

# **STUDIES ON KNOT THEORY**

Project Report submitted to

**ST.MARY'S COLLEGE(AUTONOMOUS),THOOTHUKUDI**

Affiliated to

**MANONMANIYAM SUNDARANAR UNIVERSITY,TIRUNELVELI**

In partial fulfilment of the requirement for the award of degree of

**Bachelor of science in Mathematics**

Submitted by

NAMES	REG.NO.
AMALA AMISHA.M	18AUMT01
BRINDHA.B	18AUMT09
JEMINA.R	18AUMT24
LAKSHMI PRIYA.V.S	18AUMT31
PRETHI.S	18AUMT39
SIVAGAMI.M	18AUMT49

Under the Guidance of

**Dr.G.PRISCILLA PACIFICA M.Sc.,M.Phil.,Ph.D.,SET**

Assistant Professor of Mathematics

**ST.MARY'S COLLEGE(AUTONOMOUS),THOOTHUKUDI.**



Department of Mathematics

**ST.MARY'S COLLEGE (AUTONOMOUS),**

Thoothukudi.

(2020-2021)

## CERTIFICATE

We hereby declare that the project report entitled "STUDIES ON KNOT THEORY" being submitted to "St.Mary's College (Autonomous),Thoothukudi affiliated to Manonmaniam Sundaranar University,Tirunelveli in partial fulfilment for the award of degree of Bachelor of science in Mathematics and it is a record of work done during the year 2020-2021 by the following students:

AMALA AMISHA.M

18AUMT01

BRINDHA.B

18AUMT09

JEMINA.R

18AUMT24

LAKSHMI PRIYA.V.S

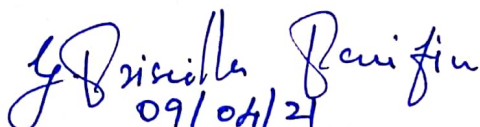
18AUMT31

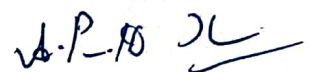
PRETHI.S

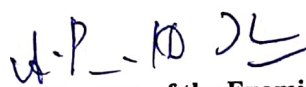
18AUMT39

SIVAGAMIL.M

18AUMT49

  
09/04/21  
Signature of the guide

  
Signature of the HOD

  
Signature of the Examiner

  
Signature of the Principal  
**St. Mary's College (Autonomous)**  
Thoothukudi - 628 001.



## DECLARATION

We hereby declare that the project report entitled “KNOT THEORY” is our original work .It has not been submitted to any university for any degree or diploma.

M. Amala Amisha  
(AMALA AMISHA.M)

B Brindha  
(BRINDHA.B)

V.S.Lakshmi Priya  
(LAKSHMI PRIYA.V.S)

R. Jemina  
(JEMINA.R)

S. Preethi  
(PRETHI.S)

M. Sivagami  
(SIVAGAMI.M)

## ACKNOWLEDGEMENT

*First of all , We thank lord Almighty for showering his blessings to undergo this project.*

*With immense pleasure, we register our deep sense of gratitude to our guide Dr.G.Priscilla Paciffica M.Sc., M.Phill., Ph.D., SET and the Head of the Department Dr.A.Punitha Tharani M.Sc.,M.Phill., Ph.D., for having imported necessary guidelines throughout the period of our studies.*

We thank our beloved Pricipal Rev.Dr.Sr.A.S.J.Lucia Rose M.Sc., PGDCA., M.Phill., Ph.D., for providing us the help to carry out our project work successfully.

Finally, We thank all those who extended their helping hands regarding this project.

# CONTENT

Chapter	Topic	Page.No
	Introduction	1
1.	PRELIMINARIES	2
2.	KAUFFMAN'S BRACKET AND JONES POLYNOMIAL	8
3.	SEIFERT MATRICES	16
4.	FUNDAMENTAL PROBLEMS OF KNOT THEORY	23
5.	APPLICATIONS	31
	CONCLUSION	38
	Bibliography	39

# Introduction

*Knots have been around prehistoric times, and remain a vital part of everyday life today. They are used by sailors, climbers, fishermen and surgeons, as well as for mundane tasks as tying a shoelace. Knot theory is one of the most active areas of research in mathematics today, and its techniques can be found in such wide ranging areas as fluid dynamics, solar physics, DNA research, and quantum computation. However, it was only as late as the twentieth century that mathematicians really began to seriously study knots.*

*In chapter 1, we discuss about the preliminary concepts of knot theory, Basic definitions and Reidemeister moves, Skein relation.*

*Chapter II deals with Kauffman Bracket, Jones Polynomial, how to find the Kauffman bracket, Jones Polynomial for trefoil knot and how Reidemeister moves hold knot invariant.*

*In chapter III, we discuss about Seifert surface, Seifert matrix, and genus of a knot.*

*Chapter IV deals with Fundamental problems of knot Theory.*



# **CHAPTER – I**

# Preliminaries

*This chapter about the basic concepts of knot theory,  
Preliminary definitions and also about the Reidemeister Moves,  
Skein relation in knot theory.*

## Definition 1:

*The curve intersections in a projection are called Precrossings.  
A Precrossing is said to be have been resolved once we have,  
selected the crossing information.*

## Definition :2

*Once all crossing information is determined in a link  
projection,the image is then a Link diagram.*

Example:

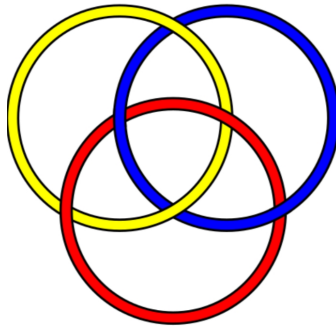


Figure 1: Link Diagram

## Definition:3

*A knot without knotting is called a Trivial Knot or Unknot*

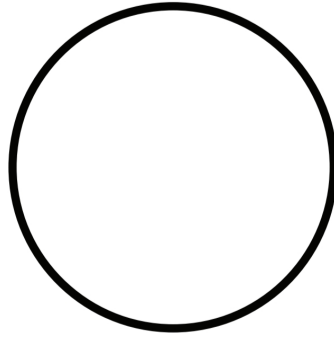


Figure 2: *Unknot*

*This is example of knot.*

**Definition :4**

*Minimum number of crossings given any knot diagram is called a Crossing number.*

*Example*

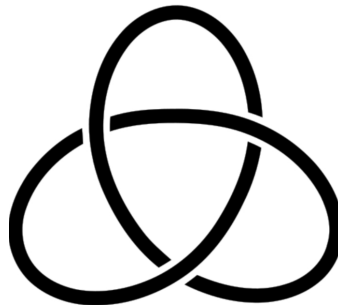


Figure 3: *Crossing number*

**Definition:5**

*The minimum number of times the knot must pass through itself before it becomes unknotted is called the Unknotting Number of knot.*

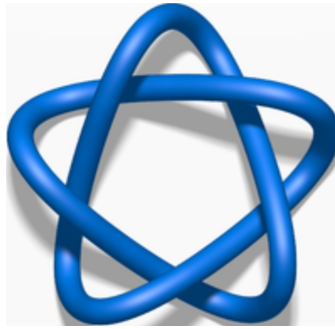


Figure 4:

*This is a example of unknotting number of knot.*

**Definition:6**

*A link is called Chiral if is not equivalent to mirror image.*

*Example: The simplest chiral is Trefoil knot*

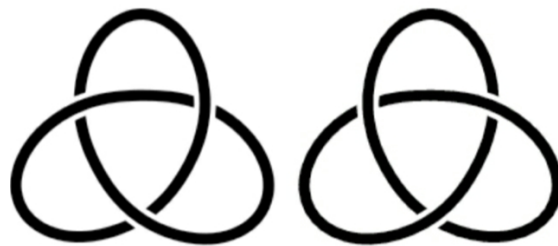


Figure 5: *Right and Left Trefoil knot*

**Definition:7**

*A link is called Amphichiral or Achiral if it is equivalent to its mirror image .*

*Example*



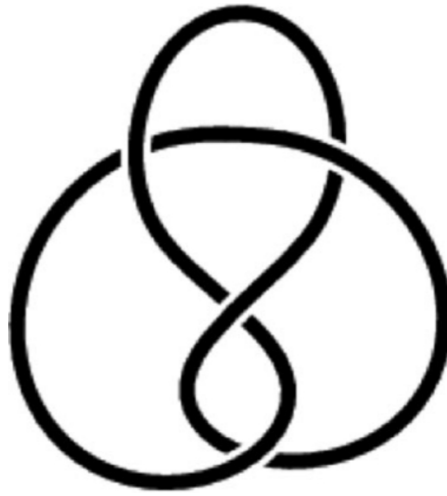


Figure 6: Amchiral or Achiral knot

### Definition:8

*A Knot invariant is a property of a knot that does not change under ambient isotopy .If two knots have different values for any knot invariant,then it is impossible.To transform one into the other ,thus they are not equivalent.*

### Definition 9:

## The Reidemeister moves

*A solitary elementary knot moves ,as might be expected,gives rise to various changes in the regular diagram.However ,it is possible to restrict ourselves to just the four moves shown in picture The Reidemeister moves Picture:*

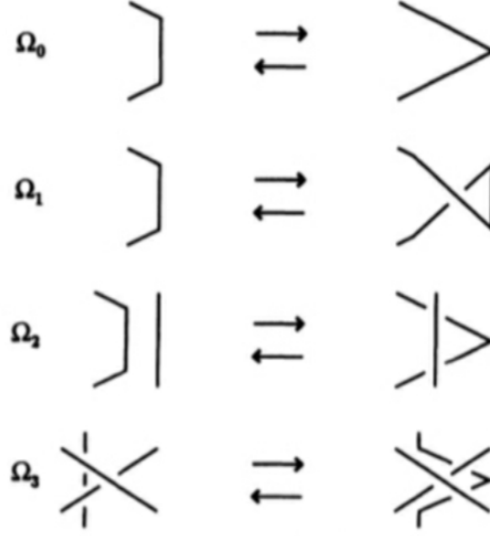


Figure 7: Reidemeister moves

### Definition 10

Given an oriented knot  $K$ , then we may assign to it a Laurent polynomial  $\nabla_k(z)$ , with a fixed indeterminate  $z$ , by means of the of the following two axioms:

Axiom 1 :If  $k$  is the trivial knot ,then we assign  $\nabla_k(z) = 1$

Axiom 2: Suppose that  $D_+, D_-, D_0$  are the regular diagrams, respectively, of the three knots (or links ),  $K_+, K_-, K_0$ . These regular diagrams are exactly the same except at a neighbourhood of one crossing point .In this neighbourhood, the regular diagrams differ in the manner shown in figure 8.

(Note: In the case of  $D_+$  ( $D_-$ ) within this neighbourhood, there exists only a positive (negative) crossing).

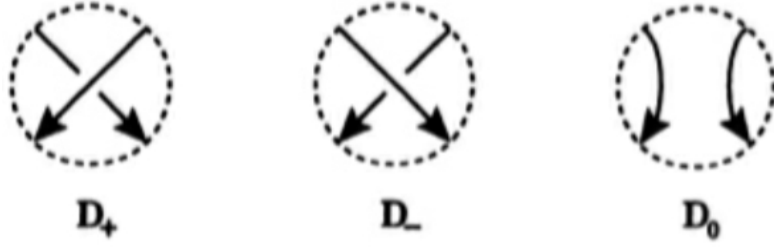


Figure 8: skein diagrams

*Then the Laurent polynomials of the three knots(or links) are related as follows:*

$$\nabla_{K_+}(z) - \nabla_{K_-}(z) = z\nabla_{K_0}(z).$$

*The three regular diagrams  $D_+, D_-, D_0$  formed as above are called skein diagrams, and the relation, between the Laurent polynomials of  $K_+, K_-, K_0$  (whose regular diagrams these are) is called the Skein relation. Also an operation that replaces one of  $D_+, D_-, D_0$  by the other two is called a skein operation.*

# **CHAPTER - II**



## 2. Kauffman Bracket and Jones Polynomial

*In this chapter we discuss about the Kauffman Bracket and Jones Polynomial and whether the Kauffman Bracket satisfies knot invariant or not*

### 2.1 Kauffman Bracket

#### Definition.2.1.1

*The Kauffman Bracket of a link diagram  $L$ , is a polynomial in integer powers of the variable  $A$ , denoted by  $\langle L \rangle$ , defined by the following rules.*

1.  $\langle \bigcirc \rangle = 1$
2.  $\langle L' \cup \bigcirc \rangle = (-A^2 - A^{-2}) \langle L' \rangle$
3.  $\langle \nearrow \searrow \rangle = A \langle \rangle \langle \rangle + A^{-1} \langle \frown \rangle$   
 $\langle \searrow \nearrow \rangle = A^{-1} \langle \rangle \langle \rangle + A \langle \smile \rangle$

Figure 1:

#### Example.2.1.1

*We compute the following bracket polynomials*

$$\langle \bigcirc \bigcirc \rangle \text{ and } \langle \bigcirc \bigcirc \rangle.$$

Figure 2:

**Sol:**

**Part.1**

*Refer Fig.3*

$$\begin{aligned}
\langle \bigcirc \bigcirc \rangle &= A \langle \bigcirc \bigcirc \rangle + A^{-1} \langle \bigcirc \rangle && \text{(Rule 3)} \\
&= A(-A^2 - A^{-2}) \langle \bigcirc \rangle + A^{-1} \langle \bigcirc \rangle && \text{(Rule 2)} \\
&= A(-A^2 - A^{-2}) + A^{-1} && \text{(Rule 1)} \\
&= -A^3
\end{aligned}$$

Figure 3:

$$\begin{aligned}
\langle \bigcirc \bigcirc \rangle &= A^{-1} \langle \bigcirc \bigcirc \rangle + A \langle \bigcirc \rangle && \text{(Rule 3)} \\
&= A^{-1}(-A^2 - A^{-2}) \langle \bigcirc \rangle + A \langle \bigcirc \rangle && \text{(Rule 2)} \\
&= A^{-1}(-A^2 - A^{-2}) + A && \text{(Rule 1)} \\
&= -A^{-3}
\end{aligned}$$

Figure 4:

## Part.2

*Refer Fig.4*

### Theorem :2.1.2

*If a diagram is changed by a type-I move its bracket polynomial changes in the following way*

*(i).Refer Fig.5*

*(ii).Refer Fig.6*

$$\langle \bigcirc \rangle = -A^3 \langle \bigcirc \rangle$$

Figure 5:

and

$$\langle \text{Diagram 6} \rangle = -A^{-3} \langle \text{Diagram 6} \rangle$$

Figure 6:

proof  
Part-1

$$\begin{aligned} \langle \text{Diagram 7} \rangle &= A \langle \text{Diagram 7} \rangle + A^{-1} \langle \text{Diagram 7} \rangle \\ &= A(-A^{-2} - A^2) \langle \text{Diagram 7} \rangle + A^{-1} \langle \text{Diagram 7} \rangle \\ &= -A^3 \langle \text{Diagram 7} \rangle \end{aligned}$$

Figure 7:

Part-2

$$\begin{aligned} \langle \text{Diagram 8} \rangle &= A \langle \text{Diagram 8} \rangle + A^{-1} \langle \text{Diagram 8} \rangle \\ &= A \langle \text{Diagram 8} \rangle + A^{-1}(-A^{-2} - A^2) \langle \text{Diagram 8} \rangle \\ &= -A^3 \langle \text{Diagram 8} \rangle \end{aligned}$$

Figure 8:

**Theorem:2.1.2**

If a Diagram  $D$  is changed by a Type-II or Type-III Reidemister move, then  $\langle D \rangle$  does not change .That is

- (i)Refer Fig.9
- (ii)Refer Fig.10

$$\langle \text{Diagram 1} \rangle = \langle \text{Diagram 2} \rangle$$

Figure 9:

and

$$\langle \text{Diagram 3} \rangle = \langle \text{Diagram 4} \rangle$$

Figure 10:

**Proof:**

**Part-1**

$$\begin{aligned} \langle \text{Diagram 1} \rangle &= A \langle \text{Diagram 5} \rangle + A^{-1} \langle \text{Diagram 6} \rangle \\ &= -A^{-2} \langle \text{Diagram 7} \rangle + \langle \text{Diagram 8} \rangle + A^{-2} \langle \text{Diagram 9} \rangle. \end{aligned}$$

Figure 11:

**Part-2**

$$\begin{aligned} \langle \text{Diagram 10} \rangle &= A \langle \text{Diagram 11} \rangle + A^{-1} \langle \text{Diagram 12} \rangle \\ &= A \langle \text{Diagram 13} \rangle + A^{-1} \langle \text{Diagram 14} \rangle \\ &= \langle \text{Diagram 15} \rangle. \end{aligned}$$

Figure 12:

**Example :2.1.2**

Show that the bracket polynomial of the following trefoil knot (Refer Fig.13) is  $A^{-7} - A^{-3} - A^5$ ,

**Sol:**

Refer Fig.14



Figure 13:

$$\begin{aligned}
 \langle \text{Figure 13} \rangle &= A \langle \text{Figure 13} \rangle + A^{-1} \langle \text{Figure 13} \rangle \\
 &= A(-A^4 - A^{-4}) + A^{-7} \\
 &= (A^{-7} - A^{-3} - A^5).
 \end{aligned}$$

Figure 14:

**Example:2.1.3**

Show that the bracket polynomial of the following simple two-component link (Refer Fig.15) is  $-A^4 - A^{-4}$

**Sol:**

Refer Fig.16

**Example :2.1.4**

The bracket polynomial of the trivial link of  $n$ -components

**Sol:**

By rule.2 we have

$$\begin{aligned}
 \langle L' \cup \bigcirc \rangle &= (-A^2 - A^{-2}) \langle L' \rangle \\
 &= (-1)(A^2 + A^{-2}) \langle L' \rangle
 \end{aligned}$$

The bracket polynomial of usual projection of the trivial link of  $n$ -components will be

$$\langle \bigcirc \cup \bigcirc \cup \dots \cup \bigcirc \rangle = (-1)^{n-1} (A^2 + A^{-2})^{n-1}$$



Figure 15:

$$\begin{aligned} \langle \text{Diagram 15} \rangle &= A \langle \text{Diagram 16a} \rangle + A^{-1} \langle \text{Diagram 16b} \rangle \\ &= (-A^4 - A^{-4}). \end{aligned}$$

Figure 16:

**Definition.2.1.2**

The writhe  $\omega(D)$  of a diagram  $D$  of an oriented link is the sum of the signs of the crossings of  $D$  where each crossings has sign  $+1$  or  $-1$

**Theorem:2.1.3**

Let  $D$  be a diagram of an oriented link  $L$ . Then the expression  $(-A)^{-3\omega(D)} \langle D \rangle$  is an invariant of the oriented link  $L$

**Proof:**

As proven the bracket polynomial is unaffected by Reidemeister moves of type-II and type-III. Further more since a type-II move replaces crossings of opposite signs with no crossings ( regardless of which orientation is chosen ),  $\omega(D)$  is also unchanged by a type-II move

similarly,  $\omega(D)$  is unchanged by a type-III move and hence the claim is true for Reidemeister moves of type-II and III

consider a diagram  $D$  and let  $D'$  be the diagram after a link has been inserted using a type-I move. Then  $\omega(D') = \omega(D) + 1$  regards of the orientation.

**Hence**

$$\begin{aligned} (-A)^{-3\omega(D')} \langle D' \rangle &= (-A)^{-3(\omega(D)+1)} \langle D \rangle \\ &= (-A)^{-3\omega(D)} (-A)^{-3} (-A)^3 \langle D \rangle \\ &= (-A)^{-3\omega(D)} \langle D \rangle \end{aligned}$$

## 2.2.Jones Polynomial

**Definition:2.2.1** Suppose  $K$  is an oriented knot (or a link) and  $D$  is a oriented regular diagram for  $K$ . Then the Jones polynomial of  $K$ ,  $V_k(t)$  can be defined from the following two axioms. The polynomial itself is a Laurent's polynomial in  $\sqrt{t}$  (i.e) it may have terms in which  $\sqrt{t}$  has a negative exponent. (we assume  $(\sqrt{t})^2 = t$ ). The polynomial  $V_k(t)$  is an invariant of  $k$ .

**Axiom.1:**

If  $K$  is a interval knot, then  $V_k(t)=1$

**Axiom.2:**

Suppose that  $D_+, D_-, D_0$  are Skein diagrams, then the following Skein relation holds.

**Theorem:2.2.1:** Suppose that  $L$  is a  $\mu$ -component(oriented) link, then  $V_L(1) = (-2)^{\mu-1}$

A consequence of this proposition is that the Jones polynomial can never be zero.

**Proof:**

Since if we substitute  $t=1$  into the skein formula we obtain  $V_{L+}(1) - V_{L-}(1) = 0$ , it follows that

$$V_{L+}(1) = V_{L-}(1) = V_{O_\mu}(1)$$

We also know that, For the trivial  $\mu$ -component link  $O_\mu$ ,  $V_{O_\mu}(t) = (-1)^{\mu-1}(\sqrt{t} + \frac{1}{\sqrt{t}})^{\mu-1}$

$$(i.e) V_{O_\mu}(1) = (-2)^{\mu-1} = V_1$$

**Theorem:2.2.2**

If  $K$  is a knot or link, then

$$V_k(-1) = (-1)^{\mu(k)-1} \Delta_k(-1)$$

where  $\mu(k)$  denotes the number of components of  $K$ .

**Proof:**

If we let  $t=1$  in Axiom(2) then  $-V_{k+}(-1) + V_{k-}(-1) = (\sqrt{-1} - \frac{1}{\sqrt{-1}})V_{k_0}(-1)$

We also know that ,

$$\Delta_{k+}(t) - \Delta_{k-}(t) = (\sqrt{-1} - \frac{1}{\sqrt{-1}})V_{k_0}(t) \quad (2)$$

Sub  $t=-1$  in (2) we obtain

$$\Delta_{k+}(-1) - \Delta_{k-}(-1) = (\sqrt{-1} - \frac{1}{\sqrt{-1}}) \Delta_{k_0}(-1) \quad (3)$$

Next, we multiply (3) by the factor  $-(-1)^{(k+)-1}$  and hence

$$\begin{aligned} & -(-1)^{\mu(k+)} - 1 \Delta_{k+}(-1) + (-1)^{\mu(k+)} - 1 \Delta_{k-}(-1) = \\ & -(-1)^{\mu(k+)} - 1 (\sqrt{-1} - \frac{1}{\sqrt{-1}}) \Delta_{k_0}(-1) \end{aligned}$$

$$\begin{aligned} & \text{Since, } \mu(k-) = \mu(k+) \text{ and } \mu(k_0) = \mu(k+) + 1 \text{ (or) } \mu(k+) - 1 \\ & -(-1)^{\mu(k+)-1} \Delta_{k_0}(-1) = (-1)^{\mu(k_0)} - 1 \Delta_{k_0}(-1) \end{aligned}$$

and

$$\begin{aligned} & -(-1)^{\mu(k+)-1} \Delta_{k+}(-1) + (-1)^{\mu(k-)-1} \Delta_{k-}(-1) = (-1)^{\mu(k_0)-1} (\sqrt{-1} - \frac{1}{\sqrt{-1}}) \Delta_{k_0}(-1) \\ & (4) \end{aligned}$$

(1) (4) are exactly the same Skein relation , the required result follows



# **CHAPTER - III**

## The Seifert Matrices

### 3.1.The Seifert surface

Let us begin with the following theorem due to L.Pontrjagin and F.FRANKL

#### Theorem 3.1.1

Given an arbitrary oriented knot(or link) $K$ ,then there exists in  $R^3$  an orientable, connected surface,  $F$ , that has as its boundary  $K$ .(that is to say, there exists an orientable connected surface that spans  $K$ .)

**proof**

Suppose that  $K$  is an oriented knot (or link)and  $D$  is a regular diagram for  $K$ . Our intention is to decompose  $D$  into several simple closed curves. The first step is to draw a small circle with one of the crossing points of  $D$  as its centre.This circle intersects  $D$  at four points, say,  $a, b, c$ , and  $d$ , Figure 3.1.1(a).As shown in Figure 3.1.1(b),let us splice this crossing point and connect  $a$  and  $d$ , and  $b$  and  $c$ .



Figure 1: 3.1.1

What we have done is to change the original segments  $ac$  and  $bd$  into new segments  $ad$  and  $bc$ .In this way we can remove the crossing point of  $D$  that lies within the circle.This operation is called splicing of a knot  $K$  (along its orientation )at a crossing point of  $D$ , then we shall remove all the crossing points from  $D$ . The end result is that  $D$  becomes decomposed into several simple closed curves, Figure 3.1.2 (b)). These curves are called seifert curves.  $D$ ,itself , has been transformed into a regular diagram of a link on the plane that possess no crossing points (i.e., the trivial link). Each of these simple closed curves may now be spanned by a disk.

In the case of Figure 3.1.2(b), by slicing we obtain three disks,  $D_1, D_2, D_3$ , Figure 3.1.2(c). The boundary of  $D_i$  is the seifert curve  $C_i$  . In the figure

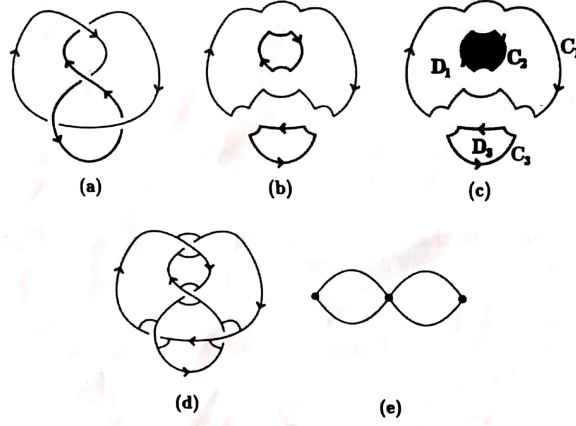


Figure 2: 3.1.2

3.1.2 (c), there is a possibility that  $D_2$  may lie on the top off  $D_1$ , or  $D_2$  may be under  $D_1$ . Finally in order to create a single surface from various disks, we need to attach to these disks small bands that have been given a single twist. To do this, firstly take a square  $abcd$  and give it a single positive or negative twist Figure 3.1.3(a) and (b), respectively; these twisted squares are the required bands.

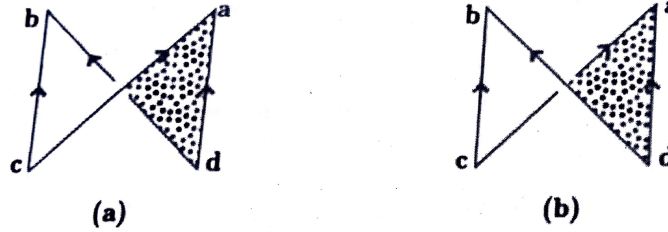


Figure 3: 3.1.3

If we attach positive (negative) bands at the places of  $D$  that corresponded to positive (negative) crossing points before they were spliced, then we obtain a connected, orientable surface  $F$ , Figure 3.1.2 (d). (In the case of a link,  $K$ , if we alter  $K$  in such a way that the projection of  $K$  is connected, then by the above method we can also obtain a connected surface.) The boundary of this surface,  $F$ , is plainly the original knot  $K$ . Further, as noted above,  $F$  is also an orientable surface.

As shown in Figure 3.1.4(a), by shading the front of the surface and dotting the back of the surface, we may distinguish between the front and back of the

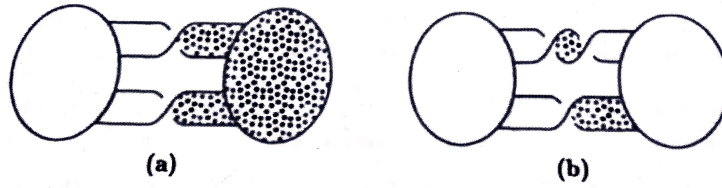


Figure 4: Figure3.1.4

surface. This allows us to assign an orientation to the surface. However, as in Figure 3.1.4(b), if one of the bands has a double twist, then it is not possible for us to distinguish front and the back.

### 3.2.The genus of a knot

The theorem states that a closed (i.e., one that is compact and without boundary) orientable surface,  $F$ , is topologically equivalent (i.e., homeomorphic) to the sphere with several handles attached to its surface. The number of these handles is called the genus of  $F$ , and is denoted

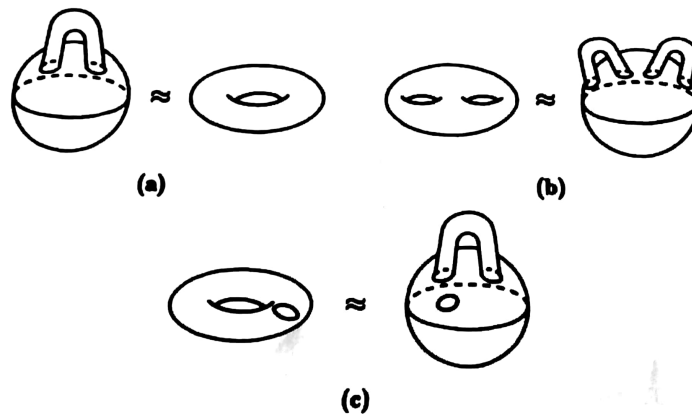


Figure 5: Figure3.2.1

#### Example 3.2.1

The surface of genus 1, shown in Figure 3.2.19(a), is called a torus, while the surface in Figure 3.2.1(b) has genus 2.

**Theorem 3.2.1.**

We may divide a closed orientable surface into  $\alpha_o$  points,  $\alpha_1$  edges, and  $\alpha_2$  faces. Let

$$\chi(F) = \alpha_o - \alpha_1 + \alpha_2$$

then  $\chi(F)$  is an integer that is independent of how we have divided  $F$ ; i.e., it is only dependent on  $F$ . This integer is called the Euler characteristic of  $F$ .

The Euler characteristic  $\chi(F)$  and the genus of  $F$ ,  $g(F)$ , are related by means of the following equation:

$$\chi(F) = 2 - 2g(F).$$

Therefore,

$$g(F) = \frac{2 - \chi(F)}{2}.$$

If  $F$  has a boundary, since the boundary is also composed of several points and edges, the above formula becomes

$$\chi(F) = 2 - \mu(F) - 2g(F),$$

where  $\mu(F)$  is the number of closed curves that make up the boundary of  $F$ .

**Example 3.2.2**

We can divide the torus with a hole in the manner shown in Figure 3.2.2, so that  $\alpha_o = 7$ ,  $\alpha_1 = 14$ , and  $\alpha_2 = 6$ . It follows from this that  $\chi(F) = -1$ , and therefore  $g(F) = 1$ .

**Exercise: 3.2.1**

Show, by suitably dividing it, that the sphere  $S^2$  has the Euler characteristic 2. (Refer figure 3.2.2)

**Solution:**

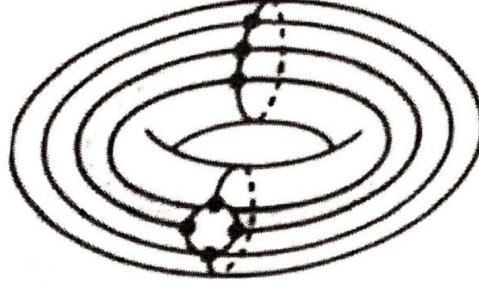


Figure 6: Figure 3.2.2

Let us now apply the above Euler characteristic, to the seifert surface constructed. We may think of the disks and bands of  $F$  as a division of  $F$ . The points of  $F$  in this division are the four vertices of each band. The edges of  $F$  are the polygonal curves that constitute the edges of bands and the boundaries of the disk between the vertex points. The faces of  $F$  are the disks and the bands.

### Exercise 3.2.2

Show that if  $d$  is the number of disks and  $b$  the number of bands, then  $\alpha_0 = 4b$ ,  $\alpha_1 = 6b$ , and  $\alpha_2 = b + d$ .

**solution:**

We know that,

$$\chi(F) = 4b - 6b + b + d$$

(from Exercise 3.2.2)

$$\chi(F) = d - b.$$

Further,  $\mu(K)$  is just the number of components of the link  $K$ . So from (3.2.2) we obtain that

$$\begin{aligned} 2g(F) &= 2 - \mu(K) - \chi(F). \\ 2g(F) &= 2 - \mu(K) - d + b, \\ 2g(F) + \mu(K) - 1 &= 1 - d + b. \end{aligned}$$

In the special case when  $K$  is a knot, since  $\mu(K) = 1$  it follows that

$$2g(F) = 1 - d + b.$$

For the rest of the section let us consider this number , i.e.,  $1 - d + b$ . Suppose  $\Gamma(D)$  is the seifert graph constructed from the seifert surface in Figure 3.2.1(e). Since  $\Gamma(D)$  is a plane graph ,  $\Gamma$  divides  $S^2$  into several domains. (We may think of the sphere  $S^2$  as  $R^2$  with the addition of the point at infinity.) In this partition of  $S^2$ , the number of points is  $d$  and the number of edges is  $b$ . Suppose that the number of faces is  $f$ ; then from Theorem 3.2.1 and Exercise 3.2.1 we obtain,

$$2 = \chi (S^2) = d - b + f .$$

Therefore,

$$f - 1 = 1 - d + b ,$$

i.e.,  $1 - d + b$  is equal to the number of faces of this division of  $S^2$ , excluding the face that contains the point at infinity,  $\infty$ .

### 3.3.The Seifert matrix

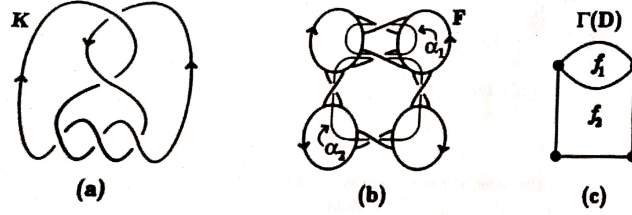


Figure 7: Figure 3.3.1

Suppose that  $F$  is a seifert surface created from the regular diagram ,  $D$ , of a Knot (or link)  $K$ , and  $\Gamma(d)$  is its seifert graph. We want to create exactly  $2g(F) + \mu(K) - 1$  closed curves that lie on  $F$ .

When  $\Gamma(D)$  partitions  $S^2$  , then we showed in the previous section that  $2g(F) + \mu(K) - 1$  (  $= 1 - d + b = f - 1$  ) is equal to the number of domains (excluding the domain that contains  $\infty$ ). The boundary of each these domains (faces) is a closed curve of  $\Gamma(D)$ . Therefore, We can form these closed curves, create the closed curves on the seifert surface.

#### Example 3.3.1.

If we transform the regular diagram of the right hand tefoil knot to the one in Figure 3.3.2(a), then it is fairly straightforward to see its seifert surface is the one in Figure 3.3.2(b) and the subsequent Seifert graph is as in

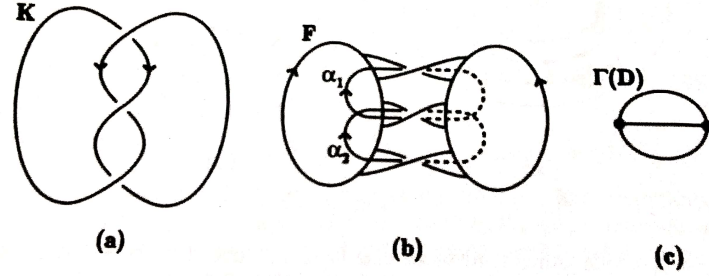


Figure 8: Figure 3.3.2

Figure 3.3.2(c).

From Figure 3.2.2 (b) it follows that there are two closed curves  $\alpha_1$  and  $\alpha_2$  on the seifert surface. The mutual relationship between  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_1^\#$  and  $\alpha_2^\#$  are shown in Figure 3.3.3(a) ~ (d).

From these four diagrams it follows that

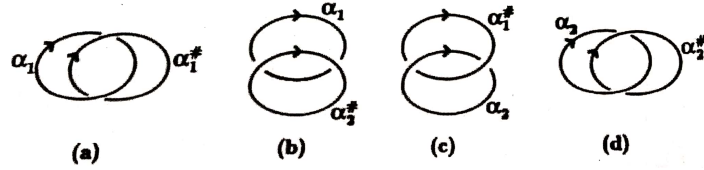


Figure 9: Figure 3.3.3.

$$lk(\alpha_1, \alpha_1^\#) = -1, \quad lk(\alpha_2, \alpha_1^\#) = 1, \quad lk(\alpha_2, \alpha_2^\#) = -1,$$

with the linking number of the other case equal to 0.

Therefore, the seifert matrix for the right-hand trefoil knot is  $M = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}$ .



# **CHAPTER-IV**

# Fundamental Problems of Knot Theory

*The problems that arise when we study the theory of knots can essentially be divided into two types. On the one hand, there are those that we shall call Global problems, while, in contrast, there are those that we shall call Local problems.*

*Global problems concern themselves with how the set of all knots behaves. As the label implies, in contraposition, Local problems are concerned with the exact nature of a given knot. As to the question which is the more important, and hence we should concentrate our attention on, the unhelpful answer is that it is impossible to say. In order to solve Global problems it is often necessary to find solutions to various Local problems. Conversely, the determination of Local problems may rely on how they fit within the Global problem.*

*In this chapter, we shall explain and give examples of these two types of problems. Problems in the theory of knots are not just limited to this bifurcation into Global and Local problems. However, in the past the above dichotomy has formed the axis around which knot theory has developed, and it is more than likely that this will substantially remain the case in the foreseeable future.*

## 1 Global Problems:-

*One of the typical classical Global problems is the classification problem.*

### (1) The Classification problem

*The classification problem, at least in definition, is very straightforward, as the name suggests we would like to create a complete knot (or link) table. What exactly we mean by a complete table is one in which, firstly, no two knots are equivalent, and, secondly, a given arbitrary knot is equivalent to some knot in this table.*

At the time of writing, a complete table in the above strict sense has been compiled only up to prime knots with 13 crossings. One future problem is to steadily expand this table. Another (sub-) problem that germinates directly from the original classification problem is to create a complete table for only certain specific types of knots, for example, for alternating knots. As we introduce other types of knots, this question of whether we classify them completely will always be in the vanguard of the questions that we will ask ourselves. In fact, in Chapters Torus Knots and Tangles and 2-Bridge Knots we shall discuss two specific knot types that have been completely classified.

## (2) A fundamental conjecture

*This conjecture can be immediately stated as follows:*

***If  $S^3 - K_1$  and  $S^3 - K_2$  , which are usually called complementary spaces, for two knots  $K_1$  and  $K_2$  , respectively, are homeomorphic, then the knots are equivalent.***

*This conjecture can readily be seen to be the converse of Theorem [If two knots  $K_1$  and  $K_2$  that lie in  $S^3$  are equivalent, then their complements  $S^3 - K_1$  and  $S^3 - K_2$  are homeomorphic. ]*

*In the late 1980s this conjecture was, in fact, proven by C. McA Gordon and J. Luecke [GL]. As a consequence of this result, the problem of knots in  $S^3$  transforms itself from what we may call a relative problem which concerned itself with the shape of a knot in  $S^3$ , into an absolute problem, which now concerns itself with the study of the complementary spaces.*

*However, much to our dismay we cannot always transform a relative problem into an absolute problem. The counterexample that immediately comes to hand is that, in fact, the above fundamental conjecture is false in the case of links.*

**Example 1:- Although the two links in Figure 1 are not equivalent, their complementary spaces are homeomorphic<sup>7</sup>.**

*In general, results that hold for knots pass through fairly readily to hold for links as well. However, as the above example shows, we cannot take this for granted.*

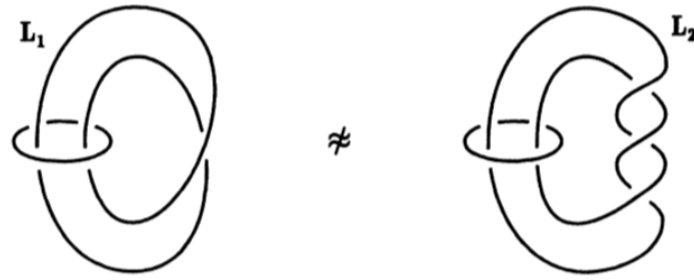


Figure 1:

### (3)Kont invariants

*As a way of determining whether two knots are equivalent, the concept of the knot invariant plays a very important rôle. The types of knot invariants are not just limited to, say, numerical quantities. These knot invariants can also depend on commonly used mathematical tools, such as groups or rings.*

*Suppose that to each knot,  $K$ , we can assign a specific quantity  $\rho(K)$ . If for two equivalent knots the assigned quantities are always equal, then we call such a quantity,  $\rho(K)$ , a knot invariant. This concept of assigning some mathematical quantity to an object under investigation is not limited just to knot theory, it can be found in many branches of mathematics. Probably the simplest analogous example occurs in group theory. The number of elements in a group, called the order of the group, is a group invariant, since for isomorphic groups their respective orders are equal.*

*We know that if a knot  $K$  and another knot  $K'$  are equivalent, then it is possible to change  $K$  into  $K'$  by applying the elementary knot moves to  $K$  a finite number of times. Therefore, for a quantity  $\rho(K)$  to be a knot invariant,  $\rho(K)$  should not change as we apply the finite number of elementary knot moves to the knot  $K$ . It follows from this, for example, that the number of edges of a knot is not a knot invariant. The reason is that the operations defined in Elementary Knot Moves.(1) and (1)' either increase or decrease the number of edges. Similarly, if we consider the operations in (2) and (2)' of the same definition, then it also follows that the size of a knot is not a knot invariant.*

*A knot invariant, in general, is unidirectional, i.e.,*

*if two knots are equivalent then their invariants are equal.*

For many cases the reverse of this arrow does not hold. In contraposition, if two knot invariants are different then the knots themselves cannot be equivalent, and so a knot invariant gives us an extremely effective way to show whether two knots are non-equivalent. The history of knot theory may be said to be an account of how the various knot invariants were discovered and their subsequent application to various problems. To find such knot invariants is by definition a Global problem. On the other hand, to actually calculate many of these knot invariants, which we shall discuss in Chapter Classical Knot Invariants, is quite difficult. Further, to find a method to calculate these invariants is also a Global problem.

## 2 Local Problems:-

To illustrate and explain the idea of a Local problem, we shall give several examples.

