy$-plane and let $M$ and $N$ be real valued functions defined on $D$. Consider an equation

$$M(x, y(x))dx + N(x, y(x))dy = 0, \ (x, y(x)) \in D. \tag{14.6.9}$$

**Definition 14.6.1 (Exact Equation)** The Equation (14.6.9) is called Exact if there exists a real valued twice continuously differentiable function $f$ such that

$$\frac{\partial f}{\partial x} = M \text{ and } \frac{\partial f}{\partial y} = N.$$

**Theorem 14.6.2** Let $M$ and $N$ be "smooth" in a region $D$. The equation (14.6.9) is exact if and only if

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}. \tag{14.6.10}$$

PROOF. Let Equation (14.6.9) be exact. Then there is a "smooth" function $f$ (defined on $D$) such that $M = \frac{\partial f}{\partial x}$ and $N = \frac{\partial f}{\partial y}$. So, $\frac{\partial M}{\partial y} = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial N}{\partial x}$ and so Equation (14.6.10) holds.

Conversely, let Equation (14.6.10) hold. We now show that Equation (14.6.10) is exact. Define $G(x, y)$ on $D$ by

$$G(x, y) = \int M(x, y)dx + g(y)$$

where $g$ is any arbitrary smooth function. Then $\frac{\partial G}{\partial x} = M(x, y)$ which shows that

$$\frac{\partial}{\partial x} \cdot \frac{\partial G}{\partial y} = \frac{\partial}{\partial y} \cdot \frac{\partial G}{\partial x} = \frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}.$$

So $\frac{\partial}{\partial x}(N - \frac{\partial G}{\partial y}) = 0$ or $N - \frac{\partial G}{\partial y}$ is independent of $x$. Let $\phi(y) = N - \frac{\partial G}{\partial y}$ or $N = \phi(y) + \frac{\partial G}{\partial y}$. Now

$$\begin{aligned} M(x, y) + N\frac{dy}{dx} &= \frac{\partial G}{\partial x} + \left[\frac{\partial G}{\partial y} + \phi(y)\right]\frac{dy}{dx} \\ &= \left[\frac{\partial G}{\partial x} + \frac{\partial G}{\partial y} \cdot \frac{dy}{dx}\right] + \frac{d}{dy}\left(\int \phi(y)dy\right)\frac{dy}{dx} \\ &= \frac{d}{dx}G(x, y(x)) + \frac{d}{dx}\left(\int \phi(y)dy\right) \qquad \text{where } y = y(x) \\ &= \frac{d}{dx}\big(f(x, y)\big) \qquad \text{where } f(x, y) = G(x, y) + \int \phi(y)dy \end{aligned}$$

□

# Index