### (1) When are a knot $K$ and its mirror image $K^*$ equivalent?

If  $K$  and  $K^*$  are, in fact, equivalent, then we say that  $K$  is an amphicheiral knot (sometimes also referred to as an **achiral knot**). For example, since the right-hand trefoil knot [Figure 2(b)] and its mirror image, the left-hand trefoil

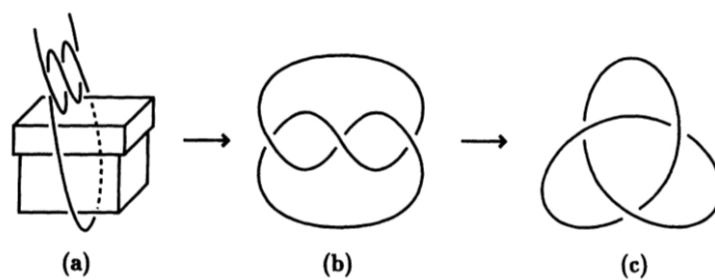


Figure 2:

knot (Figure 3), are not equivalent, the trefoil knot is not amphicheiral. On the other hand, however, the knot is amphicheiral. Due to the extremely special

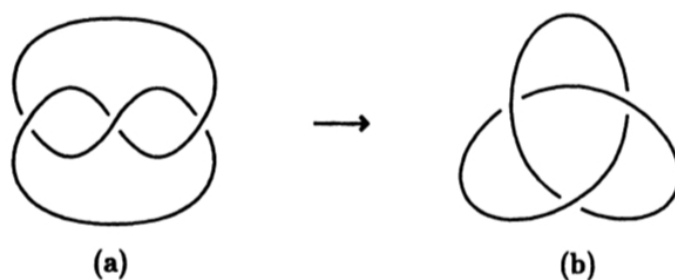


Figure 3:

*nature of amphicheiral knots, there are, in relative terms, very few of them. This (Local) problem has over the years been quite extensively studied, and for particular types of knots many amphicheiral results have been proven.*

## (2) When is a given knot prime?

*In the way described in Figure 4 , a regular diagram of  $K_1 \# K_2$ , the connected*

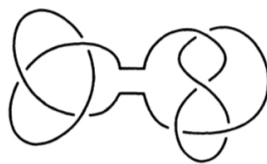


Figure 4:

*sum of  $K_1$  and  $K_2$  may be constructed by placing the regular diagrams of  $K_1$  and  $K_2$  side by side and then connecting them by means of two parallel segments.*

*Therefore, if a knot  $K$  can be decomposed into  $K_1$  and  $K_2$ , then  $K$  has a regular diagram of this type shown in Figure 4. However, although theory predicts this in practice, since most regular diagrams of non-prime knots are usually not so nicely presented, we cannot deduce from the regular diagram whether a knot is prime.*

**Example 2:- The regular diagram of the knot,  $K$ , shown in Figure 5(a) is not of the form of Figure 4, but  $K$  is not a prime knot.**

*Recently, this (Local) problem has been completely resolved in the case of alternating knots*

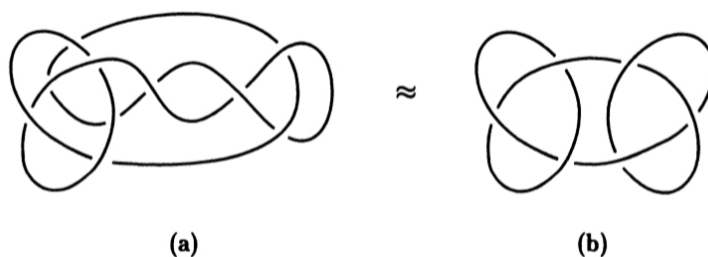


Figure 5:

### (3) When is a knot invertible?

We know that we can assign to a knot two different, opposite orientations. Let us denote one of these knots by  $K$  and the other, with the opposite orientation,  $-K$ . We would like to determine whether  $K$  and  $-K$  are equivalent. When  $K$  and  $-K$  are, in fact, equivalent, then  $K$  is said to be invertible. Knots with a relatively small number of crossing points are in general invertible. It follows from the Figure 6 that the left-hand trefoil knot is an example of an invertible knot.

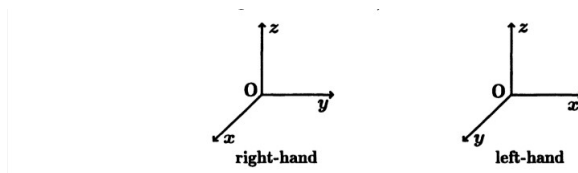


Figure 6:

That non-invertible knots do exist was first shown by H.F. Trotter in 1963<sup>8</sup>. The knot in Figure 7(a) was the example that was given by Trotter; following this discovery, many other non-invertible knots were soon found.

In contrast to 1963, it is now fair to say that almost all knots are non-invertible. We have drawn in Figure 7(b) the simplest non-invertible knot.

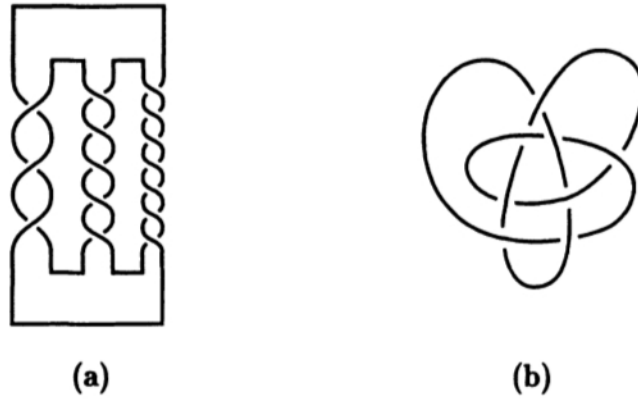


Figure 7:

#### (4) What is the period of a knot?

*If we rotate the figure 8 knot, Figure 8(a), by an angle of  $\pi$  about the  $Oz$ -axis, the figure will rotate to its original form. So, this knot may be said to have period*

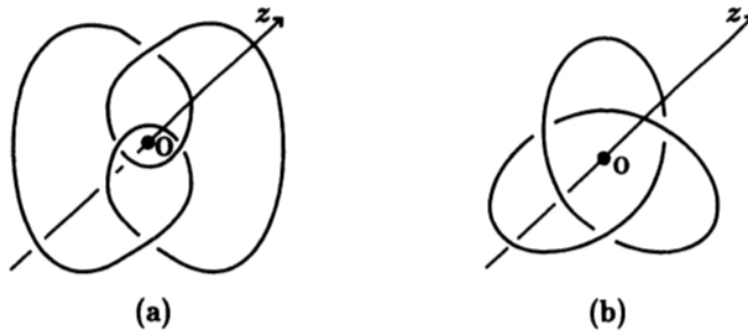


Figure 8:

*2. The left-hand trefoil knot, Figure 8(b), if it is rotated by  $\frac{2\pi}{3}$  about the  $Oz$ -axis, will also rotate to its original shape. In general, if we can rotate a knot by an angle  $\frac{2\pi}{n}$  about a certain axis so that it rotates to its original shape, then we say that this knot has period  $n$ . In this case, the (Local) problem is to determine all the periods for a given knot. This problem has, also, been extensively studied and has been completely solved for particular types of knots.*



## (5) When is a knot a slice knot?

*Of all the (Local) problems that we have so far discussed, this is probably by far the most difficult. The present state of affairs is that only several necessary conditions are known for a knot to be a slice knot. Further, effective methods to determine slice knots are also not known. Therefore, this (Local) problem seems at the moment to be quite intractable.*

*The subsequent chapters will be an exposition of knot theory, which will take their bearings from the bifurcation of knot theory problems outlined in this chapter, namely, the Global and Local problems.*

# **CHAPTER-V**

## 5.1. QUANTUM COMPUTATION

Classical computers use a bit to encode information. A bit stores either a 0 or 1 and data is encoded in strings of 0's and 1's. Logical gates act on these strings of data taking two bits as input and producing a single bit as output. Clearly, what we think of as a "computer computation" requires thousands of bits and operation. To perform a task on a computer, we need to know how many bits are needed to store the data and how many operations (such as addition or multiplication) are required to perform the computation. The number of operations required to perform the task is referred to as the computational complexity. To put this in perspective, currently, a computation on  $x$  bits requires a number of operations that can be expressed as a polynomial in the variable  $x$ .

For example, let's consider matrix multiplication of an  $n \times n$  matrix. In an  $n \times n$  matrix, there are  $n^2$  entries. How many operation does it take to multiply two  $n \times n$  matrices? We let  $A, B$  and  $C$  denote  $n \times n$  matrices and suppose that  $AB=C$ . Recall that the entry in the  $i$ -th row and  $j$ -th column of  $C$  is computed as follows

$$C_{ij} = \sum_{k=1}^n A_{ik} B_{kj} \quad (1)$$

The computation of  $C_{ij}$  in equation (1) requires  $n$  multiplication operation and  $n-1$  addition operation. A single entry in the matrix  $C$  requires roughly  $2n$  operations. There are  $n^2$  entries. computing the entire matrix  $C$  requires roughly  $2n^3$  operations. Notice that  $2x^3$  is a polynomial, so we say that we can compute matrix multiplication in polynomial time.

Now consider the bracket polynomial. A virtual link diagram  $D$ , with  $n$  classical crossings, has  $2^n$  states. Recall the definition of the bracket polynomial:

$$\langle D \rangle = \sum_{s \in S} A^{\alpha(s) - \beta(s)} d^{|s|} \quad (2)$$

There are  $2^n$  states and each state must be evaluate (count the number of total loops) and then the values are added. Taking this simplistic viewpoint, each state requires two operation: Evaluation and addition to a running total. For a virtual link diagram with  $n$  classical crossings, that is  $(2^n) \times 2$  total operations. The total number of computations is an exponential expression based on the number of classical crossings.

Algorithms or computations that require an exponential number of calculations are challenging, if not impossible to compute. Their rapid growth in size means that it is very easy to run out of computer memory.

Quantum computation is fundamentally different from classical computation. The basic unit of data is a **qubit**-a "quantum bit". A qubit is neither a 0 or a 1, but a **superposition** of the two classical states. Qubits are expressed as linear combinations:

$$|\Psi\rangle = \alpha |0\rangle + \beta |1\rangle \quad (3)$$

Where  $\alpha$  and  $\beta$  are complex number. The notation  $|\psi\rangle$  is called a ket and is part of Dirac notation. When measured, the qubit collapses to a 0 with probability  $|\alpha|^2$  and state 1 with probability  $|\beta|^2$ . Until the qubit is measured, the qubit carries information regarding both bits. Returning to the bracket polynomial, a type A smoothing could be encoded as a 0 and a type B smoothing could be encoded as a 1. For a virtual knot diagram with  $n$  classical crossing, all  $2^n$  states could be encoded using  $n$  qubits, instead of an  $n(2^n)$  classical bits.

Two famous quantum algorithms are Shor's Algorithm (factoring a number into primes) and Grover's algorithm (sorting an unstructured list). Both of these quantum algorithms are "faster" than their classical counterparts. One reason that knot theorists are interested in quantum algorithms is because there is also a quantum algorithm for the computation of the  $f$ -polynomial.

Another reason for knot theorist's in quantum computation is one of the models of quantum computation. Theoretical models of quantum computation involve three main steps:

1. Initialize the system (construct qubits)
2. Apply logical gates (operation)
3. Measure the outcome (read the results)

The first step (system initialization) consists of choosing a linear combination of qubit vectors of length  $n$  with entries in  $\mathbb{Z}_2$  and complex coefficients. The logical gates (operations) are simulated via matrix multiplication using unitary matrices. (We let  $A^*$  denote the conjugate transpose of a matrix. Then a unitary matrix satisfies the equation  $(A^*)^T = A^{-1}$  or that  $A(A^*)^T = I = (A^*)^T A$ .) One model of quantum computation is topological quantum computation. This model involves anyons particles that are arranged in a line in a 2-D plane. Through some mechanism, these particles move around each other in the plane. Adding the dimension of time, a braid emerges from this scenario. The particles involved in this scenario are called non-abelian anyons.

By braid, we mean an  $n$ -strand tangle with some specific restrictions. The  $n$ -strand **braid group**  $B_n$  has the set of generators  $(\sigma_1, \sigma_2, \dots, \sigma_{n-1})$ . The braid group also contains the  $n$ -strand identity braid as shown in the **Figure 1**. The inverse of  $\sigma_1$  is obtained by switching the positive crossing to a negative crossing. The elements of  $B_n$  consist of all  $n$  -  $n$  tangles formed by concatenation of the generators. The generators satisfy the following identities

$$\sigma_i \sigma_j = \sigma_j \sigma_i \text{ for all } |i - j| > 1 \quad (4)$$

$$\sigma_i \sigma_i \pm 1 \sigma_i = \sigma_i \pm 1 \sigma_i \sigma_i \pm 1 \quad (5)$$

An example of concatenation is shown in **figure-2**. The braid group can be mapped onto a set of  $2n \times 2n$  matrices:  $\phi: B_n \rightarrow SL_n(\mathbb{C})$ . These matrices respect

the Reidemeister moves and, when unitary, can also be viewed as simulated quantum computer. Research has shown these braid operators are universal, meaning that the entire set of logical gates can be constructed using a small set of matrices.

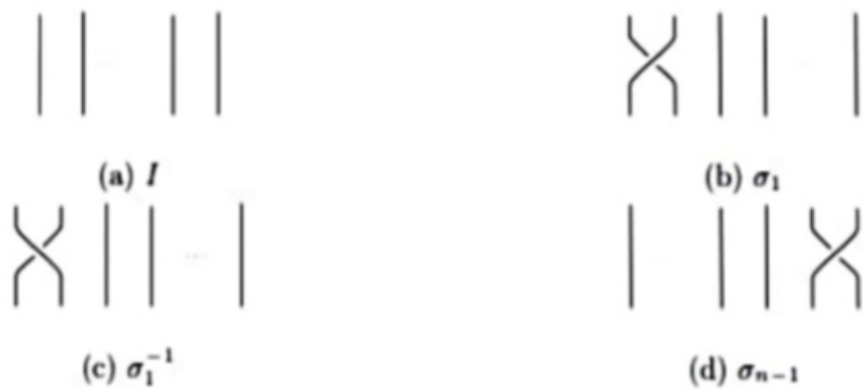


Figure 1:

**Braid Generators**

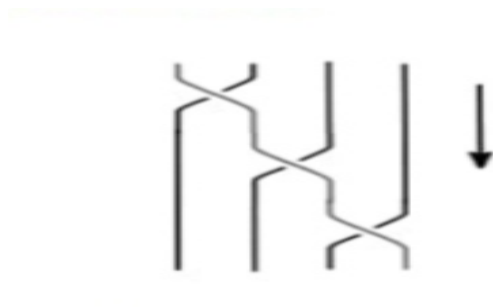


Figure 2:

**Braid:**  $\sigma_1 \sigma_2 \sigma_3$

## (II).APPLICATIONS OF KNOT THEORY

### 5.2 TEXTILES

**KNOTS** by themselves are any accidental or intentional entanglement of cord,braid,ribbon,beading,fabric or other material that will create a new shape or structure by forming loops,intertwining,and weaving of the basic **FABRIC**.The nw structure may be used to enhance or accessorize many forms of dress

**BRAIDING in TEXTILES**machine or hand method of interlacting three or more yarns or bias cut cloth strips in such a way that they cross one another and are laid together in diagonal formation,forming a narrow strip of flat or tublar fabric

Traditional woven textiles have a wrap and weft thread structure the wrap threads extend across the width of the fabric while the weft threads run the length of the fabric.The interactions between the wrap and weft threads are reular repeating along both the length and width of the fabric.Using mathematical terminology, **fabrcic** is a doubly periodic oriented plane knot diagram.Here are the several examples of woven fabric.

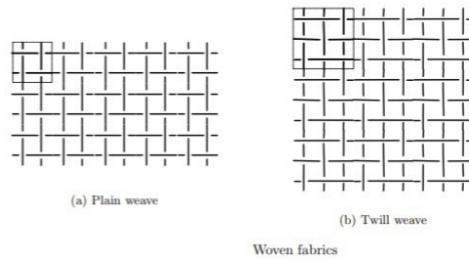


Figure 1:

The fact that the fabric cn b described using a small region of the doubly periodic pattern leads to the following definition from Vassiliev.A **fabrcic kernel** is a  $n$ - $m$  tangle with no virtual crossings as shown below.

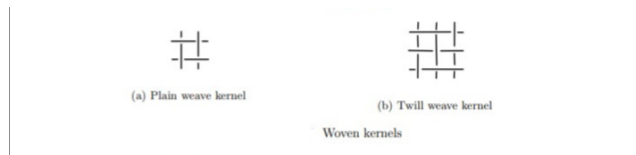


Figure 2:

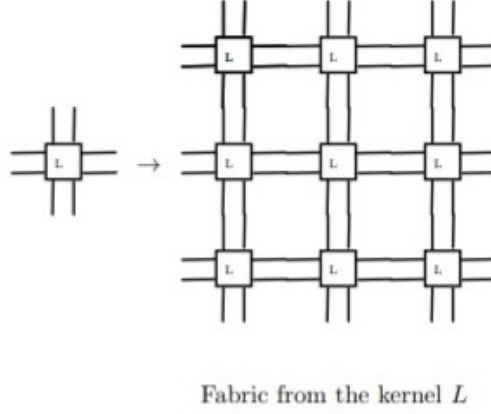


Figure 3:

Given an  $n$ - $m$  tangle  $\mathbf{L}$ , we can construct a fabric kernel as shown in the above figure 3 (fabric from the kernel  $L$ ).

The fabric kernel can be viewed in three distinct contexts: as a link on a torus, as a classical link (where components  $\mathbf{X}$  and  $\mathbf{Y}$  mark the frame of the torus), or as a virtual link diagram. Each possibility is shown in the below figure 4 (kernel visualization)

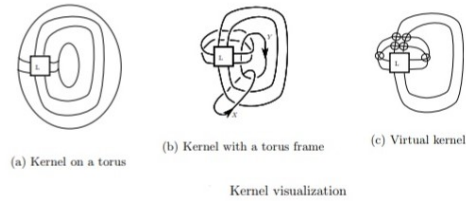
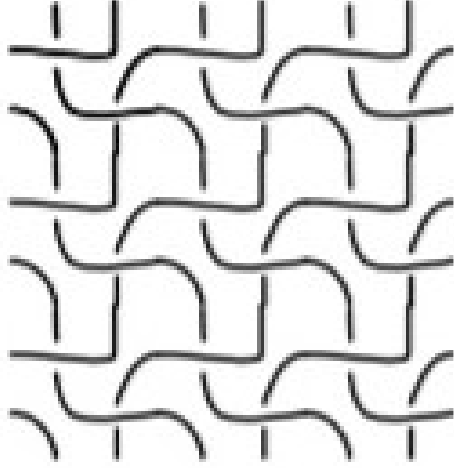


Figure 4:

The following invariants can be applied to the fabric kernel and can be used to describe or differentiate between different fabric kernels. We apply these invariants to the kernel of figure (a) (DIFFERENT FABRICS), which is shown in figure (plain weave tile).

1. The **crossing numbers** of the kernel,  $\mathbf{L}$ , denoted  $C(L)$  count the number of crossings in the kernel. There are 2 crossings which is shown in below figure.



(a) Plain weave

Figure 5:

2. The **number of components** that result when the edges of the fabric kernel are identified to form a torus is also an invariant. In the above diagram plain weave, two components are formed when the edges are identified.

3. The linking number of the components is also an invariant. We order the components of the fabric kernel after identifying the edges. Then, the linking number of the fabric kernel is defined as  $\mathbf{L}(L) = \sum_{i < j} |l(K_i, K_j)|$  where  $l(K_i, K_j)$  is the sum of the signs of the crossings where components  $i$  and  $j$  meet. There are two components in the above diagram (figure a) and  $L(L) = 0$

4. The **axial type** of each strand can also be determined by computing the repeat of an individual strand across the surface of the kernel. The repeat is then expressed in terms of  $\mathbf{X}$  and  $\mathbf{Y}$ , where  $\mathbf{X}$  denotes the oriented, vertical boundary of the fabric kernel as shown in the below figure (6)

**Plain weave tile** pays homage to the irregular qualities of silk, with embellished lustrous accents to add scale. Designed and assessed for environmental and social responsibility across material health, material reuse, renewable energy and social fitness.



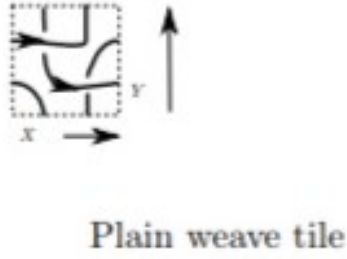


Figure 6:

Tracing the components, we see that the axial type of one component is  $\mathbf{X}-\mathbf{Y}$  and the other component is  $\mathbf{X}+\mathbf{Y}$ . Traditional fabric is either knitted (constructed from one continuous fiber) or woven (constructed from warp and weft fibers). The **number of axial components** identifies the fabric type: knitted fabric has one axial fiber and woven fabric has two axial fibers (warp and weft). However, new industrial processes and 3-D printing open up the possibility of multi-axial fabrics and fabrics that include closed components. Additional fabric structures are shown in the below diagram.

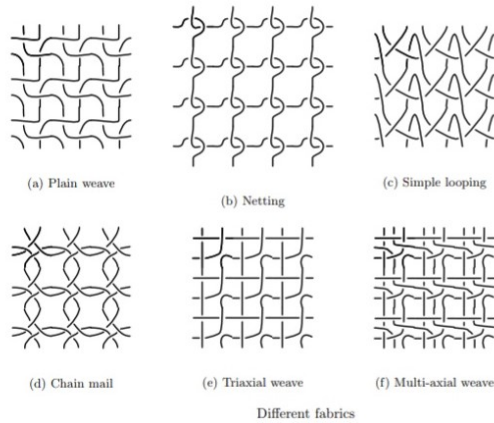


Figure 7:

In particular, chain mail fabrics (which contain closed components) have axial types 0, as the component bounds a closed region.

# CONCLUSION

*Knot theory is a subject suitable for understanding nature deeply for learning in an early age. We can watch a knot with eyes and our ability of space perception will be grown up playing with it. Knot theory remained as a beautiful mathematical theory in its own right. There are several practical applications of knot theory have come to light, including DNA knotting and other topics in biology, chemistry and physics. There are clearly a large number of areas that we could explore now, relating to a number of different subjects. It is this relation to such a wide range of subjects that makes knot theory so interesting.*

*This project deals about the basic definitions, Reidemeister moves, and skein relation in the first chapter. Kauffman bracket and Jones polynomial in second chapter. Seifert Surface, Seifert matrix, the genus of a knot in third chapter. Fundamental problems of Knot theory in fourth chapter. And we conclude Application in textiles.*

# Bibliography

- (1) *Charles Livingston , Knot Theory , 1993 , The Mathematical Association Of America*
- (2) *Heather A.Dye, An Invitition to Knot Theory Virtual and Classical , 2016 CRC Press.*
- (3) *Inga Johnson and Allison K.Henish , An Interactive Introduction to Knot Theory ,2017, Dove publications,Inc.*
- (4) *Kunio Murasugi , Knot Theory and Its Applications, 1993 in Japanese,*
- (5) *W.B Raymond Lickrish,An Introdution to Knot Theory , 1997,Springer-Verlag New York.*

# DATA ANALYTICS AND DOCUMENTATION USING MATHEMATICAL SOFTWARES

Project Report submitted to

ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI

Affiliated to

MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELI

In partial fulfilment of the requirement for the award of degree of

BACHELOR OF SCIENCE IN MATHEMATICS

Submitted by

NAMES	REG.NO
ANITHA MATHI.P	18AUMT02
JENITA ADLINDE.G	18AUMT26
JOAN SILVIYA .G	18AUMT28
LOURDUTHANGAGNANAMICHEAL .T	18AUMT34
VINISHNO ANGELIN.M	18AUMT55

Under the Guidance of

Dr. Sr. S. KULANDAI THERESE M.Sc.,B.Ed.,M.Phil.,Ph.D.,

Assistant Professor of mathematics

ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.



Department of Mathematics

St.Mary's College (Autonomous), Thoothukudi.

(2020 – 2021)

## CERTIFICATE

We hereby declare that the project report entitled "**DATA ANALYTICS AND DOCUMENTATIOIN USING MATHEMATICAL SOFTWARES**" being submitted to ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI affiliated to MANONMANIAM SUNDARANAR UNIVERSITY ,TIRUNELVELI in partial fulfilment for the award of degree of **Bachelor of science in Mathematics** and it is a record of work done during the year 2020-2021 by the following students:

**ANITHA MATHI.P**

**18AUMT02**

**JENITA ADLINDE.G**

**18AUMT26**

**JOAN SILVIYA .G**

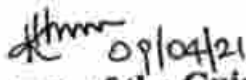
**18AUMT28**


**LOURDUTHANGAGNANAMICHEAL .T**

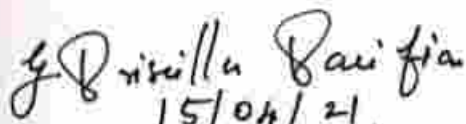
**18AUMT34**


**VINISHNO ANGELIN.M**

**18AUMT55**

  
09/04/21  
Signature of the Guide

  
Signature of the HOD

  
15/04/21  
Signature of the Examiner

  
Signature of the Principal  
St. Mary's College (Autonomous)  
Thoothukudi - 628 001.

## DECLARATION

We hereby declare that the project report entitled "DATA ANALYTICS AND DOCUMENTATION USING MATHEMATICAL SOFTWARES" is our original work. It has not been submitted to any university for any degree or diploma.

P. Anitha Mathi.  
(ANITHA MATHI.P)

Gr. Jenita Adlinde  
(JENITA ADLINDE.G)

M. Vinishno Angelin  
(VINISHNO ANGELIN.M)

Gr. Joan Silviya  
(JOAN SILVIYA.G)

T. Lourduthangagnanamicheal  
(LOURDUTHANGAGNANAMICHEAL.T)

## **ACKNOWLEDGEMENT**

First of all, we thank lord Almighty for showering his blessings to undergo this project.

With immense pleasure, we register our deep sense of gratitude to our guide Dr. Sr. S. KULANDAI THERESE M.Sc., B.Ed., M.Phil., Ph.D. and the Head of the Department Dr. A. PUNITHA THARANI M.Sc., M.Phil., Ph.D., for having imported necessary guidelines throughout the period of our studies.

We thank our beloved Principal Rev. Dr. Sr. A.S.J. LUCIA ROSE M.Sc., PGDCA., M.Phil., Ph.D., for providing us the help to carry out our project work successfully.

Finally, we thank all those who extended their helping hands regarding this project.

# Contents

Pg.No

1.Data Analytics using SPSS.....	1
1.1 Important and Benefits of SPSS in research.....	1
1.2 Statistical Methods of SPSS.....	2
1.3 Using SPSS for t-Test.....	3
1.3.1 One Sample t-Test.....	4
1.3.2 Independent Sample t-Test.....	4
2.Mathematical Plots Using Matlab.....	1
2.1 Analytic Plotting With Symbolic Math Toolbox.....	1
3.Mathematical Documentation and Computation using Online Tools.....	1
3.1. LaTeX Documentation Using Online Tools.....	1
3.2 Wolfram Alpha.....	2
4.Raw Graphs .....	1
4.1 Different types of Raw Graph.....	1
5.Interconnection Networks.....	1
5.1 Network Topology.....	2
5.2 Types of Interconnecting Networks.....	3
Conclusion	
References	



# Chapter 1 Data Analytics using SPSS

SPSS (Statistical Package for the Social Sciences), also known as IBM SPSS Statistics, is a software package used for the analysis of statistical data. Although the name of SPSS reflects its original use in the field of social sciences, its use has since expanded into other data markets. SPSS is commonly used in healthcare, marketing and education research. The types of data analysis using SPSS is widely varied. Common sources include survey results, organization customer databases, Google Analytics and scientific research results. SPSS supports both analysis and modification of many kinds of data and almost all formats of structured data. The software supports spreadsheets, plain text and relational databases such as SQL, SATA and SAS. Provides data analysis for descriptive and bivariate statistics, numerical outcome predictions and predictions for identifying groups. The software also provides data transformation, graphing and direct marketing features. The software interface displays open data similarly to a spreadsheet in its main view. With its secondary variable view, the metadata that describes the variables and data entries present in the data file are displayed. The software package was created in 1968 by SPSS Inc. and was acquired by IBM in 2009. While the software was renamed to IBM SPSS Statistics, it is still commonly referred to as just SPSS. This was last updated in October 2018 Continue Reading About SPSS (Statistical Package for the Social Sciences) Applicability of SPSS statistics for data analysis.

There are 15 modules IBM SPSS statistics for data analysis according to the research needs. SPSS Exact Tests module enables one to use small samples and still feel confident about the results forecasting module enables analysts to predict trend and develop forecasts quickly and easily-without being an expert statistician. Missing data can seriously affect your models and results. It is used by survey researchers, social scientists, data miners and market researchers to validate data.

## 1.1 Importance and benefits of SPSS in research:

Statistics is a software package used for logical batched and non-batched statistical analysis. This software is one of the most popular statistical packages which can perform highly complex data manipulation and analysis with simple instructions. SPSS can take data from almost any type of file and use them to generate tabulated reports, charts and plots of distributions and trends, descriptive statistics and conduct complex statistical analyses. This packages of program is available for both personal and mainframe computers. SPSS introduces the following four programs that help researchers with their complex data analysis needs.

### **Statistics Program:**

SPSSs statistics program gives a large amount of basic statistical functionality; some include frequencies, cross-tabulation, bivariate statistics, etc.

### **Modeler Program:**

Researchers are able to build and validate predictive models with the help of advanced statistical procedures.

### **Text Analytics for Surveys Program:**

It gives robust feedback analysis. which in turn get a vision for the actual plan.

### **Visualization Designer:**

Researchers found this visual designer data to create a wide variety of visuals like

density charts and radial box plots.

### **Features of SPSS:**

- The data from any survey collected via Survey Gizmo gets easily exported to SPSS for detailed and good analysis.
- In SPSS, data gets stored in SAV format. These data mostly comes from surveys. This makes the process of manipulating, analyzing and pulling data very simple.
- SPSS have easy access to data with different variable types. These variable data is easy to understand. SPSS helps researchers to set up model easily because most of the process is automated.
- After getting data in the magic of SPSS starts. There is no end to what we can do with this data.
- SPSS has a unique way to get data from critical data also. Trend analysis, assumptions, and predictive models are some of the characteristics of SPSS.
- SPSS is easy for you to learn, use and apply.
- It helps in to get data management system and editing tools handy.
- SPSS occurs you in-depth statistical capabilities for analysing the exact outcome
- SPSS helps us to design, plotting, reporting and presentation features for more clarity.

### **1.2 Statistical Methods of SPSS:**

- Many statistical methods can be used in SPSS, which are as follows: Prediction for a variety of data for identifying groups and including methodologies such as cluster analysis, factor analysis, etc.
- Descriptive statistics, including the methodologies of SPSS, are frequencies, cross-tabulation, and descriptive ratio statistics, which are very useful.
- Also, Bivariate statistics, including methodologies like analysis of variance (ANOVA), means, correlation, and nonparametric tests, etc. Numeral outcome prediction such as linear regression.
- It is a kind of self-descriptive tool which automatically considers that you want to open an existing file, and with that opens a dialog box to ask which file you would like to open. This approach of SPSS makes it very easy to navigate the interface and windows in SPSS if we open a file.

- Besides the statistical analysis of data, the SPSS software also provides data management features; this allows the user to do a selection, create derived data, perform file reshaping, etc. Another feature is data documentation. This feature stores a metadata dictionary along with the data file.

Types of SPSS:

It has two types of views those are:

- Variable View and
- Data View

### **Variable View**

#### **Name:**

This is a column field, which accepts the unique ID. This helps in sorting the data. For example, the different demographic parameters such as name, gender, age, educational qualification are the parameters for sorting data. The only restriction is special characters which are not allowed in this type.

#### **Label:**

The name itself suggests it gives the label. Which also gives the ability to add special characters.

#### **Type:**

This is very useful when different kind of data are getting inserted.

#### **Width:**

We can measure the length of characters.

#### **Decimal:**

while entering the percentage value, this type helps us to decide how much one needs to define the digits required after the decimal.

#### **Value:**

This helps the user to enter the value.

#### **Missing:**

This helps the user to skip unnecessary data which is not required during analysis.

#### **Align:**

Alignment, as the name suggests, helps to align left or right. But in this case, for ex. Left align.

#### **Measure:**

This helps to measure the data being entered in the tools like ordinal, cardinal, nominal. The data has to enter in the sheet named variable view. It allows us to customize the data type as required for analysing it. To analyse the data, one needs to populate the different column headings like Name, Label, Type, Width, Decimals, Values, Missing, Columns, Align, and Measures. These headings are the different attributes which, help to characterize the data accordingly.

#### **Data View**

The data view is structured as rows and columns.

### **1.3 Using SPSS for t-test**

Here we will discuss computation of t-test (independent sample and on e- sample

) with the help of SPSS .

### 1.3.1 One Sample t-Test

Select and move the variables to Test Variable(s) box by clicking the arrow button

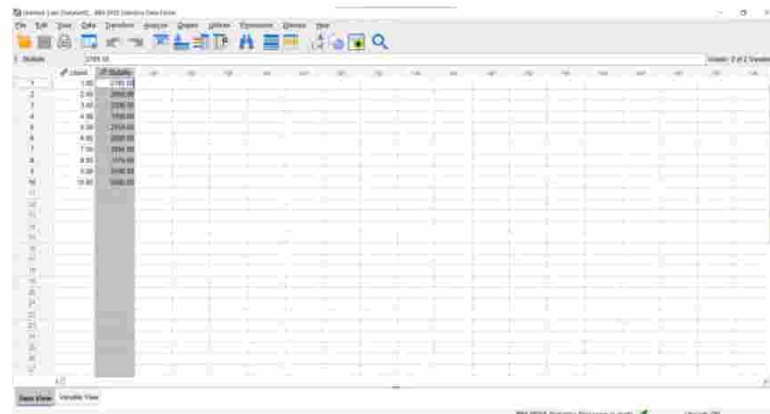


Figure 1:

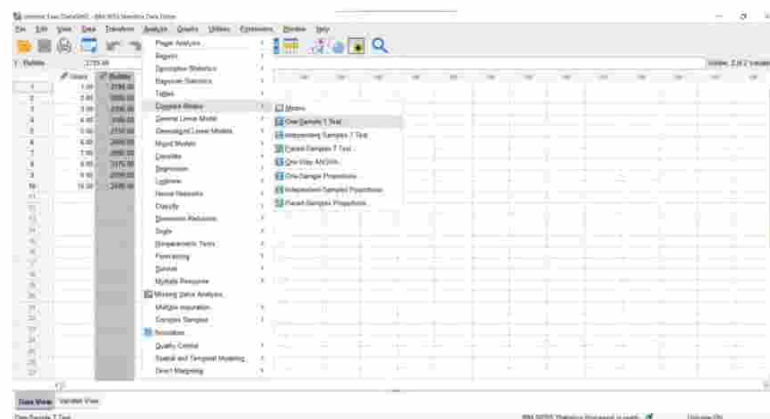


Figure 2:

.The Test Value is zero .If you want to compare the mean with a value other than zero enter that particular value. Now click Options... A sub-dialog box entitled One -Sample T test: Option will appear on the screen

Make entries pertaining to Confidence Interval and Missing Values 95 percentage is the default selection for confidence interval and Exclude cases analysis by analysis is again a default selection .To change it click radio button and then click Continue You will return back to the main dialog box. Finally click OK SPSS Viewer will appear with the output. The screen will display N, mean, SD, SE, t, degree of freedom, significance level, mean difference and lower and upper 95 percentage confidence interval of the difference.

### 1.3.2 Independent Sample t - Test

After opening the data file, click Analysis in the menu bar. Select Compare Means in the drop box and then click Independent-Sample T Test... in the submenu. A

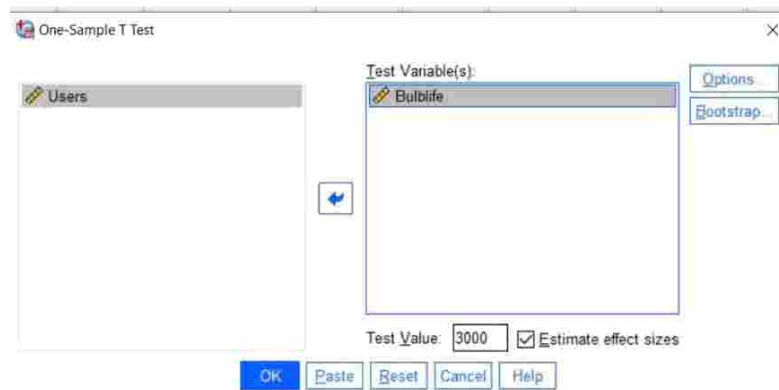


Figure 3:

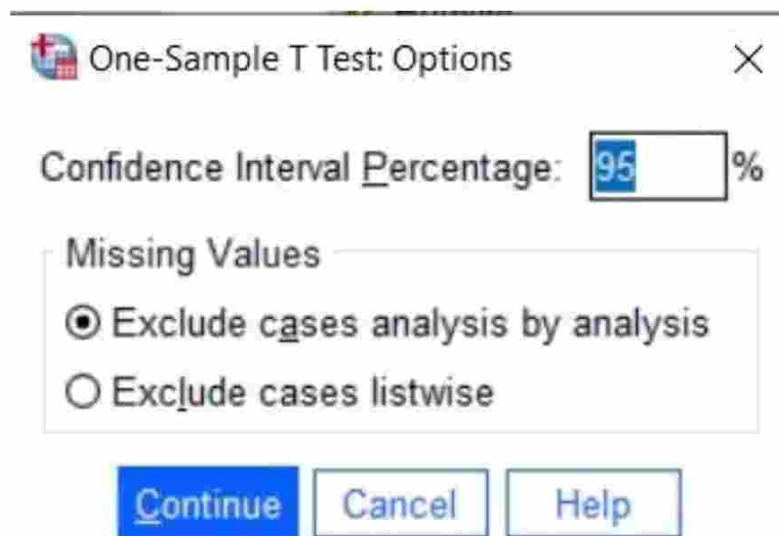


Figure 4:

dialog box entitled Independent-Samples T Test will appear on the screen All the variables will be enlisted in the left box. Select and move the variable to the Test variable(s) box by clicking the arrow button. Then select and move the grouping variable with a pair of question marks in a brackets. For example, if hard code is the grouping variable then it will appear as hand code. Click Define groups... A sub-dialog box will appear asking you to define groups 1 and 2 .You have to enter the grouping variable used to classify the data. Remember that only numeric value can be entered in these boxes. Click Continue and the sub-dialog box will disappear. Now click Options... and a sub-dialog box will appear on the screen. Then has been described. After doing the needful, click Continue to return back to menu dialog box. Finally click OK. SPSS viewer will appear on the screen with the output.

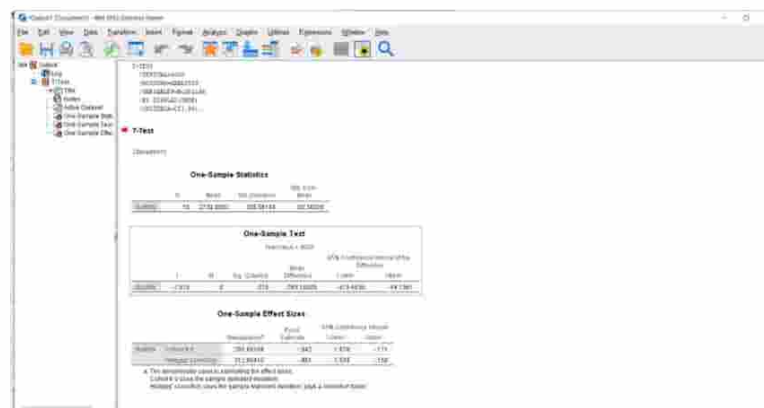


Figure 5:

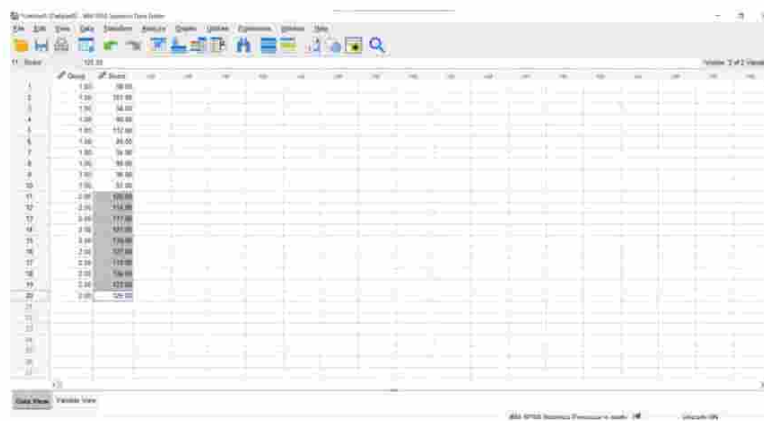


Figure 6:

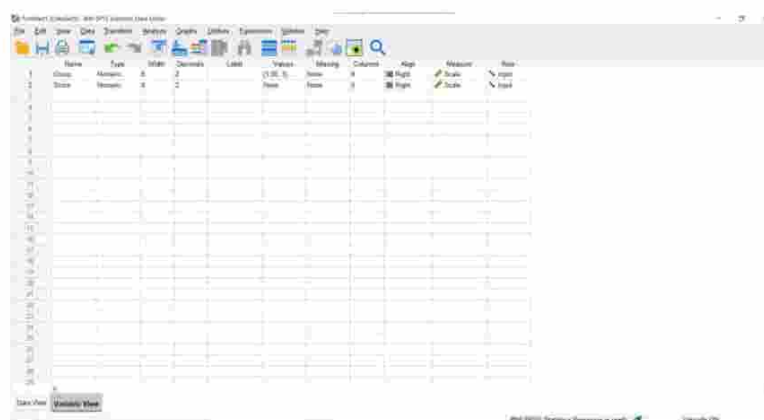


Figure 7:

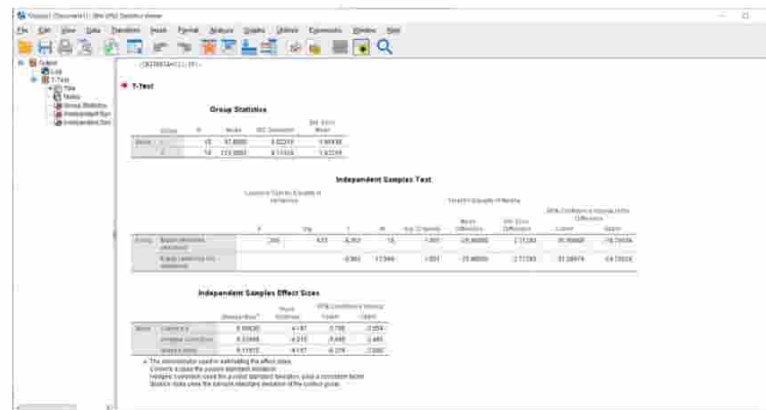


Figure 8:

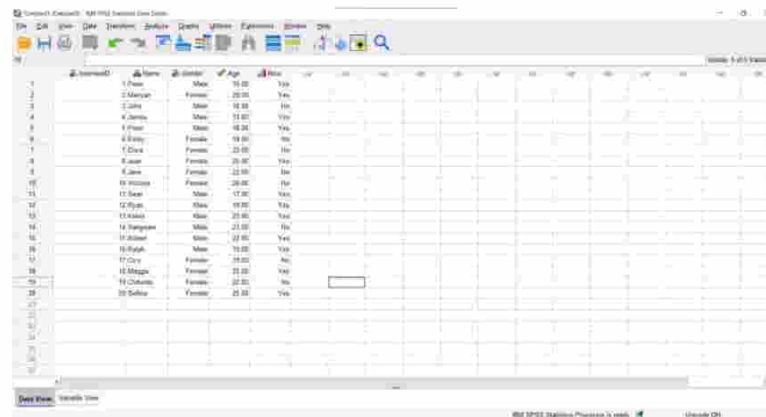


Figure 9:

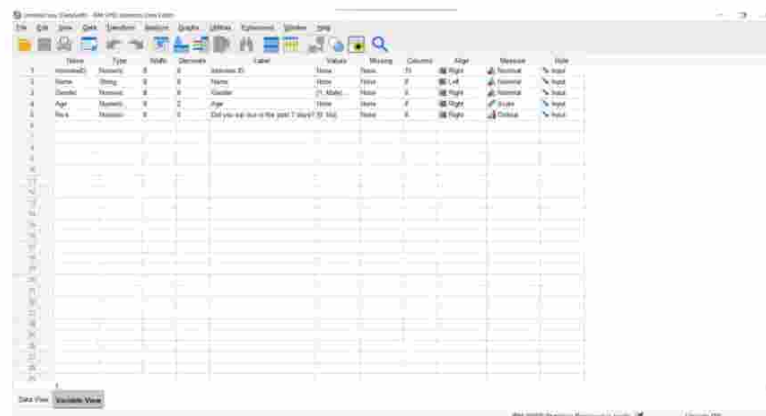


Figure 10:







# Chapter 2

## Mathematical Plots Using MATLAB

### 2.1 Analytical Plotting With Symbolic Math Toolbox

The fplot family accepts symbolic expressions and equations as inputs enabling easy analytical plotting without explicitly generating numerical data.

This example features the following functions

- fplot
- fplot3
- fsurf
- fcontour
- fmesh
- fmesh
- fimplicit
- fimplicit

Interactively plot functions of one variable

```
>> syms x
>> fplot(sin(exp(x)))
```

Figure 1.1:

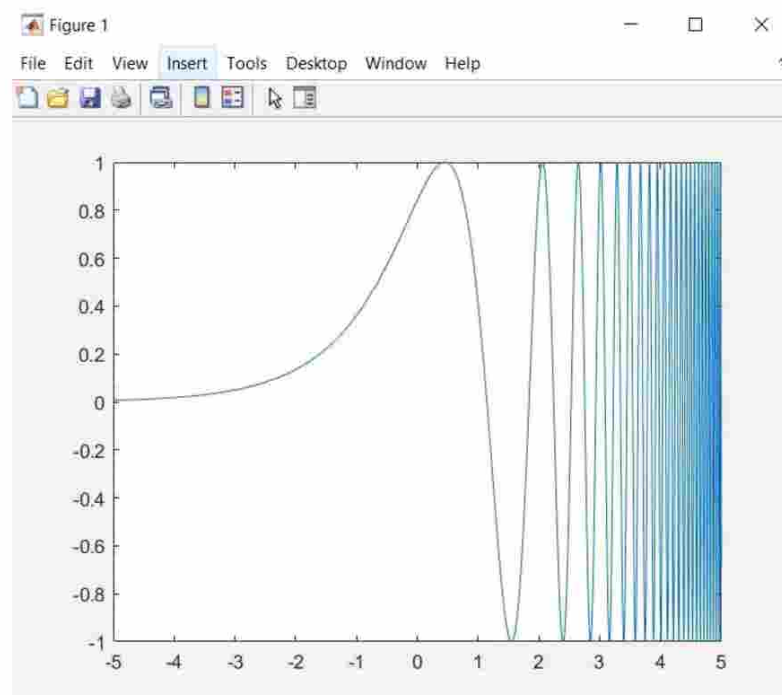


Figure 1.2:

```
>> syms x
>> fplot(sin(exp(x)))
>> fplot([sin(x), cos(x), tan(x)])
```

Figure 1.3:

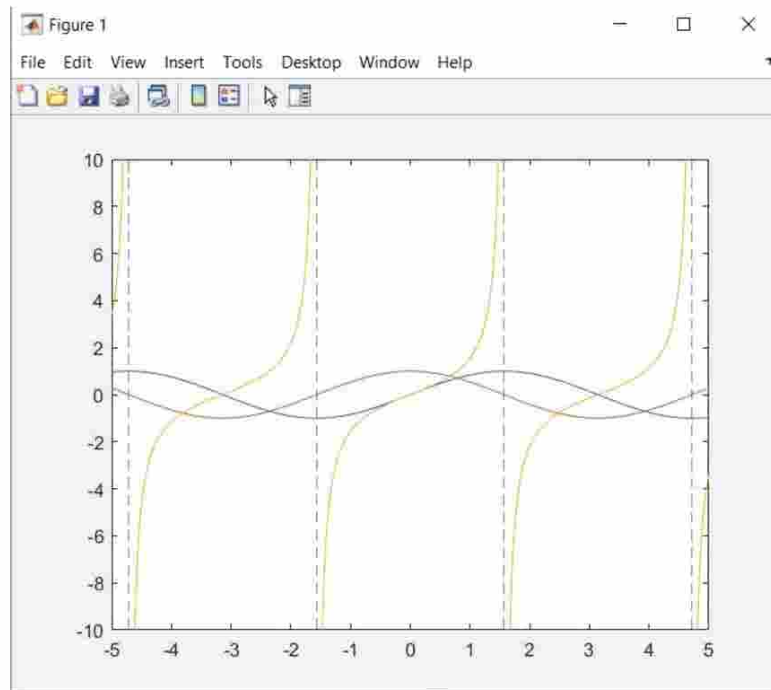


Figure 1.4:

```
>> syms x y
>> r = 1:10;
>> fimplicit(x^2+y^2 == r)
```

Figure 1.5:

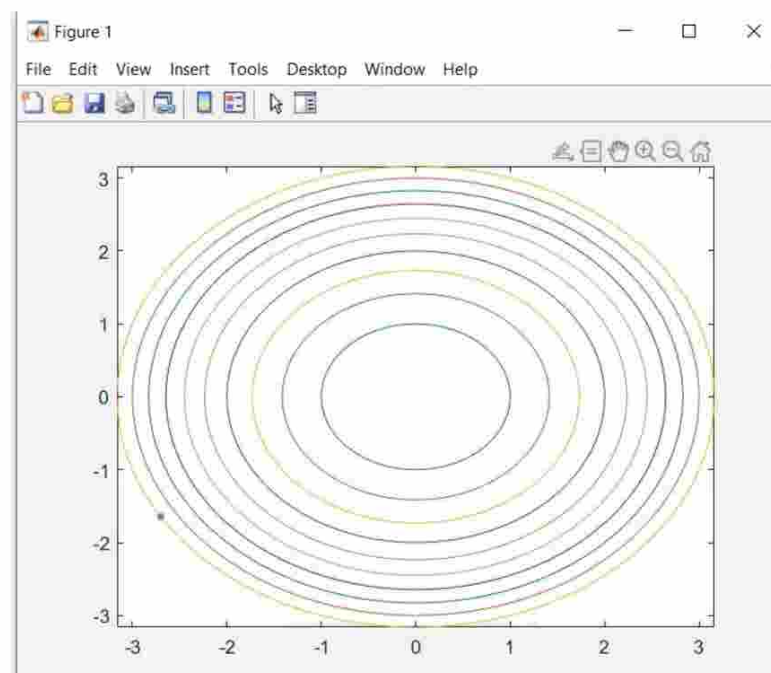


Figure 1.6:

```

Command Window

>> %Example 2.1
>> %Define values for x0, and h
>> x0 = 0; h = 1.0;
>> %Define function
>> f = @(x) - 0.15*x^4 - 0.17*x^3 - 0.25*x^2 - 0.25*x + 1.25;
>> %Define derivatives
>> fprime = @(x) -0.60*x^3 - 0.51*x^2 - 0.50*x - 0.25;
>> f2prime = @(x) -1.80*x^2 - 1.02*x - 0.50;
>> f3prime = @(x) - 3.60*x - 1.02;
>> f4prime = @(x) - 3.60;
>> f5prime = @(x) 0;
>> %Define the terms of the Taylor polynomial at x0
>> term1 = f(x0);
>> term2 = h*fprime(x0);
>> term3 = h^2/2*f2prime(x0);
>> term4 = h^3/6*f3prime(x0);
>> term5 = h^4/24*f4prime(x0);
>> term6 = h^5/120*f5prime(x0);
>> %Define the approximations of different order
>> ftaylor0 = term1;
>> ftaylor1 = term1 + term2;
>> ftaylor2 = term1 + term2 + term3;
>> ftaylor3 = term1 + term2 + term3 + term4;
>> ftaylor4 = term1 + term2 + term3 + term4 + term5;
>> ftaylor5 = term1 + term2 + term3 + term4 + term5 + term6;
>> %Define x
>> x = x0 + h;
>> %Calculate the exact answer using the function given
>> Exact = f(x);
>> %Errors for the Taylor polynomial approximations
fx >> err0 = Exact - ftaylor0;

```

Figure 1.7: Taylor's Series Expansion

```

>> err1 = Exact - ftaylor1;
>> err2 = Exact - ftaylor2;
>> err3 = Exact - ftaylor3;
>> err4 = Exact - ftaylor4;
>> err4 = Exact - ftaylor4;
>> err5 = Exact - ftaylor5;
>> %Print out values for set h value
>> T = table(ftaylor0, ftaylor1, ftaylor2, ftaylor3, ftaylor4, ftaylor5)
>> Terr = table(err0, err1, err2, err3, err4, err5)

T =

1x6 table

    ftaylor0    ftaylor1    ftaylor2    ftaylor3    ftaylor4    ftaylor5
    _____    _____    _____    _____    _____    _____
    1.25         1         0.75         0.58         0.43         0.43

Terr =

1x6 table

    err0    err1    err2    err3    err4    err5
    _____    _____    _____    _____    _____    _____
    -0.82    -0.57    -0.32    -0.15    -1.1102e-16    -1.1102e-16

```

Figure 1.8:

```

%%
>> %Example 2.1(cont.)
>> x = 0:.01:5;
>> length = size(x);
>> x0 = zeros(length);
>> %Define function
>> f = @(x) - 0.15*x.^4 - 0.17*x.^3 - 0.25*x.^2 - 0.25*x +1.25;
>> %Define derivatives
>> fprime = @(x) -0.60*x.^3- 0.51*x.^2- 0.50*x- 0.25;
>> f2prime = @(x) -1.80*x.^2 - 1.02*x - 0.50;
>> f3prime = @(x) - 3.60*x.^1.02;
>> f4prime = @(x) - 3.60;
>> f5prime = @(x) 0;
>> %Define the terms of the Taylor polynomial at x0
>> ftaylor0 = f(x);
>> ftaylor1 = ftaylor0 + (x - x0).*fprime(x0);
>> ftaylor2 = ftaylor1 + (x - x0).^2/2.*f2prime(x0);
>> ftaylor3 = ftaylor2 + (x - x0).^3/6.*f3prime(x0);
>> ftaylor4 = ftaylor3 + (x - x0).^4/24.*f4prime(x0);
>> ftaylor5 = ftaylor4 + (x - x0).^5/120.*f5prime(x0);
>> %Errors generated with the various approximations
>> err0 = f(x) - ftaylor0;
>> err1 = f(x) - ftaylor1;
>> err2 = f(x) - ftaylor2;
>> err3 = f(x) - ftaylor3;
>> err4 = f(x) - ftaylor4;
>> err5 = f(x) - ftaylor5;
>> %plot the Taylor approximations
>> figure
>> p1 = plot(x,ftaylor0,'blue',x,ftaylor3,'red--',x,ftaylor5,'black:');

```

Figure 1.9:

```

p2 =
3x1 Line array:

Line
Line
Line

>> xlabel('x-value')
>> ylabel('Errors as a function of x')
>> legend('Zero order approx:err0(x)', 'Third order approx:err3(x)', 'Fifth order approx:err5(x)', 'L
>> p2(1).LineWidth = 2;
>> p2(2).LineWidth = 2;
>> %--- 30-03-2021 20:10 --%

```

Figure 1.10:

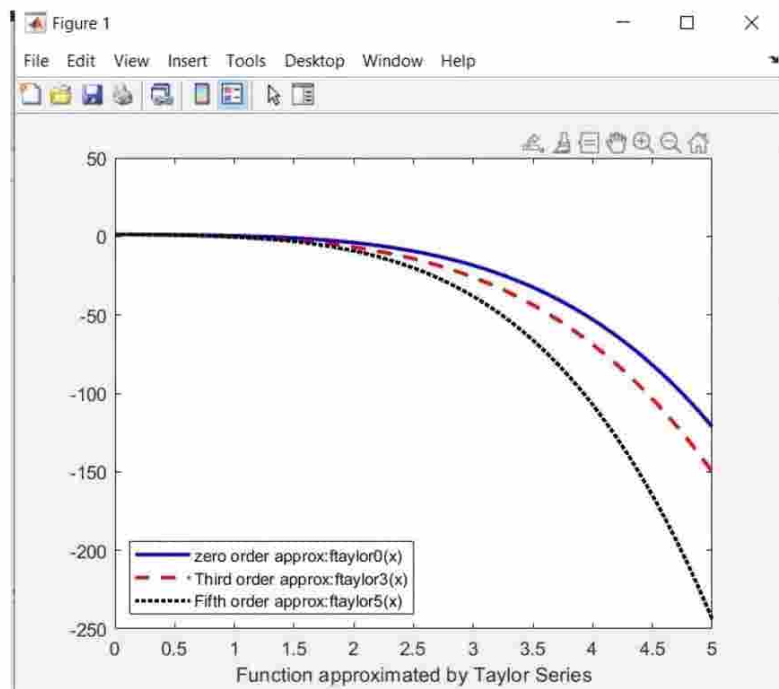


Figure 1.11:

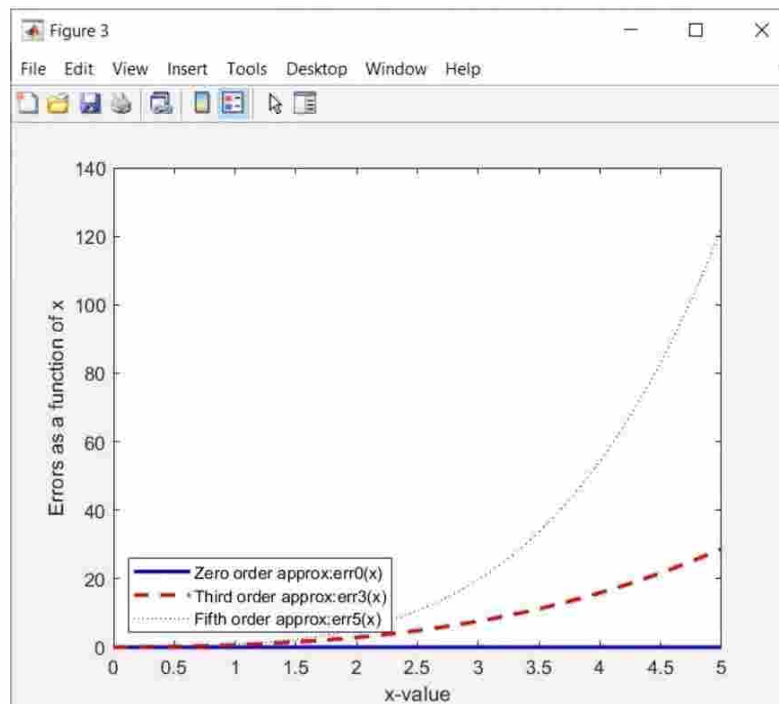


Figure 1.12:

# Chapter 3

## Mathematical Documentation And Computation Using Online Tools

### 3.1 LaTeX Documentation Using Online Tools

LaTeX is very easy when we use online tools in many ways. There are many online tools by using these online tools we can easily receive LaTeX codes for various symbols and equations. Now we see some online tools they are as follows

- Detexify
- Online LaTeX Editor
- LaTeX equation editor
- Mathpix

#### 1. Detexify

By Using Detexify LaTeX handwritten symbol recognition we can receive LaTeX codes for various symbols by drawing symbols option we can receive LaTeX codes easily.

#### 2. Online LaTeX Editor

we can receive LaTeX codes for basic formulas, maths, arrows, alphabets, fronts, colors, functions, symbols, operators by this website

#### 3. LaTeX equation editor

By using online LaTeX equation editor we can receive LaTeX codes for all types of equations and symbols.

#### **4.Mathpix**

By using snip app we can receive LaTeX codes easily by scanning the equation or any things and also by drawing a symbol or equation and scanning it, the LaTeX will be received.

#### **3.2 Wolfram Alpha**

Wolfram Alpha is a unique engine for computing answers and providing knowledge. It works by using its vast store of expert-level knowledge and algorithms to automatically answer questions, do analysis and generate reports.



# Detexify

## Symbol table

[classify](#)[symbols](#)Sort by Filter by  $\text{\texttt{!~}}$ 

textmode

 $\text{\texttt{[}}$ 

mathmode

 $\text{\texttt{]}}$ 

mathmode

 $\text{\texttt{/}}$ 

mathmode

 $\text{\texttt{\_}}$ 

textmode &amp; mathmode

 $\text{\texttt{\--}}$ 

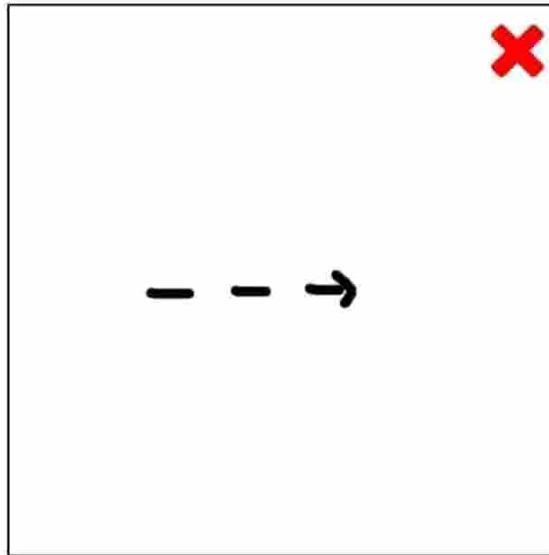
textmode

Figure 1.1: 1.Detexify

# Detexify

classify

symbols



Score: 0.0761100806632356  
`\usepackage{ amssymb }`  
`\Rrightarrow`  
mathmode



Score: 0.08548947411580572  
`\usepackage{ amssymb }`  
`\dashrightarrow`  
mathmode



Score: 0.15344422420686915  
`\not\equiv`  
mathmode



Score: 0.19662921241607895  
`\usepackage{ amssymb }`  
`\risingdotseq`  
mathmode

Figure 1.2: 1.Detexify



Figure 1.3: 2.Online LaTeX Editor



6



Figure 1.5: 3.LaTeX equation editor

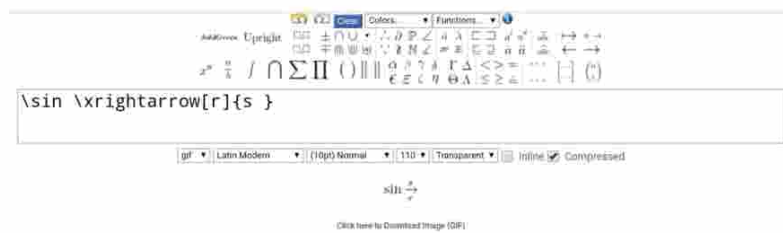


Figure 1.6: 3.LaTeX equation editor

OCR

Search

Solver

$$s^2 = 4c^2(u + c^2)$$

View URL

$$s^2 = 4c^2(u + c^2)$$

$s^2=4 c^2\left(u+c^2\right)$

$s^2=4 c^2\left(u+c^2\right)$

$s^2=4 c^2\left(u+c^2\right)$

$$s^2=4 c^2\left(u+c^2\right)$$

Confidence

Figure 1.7: 4.Mathpix

OCR

Search

Solver

View URL

$$y'' - 3y^2 + 2y - 24e - 2x = 0$$

$y^{\prime \prime}-3 y^{2}+2 y-24 e-2 x=0$

$y^{\prime \prime}-3 y^{2}+2 y-24 e-2 x \ldots$

$y^{\prime \prime}-3 y^{2}+2 y-24 e \ldots$

$\begin{equation} y^{\prime \prime}-3 y \ldots \end{equation}$

Confidence



Figure 1.8:



Figure 1.9: Wolfram Alpha

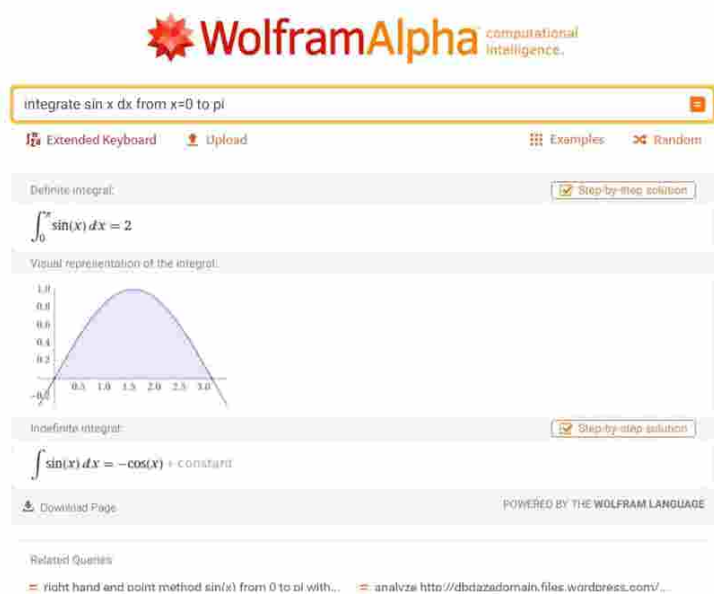


Figure 1.10: Wolfram Alpha



# Chapter 4

## Raw graphs

**Definition:** Raw graphs is a web-browser based data visualization tool that is easy to navigate for users unfamiliar with statistics and data sets. Once your data is uploaded, you can play with various charts and graphs to convey your information. A helpful feature of Raw Graphs is the descriptions under charts you select.

### 4.1 Different types:

- Alluvial Diagram
- Arc Diagram
- Bar Chart
- Multi-set bar Chart
- Stacked bar Chart
- Bees warm Plot
- Box Chart
- Bubble Chart
- Bump Chart
- Circle Packing
- Contour Plot
- Linear Dendiogram
- Hexagonal Binning

- Line Chart
- Matrix Plot
- Radar Chart

**Alluvial Diagram :** Alluvial Diagrams are a type of flow diagram originally developed to represent changes in network structure over time. In allusion to both their visual appearance and their emphasis on flow alluvial diagrams are named after alluvial fans that are naturally formed by the soil deposited from streaming water.

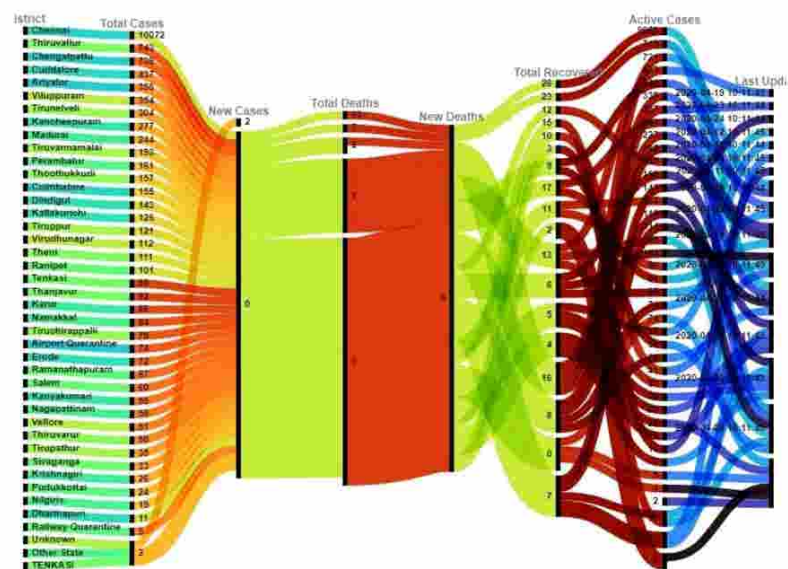


Figure 1.1:

**Arc Diagram :** In Graph drawing, an arc diagram is a style of graph drawing in which the vertices of a graph are placed along a line in the Euclidean plane, with edges being drawn as semicircles in one of the two half planes bounded by the line, or as smooth curves formed by sequences of semicircles.

**Bar Chart:** A diagram in which the numerical values of variables are represented by the height or length of lines or rectangles of equal width  
 ” the bar chart shows sales on the left and cost of sales on the right ”

### Multi-set bar chart:

Multi- set bar charts also known as clustered bar charts are a variation of a bar chart where two or more data sets are plotted side by side along a common axis. These are used to compare variables within the same category , and thus provide more information than a simple bar chart.

### Stacked Bar Chart:



Figure 1.2:

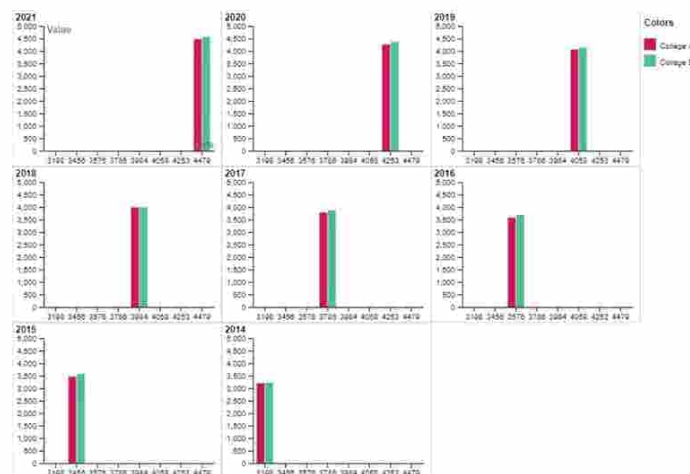


Figure 1.3: Multi-set bar chart

The stacked bar chart (aka stacked bar graph) extends the standard bar chart from looking at numeric values across one categorical variable to two each bar in a standard bar chart is divided into a number of sub-bars stacked end to end each one corresponding to a level of second categorical variable.

### Bees warm Plot:

Bees warm Chart is a one-dimensional Chart (or plot)- or in other words -a chart that shows all the information on a single-axis (usually X axis). It displays values as a collection of points similar to scatter plot.

### Box Plot :

A simple way of representing statistical data on a plot in which a rectangle is drawn to represent the second and third quartiles, usually with a vertical line inside to indicate the median value. The lower and upper quartiles are shown as horizontal lines either side of the rectangle.

### Bubble Chart :

A bubble chart is a type of chart that displays three dimensions of data, Each

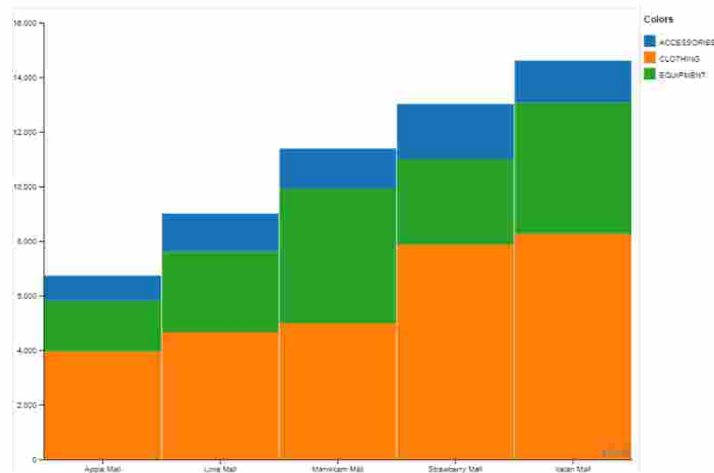


Figure 1.4: Stacked Bar Chart

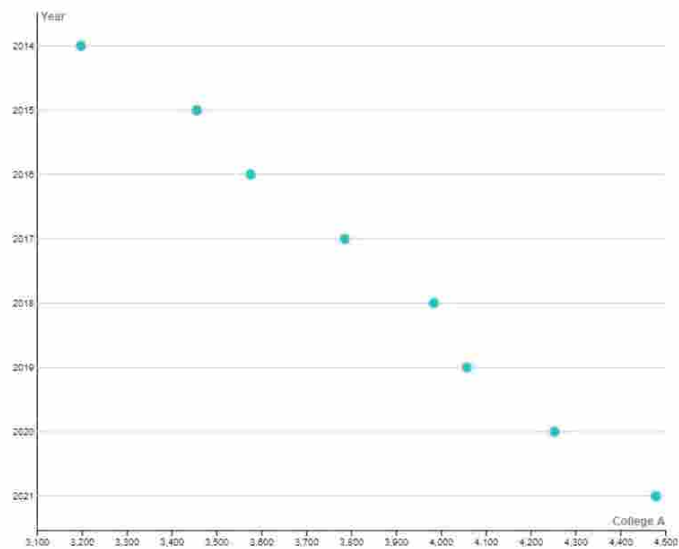


Figure 1.5: Beeswarm Plot

entity with its triplet of associated data is plotted as a disk that expresses two of the values through the disk's xy location and the third through its size.

### Bump Chart :

A bump chart is a special form of a line plot. This kind of plot is designed for exploring changes in rank order time. The focus here is usually on comparing the position or performance of multiple observations with respect to each other rather than the actual values itself.

### Circle Packing:

In geometry, circle packing is the study of the arrangements of circles on a given surface such that no overlapping occurs and so that no circle can be enlarged without creating an overlap. The associated packing density, of an arrangement is the

proportion of the surface covered by the circles.

### Circular dendrogram:

A dendrogram is an network structure. It is constituted of a root node that gives birth to several nodes connected by edges or branches.

A hierarchic dataset previous the links between nodes explicitly. Like above. The result of a clustering algorithm can be visualized as a dendrogram.

### Contour plot:

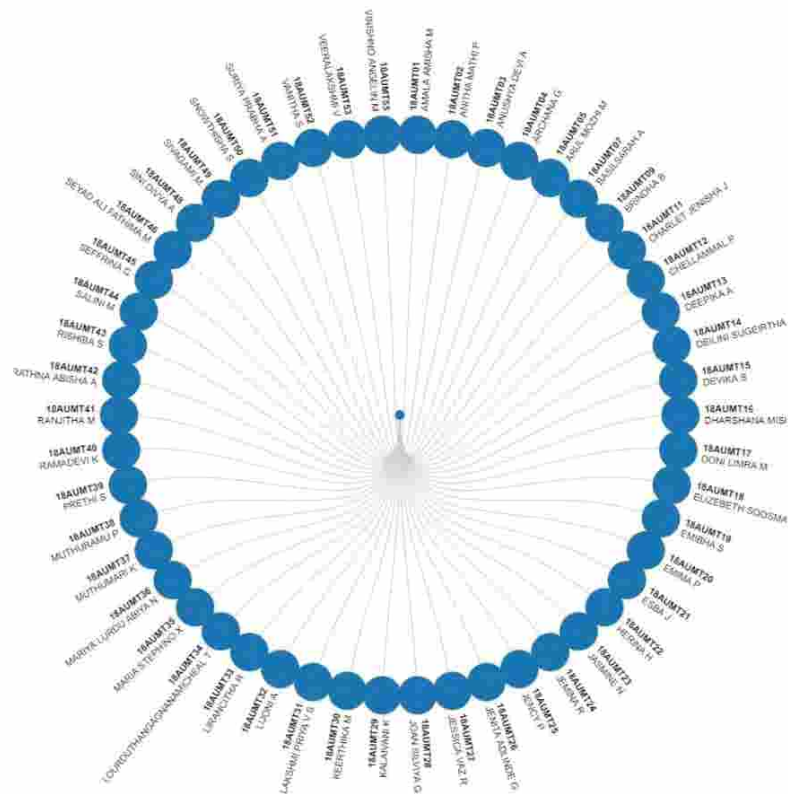


Figure 1.6: Circular dendrogram

A Contour line of a function of two variables is a curve along which the function, so that the curve joins points of equal value. It is a plane section of the three-dimensional graph of the function  $f$  parallel to the plane.

### Convex Hull :

In geometry, the convex hull or convex of a shape is the smallest convex set that contains it... For a bounded subset of the plane, the convex hull may be visualized as the shape enclosed by a rubber band stretched around the subset.

### Hexagonal binning:



Hexagonal binning is a form of bivariate histogram useful for visualizing the structure in datasets with large  $n$  the  $xy$  plane over the set  $(\text{range}(x), \text{range}(y))$  is tessellated by a regular grid of hexagons .

### Line Cart :

A line chart or line plot or line graph or curve chart is a type of chart which displays information as a series of data points called markets connected by straight line segments. It is a basic type of chart common in many fields.

### Matrix Plot :

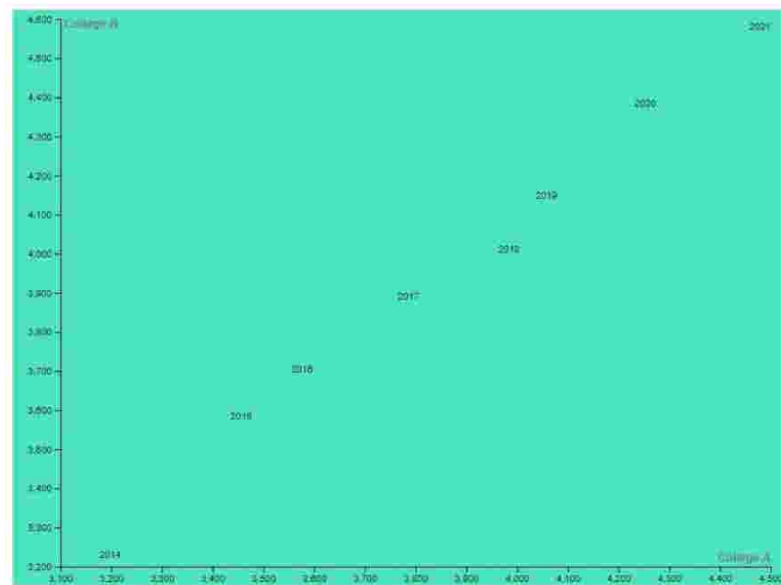


Figure 1.7: Line Chart

A matrix plot is an array of scatterplots. There are two types of matrix plots. Matrix of plots and each  $y$  varies each  $x$ . This type of matrix plot accepts  $y$ -axis and  $x$ -axis variables, then creates a plot for each possible  $xy$  combination.

**Sun Burst Diagram :** It displays hierarchically structured data and a related quantitative dimension using concentric circles. The circle in the centre represents the root node, with the hierarchies moving outward from the centre. The angle of each arc corresponds to the qualitative dimension.

**Linear Dendrogram:** It displays hierarchically structured data and a related quantitative dimension using concentric circles. The circle in the centre represents the root node, with the hierarchies moving outward from the centre. The angle of each arc corresponds to the qualitative dimension.





Figure 1.10:



Figure 1.11: Linear Dendrogram



# Chapter 5

## Interconnection Networks

### **Definition:**

Interconnection networks are composed of switching elements. Topology is the pattern to connect the individual switches to other elements, like processors, memories and other switches. A network allows exchange of data between processors in the parallel system.

Interconnection networks, also called multi-stage interconnection networks, are high-speed computer networks. They are connections between nodes where each node can be single processor or a group of processors or memory modules.

### **Example of interconnection:**

Places and the people and organisations in them are interconnected with other places in a variety of ways. Environmental and human processes, for example, the water cycle, urbanisation or human-induced environmental change, are sets of cause-and-effect interconnections that can operate between and within places.

### **Interconnection Network Benefits:**

Networking was first made use of in the 1950s in telecommunication to connect phone calls through switchboard that switches between electric connections. In the computing world, networking provides a method for fast communication between multiple computers connected to a network.

The idea behind interconnection networks is that when a computing task involving large amounts of data cannot be sufficiently handled by a single processor, the task is broken up into parallel tasks that are performed at the same time, so the processing time is vastly reduced. Efficient interconnection networks are critical for high-speed data transfer between the different elements in parallel processing.

## **5.1 Network topology:**

Network topology is the arrangement of the elements (links, nodes, etc.) of a communication network. Network topology can be used to define or describe the arrangement of various types of telecommunication networks, including commands and control radio networks, industrial fieldbusses and computer networks.

Network topology is the topological structure of a network and may be depicted physically or logically. It is an application of graph theory wherein communicating devices are modeled as nodes and the connections between the devices are modeled as links or lines between the nodes.

Examples of network topologies are found in local area network (LAN), a common computer network installation. Any given node in the LAN has one or more physical links to other devices in the network; graphically mapping these links results in a geometric shape that can be used to describe the physical topologies of the network. A wide variety of physical topologies have been used in LANs, including ring, bus, mesh and star. Conversely, mapping the data flow between the components determines the logical topology of the network. In comparison, Controller Area Networks, common in vehicals, are primarily distributed control system networks of one or more controllers interconnected with sensors and actuators over, invariably, a physical bus topology.

**There are five types of topology in computer networks:**

- Mesh Topology
- Star Topology
- Bus Topology
- Ring Topology
- Hybrid Topology

### **Mesh Topology:**

A mesh topology is a network setup where each computer and network device is interconnected with one another. This topology setup allows for most transmissions to be distributed even if one of the connections goes down. It is a topology commonly used for wireless networks.

### **Star Topology:**

A star network is an implementation of a spoke–hub distribution paradigm in computer networks. In a star network, every host is connected to a central hub. In its simplest form, one central hub acts as a conduit to transmit messages. The star network is one of the most common computer network topologies.

**Bus Topology:**

A bus network is a network topology in which nodes are directly connected to a common half-duplex link called a bus. A host on a bus network is called a station. In a bus network, every station will receive all network traffic, and the traffic generated by each station has equal transmission priority.

**Ring Topology:**

A ring network is a network topology in which each node connects to exactly two other nodes, forming a single continuous pathway for signals through each node – a ring. Data travels from node to node, with each node along the way handling every packet.

**Hybrid Topology:**

A tree network, or star-bus network, is a hybrid network topology in which star networks are interconnected via bus networks. Tree networks are hierarchical, and each node can have an arbitrary number of child nodes.

## **5.2 There are two main types of interconnection networks:**

5.2.1 static and

5.2.2 dynamic.

**Static network:**

When the connections are hard-wired, fixed, and unchangeable it is a static network.

**Dynamics network:**

Dynamic networks make use of switches and allow for reconfiguration of the network even during the execution of a parallel program.

**Difference between a dynamic and static IP(Internet Protocol) address:**

5.2.3 When a device is assigned a static IP address, the address does not change.

- Most devices use dynamic IP addresses, which are assigned by the network when they connect and change over time.

**Use of static IP address:**

A static IP address is an IP Address associated with your account that never changes and can be assigned to a specific device. Every time that you connect to the AT and T network the static IP address routes traffic to the computer or device that can be assigned an IP (such as a router or firewall).

**Use of dynamic IP address:**

A dynamic IP address is an IP address that an ISP lets you use temporarily. If a dynamic address is not in use, it can be automatically assigned to a different device. Dynamic IP addresses are assigned using either DHCP or PPPoE.

# Conclusion

Open source users get the access to superior data management, ease of use and presentation quality output that is available from IBM SPSS Statistics. SPSS Statistics users get access to a rich, ever-expanding collection of statistical analysis and graphing libraries to help them gain deeper insights from their data.

The separation of content and layout is useful for nearly every task. That's why it is used by most people who work professionally with texts. A core function is the proper display of mathematical expressions, that makes to the publishing standard in the scientific world. Another key feature for me is the possible automation. The .tex files are plain text, so you can work with them cross platform and its easy to handle them with programs.

MATLAB can be used as a tool for simulating various electrical networks but the recent developments in MATLAB make it a very competitive tool for Artificial Intelligence, Robotics, Image processing, Wireless communication, Machine learning, Data analytics and whatnot.

Semifinished vectors and data structures. RAWGraphs allows you to export visualizations as vector (SVG) or raster (PNG) images and embed them into your web page. Since RAWGraphs produces semifinished visualizations, you can even open them in your favorite vector graphics editor and improve them. Discover more RAW Graphs is open.

Interconnection networks play a central role in determining the overall performance of the multiprocessor systems. The Interconnection networks are like customary network systems consisting of nodes and edges. The nodes are switches having few input and few output (say  $n$  input and  $m$  output) lines.

# References

[1] Darren Gorge and Paul Mallery *IBM SPSS Statistics*(originally published in 12 october 2018).

[2] Andy Field *Discovering Statistics Using SPSS* (originally published in 2000).

[3] Bruce Littlefield and Duane C.hanselman *Mastering MATLAB*(originally published in 1996).

[4] Brad Wyble and David Albert Rosenbaum *MATLAB for Behavioral Scientists*(originally published in 2007).

[5] Leslie B.Lamport *Latex Document Preparation System Users*(originally published in 1984)

[6] Helmut kopka *Guide to LaTeX* (originally published in 1992).

[7] Eddie Law *Multistage interconnection Networks for Switching Fabric Design*.

[8] Raif O Onvural and Arne Nilsson *Local Area Network Interconnection*.

[9] William James Dally and Brian Towels *Principles and Practices of Interconnection Network*.

# **STUDIES BLOODSTAIN PATTERN ANALYSIS**

Project report submitted to

**ST. MARY' S COLLEGE (Autonomous), THOOTHUKUDI.**

Affiliated to

**MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELL.**

In partial fulfilment of the requirement for the award of degree of

**Bachelor of Science in Mathematics**

Submitted by

**NAMES**  
**ANUSHYA DEVI. A**  
**BASIL SARAH. A**  
**MUTHUMARI. K**  
**RANJITHA. M**  
**VANITHA. S**

**REG. NO.**  
**18AUMT03**  
**18AUMT07**  
**18AUMT37**  
**18AUMT41**  
**18AUMT52**

Under the guidance of

**DR.V. L. STELLA ARPUTHA MARY M.Sc., M. Phil., B. Ed., Ph. D.,**

Assistant Professor in Mathematics

**ST. MARY' S COLLEGE (Autonomous), THOOTHUKUDI..**



Department of Mathematics

**ST. MARY'S COLLEGE (Autonomous), THOOTHUKUDI..**

(2020 – 2021)

## CERTIFICATE

We hereby declare that the project report entitled "STUDIES ON BLOODSTAIN PATTERN ANALYSIS" being submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli in partial fulfilment for the award of the degree of Bachelor of Science in Mathematics and it is a record of work done during the year 2020 – 2021 by the following students:

ANUSHYA DEVI. A

18AUMT03

BASIL SARAH. A

18AUMT07

MUTHUMARI. K

18AUMT37

RANJITHA. M

18AUMT41

VANITHA. S

18AUMT52

*Vf. Snella Arputha Mary*  
Signature of the Guide

*J.P.D. J.L.*  
Signature of the HOD

*J.P.D. J.L.*  
Signature of the Examiner

*Lucia Rose*  
Signature of the Principal  
St. Mary's College (Autonomous)  
Thoothukudi - 628 001.



## DECLARATION

We hereby declare that the project entitled "BLOOD STAIN PATTERN ANALYSIS" is our original work. It has not been submitted to any University for any degree or diploma.

A. Anushya Devi  
(AUNSHYA DEVI.A).

A. Basil Sarah  
(BASIL SARAH.A)

K. Muthumai  
(MUTHU MARI.K).

M. Ranjitha  
(RANJITHA.M)

S. Vanitha  
(VANITHA.S)

# **Bloodstain Pattern Analysis**

# **Contents**

## **Page. No**

## **Introduction**

**3**

### **1.Introduction to blood stain pattern Analysis.**

1.1	Principles of Bloodstain Pattern Analysis.	4
1.2	Objectiyes of Bloodstain Pattern Analysis.	4
1.3	Types of Bloodstain Pattern Analysis.	6
1.4	Classification of Bloodstain Pattern Analysis.	8

### **2.Determining the point of convergence and the area of origin.**

2.1	Identify well formed stains in the pattern.	10
2.2	Identify Directionality of the stains.	12
2.3	Identify point of Convergence for the pattern.	13
2.4	Identify impact Angel for the stains.	19

### **3.Method used for Measurement of Bloodstain.**

3.1	Stain measurement.	27
3.2	Combine the Information to Establish an Area of origin.	38
3.3	Defing Area of origin with the tangent function.	35

## **Conclusion**

**40**

## **References**

**41**

## Introduction

Blood stain pattern help to find at which time the incident happens. It is every useful method in the forensic department. Many cases all over the world has been solved by the blood stain pattern technology. So it is very important to everyone to know about the blood stain. The general role of the blood stain pattern analysis in a criminal investigation is to assist in the reconstruction of those events of an alleged incident that could have created the stains and stain patterns present at a crime scene.

In some cases it may be necessary to conduct a blood stain interpretation using photographs, information that may be gained with blood stain pattern analysis include. For example, the position of the individual when the blood was deposited (sitting, standing etc..,) the relative position of individuals at the time of bloodshed the possible type of weapon used as well as possible mechanisms that could have produced the blood staining on a surface. The distinctive blood stain patterns occur because of the physical properties of the blood and how it reacts when acted upon by physical force.

The limitations of the blood stain pattern analysis include the fact it cannot recreate the entire scenario, as there are unknown variables that analysts cannot account for using scientific method. The blood may also have carried bite of skin hair or clothing with it all these materials can be recovered and analysed to provide more information about the victim the assailant. Blood stain pattern analysis has been accepted as reliable evidence by appellate courts in one state after another with little or no examination of its scientific accuracy.

# **Chapter 1**

## **I. Introduction to Bloodstain Pattern Analysis**

### **1.1 Principles of Bloodstain Pattern Analysis**

To understand how analysts interpret bloodstains, one must first understand the basic properties of blood.

Blood contains both liquid (plasma and serum) and solids (red blood cells, white blood cells, platelets and proteins). Blood is in a liquid state when inside the body, and when it exits the body, it does so as a liquid. But as anyone who has had a cut or a scrape knows, it doesn't remain a liquid for long. Except for people with hemophilia, blood will begin to clot within a few minutes, forming a dark, shiny gel-like substance that grows more solid as time progresses. The presence of blood clots in bloodstains can indicate that the attack was prolonged, or that the victim was bleeding for some time after the injury occurred.

Blood can leave the body in many different ways, depending on the type of injury inflicted. It can flow, drip, spray, spurt, gush or just ooze from wounds.

### **1.2 Objectives of Bloodstain Pattern Analysis**

BPA is a discipline that uses the fields of biology, physics, and mathematics. BPA accomplished by direct scene evaluation and/or careful study of scene photographs

(preferably color photographs with measuring device in view) in conjunction with detailed examination of clothing, weapons, and other objects regarded as physical evidence. Details of hospital records, postmortem examination, and autopsy photographs also provide useful information and should be included for evaluation and study. In cases where a scene investigation is not possible and photographs must be relied on, detailed sketches, diagrams, reports of crime scene investigation, and laboratory reports should be available for review.

Relative to the reconstruction of a crime scene, BPA may provide information to the investigator in many areas.

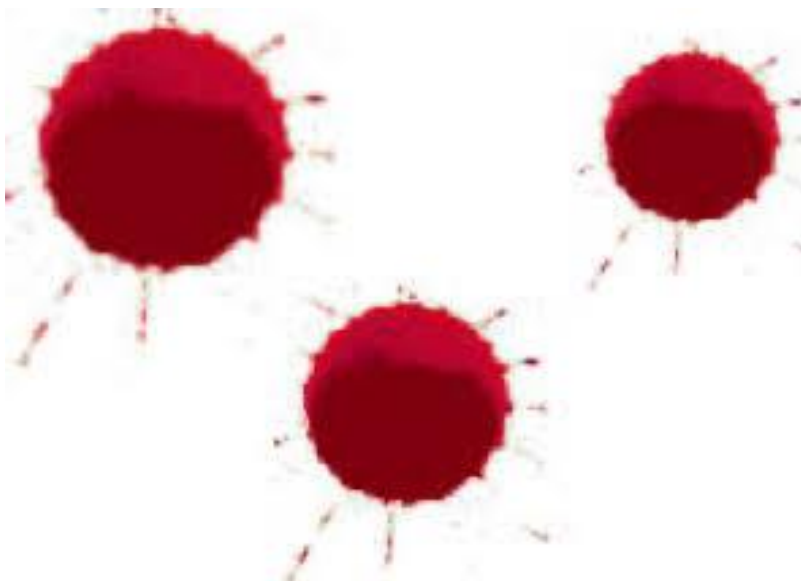
- Areas of convergence and origin of the bloodstains
- Type and direction of impact that produced bloodstains or spatter
- Mechanisms by which spatter patterns were produced
- Assistance with the understanding of how bloodstains were deposited onto items of evidence
- Possible position of victim, assailant, or objects at the scene during bloodshed
- Possible movement and direction of victim, assailant, or objects at the scene after bloodshed
- Support or contradiction of statements given by accused and/or witnesses
- Additional criteria for estimation of postmortem interval
- Correlation with other laboratory and pathology findings relevant to the investigation

The goal of the reconstruction of the crime scene using BPA is to assist the overall forensic investigation with the ultimate questions that must be addressed, which include, but are not limited to, the following.

- What event(s) occurred?
- Where did the event(s) occur?
- When and in what sequence did they occur?
- Who was there during each event?
- Who was not there during each event?
- What did not occur?

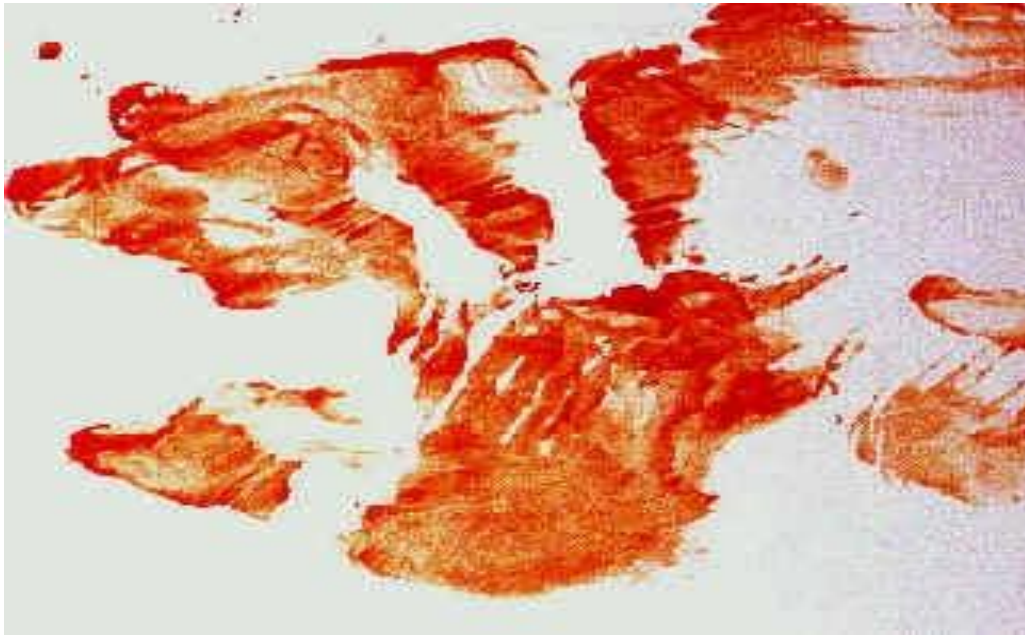
### **1.3 Types of Bloodstain Patterns:**

- (i) **Passive bloodstains:** Drops created or formed by the force of gravity acting alone.



- (ii) **Transfer bloodstain:** This is created when a wet, bloody surface comes in contact with a secondary surface. A recognizable image of all or a portion of

the original surface may be observed in the pattern, as in the case of a bloody hand or footwear.



- (iii) Projected bloodstains:** These are created when an exposed blood source is subjected to an action or force, greater than the force of gravity. The size, shape, and number of resulting stains will depend, primarily, on the amount of force utilized to strike the blood source.





## **1.4 Classification of bloodstain spatter by velocity:**

There are three basic categories of stain groups based on the idea of the size of the bloodstain compared with the amount of force propelling the bloodstain.

### **(i) Low Velocity Impact Spatter**

Low Velocity Impact Spatter is considered to be a force or energy equivalent to normal gravitational pull up to a force or energy of 5 ft/s. The resulting stain is relatively large usually 4 mm in diameter or greater. Free falling drop of blood affected only by gravity.

### **(ii) Medium Velocity Impact Spatter**

Medium velocity spatter is considered when a source of blood is subjected to a force 5 to 25 ft. Per second. The resulting stains range from 1 to 4 mm in diameter. These type of stains are usually associated with being or stabbing.

(iii) **High Velocity Impact Spatter**

High velocity spatter are created when the source of blood is subjected to a force with a velocity greater than 100 ft .per second .The resulting stain is predominantly less than 1 mm. in diameter although smaller and larger stain may be observed. These type of stains are usually associated with gunshot injuries.

# **Chapter2**

## **2.Determining the Point of Convergence and the Area of Origin**

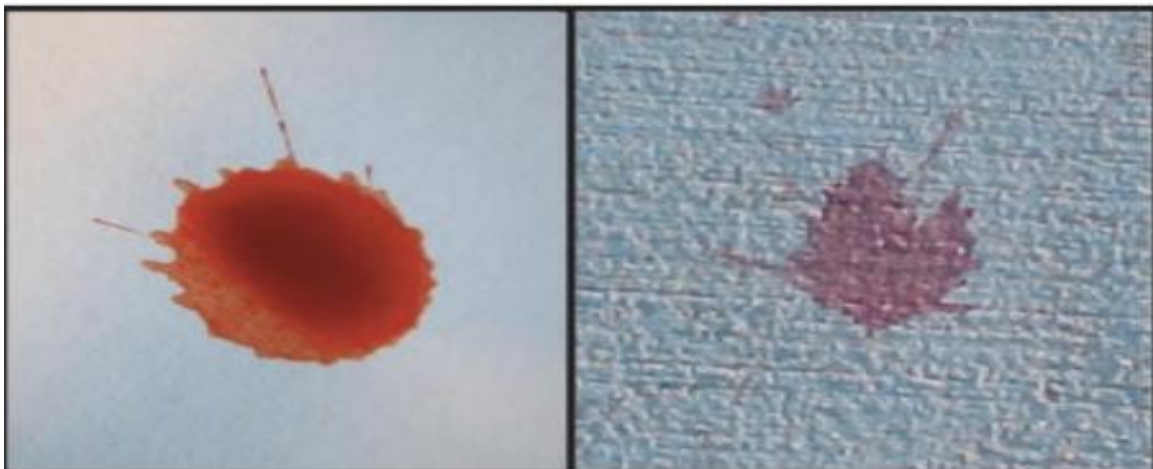
Our prior discussion of directionality and the evaluation of the path a given droplet was traveling lead us to the next step: defining from where the droplet came and determining what common convergence point several spatters may or may not have.

We can determine the area of origin for impact spatter patterns by following five steps. These steps are:

1. Identify well- formed spatter stains in the pattern.
2. Identify directionality of the stains.
3. Identify point of convergence for the pattern.
4. Identify impact angle of the stains.
5. Combine the information for area of origin.

### **2.1 Identify Well- Formed stains in the Pattern**

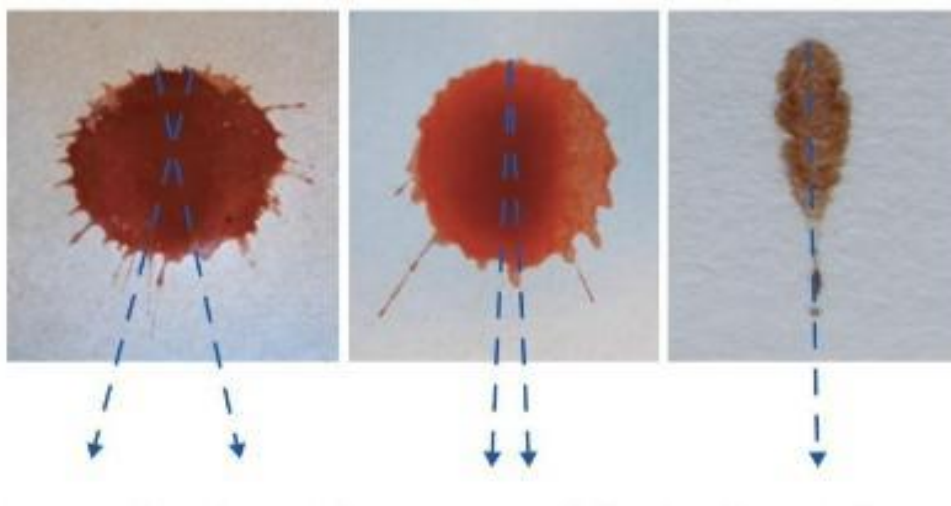
The concept of a well- formed stain is simply a stain that is symmetrical along its long and short axis. Figure (a) contrasts the difference between symmetrical and Asymmetrical stains.



Symmetrical shaped stain (a)

Asymmetrical shaped stain

**Figure(a):** The symmetry of the stain is important to the analyst for both defining directionality and impact angle. In pursuing an analysis of the area of origin, the analyst chooses stains that are symmetrical. The stain on the left is symmetrical with clearly defined margins. The stains on the right, having impacted onto a rough surface, is not symmetrical. Defining a clear length and width, or identifying a long axis in such a stain is difficult.



(b) Directionality Issue Based on Stain Shape

**Figure(b):** *Directionality of a stain becomes more defined as the stain becomes more elliptical. In the round stain on the left, it is difficult to know where the long axis of the stain is. The middle stain provides a clearer idea of this long axis, but even in this instance, different analysts might view it in slightly different ways. The elliptical stain on the right allows for a very clear understanding of where the long axis is. Different analysts viewing directionality in this stain would not vary in any significant fashion.*

## **2.2 Identify Directionality of the stains**

In spatter stains, directionality is defined by the long axis of the ellipse and the presence of scallops, satellites and tails. The ability to recognize directionality is critical to locating a convergence point and the area of origin. The shape of the spatter stain limits what we learn about directionality. Elliptical stains offer information that is more specific regarding directionality; circular stains offer less (see figure(B) ). As we discuss impact angle determination, we will find there is a similar correlation of error associated with impact angles. Both issues ( directional ambiguity and error rate associated with impact angle) functionally establish which stains the analyst will use in defining area of origin. As a general rule, stains that are generally circular ( $65^{\circ}$  to  $90^{\circ}$  impacts) should not be utilized for this evaluation. Stains that have a clear elliptical shape ( $10^{\circ}$  to  $65^{\circ}$ ) can be used for directional evaluation.

Directionality defines the path of the droplet as it struck a target. Often this is described in general terms (e.g., left to right). This direction of travel can also be defined numerically by a specific angle. The directional angle, also known as the gamma angle, described directionality as a specific angle ( between  $0^{\circ}$  and  $359^{\circ}$  ) as it relates to a reference point. Generally, this reference is north for patterns on horizontal surfaces

and up for vertical surfaces. The directionality is established, and then the long axis of the stain is measured against the reference point. See figure(C)

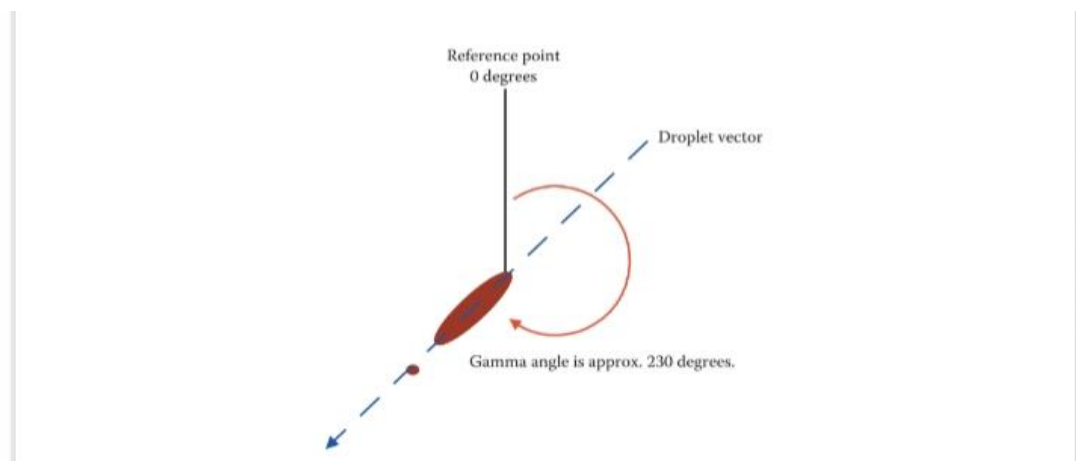


Figure (C)

**Figure(C):** Directional or gamma angle is the angle between the long axis of a stain and a standard reference point. This angle describes the directionality, but does so from standard reference point, thus allowing it to be used by forensic software.

## **2.3 Identify point of convergence for the pattern**

Depending upon the specific questions raised about a given spatter, the needs of the analysis may be different. In some instances, a point of convergence analysis (a top a view) may be sufficient to answer the questions. At other times, the analyst may require more detailed information, demanding the use of the area of the origin evaluation methods.

For a stain on a horizontal surface if we can define the stain's directionality and then draw a reverse azimuth (that is, a line that extends backward along the path the droplet was following), we can be reasonably sure the droplet originated somewhere along this avenue. Looking down on the droplet's path in the top view there are no forces that affect the flight path.

In this top view, once set in motion, assuming no ricochet event exists, a droplet will follow a straight path from its source to its destination. Gravity and air resistance affect the droplet only in the vertical plane of its parabola (a side view perspective).

For example, given Figure (D) stain #1 must originate somewhere along a reverse path as indicated by the directionality of the droplet. The only limits to the origin are the room's limitations or any intermediate obstacles. If figure (D) represented the room boundaries and the possible reverse path for stain #1 extends 13 ft, then the drop's origin must lie somewhere within those 13 ft. With only one stain the resulting parameter of possible origins is very wide.

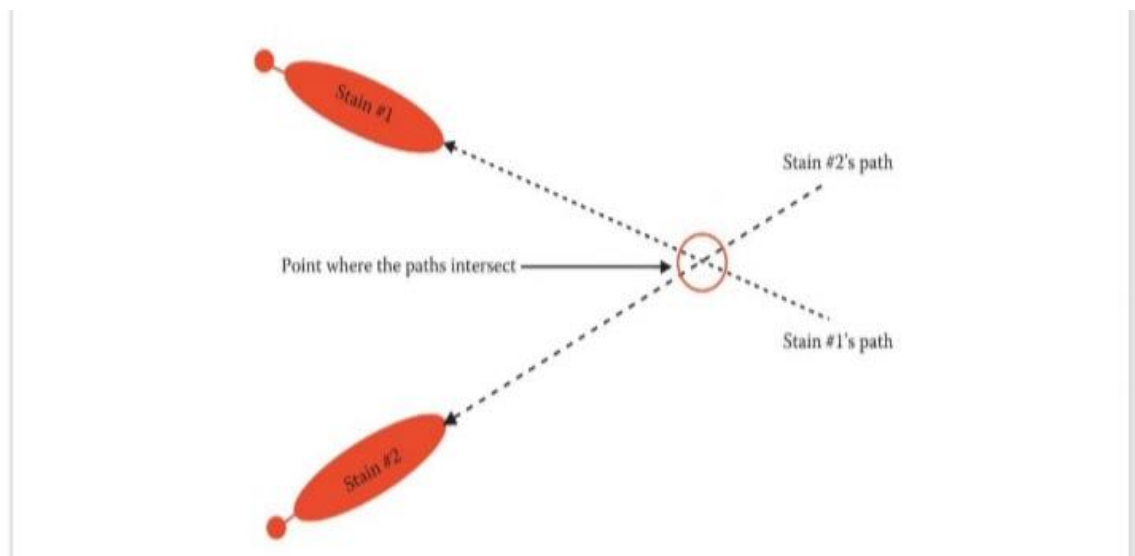
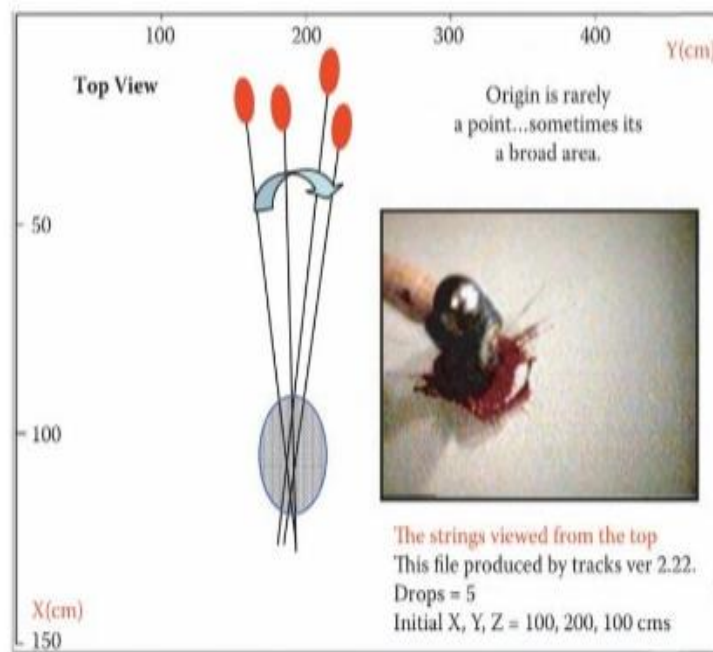


Figure (D)

**Figure(D):** By following the reverse vector of each droplet to the point where they intersect, we can establish a probable point of convergence in two dimensions.

If we introduce a second stain as in figure (D), we can then look for a point where the two paths intersect. The process is no different from the technique known as resection in map reading. By taking two unknown points and applying reverse azimuth from each we define an unknown point where the azimuth cross. In this instance, the unknown point is the likely source for both stains, their point of convergence.

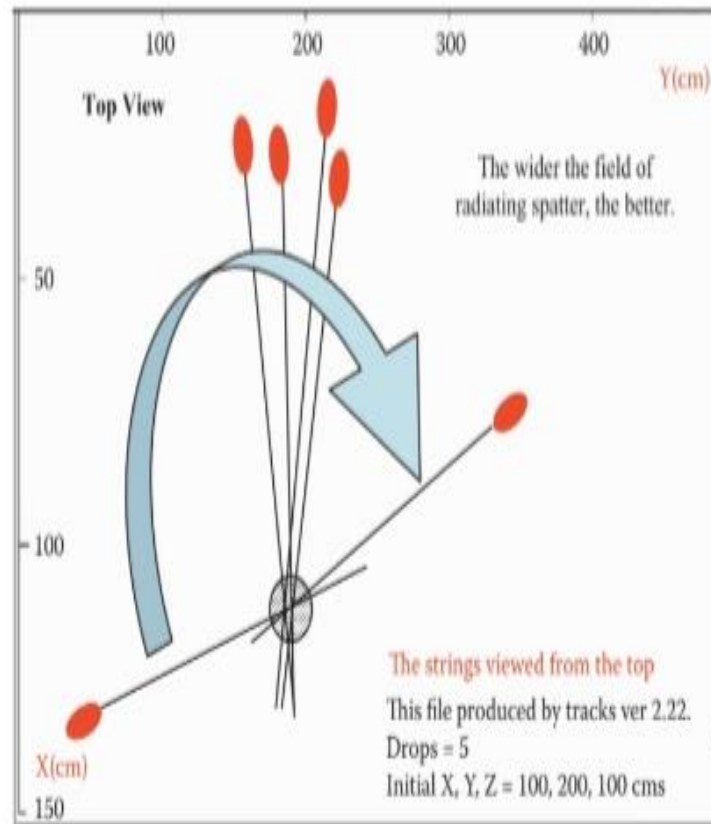
Keep in mind that the convergence "point" is likely to be an area. This individual lines created by these reverse azimuth will cross but not at a distinct point (eg., a specific XYZ position in space) (see figure E). However, the wider the radiating pattern of spatter involved, the more likely we can resolve the convergence to a more refined point (see figure 8m6).



Figure(E)



**Figure (E)** :At the convergence, the lines will rarely cross each other at a single point.  
The convergence point is often an area as well.



Figure(F)

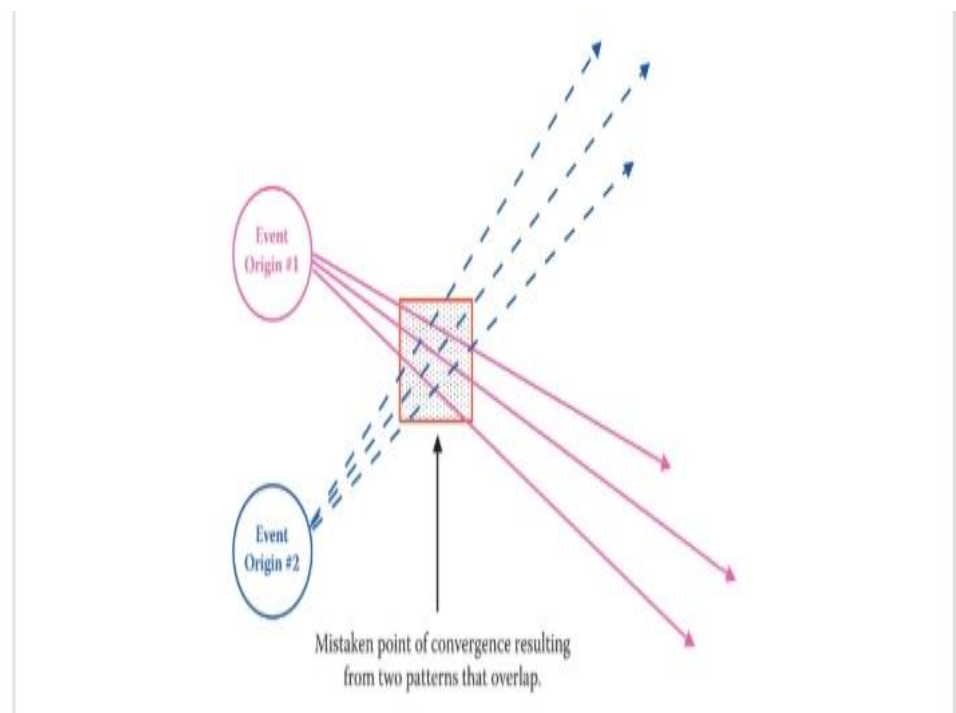
**Figure (F)**: The wider the field of radiating spatter, the more refined the convergence point becomes. Note that the inclusion of the far left stains functionality tightens the convergence point as compared to that observed in Figure€.

By limiting ourselves to this top view dimension ,we certainly gain simplicity in the evaluation. We also gain an inherent difficulty .If our circumstances limit us to only a few stains ,it is always possible they are the result of more than one event. That is even created stain #1 followed by another unrelated action that caused stain #2 .Given this

situation, the points where their reverse paths cross is simply coincidental and has no investigative significance.

Widespread within a scene and showing no common convergence point, we would easily recognize such stains as separate actions. Found in proximity to each other, the likelihood of a coincidental convergence of the flight paths increases. Unfortunately, we may choose to read this convergence as a single origin both stains. With only a few stains to work with, it may be difficult to recognize such an error.

The same is certainly true when viewing two adjacent patterns. We may see a coincidental convergence for two patterns and read this as the source for both. Figure (G) illustrates the possible error when viewing patterns from closely located impacts.



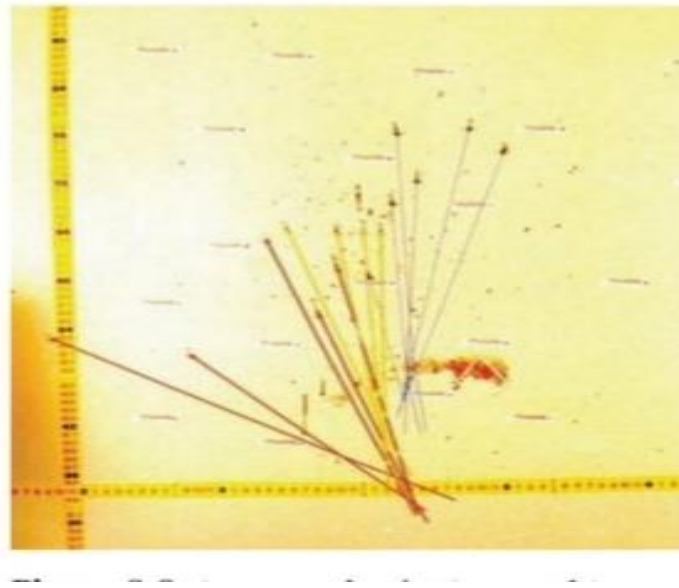
Figure(G)

**Figure(G):** *In addition to the possibility of two stains having a coincidental intersecting point, it is also possible for two patterns to overlap in this fashion. If this condition is not considered, it might well result in a mistaken point of convergence for both patterns.*

As the number of evaluated stains increases, our confidence increases as well. The more paths we find that intersect a given area, the more likely it is that we have a true point of convergence. Even in circumstances of multiple overlying the events with the spatter intermixed, the primary convergence points may still be evident. In the situation, the various paths may cross at several location, but where we find clusters of interesting paths will establish the primary convergence points of the various events.

What this top view method of analysis of the convergence point doesn't establish the three-dimensional origin of the spatter, the point above the convergence point where the spatter. This location is known as the area of origin, a position in three-dimensional space. In the top view approach, we establish a convergence in two dimensions and accept the flight paths originate somewhere above it. Once again, the possibility of multiple events always exists. The analyst may be viewing spatter from two or more events that originated at the same location in the room, but from different heights during different events ( e.g., One impact at 6 ft, another at 3 ft).

As described the top view convergence technique is effective in and of itself in some instances. It represents a functional method of portraying this convergence in the scene (see Figure (H) ). This information alone limits the possible area where the event could have originated.



Figure(H)

**Figure (H)** *An example of using graphic tape to demonstrate the point of convergence in a spatter pattern.*

On vertical surfaces, point of convergence is more problematic .In sum limited instances , we can use this method, but because gravity and air resistance affect the path of the droplet in the vertical plane, directionality is also affected.

To this point, we have narrowed our scope to finding the point of convergence along the paths of the stains of interest. The converging lines of these reverse vector establish a point of convergence for the pattern. To more effectively limit the source of the pattern, we must look at the flight paths in the vertical plane a side -view approach of the target .This requires including the stains' s impact angle in the analysis.

## **2.4 Identify impact Angles for the Stains**

A relationship exists between the length and width of the resulting stain and the angle at which the droplet impacts . The shape of the stains defines the angle of impact.

In general the terms what this means is more than the circular stain , the more perpendicular the angle at which it struck the target surface (eg.,90°) , whereas the more elliptical the shape ,the more acute the angle of impact (eg.,10°).See figure(I). With practice and experience and by observation alone, the analyst will recognize the general angle of impact based solely on the shape . This angle can also be computed to within a few degrees of the actual impact angle based on the relationship. We apply the concept the width /length ratio in conjunction with specific trigonometric function (eg., sine function).This allows the analyst to use straight line geometry techniques in defining the blood stain event.



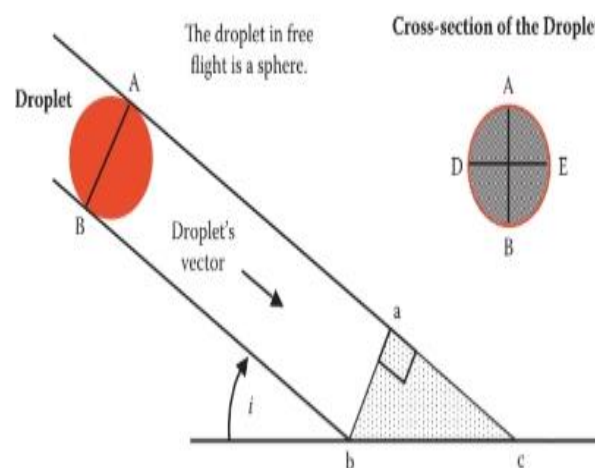
Figure(I)

**Figure (I):** Impact angle and stain shape. There is a direct correlation between the shape of a stain and the angle at which it struck a target surface. The more elliptical the stain's shape, the more acute the angle of impact. The more circular the stain, the closer it fell to 90° on the target.

Given a well formed stain where we can accurately measure the width and length we can easily establish the impact angle .

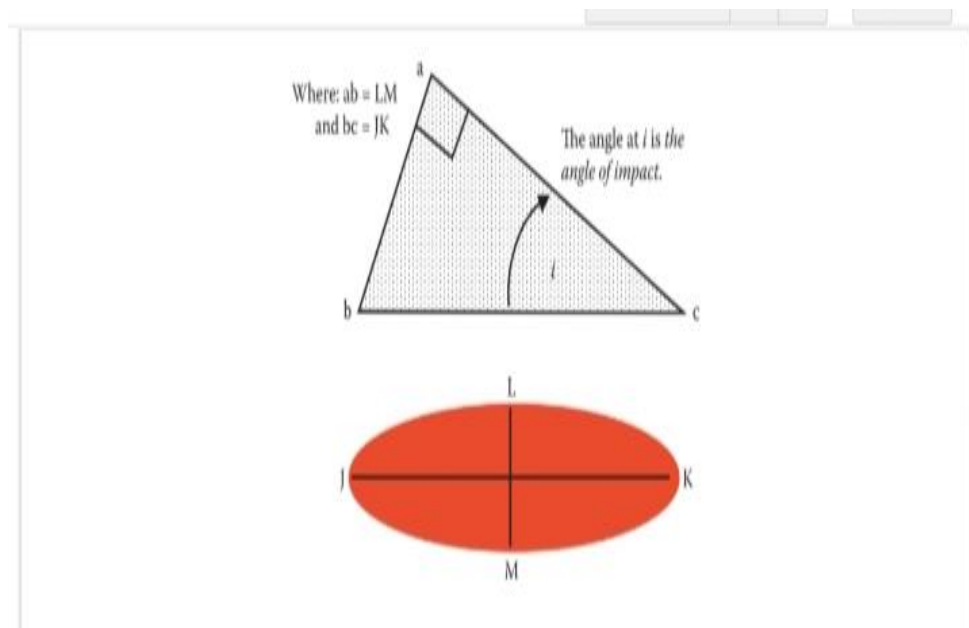
To apply the impact angle formula it is important to understand that in right triangles certain relationships exist between the angles of the triangles and the length of its sides. These relationships are trigonometric functions such as sine, cosine and tangent. These relationships are in no way dependent upon the factors found at the crime scene ; they are mathematical in nature. What we do is make an analogy to our scene using these relationships.

Imagine a right triangle formed between the droplet and the target surfaces as the droplet strikes. Figure (K) outlines how this triangle might look . A blood droplet in flight is in the shape of sphere. Therefore ,in viewing Figure (K), Line DE (the width) can be considered equal to line AB (the height) of the sphere. An analogy can then be drawn between line ab and bc and the width and length of the resulting stain. (See figure (L ). Based on the analogy line ab is represented by line LM of the stain and line bc is represented by the Line JK of the stain.



Figure(J)

**Figure (J)** The relationship of the droplet to an imagined right triangle. Using the sine function and this relationship, the analyst can establish the angle of  $i$ . This is the droplet's impact angle.



Figure(K)

**Figure (K):** We can draw an analogy between the triangle formed in figure (J) and our bloodstain. Line  $ab$  is analogous to line  $LM$ , as is  $bc$  to  $JK$ . Thus, the length and width of the stain are quantities we can apply using the sine function to determine the impact angle  $[i]$ .

As a result of the analogy, we have two known quantities from our crime scene which we can apply to a formula. By measuring the stains' length (line  $JK$ ) and width (line  $LM$ ) and applying them to following formula, the droplet's impact angle becomes evident;

$$\text{Sine } i = \text{Width (ab)} / \text{Length (bc)} \quad (1)$$

$$\text{Inverse sine (ASN) } i = \text{Impact angle} \quad (2)$$

Example:

Width = 3 mm

Length = 5mm

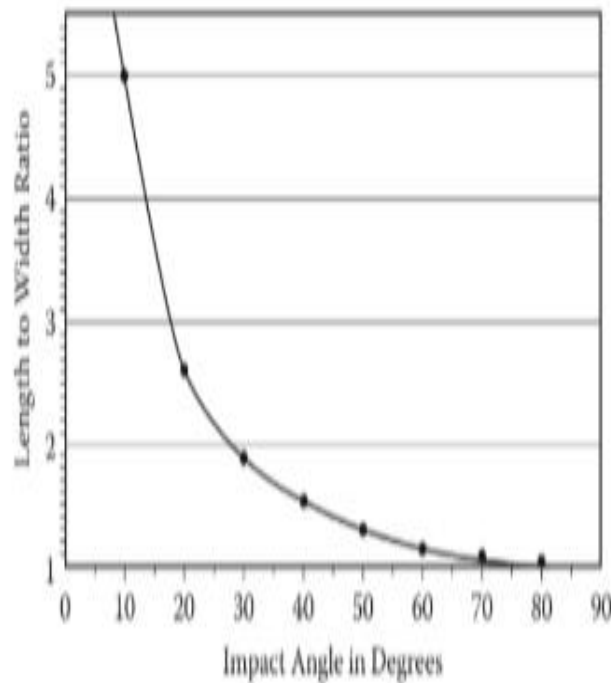
$$0.6(\text{Sine } i) = 3/5 \text{ (Width/ Length)}$$

$$\text{Inverse sine } i (0.6) = 36.8$$

Impact angle = approximately 36 to 37°

It is important to recognise that the formula provides the analyst with an estimate of the impact angle. The precision of the math should not be constructed to mean a similar precision in the definition of the angle. Issues related to the ballistic path of the droplet preclude us from accepting these angles as absolute. As a rule, impact angles are considered to be accurate to within 5° to 7°. It has always been recognised that circular shaped stains presented a greater error level. Recent studies demonstrate that when dealing with stains impact between 10° and 45°, the error rate is only 2° to 3°. This error rises to 6° to 7° for stains impacting at 60°. After 60° the error rate rises dramatically. As with the concern or directionality error rate issues demand that the analyst carefully consider what stains are utilized for area of origin determinations.





Figure(L)

**Figure(L):** A length to width ratio chart. By dividing the length of the stain by the width, we generate a number that is always greater than 1. The analyst finds this number on the vertical axis, and then reading to the right locates the corresponding point where the graph line intersects. The angle listed below this point is the approximate impact angle. For example, the L/W ratio of 1.9 equates to a 30° impact angle.

In addition to using a calculator with a sine function, there are two other related methods for determining impact angle. The first involves the length/width ratio chart (see figure M). This basic experiment is repeated in every blood stain pattern analysis class. The analyst divides the measured length of the stain by width, resulting in a

number greater than 1. If the result is less than one, we have reserved the length and width in the formula. Using the result the analyst locates the corresponding number on the vertical axis of the L/W ratio chart. Where the line intersects this point, one simply reads the angle listed on the lower axis.

Degrees	Sine	Degrees	Sine	Degrees	Sine	Degrees	Sine
1	0.0175	26	0.4384	51	0.7771	76	0.9703
2	0.0349	27	0.454	52	0.788	77	0.9744
3	0.0523	28	0.4695	53	0.7986	78	0.9781
4	0.0698	29	0.4848	54	0.809	79	0.9816
5	0.0872	30	0.5	55	0.8192	80	0.9848
6	0.1045	31	0.515	56	0.829	81	0.9877
7	0.1219	32	0.5299	57	0.8387	82	0.9903
8	0.1392	33	0.5446	58	0.848	83	0.9925
9	0.1564	34	0.5592	59	0.8572	84	0.9945
10	0.1736	35	0.5736	60	0.866	85	0.9962
11	0.1908	36	0.5878	61	0.8746	86	0.9976
12	0.2079	37	0.6018	62	0.8829	87	0.9986
13	0.225	38	0.6157	63	0.891	88	0.9994
14	0.2419	39	0.6293	64	0.8988	89	0.9998
15	0.2588	40	0.6428	65	0.9063		
16	0.2756	41	0.6561	66	0.9135		
17	0.2924	42	0.6691	67	0.9205		
18	0.309	43	0.682	68	0.9272		
19	0.3256	44	0.6947	69	0.9336		
20	0.342	45	0.7071	70	0.9397		
21	0.3584	46	0.7193	71	0.9455		
22	0.3746	47	0.7314	72	0.9511		
23	0.3907	48	0.7431	73	0.9563		
24	0.4067	49	0.7547	74	0.9613		
25	0.4226	50	0.766	75	0.9659		

Figure(M)

**Figure(M):** An abbreviated sine function table. In this method, we divide the width by the length of the stain, Which always generates a number less than 1. The analyst then looks for the closest corresponding number under the sine column of the chart. The

*number adjacent to this lists the degree of the impact angle. For example, a 0.5 W/L ratio equates to a 30° impact angle.*

The second method involves dividing the width of the stain by the length and comparing this number to a sine function table (see figure(M) ). In this instance, the result will always be less than 1. The analyst finds the corresponding number on the sine function table to determine the angle. A sine table simply eliminates the need for a scientific calculator .

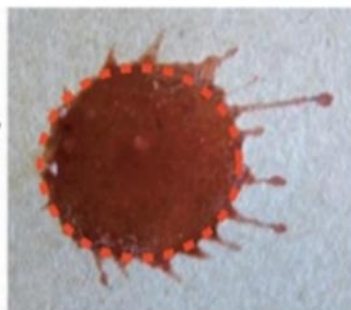
# Chapter 3

## 3.Method used for Measurement of Bloodstain Pattern.

### 3.1 Stain measurement

Measuring the stain is obviously as this provides the analyst with a length and width. In considering the measurements, the analyst measures only the main body of stain. The measurement must exclude any portion of the satellite, scallops, spine present in the stain. To accomplish this, one need simply envision a perfect ellipse superimposed on the stain. By choosing the point on the stain that naturally completes the ellipse, remaining tail portions are not subjectively drawn into the calculation (see figure (O)).

No portion of spines, tails, or satellite spatter should be considered when measuring the stain.



To eliminate these excess portions, simply imagine a perfect ellipse and compare the stain to it.



Figure(O)

**Figure (O) :** *By comparing the stain to an imagined ellipse, one can generally distinguish those portions of the stain to disregard during the Measurement process . including scallops, spines and satellite in the measurement will result in a skewed estimate of the impact angle.*

It is important to understand that the inclusion of any excess scallop or portion of the spine will change the overall length to width ratio.

Example:

Consider a stain with an actual length of 5mm,width of 4mm, and a scalloped tail of 0.5mm.

Given such a stain the following is possible:

Correct Measurement and Evaluation:

Length : 5mm plus 0.5mm scallop

Width : 4mm

Impact angle:53°

Incorrect Measurement and Skewed Evaluation:

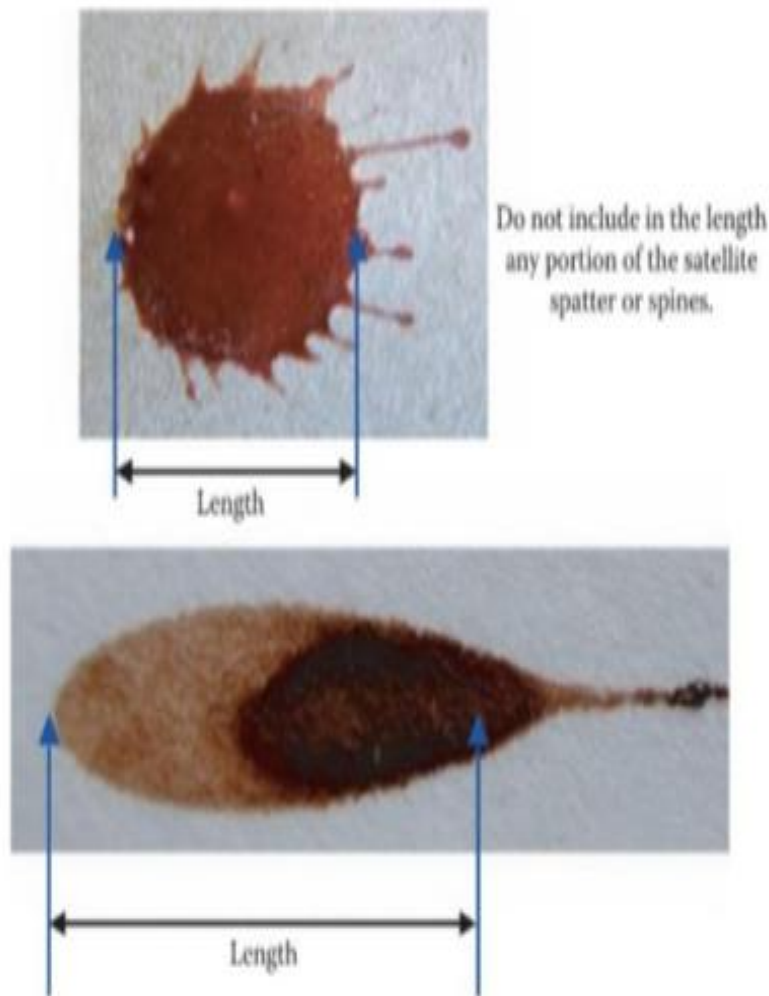
Length:5mm plus 0.5mm scallop

Width:4mm

Impact angle:46°

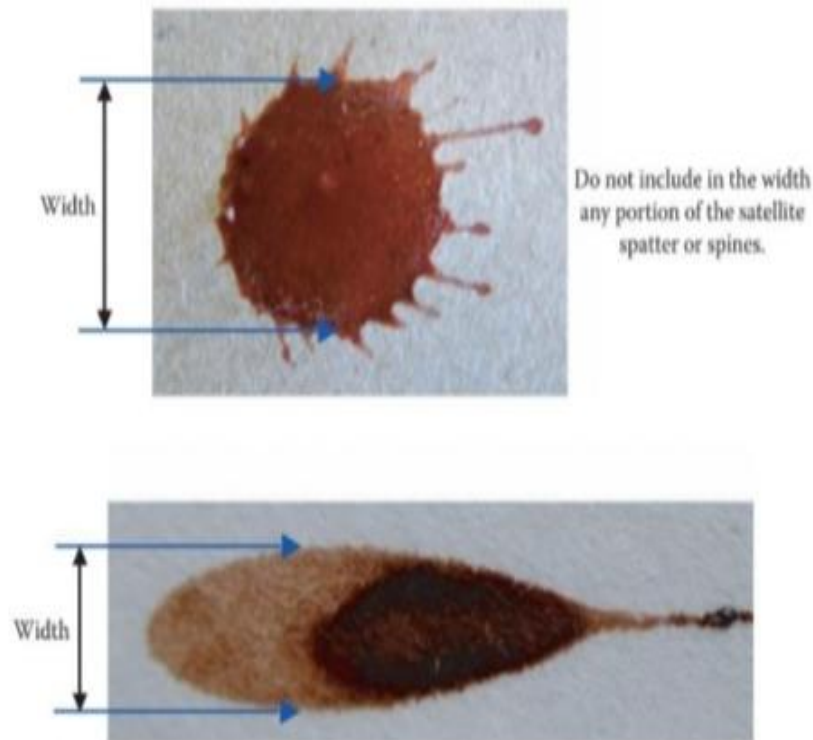
In this example,the small excess scallop adds an error of 7°to what we already accept as an estimation of the impact angle.Obviously ,by including these excess portion,the

analyst can change the calculated impact angle significantly . Unfortunately, there are no absolute rules for closing our the ellipse. The analyst must judge each stain individually and attempt to estimate these scalloped edges and the tail portions from the measurement (figure (P) and figure (Q) ).



Figure(P)

**Figure (P)** The length of the stain is measured along its major axis the analyst must exclude the measurement any spines, scallops, tail or satellite that may be present



Figure(Q)

**Figure(Q):** The width is measured along the minor axis of the stain. Although satellite spatter are less a concern when measuring this axis, scalloped along the outer edges of the stain are often encountered. Do not include the scallop in the measurement.

### **3.2 Combine the Information to Establish an Area of Origin**

Having learned to determine the impact angle , we combine the convergence point ( the top view) and the impact angle ( the side view) information in order to identify the area of origin for the spatter event. This is a location in 3-dimensional space. For many years,

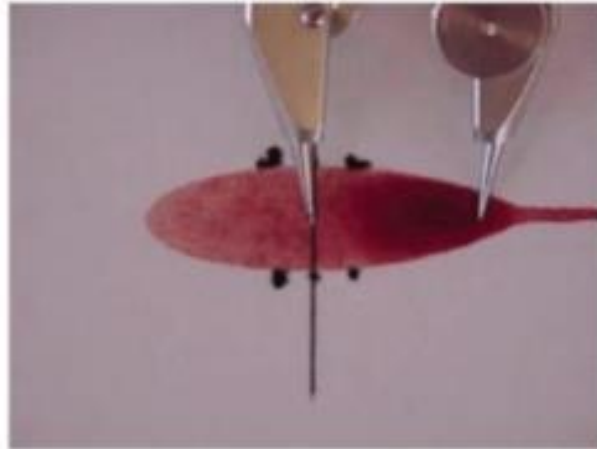
This location was referred to as the point of origin; but “point” suggests an actual point, a specific position in the X, Y, and Z planes. Because of the issues associated with defining directionality and impact angle we recognize that what we are defining is a general area, not a point. In some instances, this area may be quite specific, but in others, it may be quite large. For this reason, the “the point of origin” is now generally described as an area of origin. Both terms refer to the same thing and are synonymous.



Figure(R)

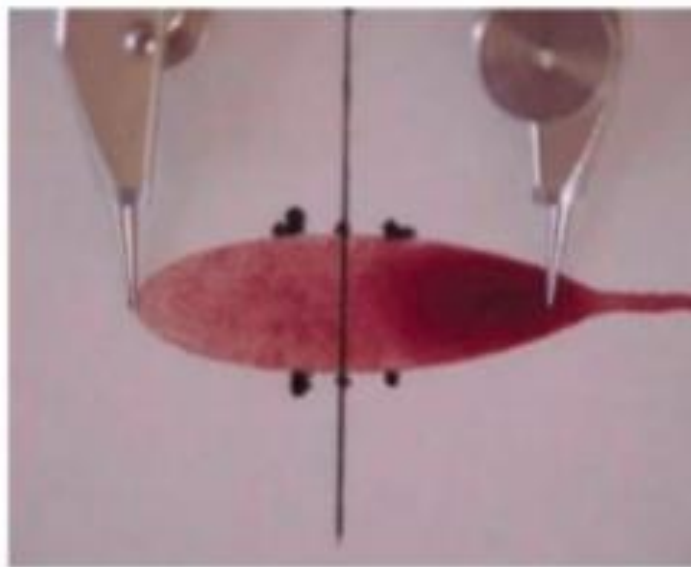
**Figure(R):** To locate the long axis length, place one arm on the center line and the other on the top of the stain. This measurement doubled identifies the length of the long axis.





Figure(S)

**Figure(S):** To visualize this point, without adjusting the arms and keeping the arm on the center point, simply rotate the other arm to the other end of the stain.



## Figure(P)

**Figure(P):** *As with the ellipse template comparison method, note that there is an area on the trailing edge (where the tail and stain meet) which is not included when measuring the stain length.*

For the sake of simplicity in the following example we will leave all stains on a single surface. See figure (U). the flight paths of the four stains appear to have a common convergence point in the scene. Our stains are “well-formed” and the analyst measures each and applies the sine formula, identifying each stain's impact angle.

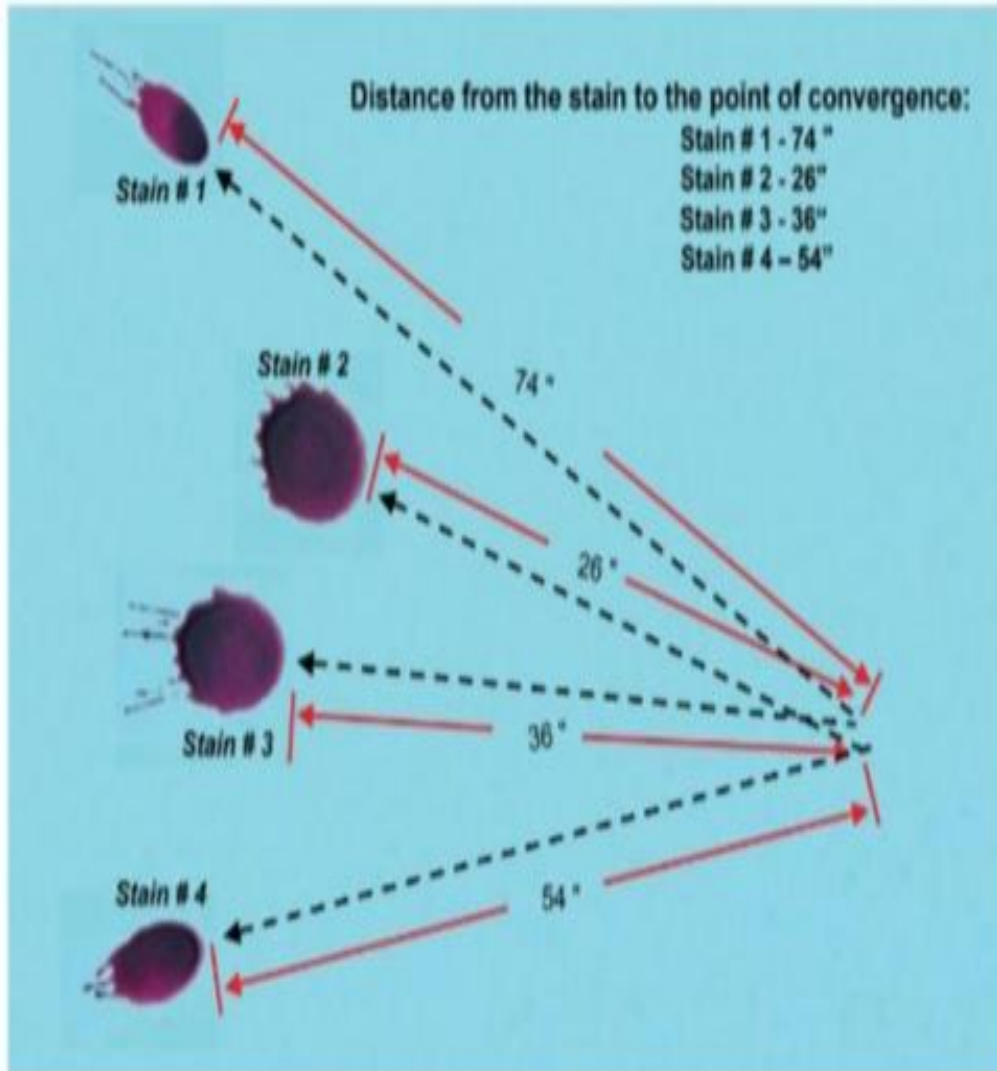
Accept for the example that based on the width/length ratio the stains impacted as follows:

Stain #1:30°

Stain #2:60°

Stain #3:50°

Stain # 4:40°



Figure(U)

**Figure (U):** An example of using point of convergence evaluations. In this instance, we have four stains with a common convergence point and we wish to determine if they share a common area of origin. We measure the distance from the base of each stain to the convergence point. We then combine that information with the impact angle in a graph ( figure (V)), which allows the analyst to visualize the area of origin for the spatter.

For purposes of the example, imagine that the analyst then measures the distance from the convergence point to the rear of each stain. In this instance, the analyst found:

Stain #1 is 74 in. From the convergence.

Stain #2 is 26 in. From the convergence.

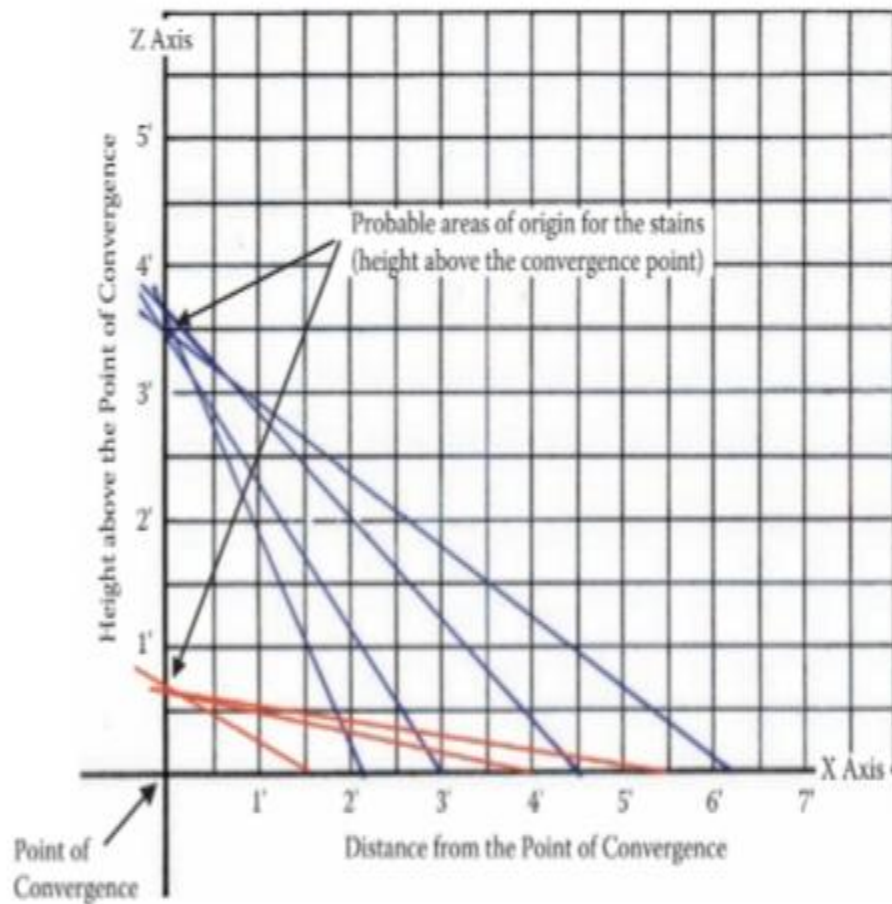
Stain #3 is 36 in. From the convergence.

Stain #4 is 54 in. From the convergence.

### **3.3 Defining area of origin with the Tangent Function**

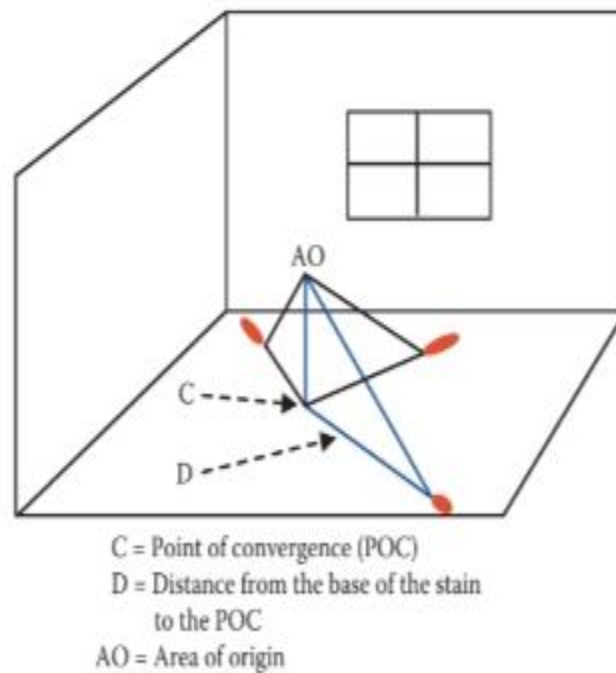
By using a scientific calculator, it is possible to forego the graphing process and simply calculate this distance above the point of convergence for each stain. The analyst does this using another relationship related to right triangles, the tangent. Figure (Y) shows the relationships of the scene to this imagined triangle.

The first step in making this determination is to identify stains that appear to have a common convergence. As I in the graphing method, the analyst measures the distance from these stains back to the point where the stains have a common intersection. The analyst also determines the impact angle of each stain.



Figure(V)

**Figure (V):** A graph indicating two areas of origin. In the scene from an overhead view, all the stains appear to share a common point of convergence. The inclusion of the impact angle information adds an additional dimension making it apparent that two separate groups are present, suggesting two separate events.



Figure(W)

**Figure (W):** The base of each stain's position, the point in two- dimensional space where their paths converge (c), and their area of origin (o) define another right triangle.

To determine area of origin, the analyst uses the following formula:

$$\tan i = H/D$$

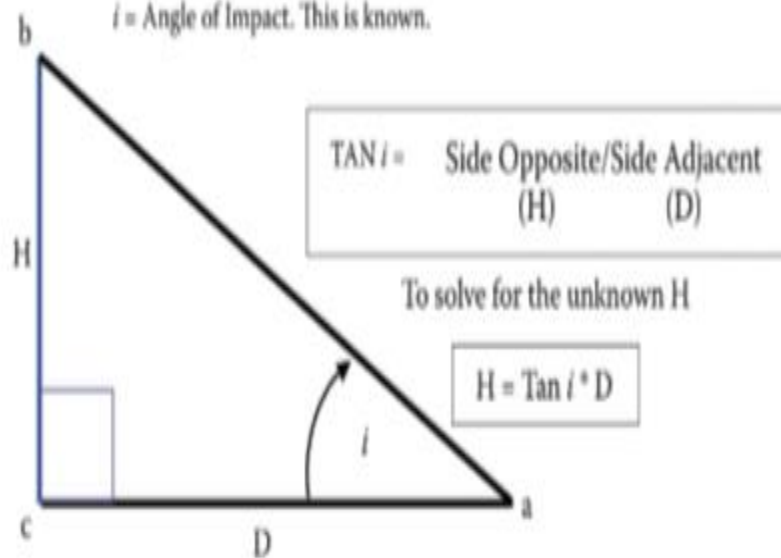
Where i equals the known impact angle, D equals the distance to the convergence point, and H equals the unknown distance above the target surface(see Figure (Z) ).For example:

Point b of the triangle is the same as AO in Figure 8.36.

Line bc = H = Height above the target. This is unknown.

Line ca = D = Distance to point of convergence.  
This is known

$i$  = Angle of Impact. This is known.



Figure(Y)

**Figure (Y):** our right triangle from Figure (Y) further defined. We know the distance from the stain to the point of convergence (ca) and we can establish the impact angle ( $i$ ). Using this information in the formula  $\tan i = H/D$ , we simply balance the equation and solve for the unknown H. Thus,  $H = \tan i \times D$

$$i = 19^\circ$$

$$D = 25 \text{ in.}$$

$$\tan 19^\circ = .344$$

$$0.344 \cdot 25 = 8$$

To solve for H we simply balanced the equation by multiplying  $\tan i$  by D, giving  $H = 8.66$  in. This procedure is convenient and provides immediate feedback at the scene. Once again, the analyst requires a calculator with trigonometric functions to use this method



## Conclusion

Mathematics play an important role in the analysis of blood spatter. The size, shape, direction and angle of impact can be analysed. Blood stain analysis can be determine the movement and direction of the blood source. When blood hits a surface at an angle other than 90, an elongated stain results. The acute angle formed between the direction of a blood drop and the plane of the surface it strikes. The direction of travel for multiple blood stains in a pattern that combined with the angle of impact determinations, is used to find the location of the blood source that was impacted to create the pattern. Surface effects on blood stain appearance are very important.

The morphology of blood stain distribution patterns at the crime scene carries vital information for a reconstruction of the events. Bloodstain pattern analysis is a forensic discipline in which, among others , the position of victims can be determined at crime scene on which blood has been shed. BPA is the forensic discipline concerned with the classification and interpretation of bloodstains and bloodstain patterns at the crime scene. BPA is also useful in forensic medicine. BPA is a valid forensic methods which belongs to the category of biological methods using trigonometric models.

# References

- 1) Stuart H. James, Paul E. Kish, T. Paulette Sutton.,  
*Principles of Bloodstain Pattern Analysis Theory and Practice.*,2005.
- 2) Anita Y. Wonder, *Bloodstain Pattern Evidence: Objective Approaches and Case Applications*, U.K., 2007.
- 3) Jay A. Seigel, Max M. Houck, *Fundamentals of Forensic Science*, U.K.,2010.



## **A STUDY ON ROBOTICS**

A project submitted to

**ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**

affiliated to

**Manonmaniam Sundaranar University, Tirunelveli.**

in partial fulfilment for the award of the degree

**Bachelor of Science in Mathematics**

Submitted by

<u>NAME</u>	<u>REGISTER NUMBER</u>
G. ARCHANA	18AUMT04
P. CHELLAMMAL	18AUMT12
A. DEEPIKA	18AUMT13
J. ESBA	18AUMT21
H. HERINA	18AUMT22
A. LIJONI	18AUMT32

Under the guidance of

**Dr. V.L. Stella Arputha Mary M.Sc.,M.Phil.,Ph.D.,**



**DEPARTMENT OF MATHEMATICS**

**St.Mary's College (Autonomous), Thoothukudi.**

**April 2021**

## CERTIFICATE

This is to certify that this project work entitled "A STUDY ON ROBOTICS" is submitted to ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI affiliated to Manonmaniam Sundaranar University, Tirunelveli, in partial fulfilment of the award of the Degree of Bachelor of Science in Mathematics and is the work done during the year 2020-2021 by the following students:

<u>NAME</u>	<u>REGISTER NUMBER</u>
G. ARCHANA	18AUMT04
P. CHELLAMMAL	18AUMT12
A. DEEPIKA	18AUMT13
J. ESBA	18AUMT21
H. HERINA	18AUMT22
A. LIJONI	18AUMT32

*V. S. Snella Arpulle Mary*  
Signature of the Guide

*V. P. A. J.*  
Signature of the HOD

*G. B. Srisilla Parithi*  
15/04/21  
Signature of the Examiner

*Lucia Rose*  
Signature of the principal  
St. Mary's College (Autonomous)  
Thoothukudi - 628 001

### DECLARATION

We hereby declare that the project reported entitled "ROBOTICS" is our original work. It has not been submitted to any university for any degree or diploma.

G. Archana  
(G.ARCHANA)

P. Chellammal  
(P.CHELLAMMAL)

A. Deepika  
(A.DEEPIKA)

J. Esba  
(J.ESBA)

H. Herina  
(H.HERINA)

A. Lijoni  
(A.LIJONI)



## ACKNOWLEDGEMENT

First of all, we thank the Almighty for showering his blessings to undergo this project.

We express our sincere gratitude and heartfelt thanks to our Principal Rev. Dr. Sr. A.S.J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., and to our Director Rev. Sr. F. Mary Joyce Baby M.A., M.Phil., for kindly permitted us to do this project.

We express our gratitude to Dr. A. Punitha Tharani, M.Sc., M.Phil., Ph. D., Hod of the Department, Department of the Mathematics for her inspirational ideas and encouragement.

We are very thankful to our guide Dr. V.L. Stella Arputha Mary M.Sc., M.Phil., Ph.D., Assistant Professor, Department of Mathematics for her efficient and effective guidance. She played a key role in preparation of this project.

We also thank our staff members for their encouragement in all our efforts. Finally, we thank all those who extended their help regarding this project.

Place : Thoothukudi

Date : 09.04.2021

## CONTENT

	Page no.
<b>Introduction</b>	<b>1</b>
<b>Preliminaries</b>	<b>2</b>
<b>Chapter 1 Stochastic State Estimation</b>	<b>5</b>
1.1 Dynamic System	5
1.2 An Example: The Mortar Shell	7
1.3 State Estimation	11
1.4 The Kalman Filter : Derivation	13
1.5 Linear System and the Kalman Filter	14
1.6 BLUE Estimators	18
<b>Chapter 2 Function Optomiation</b>	<b>21</b>
2.1 Local Minimization and Steepest Descent	22
2.2 Newton's Method	35
<b>Chapter 3 Eigen Vector</b>	<b>38</b>
3.1 Statistics	38
3.2 Spectral Embedding	40
3.3 Computation using Eigenvectors	41
3.4 Specialized property	42
3.5 Computing Eigenvalues	43
3.6 Sensitivity and Conditioning	46
<b>Conclusion</b>	<b>48</b>
<b>Reference</b>	<b>49</b>



## Introduction

Robotics research has been increasing exponentially and marking a new industrial revolution. Today, above one million robots are operating globally and the number is growing with time.

Today we have many robots with brainpower comparable or even beyond human intelligence, physical capability, perception and behaviour. And in some areas like computer-aided surgery, these intelligent machines can even surpass human capabilities.



Robotic systems play a crucial role in the world and sustainability. Robots presence and our dependencies on them are progressively growing. The following chapters brings together mathematical developments in the important fields of robotics, control and vision. Here, basic concepts about robotics in mathematics are presented.

## Preliminaries

### Dynamic System:

A dynamic system is a system whose phenomena occur over time. One often says that a system evolves over time.

Example: An electric circuit, whose input is the current in a given branch and whose output is a voltage across a pair of nodes.

### Linearity:

The mathematics becomes particularly simple when both the evolution function  $f$

and the output function  $h$  are linear. Then, the system equations become

$$x_{k+1} = F_k x_k + G_k u_k + \eta_k$$

$$y_k = H_k x_k + \xi_k$$

for the discrete case, and

$$\dot{x}(t) = F(t)x(t) + G(t)u(t) + \eta(t)$$

$$y(t) = H(t)x(t) + \xi(t)$$

for

for the continuous one.

### Kantorovich inequality:

Let  $Q$  be a positive definite, symmetric,  $n \times n$  matrix. For any vector  $y$  there holds

$$\frac{(y^T y)^2}{y^T Q^{-1} y y^T Q y} \geq \frac{4\sigma_1 \sigma_n}{(\sigma_1 + \sigma_n)^2}$$

Where  $\sigma_1$  and  $\sigma_n$  are respectively the largest and smallest value of  $Q$ .

### Eigenvalue and Eigenvector :

An Eigenvector  $\vec{x} \neq \vec{0}$  of the matrix  $A \in \mathbb{R}^{n \times n}$  is any vector satisfying  $A\vec{x} = \lambda \vec{x}$  for some  $\lambda \in \mathbb{R}$ ; the corresponding  $\lambda$  is known as an Eigenvalue. Complex Eigenvalues and Eigenvectors satisfy the same relationships with  $\lambda \in \mathbb{C}$  and  $\vec{x} \in \mathbb{C}^n$ .



### Spectrum and spectral radius:

The spectrum of  $A$  is the set of eigenvalues of  $A$ . The spectral radius  $\rho(A)$  is the eigenvalue  $\lambda$  maximizing  $|\lambda|$ .

The scale of an eigenvector is not important. In particular, scaling an eigenvector  $\vec{x}$  by  $c$  yields  $A(c\vec{x}) = cA\vec{x} = c\lambda\vec{x} = \lambda(c\vec{x})$ , so  $c\vec{x}$  is an eigenvector with the same eigenvalue. We often restrict our search by adding a constraint  $|\vec{x}|=1$ . Even this constraint does not completely relieve ambiguity, since now  $\pm\vec{x}$  are both eigenvectors with the same eigenvalue.

### Nondefective:

A matrix  $A \in \mathbb{R}^{n \times n}$  is nondefective or diagonalizable if its eigenvectors span  $\mathbb{R}^n$ .

We call such a matrix diagonalizable for the following reason:

If a matrix is diagonalizable, then it has  $n$  eigenvectors  $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^n$  with corresponding (possibly non-unique) eigenvalues  $\lambda_1, \dots, \lambda_n$ . Take the columns of  $X$  to be the vectors  $\vec{x}_i$ , and define  $D$  to be the diagonal matrix with eigenvalues  $\lambda_1, \dots, \lambda_n$  along the diagonal. Then, by definition of eigenvalues we have  $AX = XD$ ; this is simply a "stacked" version of  $A\vec{x}_i = \lambda_i\vec{x}_i$ . In other words,

$$D = X^{-1}AX$$

meaning  $A$  is diagonalized by a similarity transformation  $A \mapsto X^{-1}AX$ :

### Similar matrices:

Two matrices  $A$  and  $B$  are similar if there exists  $T$  with  $B = T^{-1}AT$ .

Similar matrices have the same eigenvalues, since if  $B\vec{x} = \lambda\vec{x}$ , then  $T^{-1}AT(T\vec{x}) = \lambda(T\vec{x})$ .

Equivalently,  $A(T\vec{x}) = (T\vec{x})\lambda$ , showing  $T\vec{x}$  is an eigenvector with eigenvalue  $\lambda$ .

### Symmetric and Positive Definite Matrices

#### Complex conjugate:

The complex conjugate of a number  $z \equiv a + bi \in \mathbb{C}$  is  $\bar{z} \equiv a - bi$ .

**Conjugate transpose:**

The conjugate transpose of  $A \in \mathbb{C}^{m \times n}$  is  $A^H \equiv A^*$ .

**Hermitian matrix:**

A matrix  $A \in \mathbb{C}^{n \times n}$  is Hermitian if  $A = A^H$ .

## CHAPTER 1

### Stochastic State Estimation

The most important part of studying a problem in robotics or vision, as well as in most other sciences, is to determine a good model for the phenomena and events that are involved. For instance, studying manipulation requires defining models for how a robot arm can move and for how it interacts with the world. Analyzing image motion implies defining models for how points move in space and how this motion projects onto the image. When motion is involved, as is very often the case, models take on frequently the form of *dynamic systems*. A dynamic system is a mathematical description of a quantity that evolves over time. The theory of dynamic systems is both rich and fascinating.

#### 1.1 Dynamic Systems

In its most general meaning, the term *system* refers to some physical entity on which some action is performed by means of an input  $u$ . The system reacts to this input and produces an output  $y$  (see figure 1.1).

A *dynamic system* is a system whose phenomena occur over time. One often says that a system evolves over time. Simple examples of a dynamic system are the following:

- An electric circuit, whose input is the current in a given branch and whose output is a voltage across a pair of nodes.
- A chemical reactor, whose inputs are the external temperature, the temperature of the gas being supplied, and the supply rate of the gas. The output can be the temperature of the reaction product.
- A mass suspended from a spring. The input is the force applied to the mass and the output is the position of the mass.



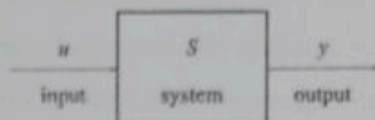


Figure 1.1: A general system.

In all these examples, what is input and what is output is a choice that depends on the application. Also, all the quantities in the examples vary continuously with time. In other cases, as for instance for switching networks and computers, it is more natural to consider time as a discrete variable. If time varies continuously, the system is said to be *continuous*; if time varies discretely, the system is said to be *discrete*.

### 1.1.1 Uncertainty

The systems are called *deterministic*, since the evolution is exactly determined once the initial state  $x$  at time 0 is known. Determinism implies that both the evolution function  $f$  and the output function  $h$  are known exactly. This is, however, an unrealistic state of affairs. In practice, the laws that govern a given physical system are known up to some uncertainty. In fact, the equations themselves are simple abstractions of a complex reality. The coefficients that appear in the equations are known only approximately, and can change over time as a result of temperature changes, component wear, and so forth. A more realistic model then allows for some inherent, unresolvable uncertainty in both  $f$  and  $h$ . This uncertainty can be represented as noise that perturbs the equations we have presented so far. A discrete system then takes on the following form:

$$x_{k+1} = f(x_k, u_k, k) + \eta_k$$

$$y_k = h(x_k, k) + \xi_k$$

and for a continuous system

$$\dot{x}(t) = f(x(t), u(t), t) + \eta(t)$$

$$y(t) = h(x(t), t) + \xi(t).$$

Without loss of generality, the noise distributions can be assumed to have zero mean, for otherwise the mean can be incorporated into the deterministic part, that is, in either  $f$  or  $h$ . The mean may not be known, but this is a different story: in general the parameters that enter into the definitions of  $f$  and  $h$  must be estimated by some method, and the mean perturbations are no different.

A common assumption, which is sometimes valid and always simplifies the mathematics, is that  $\eta$  and  $\xi$  are zero-mean Gaussian random variables with known covariance matrices  $Q$  and  $R$ , respectively.

### 1.1.2 Linearity

The mathematics becomes particularly simple when both the evolution function  $f$  and the output function  $h$  are linear. Then, the system equations become

$$\begin{aligned}x_{k+1} &= F_k x_k + G_k u_k + \eta_k \\ y_k &= H_k x_k + \xi_k\end{aligned}$$

for the discrete case, and

$$\begin{aligned}\dot{x}(t) &= F(t)x(t) + G(t)u(t) + \eta(t) \\ y(t) &= H(t)x(t) + \xi(t)\end{aligned}$$

for the continuous one. It is useful to specify the sizes of the matrices involved. We assume that the input  $u$  is a vector in  $R^p$ , the state  $x$  is in  $R^n$ , and the output  $y$  is in  $R^m$ . Then, the *state propagation matrix*  $F$  is  $n \times n$ , the *input matrix*  $G$  is  $n \times p$ , and the *output matrix*  $H$  is  $m \times n$ . The covariance matrix  $Q$  of the *system noise*  $\eta$  is  $n \times n$ , and the covariance matrix of the output noise  $\xi$  is  $m \times m$ .

## 1.2 An Example: the Mortar Shell

In this section, the example of the mortar shell will be discussed in order to see some of the technical issues involved in setting up the equations of a dynamic system. In particular, we consider discretization issues because the physical system is itself continuous, but we choose to model it as a discrete system for easier implementation on a computer.

In sections 1.3 through 1.5, we consider the *state estimation* problem: given observations of the output  $y$  over an interval of time, we want to determine the state  $x$  of the system. This is a very important task. For instance, in the case of the mortar shell, the state is the (initially unknown) position and velocity of the shell, while the



output is a set of observations made by a tracking system. Estimating the state then leads to enough knowledge about the shell to allow driving an antiaircraft gun to shoot the shell down in mid-flight.

You spotted an enemy mortar installation about thirty kilometers away, on a hill that looks about 0.5 kilometers higher than your own position. You want to track incoming projectiles with a Kalman filter so you can aim your guns accurately. You do not know the initial velocity of the projectiles, so you just guess some values: 0.6 kilometers/second for the horizontal component, 0.1 kilometers/second for the vertical component. Thus, your estimate of the initial state of the projectile is

$$\hat{x}_0 = \begin{bmatrix} \dot{d} \\ d \\ \dot{z} \\ z \end{bmatrix} = \begin{bmatrix} -0.6 \\ 30 \\ 0.1 \\ 0.5 \end{bmatrix}$$

where  $d$  is the horizontal coordinate,  $z$  is the vertical, you are at  $(0; 0)$ , and dots denote derivatives with respect to time.

From your high-school physics, you remember that the laws of motion for a ballistic trajectory are the following:

$$d(t) = d(0) + \dot{d}(0)t \quad \dots (1.1)$$

$$z(t) = z(0) + \dot{z}(0)t - \frac{1}{2}gt^2 \quad \dots (1.2)$$

where  $g$  is the gravitational acceleration, equal to  $9.8 \times 10^3$  kilometers per second squared. Since you do not trust your physics much, and you have little time to get ready, you decide to ignore air drag. Because of this, you introduce a state update covariance matrix  $Q = 0.1I_4$ , where  $I_4$  is the  $4 \times 4$  identity matrix.

All you have to track the shells is a camera pointed at the mortar that will rotate so as to keep the projectile at the center of the image, where you see a blob that increases in size as the projectile gets closer. Thus, the aiming angle of the camera gives you elevation information about the projectile's position, and the size of the blob tells you something about the distance, given that you know the actual size of the projectiles used and all the camera parameters. The projectile's elevation is

$$e = 1000 \frac{z}{d} \quad \dots (1.3)$$

when the projectile is at  $(d, z)$ . Similarly, the size of the blob in pixels is



$$s = \frac{1000}{\sqrt{d^2 + z^2}} \quad \dots (1.4)$$

You do not have very precise estimates of the noise that corrupts  $e$  and  $s$ , so you guess measurement covariances  $R_e = R_s = 1000$ , which you put along the diagonal of a  $2 \times 2$  diagonal measurement covariance matrix  $R$ .

### 1.2.1 The Dynamic System Equation

Equations (1.1) and (1.2) are continuous. Since you are taking measurements every  $dt = 0.2$  seconds, you want to discretize these equations. For the  $z$  component, equation (1.2) yields

$$\begin{aligned} z(t + dt) - z(t) &= z(0) + \dot{z}(0)(t + dt) - \frac{1}{2}g(t + dt)^2 - \left[ z(0) + \dot{z}(0)t - \frac{1}{2}gt^2 \right] \\ &= (\dot{z}(0) - gt)dt - \frac{1}{2}g(dt)^2 \\ &= \dot{z}(t)dt - \frac{1}{2}g(dt)^2 \end{aligned}$$

Since  $\dot{z}(0) - gt = \dot{z}(t)$ .

Consequently, if  $t + dt$  is time instant  $k + 1$  and  $t$  is time instant  $k$ , you have

$$z_{k+1} = z_k + \dot{z}_k dt - \frac{1}{2}g(dt)^2 \quad \dots (1.5)$$

The reasoning for the horizontal component  $d$  is the same, except that there is no acceleration:

$$d_{k+1} = d_k + \dot{d}_k dt \quad \dots (1.6)$$

Equations (1.5) and (1.6) can be rewritten as a single system update equation

$$x_{k+1} = Fx_k + Gu$$

where

$$x_k = \begin{bmatrix} d_k \\ \dot{d}_k \\ z_k \\ \dot{z}_k \end{bmatrix}$$

is the state, the  $4 \times 4$  matrix  $F$  depends on  $dt$ , the control scalar  $u$  is equal to  $-g$ , and the  $4 \times 1$  control matrix  $G$  depends on  $dt$ . The two matrices  $F$  and  $G$  are as follows:

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ dt & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & dt & 1 \end{bmatrix} \quad G = \begin{bmatrix} 0 \\ 0 \\ dt \\ -\frac{dt^2}{2} \end{bmatrix}$$

### 1.2.2 The Measurement Equation

The two nonlinear equations (1.3) and (1.4) express the available measurements as a function of the true values of the projectile coordinates  $d$  and  $z$ . We want to replace these equations with linear approximations.

To this end, we develop both equations as Taylor series around the current estimate and truncate them after the linear term. From the elevation equation (1.3), we have

$$e_k = 1000 z_d^z \approx 1000 \left[ \frac{z_k}{\hat{d}_k} + \frac{z - \hat{z}_k}{\hat{d}_k} - \frac{\hat{z}_k}{\hat{d}_k^2} (d - \hat{d}_k) \right],$$

so that after simplifying we can redefine the measurement to be the discrepancy from the estimated value:

$$e'_k = e_k - 1000 \frac{\hat{z}_k}{\hat{d}_k} \approx 1000 \left( \frac{z}{\hat{d}_k} - \frac{\hat{z}_k}{\hat{d}_k^2} d \right). \quad \dots (1.7)$$

We can proceed similarly for equation (1.4):

$$s_k = \frac{1000}{\sqrt{d^2 + z^2}} \approx \frac{1000}{\sqrt{\hat{d}_k^2 + \hat{z}_k^2}} - \frac{1000 \hat{d}_k}{(\hat{d}_k^2 + \hat{z}_k^2)^{3/2}} (d - \hat{d}_k) - \frac{1000 \hat{z}_k}{(\hat{d}_k^2 + \hat{z}_k^2)^{3/2}} (z - \hat{z}_k)$$

and after simplifying:

$$\begin{aligned} s'_k &= s_k - \frac{2000}{\sqrt{\hat{d}_k^2 + \hat{z}_k^2}} \\ &\approx -1000 \left[ \frac{\hat{d}_k}{(\hat{d}_k^2 + \hat{z}_k^2)^{3/2}} d + \frac{\hat{z}_k}{(\hat{d}_k^2 + \hat{z}_k^2)^{3/2}} z \right] \dots (1.8) \end{aligned}$$

The two measurements  $e'_k$  and  $s'_k$  just defined can be collected into a single measurement vector



$$y_k = \begin{bmatrix} e'_k \\ s'_k \end{bmatrix},$$

and the two approximate measurement equations (1.7) and (1.8) can be written in the matrix form

$$y_k = H_k X_k \quad \dots (1.9)$$

where the measurement matrix  $H_k$  depends on the current state estimate  $\hat{x}_k$ :

$$H_k = -1000 \begin{bmatrix} \hat{d}_k & \hat{z}_k \\ 0 & (\hat{d}_k^2 + \hat{z}_k^2)^{3/2} & 0 & (\hat{d}_k^2 + \hat{z}_k^2)^{3/2} \\ 0 & \hat{z}_k & 0 & -\hat{d}_k \\ & \hat{d}_k^2 & & \end{bmatrix}$$

As the shell approaches us, we frantically start studying state estimation, and in particular Kalman filtering, in the hope to build a system that lets us shoot down the shell before it hits us.

### 1.3 State Estimation

In this section, the estimation problem is defined in some more detail. Given a discrete dynamic system

$$x_{k+1} = F_k x_k + G_k u_k + \eta_k \quad \dots (1.10)$$

$$y_k = H_k x_k + \xi_k \quad \dots (1.11)$$

where the system noise  $\eta_k$  and the measurement noise  $\xi_k$  are Gaussian variables,

$$\eta_k \sim N(0, Q_k)$$

$$\xi_k \sim N(0, R_k),$$

The term "recursive" in the systems theory literature corresponds loosely to "incremental" or "iterative" in computer science as well as a (possibly completely wrong) estimate  $\hat{x}_0$  of the initial state and an initial covariance matrix  $P_0$  of the estimate  $\hat{x}_0$ , the Kalman filter computes the optimal estimate  $\hat{x}_{k|k}$  at time  $k$  given the measurements  $y_0, \dots, y_k$ . The filter also computes an estimate  $P_{k|k}$  of the covariance of  $\hat{x}_{k|k}$  given those measurements. In these expressions, the hat means that the quantity is an estimate. Also, the first  $k$  in the subscript refers to which variable is being estimated, the second to which measurements are being used for the estimate. Thus, in

general,  $\hat{x}_{i|j}$  is the estimate of the value that  $x$  assumes at time  $i$  given the first  $j+1$  measurements  $y_0, \dots, y_j$ .

### 1.3.1 Propagation

Just after arrival of the measurement  $y_k$ , both state estimate and state covariance matrix have been updated as described above. But between time  $k$  and time  $k+1$  both state and covariance may change. The state changes according to the system equation (1.10), so our estimate  $\hat{x}_{k+1|k}$  of  $x_{k+1}$  given  $y_0, \dots, y_k$  should reflect this change as well. Similarly, because of the system noise  $\eta_k$ , our uncertainty about this estimate may be somewhat greater than one time epoch ago. The system equation (1.10) essentially "dead reckons" the new state from the old, and inaccuracies in our model of how this happens lead to greater uncertainty. This increase in uncertainty depends on the system noise covariance  $Q_k$ . Thus, both state estimate and covariance must be *propagated* to the new time  $k+1$  to yield the new state estimate  $\hat{x}_{k+1|k}$  and the new covariance  $P_{k+1|k}$ . Both these changes are shown on the right in figure 1.2.

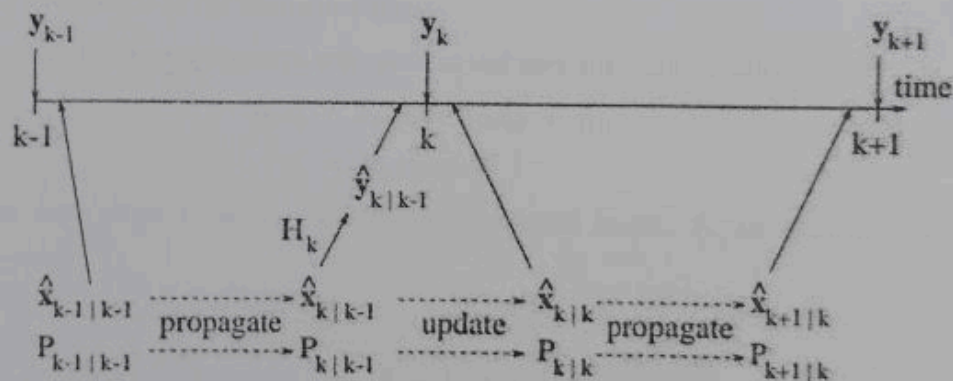


Figure 1.2: The update stage of the Kalman filter changes the estimate of the current system state  $x_k$  to make the prediction of the measurement closer to the actual measurement  $y_k$ . Propagation then accounts for the evolution of the system state, as well as the consequent growing uncertainty.

In summary, just as the state vector  $x_k$  represents all the information necessary to describe the evolution of a deterministic system, the covariance matrix  $P_{k|k}$  contains all the necessary information about the probabilistic part of the system, that is, about how both the system noise  $\eta_k$  and the measurement noise  $\xi_k$  corrupt the quality of the state estimate  $\hat{x}_{k|k}$ .



Hopefully, this intuitive introduction to Kalman filtering gives you an idea of what the filter does, and what information it needs to keep working. To turn these concepts into a quantitative algorithm we need some preliminaries on optimal estimation, which are discussed in the next section. The Kalman filter itself is derived in section 1.4.

#### 1.4 The Kalman Filter: Derivation

We now have all the components necessary to write the equations for the Kalman filter. To summarize, given a linear measurement equation

$$y = Hx + n$$

where  $n$  is a Gaussian random vector with zero mean and covariance matrix  $R$ ,

$$n \sim N(0, R),$$

the best linear unbiased estimate  $\hat{x}$  of  $x$  is

$$\hat{x} = PH^TR^{-1}y$$

where the matrix

$$P \triangleq E[(\hat{x} - x)(\hat{x} - x)^T] = (H^TR^{-1}H)^{-1}$$

is the covariance of the estimation error.

Given a dynamic system with system and measurement equations

$$x_{k+1} = F_k x_k + G_k u_k + \eta_k \quad \dots (1.19)$$

$$y_k = H_k x_k + \xi_k$$

where the system noise  $\eta_k$  and the measurement noise  $\xi_k$  are Gaussian random vectors,

$$\eta_k \sim N(0, Q_k)$$

$$\xi_k \sim N(0, R_k),$$

as well as the best, linear, unbiased estimate  $\hat{x}_0$  of the initial state with an error covariance matrix  $P_0$ , the Kalman filter computes the best, linear, unbiased estimate  $\hat{x}_{k|k}$  at time  $k$  given the measurements  $y_0, \dots, y_k$ . The filter also computes the covariance  $P_{k|k}$  of the error  $\hat{x}_{k|k} - x_k$  given those measurements. Computation occurs according to the phases of update and propagation illustrated in figure 1.2. We now apply the results from optimal estimation to the problem of updating and propagating the state estimates and their error covariances.

### 1.4.1 Propagation

Propagation is even simpler. Since the new state is related to the old through the system equation 1.19, and the noise term  $u_k$  is zero mean, unbiasedness requires

$$\hat{x}_{k+1|k} = F_k \hat{x}_{k|k} + G_k u_k,$$

which is the state estimate propagation equation of the Kalman filter. The error covariance matrix is easily propagated thanks to the linearity of the expectation operator:

$$\begin{aligned} P_{k+1|k} &= E[(\hat{x}_{k+1|k} - x_{k+1})(\hat{x}_{k+1|k} - x_{k+1})^T] \\ &= E[(F_k(\hat{x}_{k|k} - x_k) - \eta_k)(F_k(\hat{x}_{k|k} - x_k) - \eta_k)^T] \\ &= F_k E[(\hat{x}_{k|k} - x_k)(\hat{x}_{k|k} - x_k)^T] F_k^T + E[\eta_k \eta_k^T] \\ &= F_k P_{k|k} F_k^T + Q_k \end{aligned}$$

where the system noise  $\eta_k$  and the previous estimation error  $(\hat{x}_{k|k} - x_k)$  were assumed to be uncorrelated.

### 1.4.2 Kalman Filter Equations

In summary, the Kalman filter evolves an initial estimate and an initial error covariance matrix,

$$\hat{x}_{0|-1} \triangleq \hat{x}_0 \quad \text{and} \quad P_{0|-1} \triangleq P_0,$$

both assumed to be given, by the update equations

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k (y_k - H_k \hat{x}_{k|k-1}) \\ P_{k|k}^{-1} &= P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k \end{aligned}$$

where the Kalman gain is defined as

$$K_k = P_{k|k} H_k^T R_k^{-1}$$

and by the propagation equations

$$\begin{aligned} \hat{x}_{k+1|k} &= F_k \hat{x}_{k|k} + G_k u_k \\ P_{k+1|k} &= F_k P_{k|k} F_k^T + Q_k. \end{aligned}$$

## 1.5 Linear Systems and the Kalman Filter

In order to connect the theory of state estimation with what we have learned so far about linear systems, we now show that estimating the initial state  $x_0$  from the first

$k+1$  measurements, that is, obtaining  $\hat{x}_{0|k}$ , amounts to solving a linear system of equations with suitable weights for its rows.

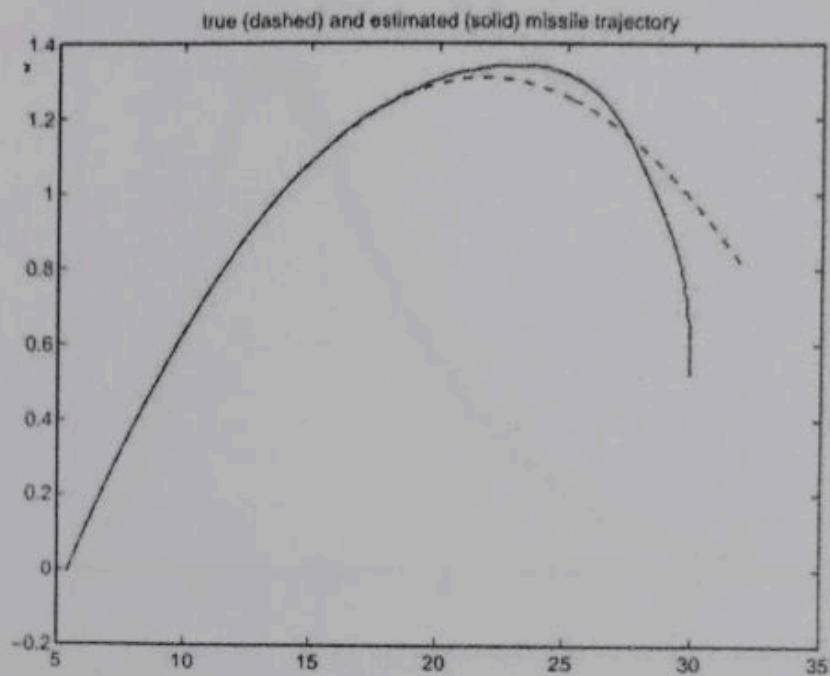


Figure 1.3: The true and estimated trajectories get closer to one another. Trajectories start on the right.

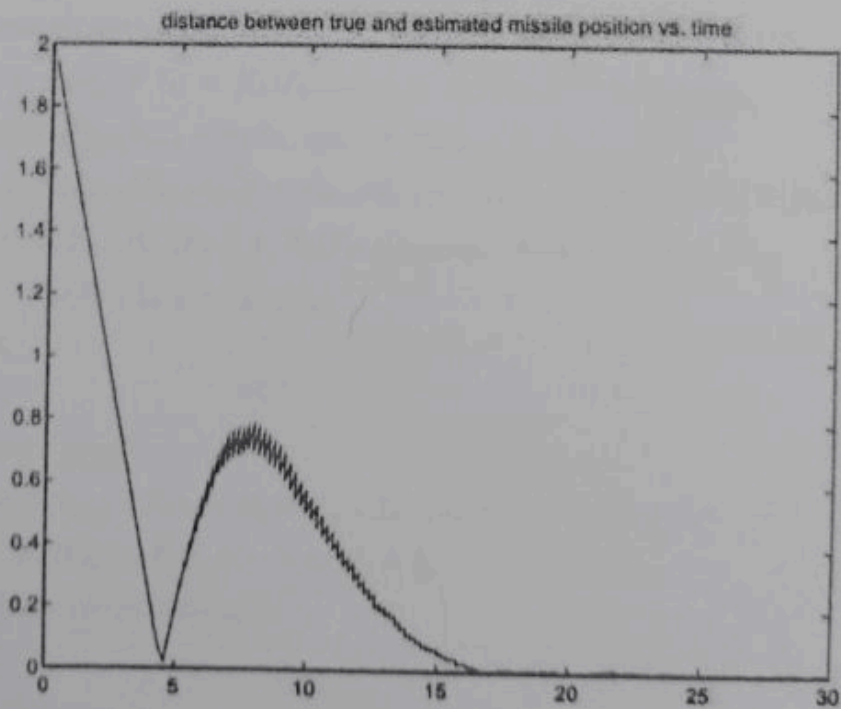




Figure 1.4: The estimate actually closes in towards the target.

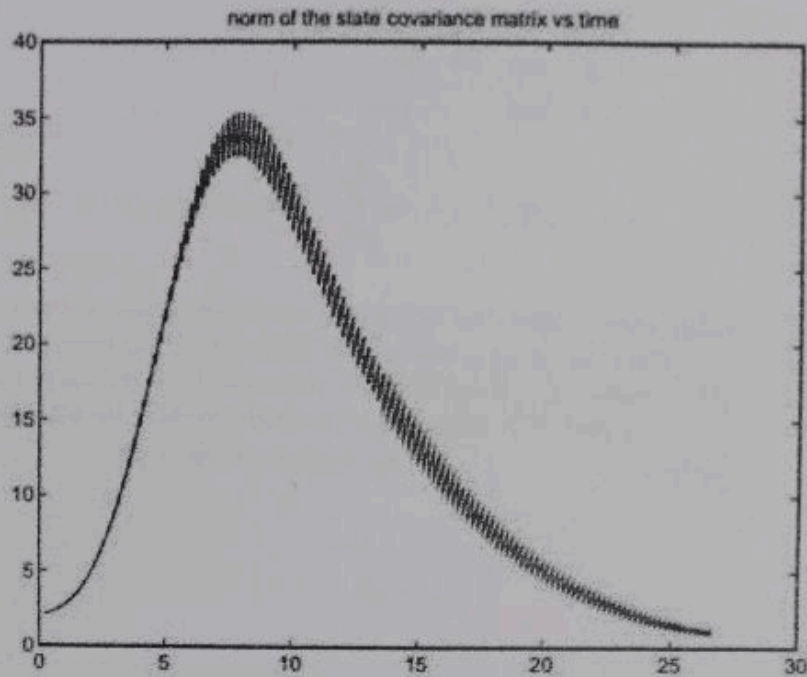


Figure 1.5: After an initial increase in uncertainty, the norm of the state covariance matrix converges to zero. Upwards segments correspond to state propagation, downwards ones to state update.

The basic recurrence equations (1.10) and (1.11) can be expanded as follows:

$$\begin{aligned}
 y_k &= H_k x_k + \xi_k = H_k (F_{k-1} x_{k-1} + G_{k-1} u_{k-1} + \eta_{k-1}) + \xi_k \\
 &= H_k F_{k-1} x_{k-1} + H_k G_{k-1} u_{k-1} + H_k \eta_{k-1} + \xi_k \\
 &= H_k F_{k-1} (F_{k-2} x_{k-2} + G_{k-2} u_{k-2} + \eta_{k-2}) + H_k G_{k-1} u_{k-1} + H_k \eta_{k-1} + \xi_k \\
 &= H_k F_{k-1} F_{k-2} x_{k-2} + H_k (F_{k-1} G_{k-2} u_{k-2} + G_{k-1} u_{k-1}) + \\
 &\quad H_k (F_{k-1} \eta_{k-2} + \eta_{k-1}) \xi_k \\
 &\vdots \\
 &= H_k F_{k-1} \dots F_0 x_0 + H_k (F_{k-1} \dots F_1 G_0 u_0 + \dots + G_{k-1} u_{k-1}) + \\
 &\quad H_k (F_{k-1} \dots F_1 \eta_0 + \dots + \eta_{k-1}) + \xi_k
 \end{aligned}$$

or in a more compact form,



$$y_k = H_k \phi(k-1, 0) x_0 + H_k \sum_{j=1}^k \phi(k-1, j) G_{j-1} u_{j-1} + v_k \quad \dots (1.20)$$

where

$$\phi(l, j) = \begin{cases} F_l \dots F_j & \text{for } l \geq j \\ 1 & \text{for } l < j \end{cases}$$

and the term

$$v_k = H_k \sum_{j=1}^k \phi(k-1, j) \eta_{j-1} + \xi_k$$

is noise.

The key thing to notice about this somewhat intimidating expression is that for any  $k$  it is a linear system in  $x_0$ , the initial state of the system. We can write one system like the one in equation (1.20) for every value of  $k = 0, \dots, K$ , where  $K$  is the last time instant considered, and we obtain a large system of the form

$$z_K = \psi_K x_0 + g_K + n_K \quad \dots (1.21)$$

where

$$z_K = \begin{bmatrix} y_0 \\ \vdots \\ y_K \end{bmatrix}$$

$$\psi_K = \begin{bmatrix} 0 \\ H_1 F_0 u_0 \\ \vdots \\ H_K \phi(K-1, 0) \end{bmatrix}$$

$$g_K = \begin{bmatrix} H_0 \\ H_1 F_0 G_0 \\ \vdots \\ H_K (\phi(K-1, 1) u_0 + \dots + G_{K-1} u_{K-1}) \end{bmatrix}$$

$$n_K = \begin{bmatrix} v_0 \\ \vdots \\ v_K \end{bmatrix}$$

Without knowing anything about the statistics of the noise vector  $n_K$  in equation (1.21), the best we can do is to solve the system

$$z_K = \psi_K x_0 + g_K$$

in the sense of least squares, to obtain an estimate of  $x_0$  from the measurements  $y_0, \dots, y_K$ :

$$\hat{x}_{0|K} = \psi_K^\dagger (z_K - g_K)$$

where  $\psi_K^\dagger$  is the pseudoinverse of  $\psi_K$ . We know that if  $\psi_K$  has full rank, the result with the pseudoinverse is the same as we would obtain by solving the normal equations, so that

$$\psi_K^\dagger = (\psi_K^T \psi_K)^{-1} \psi_K^T .$$

## 1.6 BLUE Estimators

In what sense does the Kalman filter use covariance information to produce better estimates of the state? As we will see later, the Kalman filter computes the *Best Linear Unbiased Estimate (BLUE)* of the state. In this section, we see what this means, starting with the definition of a linear estimation problem, and then considering the attributes "best" and "unbiased" in turn.

### 1.6.1 Unbiased

In addition to requiring our estimator to be linear and with minimum covariance, we also want it to be unbiased, in the sense that if repeat the same estimation experiment many times we neither consistently overestimate nor consistently underestimate  $x$ . Mathematically, this translates into the following requirement:

$$E[x - \hat{x}] = 0 \quad \text{and} \quad E[\hat{x}] = E[x] .$$

### 1.6.2 The BLUE

We now address the problem of finding the Best Linear Unbiased Estimator (BLUE)

$$\hat{x} = Ly$$

of  $x$  given that  $y$  depends on  $x$  according to the model  $y = Hx$ , which is repeated here for convenience:

$$y = Hx + n \quad \dots (1.22)$$

First, we give a necessary and sufficient condition for  $L$  to be unbiased.

**Lemma .1** *Let  $n$  in equation (1.22) be zero mean. Then the linear estimator  $L$  is unbiased if and only if*

$$LH = I ,$$

*the identity matrix.*

Proof.

$$\begin{aligned} E[x - \hat{x}] &= E[x - Ly] = E[x - L(Hx + n)] \\ &= E[(I - LH)x] - E[Ln] = (I - HL)E[x] \end{aligned}$$

since  $E[Ln] = LE[n]$  and  $E[n] = 0$ . For this to hold for all  $x$  we need  $I - LH = 0$ .  $\Delta$

And now the main result.



**Theorem 1:** The Best Linear Unbiased Estimator (BLUE)

$$\hat{x} = Ly$$

for the measurement model

$$y = Hx + n$$

where the noise vector  $n$  has zero mean and covariance  $R$  is given by

$$L = (H^T R^{-1} H)^{-1} H^T R^{-1}$$

and the covariance of the estimate  $\hat{x}$  is

$$P = E[(x - \hat{x})(x - \hat{x})^T] = (H^T R^{-1} H)^{-1}. \quad \dots (1.23)$$

**Proof.** We can write

$$\begin{aligned} P &= E[(x - \hat{x})(x - \hat{x})^T] = E[(x - Ly)(x - Ly)^T] \\ &= E[(x - LHx - Ln)(x - LHx - Ln)^T] \\ &= E[(I - LH)x - Ln][(I - LH)x - Ln]^T \\ &= E[Lnn^T L^T] \\ &= LE[nn^T]L^T \\ &= LRL^T \end{aligned}$$

because  $L$  is unbiased, so that  $LH = I$ .

To show that

$$L_0 = (H^T R^{-1} H)^{-1} H^T R^{-1} \quad \dots (1.24)$$

is the best choice, let  $L$  be any (other) linear unbiased estimator. We can trivially write

$$L = L_0 + (L - L_0)$$

and

$$\begin{aligned} P &= LRL^T \\ &= [L_0 + (L - L_0)]R[L_0 + (L - L_0)]^T \\ &= L_0RL_0^T + (L - L_0)RL_0^T + L_0R(L - L_0)^T + (L - L_0)R(L - L_0)^T \end{aligned}$$

From (1.24) we obtain

$$RL_0^T = RR^{-1}H(H^T R^{-1}H)^{-1} = H(H^T R^{-1}H)^{-1}$$

so that

$$(L - L_0)RL_0^T = (L - L_0)H(H^T R^{-1}H)^{-1} = (LH - L_0H)(H^T R^{-1}H)^{-1}$$

But  $L$  and  $L_0$  are unbiased, so  $LH = L_0H = I$ , and

$$(L - L_0)RL_0^T = 0.$$

The term  $L_0R(L - L_0)^T$  is the transpose of this, so it is zero as well. In conclusion,

$$P = L_0RL_0^T + (L - L_0)R(L - L_0)^T,$$

the sum of two positive definite or at least semidefinite matrices. For such matrices, the norm of the sum is greater or equal to either norm, so this expression is minimized when the second term vanishes, that is, when  $L = L_0$ .

✓ This proves that the estimator given by (1.24) is the best, that is, that it has minimum covariance. To prove that the covariance  $P$  of  $\hat{x}$  is given by equation (1.23), we simply substitute  $L_0$  for  $L$  in  $P = LRL^T$  :

$$\begin{aligned} P &= L_0 R L_0^T \\ &= (H^T R^{-1} H)^{-1} H^T R^{-1} R R^{-1} H (H^T R^{-1} H)^{-1} \\ &= (H^T R^{-1} H)^{-1} H^T R^{-1} H (H^T R^{-1} H)^{-1} \\ &= (H^T R^{-1} H)^{-1} \end{aligned}$$

as promised.

Δ



## Chapter 2

### Function Optimization

There are three main reasons why most problems in robotics, vision, and arguably every other science or endeavor take on the form of optimization problems. One is that the desired goal may not be achievable, and so we try to get as close as possible to it. The second reason is that there may be more ways to achieve the goal, and so we can choose one by assigning a quality to all the solutions and selecting the best one. The third reason is that we may not know how to solve the system of equations  $f(x) = 0$ , so instead we minimize the norm  $\|f(x)\|$ , which is a scalar function of the unknown vector  $x$ .

We have encountered the first two situations when talking about linear systems. The case in which a linear system admits exactly one exact solution is simple but rare. More often, the system at hand is either incompatible (some say over constrained) or, at the opposite end, underdetermined. In fact, some problems are both, in a sense. While these problems admit no exact solution, they often admit a multitude of approximate solutions. In addition, many problems lead to nonlinear equations.

Consider, for instance, the problem of Structure from Motion (SFM) in computer vision. Nonlinear equations describe how points in the world project onto the images taken by cameras at given positions in space. Structure from motion goes the other way around, and attempts to solve these equations: image points are given, and one wants to determine where the points in the world and the cameras are. Because image points come from noisy measurements, they are not exact, and the resulting system is usually incompatible. SFM is then cast as an optimization problem. On the other hand, the exact system (the one with perfect coefficients) is often close to being underdetermined. For instance, the images may be insufficient to recover a certain shape under a certain motion. Then, an additional criterion must be added to dense what a "good" solution is. In these cases, the noisy system admits no exact solutions, but has many approximate ones.

The term "optimization" is meant to subsume both minimization and maximization. However, maximizing the scalar function  $f(x)$  is the same as minimizing  $-f(x)$ , so we consider optimization and minimization to be essentially synonyms. Usually, one is after

global minima. However, global minima are hard to find, since they involve a universal quantifier:  $x^*$  is a global minimum of  $f$  if for every other  $x$  we have  $f(x) \geq f(x^*)$ . Global minimization techniques like simulated annealing have been proposed, but their convergence properties depend very strongly on the problem at hand. In this chapter, we consider local minimization: we pick starting point  $x_0$ , and we descend in the landscape of  $f(x)$  until we cannot go down any further. The bottom of the valley is a local minimum.

Local minimization is appropriate if we know how to pick  $x_0$  that is close to  $x^*$ . This occurs frequently in feedback systems. In these systems, we start at a local (or even a global) minimum. The system then evolves and escapes from the minimum. As soon as this occurs, a control signal is generated to bring the system back to the minimum. Because of this immediate reaction, the old minimum can often be used as a starting point  $x_0$  when looking for the new minimum, that is, when computing the required control signal. More formally, we reach the correct minimum  $x^*$  as long as the initial point  $x_0$  is in the basin of attraction of  $x^*$ , defined as the largest neighborhood of  $x^*$  in which  $f(x)$  is convex.

Good references for the discussion in this chapter are Matrix Computations, Practical Optimization, and Numerical Recipes in C, all of which are listed with full citations in section 2.5.

## 2.1 Local Minimization and Steepest Descent

Suppose that we want to find a local minimum for the scalar function  $f$  of the vector variable  $x$ , starting from an initial point  $x_0$ . Picking an appropriate  $x_0$  is crucial, but also very problem-dependent. We start from  $x_0$ , and we go downhill. At every step of the way, we must make the following decisions:

- Whether to stop.
- In what direction to proceed.
- How long a step to take.

In fact, most minimization algorithms have the following structure:

$$k = 0$$



while  $x_k$  is not a minimum

compute step direction  $p_k$  with  $\|p_k\|=1$

compute step size  $\alpha_k$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$k = k + 1$$

end.

Different algorithms differ in how each of these instructions is performed.

It is intuitively clear that the choice of the step size  $\alpha_k$  is important. Too small a step leads to slow convergence, or even to lack of convergence altogether. Too large a step causes overshooting, that is, leaping past the solution. The most disastrous consequence of this is that we may leave the basin of attraction, or that we oscillate back and forth with increasing amplitudes, leading to instability. Even when oscillations decrease, they can slow down convergence considerably.

What is less obvious is that the best direction of descent is not necessarily, and in fact is quite rarely, the direction of steepest descent, as we now show. Consider a simple but important case,

$$f(x) = c + a^T x + \frac{1}{2} x^T Q x \quad (2.1)$$

where  $Q$  is a symmetric, positive definite matrix. *Positive definite* means that for every nonzero  $x$  the quantity  $x^T Q x$  is positive. In this case, the graph of  $f(x)$  is a plane  $a^T x$  plus a paraboloid.

Of course, if  $f$  were this simple, no descent methods would be necessary. In fact the minimum of  $f$  can be found by setting its gradient to zero:

$$\frac{\partial f}{\partial x} = a + Q x = 0$$

so that the minimum  $x^*$  is the solution to the linear system

$$Qx = -a \quad (2.2)$$

Since  $Q$  is positive definite, it is also invertible (why?), and the solution  $x^*$  is unique. However, understanding the behavior of minimization algorithms in this simple case is crucial in order to establish the convergence properties of these algorithms for more general functions. In fact, all smooth functions can be approximated by paraboloids in a sufficiently small neighborhood of any point.

Let us therefore assume that we minimize  $f$  as given in equation (2.1), and that at every step we choose the direction of steepest descent. In order to simplify the mathematics, we observe that if we let

$$\tilde{e}(x) = \frac{1}{2}(x - x^*)^T Q(x - x^*)$$

then we have

$$\tilde{e} = f(x) - c + \frac{1}{2}x^{*T}Qx^* = f(x) - f(x^*) \quad (2.3)$$

so that  $\tilde{e}$  and  $f$  differ only by a constant. In fact,

$$\tilde{e}(x) = \frac{1}{2}(x^T Qx + x^{*T} Qx - 2x^T Qx^*) = \frac{1}{2}x^T Qx + a^T x + \frac{1}{2}x^{*T} Qx^* = f(x) - c + \frac{1}{2}x^{*T} Qx^*$$

And from the equation we obtain

$$f(x^*) = c + a^T x^* + \frac{1}{2}x^{*T} Qx^* = c - x^{*T} Qx^* + \frac{1}{2}x^{*T} Qx^* = c - \frac{1}{2}x^{*T} Qx^*$$

Since  $\tilde{e}$  is simpler, we consider that we are minimizing  $\tilde{e}$  rather than  $f$ . In addition, we can let

$$y = x - x^*$$



that is, we can shift the origin of the domain to  $x^*$  and study the function

$$e(y) = \frac{1}{2} y^T Q y$$

instead of  $f$  or  $\tilde{e}$  without loss of generality. We will transform everything back to  $x$  once we are done. Of course, by construction, the new minimum is at

$$y^* = 0$$

where  $e$  reaches a value of zero:

$$e(y^*) = e(0) = 0:$$

However, we let our steepest descent algorithm find this minimum by starting from the initial point

$$y_0 = x_0 - x^*.$$

At every iteration  $k$ , the algorithm chooses the direction of steepest descent, which is in the direction

$$p_k = -\frac{g_k}{\|g_k\|}$$

opposite to the gradient of  $e$  evaluated at  $y_k$

$$g_k = g(y_k) = \left. \frac{\partial e}{\partial y} \right|_{y=y_k} = Q y_k$$

We select for the algorithm the most favorable step size, that is, the one that takes us from  $y_k$  to the lowest point in the direction of  $p_k$ . This can be found by differentiating the function

$$e(y_k + \alpha p_k) = \frac{1}{2} (y_k + \alpha p_k)^T Q (y_k + \alpha p_k)$$

with respect to  $\alpha$ , and setting the derivative to zero to obtain the optimal step  $\alpha_k$ . We have

$$\frac{\partial e(y_k + \alpha p_k)}{\partial \alpha} = (y_k + \alpha p_k)^T Q p_k$$

and setting this to zero yields

$$\alpha_k = - \frac{(Q y_k)^T p_k}{p_k^T Q p_k} = - \frac{g_k^T p_k}{p_k^T Q p_k} = \|g_k\| \frac{p_k^T p_k}{p_k^T Q p_k} = \|g_k\| \frac{g_k^T g_k}{g_k^T Q g_k} \quad (2.4)$$

Thus, the basic step of our steepest descent can be written as follows

$$y_{k+1} = y_k + \|g_k\| \frac{g_k^T g_k}{g_k^T Q g_k} p_k$$

$$y_{k+1} = y_k + \|g_k\| \frac{g_k^T g_k}{g_k^T Q g_k} g_k \quad (2.5)$$

How much closer did this step bring us to the solution  $y^* = 0$ ? In other words, how much smaller is  $e(y_{k+1})$ , relative to the value  $e(y_k)$  at the previous step? The answer is, often not much, as we shall now prove. The arguments and proofs below are adapted from D. G. Luenberger Introduction to Linear and Nonlinear Programming, Addison-Wesley, 1973. From the definition of  $e$  and from equation we obtain

$$\frac{e(y_k) - e(y_{k+1})}{e(y_k)} = \frac{y_k^T Q y_k - y_{k+1}^T Q y_{k+1}}{y_k^T Q y_k}$$

$$= \frac{y_k^T Q y_k - (y_k - \frac{g_k^T g_k}{g_k^T Q g_k} g_k)^T Q (y_k - \frac{g_k^T g_k}{g_k^T Q g_k} g_k)}{y_k^T Q y_k}$$

$$\begin{aligned}
&= 2 \frac{g_k^T g_k}{g_k^T Q g_k} g_k^T Q y_k - \left( \frac{g_k^T g_k}{g_k^T Q g_k} \right)^2 g_k^T Q g_k \\
&\quad y_k^T Q y_k \\
&= \frac{2 g_k^T g_k g_k^T Q g_k - (g_k^T g_k)^2}{y_k^T Q y_k g_k^T Q g_k}
\end{aligned}$$

Since  $Q$  is invertible we have

$$\begin{aligned}
g_k &= Q y_k y_k^T Q^{-1} g_k \\
y_k^T Q y_k &= g_k^T Q^{-1} g_k
\end{aligned}$$

so that

$$\frac{e(y_k) - e(y_{k+1})}{e(y_k)} = \frac{(g_k^T g_k)^2}{g_k^T Q^{-1} g_k g_k^T Q g_k}$$

This can be rewritten as follows by rearranging terms:

$$e(y_{k+1}) = \left( 1 - \frac{(g_k^T g_k)^2}{g_k^T Q g_k} \right) e(y_k) \quad (2.6)$$

so if we can bound the expression in parentheses we have a bound on the rate of convergence of steepest descent. To this end, we introduce the following result.

**Lemma 2.1 (Kantorovich inequality)** Let  $Q$  be a positive definite, symmetric,  $n \times n$  matrix. For any vector  $y$  there holds

$$\frac{(y^T y)^2}{y^T Q^{-1} y y^T Q y} \geq \frac{4\sigma_1 \sigma_n}{(\sigma_1 + \sigma_n)^2}$$

where  $\sigma_1$  and  $\sigma_n$  are respectively, the largest and smallest singular values of  $Q$ .



**Proof.** Let

$$Q = U \Sigma U^T$$

be the singular value decomposition of the symmetric (hence  $V = U$ ) matrix  $Q$ . Because  $Q$  is positive definite, all its singular values are strictly positive, since the smallest of them satisfies

$$\sigma_n = \min_{\|y\|=1} y^T Q y > 0$$

by the definition of positive definiteness. If we let

$$z = U^T y$$

we have

$$\begin{aligned} \frac{(y^T Q^{-1} y)^2}{y^T Q^{-1} y y^T Q y} &= \frac{(y^T U^T U y)^2}{y^T U \Sigma^{-1} U^T y y^T U \Sigma U^T y} = \frac{(z^T z)^2}{z^T \Sigma^{-1} z z^T \Sigma z} \\ &= \frac{\sum_{i=1}^n \theta_i \sigma_i}{\sum_{i=1}^n \theta_i / \sigma_i} = \frac{\phi(\sigma)}{\varphi(\sigma)} \quad (2.7) \end{aligned}$$

Where the coefficients

$$\theta_i = \frac{z_i^2}{\|z\|^2}$$

Add up to one,

$$\sigma = \sum_{i=1}^n \theta_i \sigma_i \quad (2.8)$$

then the numerator  $\phi(\sigma)$  is  $1/\sigma$ . Of course, there are many ways to choose the coefficients  $\theta_i$  to obtain a particular value of  $\sigma$ . However, each of the singular values  $\sigma_j$  can be obtained by letting  $\theta_j = 1$  and all other  $\theta_i$  to zero. Thus, the values  $1/\sigma_j$  for  $j=1, \dots, n$  are all on the curve  $1/\sigma$ . The denominator  $\varphi(\sigma)$  is a convex combination of points on this curve. Since  $1/\sigma$  is a convex function of  $\sigma$ , the values of the denominator must be in  $\varphi(\sigma)$  the shaded area in figure. This area is delimited from above by the straight line that connects point  $(\sigma_1, 1/\sigma_1)$  with point  $(\sigma_n, 1/\sigma_n)$ , that is, by the line with ordinate

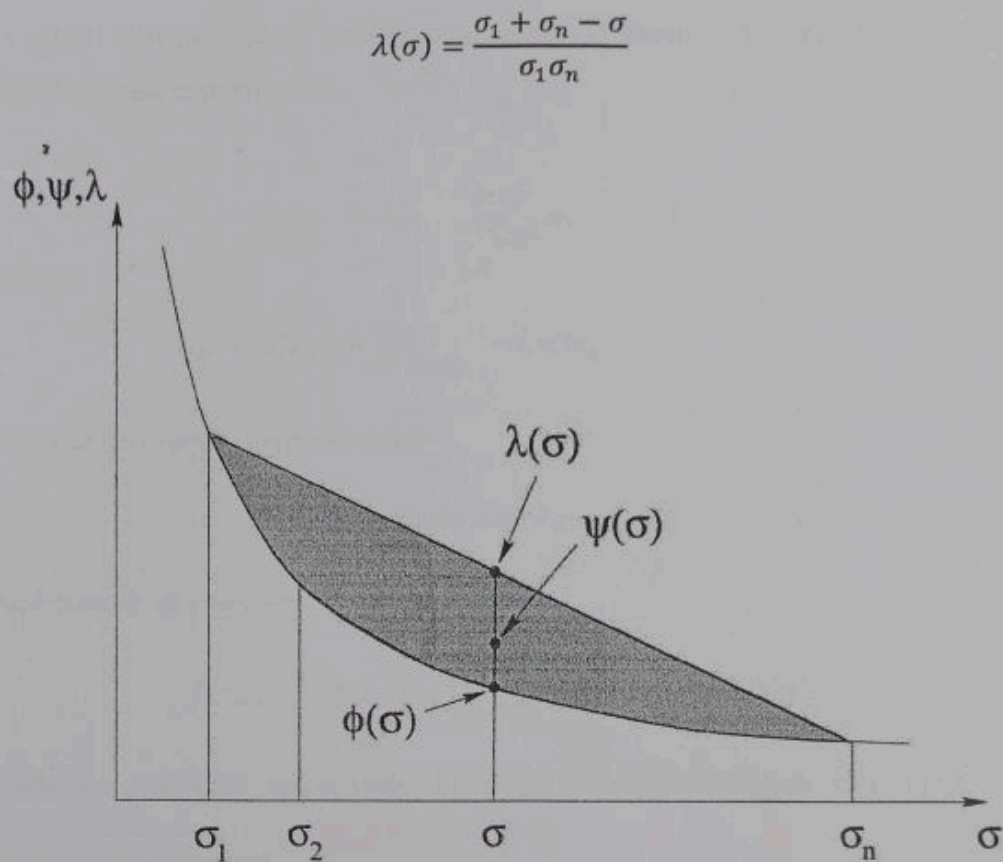


Figure 1.1 Kantorovich inequality.

For the same vector of coefficients  $\theta_i$  the values of  $\phi(\sigma)$ ,  $\psi(\sigma)$  and  $\lambda(\sigma)$  are on the vertical line corresponding to the value of given by (4.8). Thus an appropriate bound is

$$\frac{\phi(\sigma)}{\psi(\sigma)} \geq \min_{\sigma_1 \leq \sigma \leq \sigma_n} \frac{\phi(\sigma)}{\lambda(\sigma)} = \min_{\sigma_1 \leq \sigma \leq \sigma_n} \frac{1/\sigma}{(\sigma_1 + \sigma_n - \sigma)/(\sigma_1 \sigma_n)}$$

The minimum is achieved at  $\sigma = (\sigma_1 + \sigma_n)/2$ , yielding the desired result.

Thanks to this lemma, we can state the main result on the convergence of the method of steepest descent.

**Theorem 2.2** Let

$$f(x) = c + a^T x + \frac{1}{2} x^T Q x$$

be a quadratic function of  $x$ , with  $Q$  symmetric and positive definite. For any  $x_0$ , the method of steepest descent

$$x_{k+1} = x_k - \frac{g_k^T g_k}{g_k^T Q g_k} g_k \quad (2.9)$$

Where  $g_k = g(x_k) = \frac{\partial f}{\partial x} \Big|_{x=x_k} = a + Qx_k$

converges to the unique minimum point

$$x^* = -Q^{-1}a$$

of  $f$ . Furthermore, at every step  $k$  there hold

$$f(x_{k+1}) - f(x^*) \leq \left( \frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n} \right)^2 (f(x_k) - f(x^*))$$

where  $\sigma_1$  and  $\sigma_n$  are, respectively, the largest and smallest singular value of  $Q$ .

**Proof.** From the definition

$$y = x - x^* \quad \text{and} \quad e(y) = \frac{1}{2} y^T Q y \quad (2.10)$$

we immediately obtain the expression for steepest descent in terms of  $f$  and  $x$ . By equations (2.3) and (2.6) and the Kantorovich inequality we obtain

$$f(x_{k+1}) - f(x^*) = e(y_{k+1}) = \left( 1 - \frac{(g_k^T g_k)^2}{g_k^T Q^{-1} g_k g_k^T Q g_k} \right) e(y_k) \leq \left( 1 - \frac{4\sigma_1 \sigma_n}{(\sigma_1 + \sigma_n)^2} \right) e(y_k) \quad (2.11)$$

$$= \left( \frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n} \right)^2 (f(x_k) - f(x^*)) \quad (2.12)$$



Since the ratio in the last term is smaller than one, it follows immediately that  $f(x_k) - f(x^*) \rightarrow 0$  since the minimum of  $f$  is unique, that  $x_k \rightarrow x^*$ .

The ratio  $K(Q) = \sigma_1/\sigma_n$  is called the condition number of  $Q$ . The larger the condition number, the closer the fraction  $\sigma_1 - \sigma_n / \sigma_1 + \sigma_n$  is to unity, and the slower convergence. It is easily seen why this happens in the case in which  $x$  is a two-dimensional vector, as in figure. which shows the trajectory  $x_k$  superimposed on a set of isocontour of  $f(x)$ .

There is one good, but very precarious case, namely, when the starting point  $x_0$  is at one apex (tip of either axis) of an isocontour ellipse. In that case, one iteration will lead to the minimum  $x^*$ . In all other cases, the line in the direction  $p_k$  of steepest descent, which is orthogonal to the isocontour at  $x_k$  will not pass through  $x^*$ . The minimum of  $f$  along that line is tangent to some other, lower isocontour. The next step is orthogonal to the latter isocontour (that is, parallel to the gradient). Thus, at every step the steepest descent trajectory is forced to make a ninety-degree turn. If isocontours were circles ( $1 = n$ ) centered at  $x^*$ , then the first turn would make the new direction point to  $x^*$ , and minimization would get there in just one more step. This case, in which  $K(Q) = 1$ , is consistent with our analysis, because then

$$\frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n} = 0$$

The more elongated the isocontours, that is, the greater the condition number ( $Q$ ), the farther away a line orthogonal to an isocontour passes from  $x^*$ , and the more steps are required for convergence.



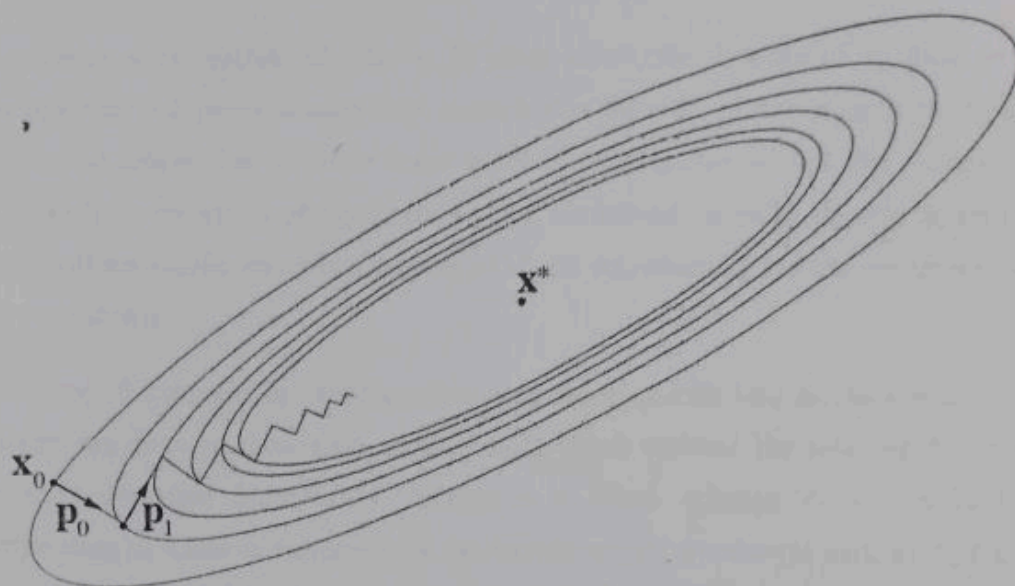


Figure 1.2: Trajectory of steepest descent.

For general (that is, non-quadratic)  $f$ , the analysis above applies once  $x_k$  gets close enough to the minimum, so that  $f$  is well approximated by a paraboloid. In this case,  $Q$  is the matrix of second derivatives of  $f$  with respect to  $x$ , and is called the Hessian of  $f$ . In summary, steepest descent is good for functions that have a well conditioned Hessian near the minimum, but can become arbitrarily slow for poorly conditioned Hessians. To characterize the speed of convergence of different minimization algorithms, we introduce the notion of the order of convergence. This is defined as the largest value of  $q$  for which the

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}$$

is definite. If  $\beta$  is this limit, then close to the solution (that is, for large values of  $k$ ) we have

$$\|x_{k+1} - x^*\| \approx \beta \|x_k - x^*\|^q$$

for a minimization method of order  $q$ . In other words, the distance of  $x_k$  from  $x^*$  is reduced by the  $q$ -th power at every step, so the higher the order of convergence, the better. The steepest descent has at best a linear order of convergence. In fact, the residuals  $|f(x_k) - f(x^*)|$  in the values of the function being minimized converge linearly. Since the gradient approaches zero when  $x_k$  tends to  $x^*$ , the arguments  $x_k$  to  $f$  can converge to  $x^*$  even more slowly.

To complete the steepest descent algorithm we need to specify how to check whether a minimum has been reached. One criterion is to check whether the value of  $f(x_k)$  has significantly decreased from  $f(x_{k-1})$ . Another is to check whether  $x_k$  is significantly different from  $x_{k-1}$ . Close to the minimum, the derivatives of  $f$  are close to zero, so if  $|f(x_k) - f(x_{k-1})|$  may be very small but  $\|x_k - x_{k-1}\|$  may still be relatively large. Thus, the check on  $x_k$  is more stringent, and therefore preferable in most cases. In fact, usually one is interested in the value of  $x^*$ , rather than in that of  $f(x^*)$ . In summary, the steepest descent algorithm can be stopped when

$$\|x_k - x^*\| < \epsilon$$

where the positive constant  $\epsilon$  is provided by the user.

In our analysis of steepest descent, we used the Hessian  $Q$  in order to compute the optimal step size. We used  $Q$  because it was available, but its computation during steepest descent would in general be overkill. In fact, only gradient information is necessary to find  $p_k$ , and a line search in the direction of  $p_k$  can be used to determine the step size  $\alpha_k$ . In contrast, the Hessian of  $f(x)$  requires computing  $\frac{n^2}{2}$  second derivatives if  $x$  is an  $n$ -dimensional vector.

Using line search to find  $\alpha_k$  guarantees that a minimum in the direction  $p_k$  is actually reached even when the parabolic approximation is inadequate. Here is how line search works.



Let

$$h(\alpha) = f(x_k + \alpha p_k) \quad (2.13)$$

be the scalar function of one variable that is obtained by restricting the function  $f$  to the line through the current point  $x_k$  in the direction of  $p_k$ . Line search first determines two points  $a, c$  that bracket the desired minimum  $\alpha_k$ , in the sense that  $a \leq \alpha_k \leq c$ , and then picks a point between  $a$  and  $c$ , say,  $b = (a+c)/2$ . The only difficulty here is to find  $c$ . In fact, we can set  $a = 0$ , corresponding through equation (2.13) to the starting point. A point  $c$  that is on the opposite side of the minimum with respect to  $a$  can be found by increasing

through values  $\alpha_1 = a, \alpha_2, \dots$  until  $\alpha_i$  is greater than  $\alpha_{i-1}$ . Then, if we can assume that  $h$  is convex between  $\alpha_1$  and  $\alpha_i$ , we can set  $c = \alpha_i$ . In fact, the derivative of  $h$  at  $a$  is negative, so the function is initially decreasing, but it is increasing between  $\alpha_{i-1}$  and  $\alpha_i = c$ , so the minimum must be somewhere between  $a$  and  $c$ . Of course, if we cannot assume convexity, we may find the wrong minimum, but there is no general-purpose x to this problem.

Line search now proceeds by shrinking the bracketing triple  $(a, b, c)$  until  $c-a$  is smaller than the desired accuracy in determining  $\alpha_k$ . Shrinking works as follows:

if  $b - a > c - b$

$u = (a + b)/2$

if  $f(u) > f(b)$

$(a, b, c) = (u, b, c)$

otherwise

$(a, b, c) = (a, u, b)$

end

otherwise

$$u = (b + c)/2$$

$$\text{if } f(u) > f(b)$$

$$(a, b, c) = (a, b, u)$$

otherwise

$$(a, b, c) = (b, u, c)$$

end

end.

It is easy to see that in each case the bracketing triple  $(a, b, c)$  preserves the property that  $f(b) \leq f(a)$  and  $f(b) \leq f(c)$ , and therefore the minimum is somewhere between  $a$  and  $c$ . In addition, at every step the interval  $(a, c)$  shrinks to  $3/4$  of its previous size, so line search will find the minimum in a number of steps that is logarithmic in the desired accuracy.

## 2.2 Newton's Method

If a function can be well approximated by a paraboloid in the region in which minimization is performed, the analysis in the previous section suggests a straight-forward  $x$  to the slow convergence of steepest descent. In fact, equation (2.2) tells us how to jump in one step from the starting point  $x_0$  to the minimum  $x^*$ . Of course, when  $f(x)$  is not exactly a paraboloid, the new value  $x_1$  will be different from  $x^*$ . Consequently, iterations are needed, but convergence can be expected to be faster. This is the idea of Newton's method, which we now summarize. Let

$$f(x_k + \Delta x) \approx f(x_k) + g_k^T \Delta x + \frac{1}{2} \Delta x^T Q_k \Delta x_k \quad (2.14)$$

be the first terms of the Taylor series expansion of  $f$  about the current point  $x_k$ , where

$$g_k = g(x_k) = \left. \frac{\partial f}{\partial x} \right|_{x=x_k}$$



$$Q_k = Q(x_k) = \left. \frac{\partial^2 f}{\partial x \partial x^T} \right|_{x=x_k} = \begin{Bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{Bmatrix}$$

are the gradient and Hessian of  $f$  evaluated at the current point  $x_k$ . Notice that even when  $f$  is a paraboloid, the gradient  $g_k$  is different from  $a$  as used in equation (2.1). In fact,  $a$  and  $Q$  are the coefficient of the Taylor expansion of  $f$  around point  $x = 0$ , while  $g_k$  and  $Q_k$  are the coefficients of the Taylor expansion of  $f$  around the current point  $x_k$ . In other words, gradient and Hessian are constantly reevaluated in Newton's method.

To the extent that approximation is valid, we can set the derivatives of  $f(x_k + \Delta x)$  with respect to  $x$  to zero, and obtain, analogously to equation, the linear system

$$Q_k \Delta x = -g_k \quad (2.15)$$

whose solution  $\Delta x_k = \alpha_k p_k$  yields at the same time the step direction  $p_k = \Delta x_k / \|\Delta x_k\|$  and the step size  $\alpha_k = \|\Delta x_k\|$ . The direction is of course undened once the algorithm has reached a minimum, that is, when  $\alpha_k = 0$ .

A minimization algorithm in which the step direction  $p_k$  and size  $\alpha_k$  are defined in this manner is called *Newton's method*. The corresponding  $p_k$  is termed the Newton direction, and the step defined by equation (2.15) is the Newton step.

The greater speed of Newton's method over steepest descent is borne out by analysis: while steepest descent has a linear order of convergence, Newton's method is quadratic. In fact, let

$$y(x) = x - Qx^{-1}g(x)$$

be the place reached by a Newton step starting at  $x$  (see equation (2.15)) and suppose that at the minimum  $x^*$  the Hessian  $Q(x^*)$  is nonsingular. Then

$$y(x^*) = x^*$$

because  $g(x^*) = 0$ , and

$$x_{k+1} - x^* = y(x_k) - y(x^*) = y(x_k) - y(x^*)$$

From the mean-value theorem, we have

$$\|x_{k+1} - x^*\| = \|y(x_k) - y(x^*)\| \leq \left\| \left[ \frac{\partial y}{\partial x^T} \right] (x_k - x^*) + \frac{1}{2} \left\| \frac{\partial^2 y}{\partial x \partial x^T} \right\| \|x_k - x^*\|^2 \right\|$$

where  $x$  is some point on the line between  $x^*$  and  $x_k$ . Since  $y(x^*) = x^*$ , the first derivatives of  $y$  at  $x^*$  are zero, so that the first term in the right-hand side above vanishes, and

$$\|x_{k+1} - x^*\| \leq c \|x_k - x^*\|^2$$

where  $c$  depends on third-order derivatives of  $f$  near  $x^*$ . Thus, the convergence rate of Newton's method is of order at least two.



## CHAPTER 3

### Eigenvectors

Eigenvectors  $\vec{x}$  and their corresponding eigenvalues  $\lambda$  of a square matrix  $A$  are determined by the equation  $A \vec{x} = \lambda \vec{x}$ . There are many ways to see that this problem is nonlinear. For instance, there is a product of unknowns  $\lambda$  and  $\vec{x}$ , and to avoid the trivial solution  $\vec{x} = \vec{0}$  we constrain  $\|\vec{x}\| = 1$ ; this constraint is circular rather than linear. Thanks to this structure, our methods for finding eigenspaces will be considerably different from techniques for solving and analyzing linear systems of equations.

#### 3.1 Statistics

Suppose we have machinery for collecting several statistical observations about a collection of items. For instance, in a medical study we may collect the age, weight, blood pressure, and heart rate of 100 patients. Then, each patient  $i$  can be represented by a point  $\vec{x}_i$  in  $\mathbb{R}^4$  storing these four values.

Of course, such statistics may exhibit strong correlation. For instance, patients with higher blood pressure may be likely to have higher weights or heart rates. For this reason, although we collected our data in  $\mathbb{R}^4$ , in reality it may—to some approximate degree—live in a lower dimensional space better capturing the relationships between the different variables.

For now, suppose that in fact there exists a one-dimensional space approximating our dataset. Then, we expect all the data points to be nearly parallel to some vector  $\vec{v}$ , so that each can be written as  $\vec{x}_i \approx c_i \vec{v}$  for different  $c_i \in \mathbb{R}$ . From before, we know that the best approximation of  $\vec{x}_i$  parallel to  $\vec{v}$  is  $\text{proj}_{\vec{v}} \vec{x}_i$ :

$$\begin{aligned} \text{proj}_{\vec{v}} \vec{x}_i &= \frac{\vec{x}_i \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v} \text{ by definition} \\ &= (\vec{x}_i \cdot \hat{v}) \hat{v} \text{ since } \vec{v} \cdot \vec{v} = \|\vec{v}\|^2 \end{aligned}$$



Here, we define  $\hat{v} = \vec{v} / \|\vec{v}\|$ . Of course, the magnitude of  $\vec{v}$  does not matter for the problem at hand, so it is reasonable to search over the space of unit vectors  $v$ .

Following the pattern of least squares, we have a new optimization problem:

$$\begin{aligned} &\text{minimize } \sum_i \|\vec{x}_i - \text{proj}_{\hat{v}} \vec{x}_i\|^2 \\ &\text{such that } \|\hat{v}\| = 1 \end{aligned}$$

We can simplify our optimization objective a bit:

$$\begin{aligned} \sum_i \|\vec{x}_i - \text{proj}_{\hat{v}} \vec{x}_i\|^2 &= \sum_i \|\vec{x}_i - (\vec{x}_i \cdot \hat{v}) \hat{v}\|^2 \quad \text{by definition of projection} \\ &= \sum_i (\|\vec{x}_i\|^2 - (\vec{x}_i \cdot \hat{v})^2) \quad \text{since } \|\hat{v}\| = 1 \text{ and } \|\vec{w}\|^2 \\ &\quad = \vec{w} \cdot \vec{w} \\ &= \text{const.} - \sum_i (\vec{x}_i \cdot \hat{v})^2 \end{aligned}$$

This derivation shows that we can solve an equivalent optimization problem:

$$\begin{aligned} &\text{maximize } \|X^T \hat{v}\|^2 \\ &\text{such that } \|\hat{v}\|^2 = 1, \end{aligned}$$

where the columns of  $X$  are the vectors  $\vec{x}_i$ . Notice that  $\|X^T \hat{v}\|^2 = \hat{v}^T X X^T \hat{v}$ , so by Example

the vector  $\hat{v}$  corresponds to the eigenvector of  $XX^T$  with the highest eigenvalue. The vector  $\hat{v}$  is known as the first principal component of the dataset.

### 3.1.1 Differential Equations

Many physical forces can be written as functions of position. For instance, the force between two particles at positions  $\vec{x}$  and  $\vec{y}$  in  $\mathbb{R}^3$  exerted by a spring can be written as  $k(\vec{x} - \vec{y})$  by Hooke's Law; such spring forces are used to approximate forces holding cloth together in many simulation systems. Although these forces are not necessarily linear in position, we often approximate them in a linear fashion. In particular, in a physical system with  $n$  particles encode the positions of all the particles simultaneously in a vector  $X \in \mathbb{R}^{3n}$ . Then, if we assume such an approximation we can write that the forces in the system are approximately  $\vec{F} \approx A \vec{X}$  for some matrix  $A$ .

Recall Newton's second law of motion  $F = ma$ , or force equals mass times acceleration. In our context, we can write a diagonal mass matrix  $M \in \mathbb{R}^{3n \times 3n}$  containing the mass of each particle in the system. Then, we know  $\vec{F} = M\vec{X}''$ , where prime denotes differentiation in time. Of course,  $\vec{X}'' = (\vec{X}')'$ , so in the end we have a first-order system of equations:

$$\frac{d}{dt} \begin{pmatrix} \vec{X} \\ \vec{V} \end{pmatrix} = \begin{pmatrix} 0 & I_{3n \times 3n} \\ M^{-1}A & 0 \end{pmatrix} \begin{pmatrix} \vec{X} \\ \vec{V} \end{pmatrix}$$

Here, we simultaneously compute both positions in  $\vec{X} \in \mathbb{R}^{3n}$  and velocities  $\vec{V} \in \mathbb{R}^{3n}$  of all  $n$  particles as functions of time.

More generally, differential equations of the form  $\vec{x}' = A\vec{x}$  appear in many contexts, including simulation of cloth, springs, heat, waves, and other phenomena. Suppose we know eigenvectors  $\vec{x}_1, \dots, \vec{x}_k$  of  $A$ , such that  $A\vec{x}_i = \lambda_i\vec{x}_i$ . If we write the initial condition of the differential equation in terms of the eigenvectors, as

$$\vec{x}(0) = c_1\vec{x}_1 + \dots + c_k\vec{x}_k,$$

then the solution to the equation can be written in closed form:

$$\vec{x}(t) = c_1 e^{\lambda_1 t} \vec{x}_1 + \dots + c_k e^{\lambda_k t} \vec{x}_k,$$

This solution is easy to check by hand. That is, if we write the initial conditions of this differential equation in terms of the eigenvectors of  $A$ , then we know its solution for all times  $t \geq 0$  for free. Of course, this formula is not the end of the story for simulation: Finding the complete set of eigenvectors of  $A$  is expensive, and  $A$  may change over time.

### 3.2 Spectral Embedding

Suppose we have a collection of  $n$  items in a dataset and a measure  $w_{ij} \geq 0$  of how similar each pair of elements  $i$  and  $j$  are; we will assume  $w_{ij} = w_{ji}$ . For instance, maybe we are given a collection of photographs and use  $w_{ij}$  to compare the similarity of their color distributions. We might wish to sort the photographs based on their similarity to simplify viewing and exploring the collection.

One model for ordering the collection might be to assign a number  $x_i$  for each item  $i$ , asking that similar objects are assigned similar numbers. We can measure how well an assignment groups similar objects by using the energy



$$E(\vec{x}) = \sum_{ij} w_{ij} (x_i - x_j)^2.$$

That is,  $E(\vec{x})$  asks that items  $i$  and  $j$  with high similarity scores  $w_{ij}$  get mapped to nearby values.

Of course, minimizing  $E(\vec{x})$  with no constraints gives an obvious minimum:  $x_i = \text{const.}$  For all  $i$ . Adding a constraint  $\|\vec{x}\| = 1$  does not remove this constant solution! In particular, taking  $x_i = 1/\sqrt{n}$  for all  $i$  gives  $\|\vec{x}\| = 1$  and  $E(\vec{x}) = 0$  in an uninteresting way. Thus, we must remove this case as well:

$$\text{minimize } E(\vec{x})$$

$$\text{such that } \|\vec{x}\|^2 = 1$$

$$\vec{1} \cdot \vec{x} = 0$$

Notice that our second constraint asks that the sum of  $\vec{x}$  is zero.

Once again we can simplify the energy:

$$E(\vec{x}) = \sum_{ij} w_{ij} (x_i - x_j)^2$$

$$= \sum_{ij} w_{ij} (x_i^2 - 2x_i x_j + x_j^2)$$

$$= \sum_i a_i x_i^2 - 2 \sum_{ij} w_{ij} x_i x_j + \sum_j b_j x_j^2$$

$$\text{for } a_i \equiv \sum_j w_{ij} \text{ and } b_j \equiv \sum_i w_{ij}$$

$$= \vec{x}^T (A - 2W + B) \vec{x} \text{ where } \text{diag}(A) = \vec{a} \quad \text{and}$$

$$\text{diag}(B) = \vec{b}$$

$$= \vec{x}^T (2A - 2W) \vec{x} \text{ by symmetry of } W$$

It is easy to check that  $\vec{1}$  is an eigenvector of  $(2A - 2W)$  with eigenvalue 0. More interestingly, the eigenvector corresponding to the second-smallest eigenvalue corresponds to the solution of our minimization objective above!

### 3.3 Computation using Eigenvectors

Take  $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^n$  to be the unit-length eigenvectors of symmetric matrix  $A \in \mathbb{R}^{n \times n}$ . Suppose we wish to solve  $A\vec{y} = \vec{b}$ . We can write

$$\vec{b} = c_1 \vec{x}_1 + \dots + c_n \vec{x}_n,$$

where  $c_i = \vec{b} \cdot \vec{x}_i$  by orthonormality. It is easy to guess the following solution:

$$\vec{y} = \frac{c_1}{\lambda_1} \vec{x}_1 + \cdots + \frac{c_n}{\lambda_n} \vec{x}_n.$$

In particular, we find:

$$\begin{aligned} A\vec{y} &= A \left( \frac{c_1}{\lambda_1} \vec{x}_1 + \cdots + \frac{c_n}{\lambda_n} \vec{x}_n \right) \\ &= \frac{c_1}{\lambda_1} A \vec{x}_1 + \cdots + \frac{c_n}{\lambda_n} A \vec{x}_n \\ &= c_1 \vec{x}_1 + \cdots + c_n \vec{x}_n \\ &= \vec{b}, \text{ as desired.} \end{aligned}$$

The calculation above is both a positive and negative result. It shows that given the eigenvectors of symmetric  $A$ , operations like inversion are straightforward. On the flip side, this means that finding the full set of eigenvectors of a symmetric matrix  $A$  is "at least" as difficult as solving  $\vec{x} = \vec{b}$ .

### 3.4 Specialized Properties

#### Characteristic Polynomial

Recall that the determinant of a matrix  $\det A$  satisfies the relationship that  $\det A \neq 0$  if and only if  $A$  is invertible. Thus, one way to find eigenvalues of a matrix is to find roots of the characteristic polynomial

$$p_A(\lambda) = \det A (A - \lambda I_{n \times n}).$$

We will not define determinants in our discussion here, but simplifying  $p_A$  reveals that it is an  $n$ -th degree polynomial in  $\lambda$ . This provides an alternative reason why there are at most  $n$  distinct eigenvalues, since there are at most  $n$  roots of this function.

#### Jordan Normal Form

We can only diagonalize a matrix when it has a full eigenspace. All matrices, however, are similar to a matrix in Jordan normal form, which has the following form:

- Nonzero values are on the diagonal entries  $a_{ii}$  and on the "superdiagonal"  $a_{i(i+1)}$ .



- Diagonal values are eigenvalues repeated as many times as their multiplicity; the matrix is block diagonal about these clusters.
- Off-diagonal values are 1 or 0.

Thus, the shape looks something like the following

$$\begin{pmatrix} \lambda_1 & 1 & & & \\ & \lambda_1 & 1 & & \\ & & \lambda_1 & & \\ & & & \lambda_2 & 1 \\ & & & & \lambda_2 \\ & & & & & \lambda_3 \\ & & & & & & \ddots \end{pmatrix}$$

Jordan normal form is attractive theoretically because it always exists, but the 1/0 structure is discrete and unstable under numerical perturbation.

### 3.5 Computing Eigenvalues

The computation and estimation of the eigenvalues of a matrix is a well-studied problem with many potential solutions. Each solution is tuned for a different situation, and achieving maximum conditioning or speed requires experimentation with several techniques. Here, we cover a few of the most popular and straightforward solutions to the eigenvalue problem frequently encountered in practice.

#### 3.5.1 Power Iteration

For now, suppose that  $A \in \mathbb{R}^{n \times n}$  is symmetric. Then, by the spectral theorem we can write eigenvectors  $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^n$ ; we sort them such that their corresponding eigenvalues satisfy  $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$ .

Suppose we take an arbitrary vector  $\vec{v}$ . Since the eigenvectors of  $A$  span  $\mathbb{R}^n$ , we can write:

$$\vec{v} = c_1 \vec{x}_1 + \dots + c_n \vec{x}_n.$$

Then,

$$\begin{aligned} A\vec{v} &= c_1 A \vec{x}_1 + \dots + c_n A \vec{x}_n \\ &= c_1 \lambda_1 \vec{x}_1 + \dots + c_n \lambda_n \vec{x}_n \text{ since } A\vec{x}_i = \lambda_i \vec{x}_i \\ &= \lambda_1 \left( c_1 \vec{x}_1 + \frac{\lambda_2}{\lambda_1} c_2 \vec{x}_2 + \dots + \frac{\lambda_n}{\lambda_1} c_n \vec{x}_n \right) \end{aligned}$$

$$A^2 \vec{v} = \lambda_1^2 \left( c_1 \vec{x}_1 + \left( \frac{\lambda_2}{\lambda_1} \right)^2 c_2 \vec{x}_2 + \dots + \left( \frac{\lambda_n}{\lambda_1} \right)^2 c_n \vec{x}_n \right)$$

$\vdots$

$$A^k \vec{v} = \lambda_1^k \left( c_1 \vec{x}_1 + \left( \frac{\lambda_2}{\lambda_1} \right)^k c_2 \vec{x}_2 + \dots + \left( \frac{\lambda_n}{\lambda_1} \right)^k c_n \vec{x}_n \right)$$

Notice that as  $k \rightarrow \infty$ , the ratio  $(\lambda_i/\lambda_1)^k \rightarrow 0$  unless  $\lambda_i = \lambda_1$ , since  $\lambda_1$  has the largest magnitude of any eigenvalue by definition. Thus, if  $\vec{x}$  is the projection of  $\vec{v}$  onto the space of eigenvectors with eigenvalues  $\lambda_1$ , then as  $k \rightarrow \infty$  the following approximation holds more and more exactly:

$$A^k \vec{v} \approx \lambda_1^k \vec{x}.$$

This observation leads to an exceedingly simple algorithm for computing an eigenvector  $\vec{x}$  of  $A$  corresponding to the largest eigenvalue  $\lambda_1$ :

1. Take  $\vec{v}_1 \in \mathbb{R}^n$  to be an arbitrary nonzero vector.
2. Iterate until convergence for increasing  $k$ :

$$\vec{v}_k = A \vec{v}_{k-1}$$

This algorithm, known as power iteration, will produce vectors  $\vec{v}_k$  more and more parallel to the desired  $\vec{x}_1$ . It is guaranteed to converge, even when  $A$  is asymmetric, although the proof of this fact is more involved than the derivation above. The one time that this technique may fail is if we accidentally choose  $\vec{v}_1$  such that  $c_1 = 0$ , but the odds of this occurring are slim to none.

Of course, if  $|\lambda_1| > 1$ , then  $\|\vec{v}_k\| \rightarrow \infty$  as  $k \rightarrow \infty$ , an undesirable property for floating point arithmetic. Recall that we only care about the direction of the eigenvector rather than its magnitude, so scaling has no effect on the quality of our solution. Thus, to avoid this divergence situation we can simply normalize at each step, producing the normalized power iteration algorithm:

$$\vec{w}_k = A \vec{v}_{k-1}$$

$$\vec{v}_k = \frac{\vec{w}_k}{\|\vec{w}_k\|}$$

Notice that we did not decorate the norm  $\|\cdot\|$  with a particular subscript. Mathematically, any norm will suffice for preventing the divergence issue, since we have shown that all norms on  $\mathbb{R}^n$  are equivalent. In practice, we often use the infinity norm  $\|\cdot\|_\infty$ ; in this case it is easy to check that  $\|\vec{w}_k\| \rightarrow |\lambda_1|$ .

### 3.5.2 Inverse Iteration

We now have a strategy for finding the largest-magnitude eigenvalue  $\lambda_1$ . Suppose  $A$  is invertible, so that we can evaluate  $\vec{y} = A^{-1}\vec{v}$  by solving  $A\vec{y} = \vec{v}$ .

If  $A\vec{x} = \lambda\vec{x}$ , then  $\vec{x} = \lambda A^{-1}\vec{x}$ , or equivalently

$$A^{-1}\vec{x} = \frac{1}{\lambda}\vec{x}$$

Thus, we have shown that  $\frac{1}{\lambda}$  is an eigenvalue of  $A^{-1}$  with eigenvector  $\vec{x}$ . Notice that if  $|a| \geq |b|$  then  $|b|^{-1} \geq |a|^{-1}$  for any  $a, b \in \mathbb{R}$ , so the smallest-magnitude eigenvalue of  $A$  is the largest-magnitude eigenvector of  $A^{-1}$ . This observation yields a strategy for finding  $\lambda_n$  rather than  $\lambda_1$  called inverse power iteration:

1. Take  $\vec{v}_1 \in \mathbb{R}^n$  to be an arbitrary nonzero vector.

2. Iterate until convergence for increasing  $k$ :

(a) Solve for  $\vec{w}_k : A\vec{w}_k = \vec{v}_{k-1}$

(b) Normalize:  $\vec{v}_k = \frac{\vec{w}_k}{\|\vec{w}_k\|}$

We repeatedly are solving systems of equations using the same matrix  $A$ , which is a perfect application of factorization techniques. For instance, if we write  $A = LU$ , then we could formulate an equivalent but considerably more efficient version of inverse power iteration:

1. Factor  $A = LU$

2. Take  $\vec{v}_1 \in \mathbb{R}^n$  to be an arbitrary nonzero vector.

3. Iterate until convergence for increasing  $k$ :

(a) Solve for  $\vec{y}_k$  by forward substitution:  $L\vec{y}_k = \vec{v}_{k-1}$

(b) Solve for  $\vec{w}_k$  by back substitution:  $U\vec{w}_k = \vec{y}_k$



(c) Normalize:  $\vec{v}_k = \frac{\vec{w}_k}{\|\vec{w}_k\|}$

### 3.6 Sensitivity and Conditioning

As warned, we have only outlined a few eigenvalue techniques out of a rich and long-standing literature. Almost any algorithmic technique has been experimented with for finding spectra, from iterative methods to root-finding on the characteristic polynomial to methods that divide matrices into blocks for parallel processing.

Just as in linear solvers, we can evaluate the conditioning of an eigenvalue problem independently of the solution technique. This analysis can help understand whether a simplistic iterative scheme will be successful for finding the eigenvectors of a given matrix or if more complex methods are necessary; it is important to note that the conditioning of an eigenvalue problem is not the same as the condition number of the matrix for solving systems, since these are separate problems.

Suppose a matrix  $A$  has an eigenvector  $\vec{x}$  with eigenvalue  $\lambda$ . Analyzing the conditioning of the eigenvalue problem involves analyzing the stability of  $\vec{x}$  and  $\lambda$  to perturbations in  $A$ . To this end, we might perturb  $A$  by a small matrix  $\delta A$ , thus changing the set of eigenvectors. In particular, we can write eigenvectors of  $A + \delta A$  as perturbations of eigenvectors of  $A$  by solving the problem

$$(A + \delta A)(\vec{x} + \delta \vec{x}) = (\lambda + \delta \lambda)(\vec{x} + \delta \vec{x}).$$

Expanding both sides yields:

$$A\vec{x} + A\delta\vec{x} + \delta A \cdot \vec{x} + \delta A \cdot \delta\vec{x} = \lambda\vec{x} + \lambda\delta\vec{x} + \delta\lambda \cdot \vec{x} + \delta\lambda \cdot \delta\vec{x}$$

Assuming  $\delta A$  is small, we will assume that  $\delta\vec{x}$  and  $\delta\lambda$  also are small. Products between these variables then are negligible, yielding the following approximation:

$$A\vec{x} + A\delta\vec{x} + \delta A \cdot \vec{x} \approx \lambda\vec{x} + \lambda\delta\vec{x} + \delta\lambda \cdot \vec{x}$$

Since  $A\vec{x} = \lambda\vec{x}$ , we can subtract this value from both sides to find:

$$A\delta\vec{x} + \delta A \cdot \vec{x} \approx \lambda\delta\vec{x} + \delta\lambda \cdot \vec{x}$$

We now apply an analytical trick to complete our derivation. Since  $A\vec{x} = \lambda\vec{x}$ , we know  $(A - \lambda I_{n \times n})\vec{x} = 0$ , so  $A - \lambda I_{n \times n}$  is not full rank. The transpose of a matrix is full-rank only if the matrix is full-rank, so we know  $(A - \lambda I_{n \times n})^T = A^T - \lambda I_{n \times n}$  also has a null space vector  $\vec{y}$ . Thus  $A^T\vec{y} = \lambda\vec{y}$  we can call  $\vec{y}$  the left eigenvector corresponding to  $\vec{x}$ . We can left-multiply our perturbation estimate above by  $\vec{y}^T$ :

$$\vec{y}^T(A\delta\vec{x} + \delta A \cdot \vec{x}) \approx \vec{y}^T(\lambda\delta\vec{x} + \delta\lambda \cdot \vec{x})$$

Since  $A^T\vec{y} = \lambda\vec{y}$ , we can simplify:

$$\vec{y}^T \delta A \cdot \vec{x} \approx \delta\lambda \vec{y}^T \vec{x}$$

Rearranging yields:

$$\delta\lambda \approx \frac{\vec{y}^T \delta A \cdot \vec{x}}{\vec{y}^T \vec{x}}$$

Assume  $\|\vec{x}\| = 1$  and  $\|\vec{y}\| = 1$ . Then, if we take norms on both sides we find:

$$|\delta\lambda| \approx \frac{\|\delta A\|_2}{|\vec{y} \cdot \vec{x}|}$$

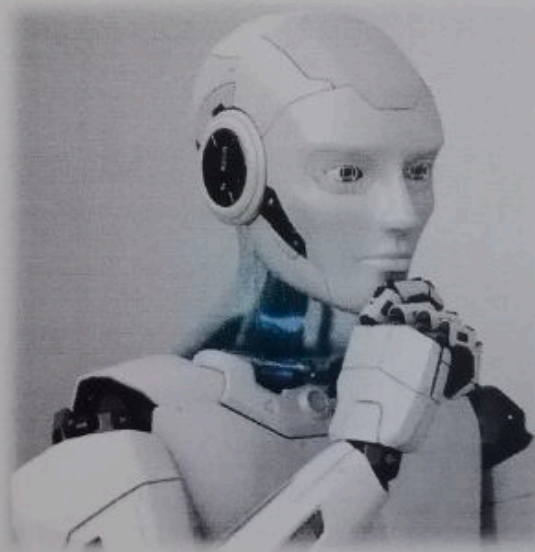
So in general conditioning of the eigenvalue problem depends on the size of the perturbation  $\delta A$  – as expected – and the angle between the left and right eigenvectors  $\vec{x}$  and  $\vec{y}$ . We can use  $1/|\vec{y} \cdot \vec{x}|$  as an approximate condition number. Notice that  $\vec{x} = \vec{y}$  when  $A$  is symmetric, yielding a condition number of 1; this reflects the fact that the eigenvectors of symmetric matrices are orthogonal and thus maximally separated.



## Conclusion

Mathematics perform a significant part in robotic modelling, planning and execution. The robotic scientists know how to deal with these complex equations and form a software framework to create more advance and functional Robotics of the century. It's essential to invest in research and development in robotics to have the most growth in the economy as it is estimated that 10 to 15 years from now, the world is going to be all Robotic.

Learning mathematics with robots helps us to visualize challenging real world applications and supports multiple representations of a problem. While we apply our knowledge of math concepts in solving real world problems with the help of robots, they develop a lasting hands on experience in a social context and a better attitude towards math education.



New ideas will evolve and innovative lessons will be developed in future that effectively use robot in the math instruction. Considering mathematical aspects in robotics the highly interdisciplinary character can be seen as well. Robotics is a field of research which yields practical results that may have a positive impact and benefit on society.

## Reference

- 1 . CS 205 , Mathematical Methods for Robotics and Vision ,Carlo Tomasi , Stanford University , Fall 1997 ,
- 2 . Mathematical Methods for Computer Vision , Robotics , and Graphics , Course Notes for CS 205A , Fall 2013 , Justin Solomon , Department of Computer Science, Stanford University .

# Queueing Theory

Project report submitted to

**ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**

Affiliated to

**MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELI**

In partial fulfillment of the requirement for the award of degree of

**Bachelor of science in Mathematics**

Submitted by

**NAMES**

**REG.NO.**

**ARUL MOZHI. M**

**18AUMT05**

**KALAIVANI. K**

**18AUMT29**

**LIRANCITHA. R**

**18AUMT33**

**SALINI. M**

**18AUMT44**

**SINI DIVYA. A**

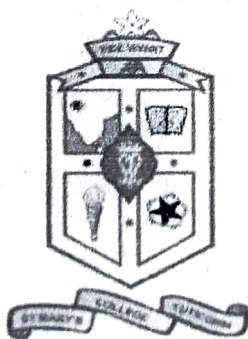
**18AUMT48**

Under the Guidance of

**Dr. A. PUNITHA THARANI, M.Sc., M.Phil., Ph.D.,**

Head and Associate Professor of Department of Mathematics

**ST. MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**



Department of Mathematics

**ST.MARY'S COLLEGE (AUTONOMOUS),**

Thoothukudi.

(2020-2021)

## CERTIFICATE

We hereby declare that the project report entitled "QUEUEING THEORY" being submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli in partial fulfillment for the award of degree of Bachelor of science in Mathematics and it is a record of work done during the year 2020-2021 by the following students:

ARUL MOZHI. M

18AUMT05

KALAIVANI. K

18AUMT29

LIRANCITHA . R

18AUMT33

SALINI. M

18AUMT44

SINI DIVYA. A

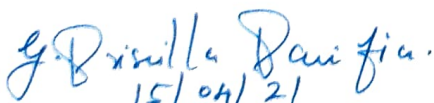
18AUMT48



Signature of the Guide



Signature of the HOD



Signature of the Examiner



Signature of the Principal

St. Mary's College (Autonomous)  
Thoothukudi - 628 001.



## DECLARATION

We hereby declare that the project report entitled "QUEUEING THEORY" is our original work. It has not been submitted to any university for any degree or diploma.

M. Arul Mozhi

(ARUL MOZHI. M)

K. Kalai Vani

(KALAIVANI. K)

R. Lirancitha

(LIRANCITHA. R)

M. Salini

(SALINI. M)

A. Sini Divya

(SINI DIVYA. A)

## **ACKNOWLEDGEMENT**

**First of all, we thank lord Almighty for showering his blessings to undergo this project.**

**With immense pleasure, we register our deep sense of gratitude to our guide Dr. A. Punitha Tharani M.Sc., M.Phil., Ph.D., Head of the department and Associate Professor of Mathematics, for having imported necessary guidelines throughout the period of our studies.**

**We thank our beloved Principal Rev. Dr. Sr. A.S.J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., for providing us the help to carry out our project work successfully.**

**Finally, we thank all those who extended their helping hands regarding this project.**

# INDEX

<i>Introduction.....</i>	<i>2 – 3</i>
<i>1.Queueing System.....</i>	<i>4–11</i>
1.1.Elements of a Queueing System	
1.1.2.Queue discipline	
1.1.3.Service mechanism	
1.1.4.Capacity of the system	
1.2.Operating characteristics	
1.3.Deterministic of Queueing system.	
<i>2.Probability Distribution in Queueing systems.....</i>	<i>12–18</i>
2.1.Distribution of Arrivals(Pure Birth Process)	
2.3.Distribution of Inter – Arrivals(Exponential Process)	
2.3.Distribution of Departures(Pure Death Process)	
2.4.Classification of Queueing Models	
2.5.Definition of Transient and Steady States.	
<i>3.Poisson Queueing systems.....</i>	<i>19–30</i>
3.1.Model $I(M/M/1):(\infty/FIFO)$	
3.1.1.Characteristics of Model I	
3.1.2.Waiting Time Distribution for Model I	
3.1.3.Characteristics of Waiting Time Distribution	
3.2.Relationships among Operating Characteristics	
Sample Problems	
<i>4.Non – Poisson Queueing Systems.....</i>	<i>31–38</i>
4.1.Erlangian Service Time Distribution with $k$ – phases	
4.2.Model $1\{M/E_k/1\}:(\infty/FIFO)$	
4.2.1.Characteristics of Model I	
Sample Problems	
<i>5.Cost Models in Queueing.....</i>	<i>39 – 43</i>
5.1.Optimum Service rate	
5.3.Queueing Control	
<i>Conclusion.....</i>	<i>44</i>
<i>Reference.....</i>	<i>45</i>

## 0. INTRODUCTION

A flow of customers from infinite/finite population towards the service facility forms a queue (waiting line) on account of lack of capability to serve them all at a time. The queues may be of persons waiting at a doctor's clinic or at railway booking office, these may be of machines waiting to be repaired or of ships in the harbour waiting to be unloaded or of letters arriving at a typist's desk. In the absence of a perfect balance between the service facilities and the customers, waiting is required either of the service facilities or for the customer's arrival.

By the term '*customer*' we mean the arriving unit that requires some service to be performed. The customer may be of persons, machines, vehicles, parts, etc.

***Queues (waiting line)*** stands for a number of customers waiting to be serviced. The queue does not include the customer to be serviced. The process or system that performs the services to the customer is termed by *service channel or service facility*.

The subject of queueing is not directly concerned with optimization (maximization or minimization). Rather, it attempts to explore, understand, and compare various queueing situations and thus indirectly achieves optimization approximately.

Queueing Problem Examples

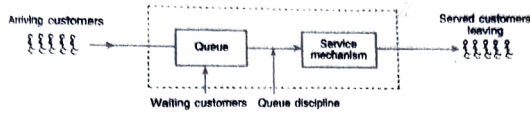


Figure 1:

Problem	Customers	Service facility
Determining the number of attendants required at a petrol station	Auto mobiles	Petrol station attendants
Scheduling of patients in a clinic	Patients	Doctors
Determining the number of runways at an airport	Aircrafts	Runways
Determining the size of a parking lot	Automobiles	Parking spaces
Determining the number of taxi cabs for a fleet	Public	Taxi cabs
Determining the capacity of a motel	Motorists	Lodging facilities
Determining the service rate for a harbour	Ships	Harbour

\*\*\*\*\*



## 1. QUEUEING SYSTEM

The mechanism of a queueing process is very simple. Customers arrive at a service counter and are attended to by one or more of the servers. As soon as a customer is served, it departs from the system. Thus a queueing system can be described as consisting of customers arriving for service, waiting for service if it is not immediate, and leaving the system after being served.

The general framework of a queueing system is shown in figure 1.

### 1.1 ELEMENTS OF A QUEUEING SYSTEM

The basic elements of a queueing system are as follows :

#### 1.1.1 Input (or Arrival) process.

This element of queueing system is concerned with the pattern in which the customers arrive for service. Input source can be described by following three factors:

##### (a)Size of the queue.

If the total number of potential customers requiring service are only few, then size of the input source is said to be *finite*. On the other hand, if potential customers requiring service are sufficiently large in number, then the input source is considered to be *infinite*.

Also, the customers may arrive at the service facility in batches of fixed size or of variable size or one by one. In the case when more than one arrival is allowed to enter the system simultaneously (entering the system does not necessarily mean entering into service), the input is said to occur in *bulk* or in *batches*. Ships discharging cargo at a dock, families visiting restaurants, etc. are the examples of bulk arrivals.

##### (b)Pattern of arrivals.

Customers may arrive in the system at known (regular or otherwise) times, or they may arrive in a random way. In case the arrival times are known with certainty, the queueing problems are categorized as deterministic models. On the other hand, if the time between



successive arrivals (inter-arrival times) is uncertain, the arrival pattern is measured by either mean arrival rate or inter-arrival time. These are characterised by the probability distribution associated with this random process. The most common stochastic queueing models assume that arrival rate follow a Poission distribution and/or the inter-arrival times follow an exponential distribution.

**(c) Customers' behaviour.**

It is also necessary to know the reaction of a customer upon entering the system. A customer may decide to wait no matter how long the queue becomes (patient customer), or if the queue is too long to suit him, may decide not to enter it (impatient customer). Machines arriving at the maintenance shop in a plant are examples of patient customers. For impatient customers,

(i) if a customer decides not to enter the queue because of its length, he is said to have *balked*.

(ii) if a customer enters the queue, but after some time loses patience and decides to leave, then he is said to have *renege*d.

(iii) if a customer moves from one queue to another (providing similar/different services) for his personal economic gains, then he is said to have *jockeyed* for position.

The final factor to be considered regarding the input process is the manner in which the arrival pattern changes with time. The input process which does not change with time is called a *stationary* input process. If it is time dependent then the process is termed as *transient*.

**1.1.2. Queue Discipline.**

It is a rule according to which customers are selected for service when a queue has been formed. The most common queue discipline is the "first come, first served" (FCFS), or the "first in, first out" (FIFO) rule under which the customers are serviced in the strict order of their arrivals. Other queue discipline include : "last in, first

out" (LIFO) rule according to which the last arrival in the system is serviced first.

This discipline is practised in most cargo handling situations where the last item loaded is removed first. Another example may be from the production process, where items arrive at a workplace and are stacked one on top of the other. Item on the top of the stack is taken first for processing which is the last one to have arrived for service. Besides these, other disciplines are : "selection for service in random order" (SIRO) rule according to which the arrivals are serviced randomly irrespective of their arrivals in the systems; and a variety of *priority* schemes-according to which a customer's service is done in preference over some other customer.

Under *priority* discipline, the service is of two types: (i) *Pre-emptive priority*. Under this rule, the customers of high priority are given service over the low priority customers. That is, lower priority customer's service is interrupted (pre-empted) to start service for a priority customer. The interrupted service is resumed again as soon as the highest priority customer has been served.

(ii) *Non pre-emptive priority*. In this case the highest priority customer goes ahead in the queue, but his service is started only after the completion of the service of the currently being served customer.

### 1.1.3. Service mechanism.

The service mechanism is concerned with service time and service facilities. Service time is the time interval from the commencement of service to the completion of service. If there are infinite number of servers then all the customers are served instantaneously on arrival and there will be no queue.

If the number of servers is finite, then the customers are served according to a specific order. the customers may be served in batches of fixed size or of variable size rather than by the same server, such as a computer with parallel processing or people boarding a bus. The service system in this case is termed as *bulk service system*.

In the case of parallel channels "fastest server rule" (FSR) is adopted. For its discussion we suppose that the customers arrive before parallel service channels. But it will be more efficient to assume incoming customer is to be assigned a server of largest service rate among the free ones.

Services facilities can be of the following types:

(a) **single queue-one server**, i.e., one queue-one service channel, wherein the customers waits till service point is ready to take him in for servicing.

(b) **several queue-several servers** wherein the customers wait in a single queue until one of the services channels is ready to take them in for servicing.

(c) **several queues-one server** wherein there are several queues and the customer may join any of these but there is only one service channel.

(d) **several servers** when there are several service channels available to provide service, much depends upon their arrangements. they may be arranged in *parallel* or in *series* or a more complex of both, depending on the design of the systems's service mechanism.

For *series channels*, a customer must pass through all the service channels in sequence before service is completed. The situations may be seen in public offices where parts of the service are done at different service counters.

**1.1.4. capacity of the system.** The source from which customers are generated may be finite or infinite. A *finite source* limits the customers arriving for service, i.e., there is a finite limit to the minimum queue size. The queue can also be viewed as one with forced balking where a customer is to balk if he arrives at a time when queue size is at its limit. Alternatively, an *infinite service* "abundant" as in the case of telephone exchange.

## **1.2 OPERATING CHARACTERISTICS OF A QUEUEING SYSTEM**

Of the operational characteristics of a queueing system, that are of general interest for the performance of an existing queueing system and to design a new system are as follows:

1. *expected number of customers in the system* denoted by  $E(n)$  or  $L$  is the average number of customers in the system, both waiting and in service. Here,  $n$  stands for the number of customers in queueing system.
2. *expected number of customers in the queue* denoted by  $E(m)$  or  $L$  is the average number of customers waiting in the queue. Here  $m = n-1$ , i.e., excluding the customer being served.
3. *expected waiting time in the system* denoted by  $E(v)$  or  $W$  is the average total time spent by a in the system. It is generally taken to be the waiting time plus servicing time.



4. *expected waiting time in queue* denoted by  $E(w)$  or  $W$  is the average time spent by a customer in the queue before the commencement of his service.

5. *The server utilization factor* (or busy period) denoted by  $p$  is the proportion of time that a server actually spends with the customers. Here,  $\lambda$  stands for the average number of customers arriving per unit of time and  $\mu$  stands for the average number of customers completing service per unit of time.

The server utilization factor is also known as *traffic intensity* or the clearing ratio.

*Line Length* (queue size):

The total number of customer in the system who are actually waiting in the line and not being serviced.

*Queue Length:*

The queue length may be defined as the line length plus number of customers being served.

Notations:

The notations used in the analysis of a queueing system are as follows;  $n$  = number of customers in the system (waiting and in service)

$P_n$  = probability of  $n$  customers in the system.

$\lambda$  = mean customer arrival rate or average number of arrivals in the queueing system per unit time.

$$\frac{\lambda}{\mu} \text{ or } \rho = \frac{\text{Average service completion time } (\frac{1}{\mu})}{\text{Average inter-arrival time } (\frac{1}{\lambda})}$$

= traffic intensity or server utilization factor (the expected fraction of time in which server is busy).

### 1.3. DETERMINISTIC QUEUEING SYSTEM

A queueing system wherein the customers arrive at regular inter-

vals and the service time for each customer is known and constant, is known as a *deterministic* queueing system.

Let the customers come at the teller counter of a bank for withdrawal every 3 minutes. Thus the interval between the arrival of any two successive customers, that is the inter-arrival time, is exactly 3 minutes. Further, suppose that the in-charge of that particular teller takes exactly 3 minutes to serve a customer. This implies that the arrival and service rates are both equal to 20 customers per hour. In this situation there shall be a queue and the in-charge of the teller shall always be busy with serving work.

Now suppose instead, that the in-charge of the teller can serve 30 customers per hour, i.e., he takes 2 minutes to serve a customer and then has to wait for one minute for the next customer to come for service. Here, also, there would be no queue, but the teller is not always busy.

Further, suppose instead, that the in-charge of the teller can serve only 15 customers per hour, i.e., he takes 4 minutes to serve a customer, clearly, in this situation he would be always busy and the queue length will increase continuously without limit with the passage of time. This implies that when the service rate is less than the arrival rate, service facility cannot cope with all the arrivals and eventually the system leads to an *explosive* situation. In such situation, the problem can be resolved by providing additional service facilities, like opening parallel counters. We can summarize the above as follows :

Let the arrival rate be  $\lambda$  customers per unit time and the service rate be  $\mu$  customers per unit time Then,

(i) if  $\lambda > \mu$ , the waiting line (queue) shall be formed and will increase indefinitely; the service facility would always be busy and the service system will eventually fail.

(ii) if  $\lambda < \mu$ , there shall be no queue and hence no waiting time; the proportion of time for service facility would be idle is



However, it is easy to visualize that the condition of uniform arrival and uniform service rates has a very limited practicability. Generally, the arrivals and servicing time are both variable and uncertain. Thus, variable arrival rates and servicing times are the more realistic assumptions. The probability queueing models are based on these assumptions.

\*\*\*\*\*

## 2. PROBABILITY DISTRIBUTIONS IN QUEUEING SYSTEMS

It is assumed that customers joining the queueing system arrive in a random manner and follow a *Poisson distribution* or equivalently the inter-arrival times obey *exponential distribution*. In most of the cases, services times are also assumed to be exponentially distributed. It implies that the probability of service completion in any short time period is constant and independent of the length of time that the service has been in progress.

In this section, the arrival and service distributions for Poisson queues are derived. The basic assumptions (axioms) governing this type of queues are stated below:

**Axiom 1.** The number of arrivals in non-overlapping intervals are statistically independent, that is, the process has independent increments.

**Axiom 2.** The probability of more than one arrival between time  $t$  and time  $t + \Delta t$ ; that is the probability of two or more arrivals during the small time interval  $\Delta t$  is negligible.

Thus,  $P_0(\Delta t) + P_1(\Delta t) + o(\Delta t) = 1$

**Axiom 3.** The Probability that an arrival occurs between time  $t$  and time  $t + \Delta t$  is equal to  $\lambda \Delta t + o(\Delta t)$ .

Thus,  $P_1(\Delta t) = \lambda \Delta t + o(\Delta t)$ ,

where  $\lambda$  is a constant and is independent of the total number of arrivals upto  $t$ ,  $\Delta t$  is an incremental element, and  $o(\Delta t)$  represents the terms such that  $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$ .

### 2.1. Distribution of Arrivals (Pure Birth Process)

The model in which only arrivals are counted and no departure takes places are called *pure birth models*. Stated in terms of queueing, birth-death process usually arise when an additional customer increases the arrival (referred as birth) in the system and decreases by departure (referred as death) of a served customer from the system.

Let  $P_n(t)$  denote the probability of  $n$  arrivals in a time interval of length  $t$  (both waiting and in service), where  $n \geq 0$  is an integer. Then  $P_n(t + \Delta t)$  being the probability of  $n$  arrivals in a time interval of length  $t + \Delta t$  (making use of *axiom 1*) is as follows:

$P_n(t + \Delta t) = P(n \text{ arrivals in time } t \text{ and one arrival in time } \Delta t)$

$$\begin{aligned}
& + P((n-1) \text{ arrivals in time } t \text{ and one arrival time } \Delta t) \\
& + P((n-2) \text{ arrivals in time } t \text{ and two arrivals time } \Delta t) \\
& + \dots + P(\text{no arrival in time } t \text{ and two arrivals time } \Delta t),
\end{aligned}$$

for  $n \geq 1$ .

Making use of *axiom 2* and *axiom 3*, this difference equation reduces to

$$\begin{aligned}
P_n(t + \Delta t) &= P_n(t)P_0(\Delta t) + P_{n-1}(t)P_1(\Delta t) + o(\Delta t) \\
&= P_n[1 - \lambda\Delta t - o(\Delta t)] + P_{n-1}(t)(\lambda\Delta t + o(\Delta t)) + o(\Delta t)
\end{aligned}$$

where the last term,  $o(\Delta t)$ , represents the terms

$$P[(n-k) \text{ arrivals in time } t \text{ and } k \text{ arrivals in time } \Delta t]$$

$$2 \leq k \leq n$$

The above equation can be re-written as

$$P_n(t + \Delta t) - P_t = -\lambda\Delta t.P_n(t) + \lambda\Delta t.P_{n-1}(t) + o(\Delta t)$$

Dividing it by  $\Delta t$  on both sides and then taking the limit as  $\Delta t \rightarrow 0$ , the equation reduces to

$$\frac{d}{dt}P_n(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n \geq 1. \quad \dots (A)$$

For the case when  $n = 0$ ,

$$P_0(t + \Delta t) = P_0(t)P_0(\Delta t) = P_0(t)[1 - \lambda\Delta t - o(\Delta t)]$$

Rearranging the terms and the dividing on both sides by  $\Delta t$ , taking the limit as  $\Delta t \rightarrow 0$ , we have

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) \quad \dots (B)$$

To solve the  $n+1$  differential-difference equations given in (A) and (B), we make use of the generating functions

$$\phi(z, t) = \sum_{n=0}^{\infty} P_n(t) \cdot z^n,$$

in the unit circle  $|z| \leq 1$ .

Now multiplying the differential-difference equations given in (B) and (A) by  $z^0, z^1, z^2, \dots, z^n$  respectively and then taking summation over  $n$  from 0 to  $\infty$ , we get

$$\sum_{n=0}^{\infty} \frac{d}{dt} P_n(t) z^n = -\lambda \phi(z, t) + \lambda z \phi(z, t).$$

This can also be written as

$$\frac{d}{dt} \phi(z, t) = \lambda(z-1)\phi(z, t).$$

An obvious solution of this differential equation is

$$\phi(z, t) = C e^{\lambda(z-1)t}$$

Where  $C$  is an arbitrary constant.

To determine the value of  $C$ , we use the initial condition that there is no arrival by time  $t=0$  and this gives

$$\phi(z, t) = P_0(0) z^n = 1$$

Now,  $P_n(0) = 0$  for  $n \geq 1$ . Therefore,  $C = 1$ .

Hence,

$$\phi(z, t) = e^{\lambda(z-1)t} \quad \dots(C)$$

Now,

$$\begin{aligned} \frac{d}{dz} \phi(z, t) |_{z=0} &= P_1(t), \\ \frac{d^2}{dz^2} \phi(z, t) |_{z=0} &= 2! P_2, \dots, \\ \frac{d^n}{dz^n} \phi(z, t) |_{z=0} &= n! P_n(t) \end{aligned}$$

Using the value of  $\phi(z, t)$  as given in equation (C), we get

$$\begin{aligned} P_0(t) &= e^{-\lambda t} \\ P_1(t) &= (\lambda t) e^{-\lambda t} \\ P_2(t) &= \frac{1}{2!} (\lambda t)^2 e^{-\lambda t} \\ P_n(t) &= \frac{1}{n!} (\lambda t)^n e^{-\lambda t} \end{aligned}$$

The general formula, therefore, is

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \text{ for } n \geq 0$$

which is the well-known *Poisson probability law* with mean  $\lambda t$ . Thus the random variable defined as the number of arrivals to a system in time  $t$ , has the *Poisson distribution* with a mean of  $\lambda t$  arrivals, or a mean arrival rate of  $\lambda$ .

## 2.2. Distribution of Inter-arrival Times(Exponential Process)

Inter-arrival times are defined as the time intervals between two successive arrivals. Here, we shall show that if the arrival process follows, the Poisson distribution an associated random variable defined as the time between successive arrivals(inter-arrival time) follows the exponential distribution  $f(t) = \lambda e^{-\lambda t}$  and vice-versa. Let the random variable  $T$  be the time between successive arrivals: then

$$P(T > t) = P(\text{no arrival in time } t) = P_o(t) = e^{-\lambda t}$$

The cumulative distribution function of  $T$  denoted by  $F(t)$  is given by

$$\begin{aligned} F(t) &= P(T \leq t) = 1 - P(T > t) \\ &= 1 - P_o(t) = 1 - e^{-\lambda t}, t > 0 \end{aligned}$$

The density function  $f(t)$  for inter-arrival times, therefore, is

$$f(t) = \frac{d}{dt}F(t) = \lambda e^{-\lambda t}, t > 0$$

The expected(or mean) inter-arrival time is given by

$$E(t) = \int_0^{\infty} t.f(t)dt = \int_0^{\infty} \lambda t e^{-\lambda t} dt = \frac{1}{\lambda},$$

where  $\lambda$  is the mean arrival rate.

Thus,  $T$  has the *exponential distribution* with mean  $1/\lambda$ . We would intuitively expect that, if the mean arrival rate is  $\lambda$ , then the mean time between arrivals is  $1/\lambda$ . Conversely, we can also show that if the inter-arrival times are independent and have the same exponential distribution then the arrival rate follows the Poisson distribution.



### 2.3. Distribution of Departures(Pure Death Process)

The model in which only departures are counted and no other arrivals allowed are called *pure death models*. The queueing system starts with  $N$  customers at time  $t = 0$ , where  $N \geq 1$ . Departures occur at the rate of  $\mu$  customers per unit time. To develop the differential-difference equations for the probability of  $n$  customers remaining after ' $t$ ' times units,  $P_n(t)$ , we make use of similar assumptions as was done for arrivals. Let the three axioms, given at the beginning of this section, be changed by using the word service instead of arrival and condition the probability statements by requiring the system to be non-empty. Let us define

$\mu\Delta t$  = Probability that a customer in service at time  $t$  will complete service during time  $\Delta t$ .

For small time interval  $\Delta t > 0$ ,  $\mu\Delta t$  gives probability of one departure during  $\Delta t$ . Using the same arguments as in pure birth process case, the differential-difference equations for this can easily be obtained:

$$P_n(t + \Delta t) = P_n[1 - \mu\Delta t - o(\Delta t)] + P_{n+1}(t) \cdot (\mu\Delta t + o(\Delta t)),$$

$$1 \leq n \leq N-1$$

$$P_0(t + \Delta t) = P_0 + P_1(t) \cdot (\mu\Delta t + o(\Delta t)), \quad n = 0$$

$$P_N(t + \Delta t) = P_N[1 - \mu\Delta t + o(\Delta t)], \quad n = N$$

Re-arranging the above equations then by  $\Delta t$  on both sides and then taking the limits as  $\Delta t \rightarrow 0$ , we get

$$\begin{aligned} \frac{d}{dt} P_n(t) &= -\mu P_n(t) + \mu P_{n+1}(t), \quad 1 \leq n \leq N-1, \quad t > 0 \\ \frac{d}{dt} P_0(t) &= \mu P_1(t); \quad n=0, \quad t \geq 0. \\ \frac{d}{dt} P_N(t) &= -\mu P_N(t); \quad n=N, \quad t \geq 0. \end{aligned}$$

The solution of these equations with initial conditions:



$$\begin{aligned} P_n(0) &= 1; & n &= N \neq 0 \\ P_n(0) &= 0; & n &\neq N \end{aligned}$$

can easily be obtained as earlier. The general solution to the above equation so obtained is

$$P_n(t) = \frac{(\mu t^{N-n} e^{-\mu t})}{dn - n!}; 1 \leq n \leq N \text{ and } P_o(t) = 1 - \sum_{n=1}^N P_n(t)$$

which is known as *truncated Poisson law*.

**Distribution of Service Times** Making similar assumption as done above for arrivals, once could utilize the same type of process to describe the service pattern. Let the three axioms be changed by using the word *service* instead of *arrival* and condition the probability statements by requiring the system to be *non-empty*. Then we can easily show that, the time  $t$  to complete on a customer as follows the exponential distribution:

$$\begin{aligned} s(t) &= \mu e^{-\mu t}, & t &> 0 \\ s(t) &= 0; & t &< 0 \end{aligned}$$

Where  $\mu$  is the mean service rate for a particular service channel. This shows that service times follows exponential distribution with mean  $2/\mu$ . The number,  $n$ , of potential services in time  $T$  will follow the *Poisson distribution* given by

$$\begin{aligned} \phi(n) &= P[n \text{ service in time } T, \text{ if servicing is going on throughout} \\ &T] = \frac{(\mu T)^n}{n!} e^{-\mu T} \end{aligned}$$

Consequently, we can also show that

$$\begin{aligned} P[\text{no service in } \Delta t] &= 1 - \mu \Delta t + o(\Delta t) \\ &\text{and} \\ P[\text{one service in } \Delta t] &= \mu \Delta t + o(\Delta t) \end{aligned}$$

**2.4. Classification of Queueing Models** Generally queueing models may be completely specified in the following symbolic form:

$$(a/b/c) : (d/e).$$

a = arrivals distribution  
b = service time(or departures) distribution  
c = number of service channels(servers)  
d = maximum number of customers allowed in the system(in queue plus in service)  
e = queue(or service) discipline.

The first and second symbols denote the type of distributions of inter-arrival times and of inter-service times, respectively. Third symbol specifies the number serves, whereas fourth symbol stands for the capacity of the system and the last symbol denotes the queue discipline.

If we specify the following Letter as:

M  $\equiv$  Poisson arrival or departure distribution,

$E_k$   $\equiv$  Erlangian or Gamma inter=arrivals follow Poisson distribution,

GI  $\equiv$  General input distribution,

G  $\equiv$  General service time distribution,

then (M/ $E_k$ /1) : ( $\infty$ /FIFO) defines a queueing system in which arrivals follow Poisson distribution service times are Erlangian, single server, infinite capacity and "first in, first out" queue discipline.

## 2.5. Definition of Transient and Steady States

A queueing system is said to be in *transient state* when its operating characteristic(like input, output, mean, queueing length, etc.) are dependent upon time.

If the characteristic of the queueing system becomes independent of time, then the *steady-state* condition is said to prevail.

If  $P_n(t)$  denotes the probability that there are  $n$  customers in the system at time  $t$ , then in the steady-state case, we have

$$\lim_{t \rightarrow \infty} P_n(t) = P_n \text{ (independent of } t\text{)}.$$

Due to practical viewpoint of the steady-state behaviour of the systems, the present chapter is amply focused on studying queueing systems under the existence of steady-state conditions. However, the differential-difference equations which can be used for deriving transient solutions will be presented.

\*\*\*\*\*

### 3. POISSON QUEUEING SYSTEMS

Queues that follow the Poisson arrivals (exponential inter-arrival time) and Poisson service (exponential service time) are called *Poisson queues*. In this section, we shall study a number Poisson queues with different characteristics.

#### 3.1. MODEL I(M/M/1):(∞/FIFO)

This model deals with a queueing system having single service channel. Poisson input, Exponential service and there is no limit on the system capacity while the customers are served on a "first in, first out" basis.

The solution procedure of this queueing model can be summarized in the following three steps:

##### Step 1.

Construction of Differential - Difference Equations. Let  $P_n(t)$  be the probability that there are  $n$  customers in the system at time  $t$ . The probability that the system has  $n$  customers at time  $(t + \Delta t)$  can be expressed as the sum of the joint probabilities of the four mutually exclusive and collectively exhaustive events as follows :

$$P_n(t + \Delta t) = P_n(t) \cdot P[\text{no arrival in } \Delta t] \cdot P[\text{no service completion in } \Delta t] + P_n(t) \cdot P[\text{one arrival in } \Delta t] \cdot P[\text{one service completed in } \Delta t] + P_{n+1}(t) \cdot P[\text{no arrival in } \Delta t] \cdot P[\text{one service completed in } \Delta t] + P_{n-1}(t) \cdot P[\text{one arrival in } \Delta t] \cdot P[\text{no service completion in } \Delta t]$$

This is re-written as :

$$P_n(t+\Delta t) = P_n(t)[1-\lambda\Delta t + O(\Delta t)][1-\mu\Delta t + O(\Delta t)] \\ + P_n(t)[\lambda\Delta t][\mu\Delta t] + P_{n+1}(t)[1-\lambda\Delta t + O(\Delta t)][\mu\Delta t + O(\Delta t)] \\ + P_{n-1}(t)[\lambda\Delta t + O(\Delta t)][1-\mu\Delta t + O(\Delta t)]$$

or

$$P_n(t + \Delta t) - P_n(t) = -(\lambda + \mu) \Delta t P_n(t) + \mu\Delta t P_{n+1}(t) \\ + \lambda\Delta t P_{n-1}(t) + O(\Delta t)$$

Since  $\Delta t$  is very small, terms involving  $(\Delta t)^2$  can be neglected. Dividing the above equation by  $\Delta t$  on both sides and then taking limit as  $\Delta t \rightarrow 0$ , we get

$$\frac{d}{dt}P_n(t) = -(\lambda + \mu)P_n(t) + \mu P_{n+1}(t) + \lambda P_{n-1}(t); n \geq 1.$$

Similarly, if there is no customer in the system at time  $(t + \Delta t)$ , there will be no service completion during  $\Delta t$ . Thus for  $n=0$  and  $t \geq 0$ , we have only two probabilities instead of four. The resulting equation is

$$P_0(t + \Delta t) = P_0(t)(1-\lambda\Delta t + O(\Delta t)) + P_1(t)(\mu\Delta t + O(\Delta t)) \\ (1-\lambda\Delta t + O(\Delta t))$$

or

$$P_0(t + \Delta t) - P_0(t) = -\lambda\Delta t P_0(t) + \mu\Delta t P_1(t) + O(\Delta t)$$

Dividing both sides of this equation by  $\Delta t$  and then taking limit as  $\Delta t \rightarrow 0$ , we get

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) + \mu P_1(t); n=0$$



### Step 2.

Deriving the Steady -State Difference Equations. In the steady - state ,  $P_n(t)$  is independent of time  $t$  and  $\lambda < \mu$  when  $t \rightarrow \infty$ . Thus  $P_n$  and

$$\frac{d}{dt} P_n(t) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Consequently the differential-difference equations obtained in Step 1 reduce to

$$0 = -(\lambda + \mu)P_n + \mu P_{n+1} + \lambda P_{n-1}; n \geq 1$$

and

$$0 = -\lambda P_n + \mu P_1; n = 0.$$

### Step 3.

*Solution of the Steady-State Difference Equations.* For the solution of the above difference equations there exist three methods, namely, the iterative method, use of generating functions and the use of linear operators. Out of these three the first one is the most straightforward and therefore the solution of the above equations will be obtained here by using the iterative method.

Using iteratively, the difference-equations yield

$$\begin{aligned} P_1 &= \frac{\lambda}{\mu} P_0, P_2 = \frac{(\lambda + \mu)}{\mu} P_1 - \frac{\lambda}{\mu} P_0 = \left(\frac{\lambda}{\mu}\right)^2 P_0 \\ P_3 &= \frac{(\lambda + \mu)}{\mu} P_2 - \frac{\lambda}{\mu} P_1 = \left(\frac{\lambda}{\mu}\right)^3 P_0 \text{ and in general } P_n = \left(\frac{\lambda}{\mu}\right)^n P_0 \\ \text{Now, } P_{n+1} &= \frac{(\lambda + \mu)}{\mu} P_n - \frac{\lambda}{\mu} P_{n-1}, n \geq 1. \end{aligned}$$

Substituting the values of  $P_n$  and  $P_{n-1}$ , the equation yields

$$P_{n+1} = \frac{(\lambda + \mu)}{\mu} \left(\frac{\lambda}{\mu}\right)^n P_0 - \frac{\lambda}{\mu} \left(\frac{\lambda}{\mu}\right)^{n-1} P_0 = \left(\frac{\lambda}{\mu}\right)^{n+1} P_0.$$

Thus, by principle of mathematical induction, the general formulae for  $P_n$ , is valid for  $n \geq 0$ . To obtain the value of  $P_0$ , we make use of the boundary condition  $\sum_{n=0}^{\infty} P_n = 1$ .

Therefore

$$1 = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n P_o = P_o \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n;$$

$$\text{Since } P_n = \left(\frac{\lambda}{\mu}\right)^n P_o$$

$$= P_o \frac{1}{1-\frac{\lambda}{\mu}}, \text{ Since } \left(\frac{\lambda}{\mu}\right) < 1$$

This gives  $P_o = 1 - \left(\frac{\lambda}{\mu}\right)$

Hence, the steady - state solution is

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) = \rho^n (1 - \rho); \rho = \left(\frac{\lambda}{\mu}\right) < 1, \text{ and } n \geq 0.$$

This expression gives us the probability distribution of queue length.

### 3.1.1. Characteristics of Model I

(i) Probability of queue size being greater than or equal to k, the number of customers is given by

$$P(n \geq k) = \sum_{k=n}^{\infty} P_k = \sum_{k=n}^{\infty} (1-\rho) \rho^k = (1-\rho) \rho^n \sum_{k=n}^{\infty} \rho^{k-n} = (1-\rho) \rho^n \sum_{k-n=0}^{\infty} \rho^{k-n}$$

$$= (1-\rho) \rho^n \sum_{x=0}^{\infty} \rho^x = \frac{(1-\rho) \rho^n}{1-\rho} = \rho^n.$$

(ii) Average number of customers in the system is given by

$$E(n) = \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n (1-\rho) \rho^n = (1-\rho) \sum_{n=0}^{\infty} n \rho^n = \rho (1-\rho) \sum_{n=0}^{\infty} n \rho^{n-1}$$

$$= \rho (1-\rho) \sum_{n=0}^{\infty} \frac{d}{d\rho} \rho^n = \rho (1-\rho) \frac{d}{d\rho} \sum_{n=0}^{\infty} \rho^n, \text{ Since } \rho < 1$$

$$= \rho (1-\rho) \frac{1}{(1-\rho)^2} = \frac{\rho}{(1-\rho)} = \frac{\lambda}{\mu - \lambda}.$$

(iii) Average queue length is given by

$$E(m) = \sum_{m=0}^{\infty} m P_n, \text{ Where } m = n - 1$$

being the number of customers in the queue, excluding the customer which is in service.

$$E(m) = \sum_{n=1}^{\infty} (n-1) P_n = \sum_{n=1}^{\infty} n P_n - \sum_{n=1}^{\infty} P_n = \sum_{n=0}^{\infty} n P_n - \left[ \sum_{n=0}^{\infty} P_n - P_o \right]$$



$$= \frac{\rho}{1-\rho} - [1 - (1-\rho)] = \frac{\rho}{1-\rho} - \rho$$

$$\frac{\rho^2}{(1-\rho)^2} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

(iv) Average length of non - empty queue is given by

$$E(m|m > 0) = \frac{E(m)}{P(m > 0)} = \frac{\lambda^2}{\mu(\mu-\lambda)} \frac{1}{\left(\frac{\lambda}{\mu}\right)^2} = \frac{\mu}{\mu-\lambda}$$

$$\text{Since } P(m > 0) = P(m > 1) = \sum_{n=0}^{\infty} P_n - P_0 = P_1 = \left(\frac{\lambda}{\mu}\right)^2$$

(v) The fluctuation (variance) of queue length is given by

$$V(n) = \sum_{n=0}^{\infty} [n - E(n)]^2 P_n = \sum_{n=0}^{\infty} n^2 P_n - [E(n)]^2$$

Using some algebraic transformations and the value of  $P_n$ , the result reduces to

$$V(n) = (1-\rho) \frac{\rho + \rho^2}{(1-\rho)^3} - \left[ \frac{\rho}{(1-\rho)} \right]^2 = \frac{\rho}{(1-\rho)^2} = \frac{\lambda\mu}{(\mu-\lambda)^2}$$

### 3.1.2. Waiting Time Distribution for Model I.

The waiting time of a customer in the system is, for the most part, a continuous random variable except that there is a non-zero probability that the delay will be zero, that is a customer entering service immediately upon arrival. Therefore, if we denote the time spent in the queue by  $w$  and  $\psi_w(t)$  denotes its cumulative probability distribution then from the complete randomness of the Poisson distribution, we have

$$\psi_w(0) = P(w = 0) \text{ (No customer in the system upon arrival)}$$

$$= P_0 = (1-\rho)$$

It is now required to find  $\psi_w(t)$  for,  $t > 0$

Let there be  $n$  customers in the system upon arrival, then in order for a customer to go into service at a time between 0 and  $t$ , all the  $n$  customers must have been served by time  $t$ . Let  $s_1, s_2, \dots, s_n$  denote service times of  $n$  customers respectively. Then

$$w = \sum_{i=1}^n s_i, (n > 1) \text{ and } w=0 (n=0).$$

The distribution function of waiting time,  $w$ , for a customer who has to wait is given by

$$P(w \leq t) = P[\sum_{i=1}^n s_i \leq t]; n \geq 1 \text{ and } t > 0$$

Since, the service time for each customer is independent and identically distributed, therefore its probability density function is given by  $\mu e^{-\mu t} (t > 0)$ , where  $\mu$  is the mean service rate. Thus

$$\psi_n(t) = \sum_{n=1}^{\infty} P_n P(n-1 \text{ customers are served at time } t) \cdot P(1 \text{ customer is served in time } \Delta t)$$

$$= \sum_{n=1}^{\infty} (1 - \frac{\lambda}{\mu}) \left(\frac{\lambda}{\mu}\right)^n \frac{(\mu t)^{n-1} e^{-\mu t}}{(n-1)!} \cdot \mu \Delta t.$$

The expression for  $\psi_w(t)$ , therefore, can be written as

$$\begin{aligned} \psi_w(t) &= P(w \leq t) = \sum_{n=1}^{\infty} P_n \int_0^t \psi_n(t) dt \\ &= \sum_{n=1}^{\infty} (1-\rho) \rho^n \int_0^t \frac{(\mu t)^{n-1}}{(n-1)!} e^{-\mu t} \cdot \mu dt = (1-\rho) \rho \int_0^t \mu e^{-\mu t} \sum_{n=1}^{\infty} \frac{(\mu t \rho)^{n-1}}{(n-1)!} dt \\ &= (1-\rho) \rho \int_0^t \mu e^{-\mu t(1-\rho)} dt \end{aligned}$$

Hence, the waiting time of a customer who has to wait is given by

$$\psi_w = \frac{d}{dt} [\psi_w(t)] = \rho(1-\rho) \cdot \mu e^{-\mu t(1-\rho)} = \lambda(1 - \frac{\lambda}{\mu}) e^{-(\mu-\lambda)t}$$

### 3.1.2. Characteristic of Waiting Time Distribution

(i) Average waiting time of a customer (in the queue) is given by

$$\begin{aligned} E(W) &= \int_0^{\infty} t \cdot \psi_w(t) dt = \int_0^{\infty} t \cdot \rho \mu (1-\rho) e^{-\mu t(1-\rho)} dt \\ &= \rho \int_0^{\infty} \frac{x e^{-x}}{\mu(1-\rho)} dx, \text{ for } \mu(1-\rho)t = x \\ &\quad \frac{\rho}{\mu(1-\rho)} = \frac{\lambda}{\mu(\mu-\lambda)} \end{aligned}$$

(ii) Average waiting time of an arrival who has to wait is given by

$$E(w|w > 0) = \frac{E(w)}{P(w > 0)} = \frac{\mu(\frac{\lambda}{\mu-\lambda})}{(\frac{\lambda}{\mu})} = \frac{1}{\mu-\lambda}$$

$$P(w > 0) = 1 - P(w = 0) = 1 - (1 - \rho) = \rho$$

For the busy period distribution, let the random variable  $v$  denote the total time the customer has to spend in the system including service. Then the probability density of its cumulative density function is given by

$$\begin{aligned} \psi(w|w > 0) &= \frac{\psi_w}{P(w > 0)} = \frac{[\lambda(1 - \frac{\lambda}{\mu})e^{-(\mu-\lambda)t}]^{\frac{\lambda}{\mu}}}{\frac{\lambda}{\mu}} \\ &= (\mu - \lambda)e^{-(\mu-\lambda)t}, t > 0. \end{aligned}$$

The average waiting time that a customer spends in the system including service is given by

$$\begin{aligned} E(v) &= \int_0^\infty t \cdot \psi(w|w > 0) dt = \int_0^\infty t \cdot (\mu - \lambda) e^{-(\mu-\lambda)t} dt \\ &= \frac{1}{\mu-\lambda} \int_0^\infty x e^{-x} dx, \text{ for } (\mu - \lambda)t = x \\ &= \frac{1}{\mu-\lambda} \end{aligned}$$

## Relationships among Operating Characteristics

We derived the following important characteristics of an M/M/1 system:

$$E(m) = \frac{\lambda^2}{\mu(\mu-\lambda)}, \quad E(w) = \frac{\lambda}{\mu(\mu-\lambda)}, \quad \text{and} \quad E(v) = \frac{1}{\mu-\lambda}$$

From these expressions, we observe some general relationships between average system characteristics as follows:

The total number of customers in the system is equal to the number of customers the queue plus a customer currently in service, i.e.,

$$E(n) = E(m) + \frac{\lambda}{\mu}$$

The total waiting time of a service customer in the system is equal to the expected waiting time the queue plus the expected service time of a customer in service, i.e.,

$$E(v) = E(m) + \frac{1}{\mu}$$

(iii) Expected number of customers in the system is equal to the average number of arrivals per unit of time multiplied by the average time spent by the customer in system, i.e.,

$$E(n) = \lambda E(v)$$

(iv) Expected number of customers in the queue is equal to the average number of arrivals per unit of time multiplied by the average time spent by a customer in the queue, i.e.,

$$E(m) = \lambda E(w)$$

**Note.** Relations between Average Queue Length and Average Waiting Time are known as *Little's Formulae*.

### SAMPLE PROBLEMS

1. A TV repairman finds that the time spent on his jobs has an Exponential distribution with mean 30 minutes. If he repairs sets in the order in which they come in, and if the arrival of sets is approximately Poisson with an average rate of 10 per 8-hour day, what is repairman's expected idle time each day? How many jobs are ahead of the average set just brought-in?

**Solution.**

We are given,

$\lambda = 10$  sets per day, and  $\mu = 16$  sets per day.

Therefore  $\rho = \frac{\lambda}{\mu} = \frac{10}{16} = 0.625$

The probability for the repairman to be idle is

$P_o = 1 - \rho = 1 - 0.625 = 0.375$

(i) Expected idle time per day =  $(8)(0.375) = 3$  hours.

(ii) Expected (or average) number of T.V. sets in the system

$$E(n) = \frac{\rho}{1-\rho} = \frac{0.625}{1-0.625} = \frac{5}{3} = 2(\text{approx.}) T.V. \text{ sets}$$

**Problem 2.** On an average 96 patients per 24-hour day require the service of an emergency clinic. Also on an average, a patient requires 10 minutes of active attention. Assume that the facility can handle only one emergency at a time. Suppose that it costs the clinic Rs.100 per patient treated to obtain an average servicing time of 10 minutes, and that each minute of decrease in this average time would cost Rs.10 per patient treated. How much would have to be budgeted by the clinic to decrease the average size of the queue from  $1\frac{1}{3}$  patients to  $\frac{1}{2}$  a patient.

**Solution.**

Here,

$$\lambda = \frac{96}{24 \times 60} = \frac{1}{15} \text{ and } \mu = \frac{1}{10} \text{ patients per minute.}$$

$$\begin{aligned} \rho &= \frac{\lambda}{\mu} \\ &= \frac{2}{3}. \end{aligned}$$

Average number of patients in the queue are given by,

$$\begin{aligned} E(m) &= \frac{\rho^2}{1-\rho} \\ &= \frac{(2/3)^2}{1-2/3} \\ &= \frac{4}{3} \end{aligned}$$

Fraction of the time for which there are no patients is given by,

$$\begin{aligned} \rho_0 &= 1 - \rho \\ &= 1 - 2/3 \\ &= 1/3 \end{aligned}$$

Now, when the average queue size is decreased from  $4/3$  patients to  $1/2$  patients, we are to determine the value of  $\mu$ . So, we have



$$E(m) = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

$$\frac{1}{2} = \frac{1/15}{\mu(\mu-1/15)}$$

i.e.,  $\mu = 2/15$  patients per minute.

**Problem 3.** Average rate of treatment required  $= 1/\mu = 15/2 = 7.5$  minutes.

i.e., a decrease in the average rate of treatment is  $(10-7.5)$  minutes or 2.5 minutes.

Budget per patient  $= \text{Rs.}(100+2.5 \times 10) = \text{Rs.}125$ .

Hence, in order to get the required size of the queue, the budget should be increased from Rs. 100 per patient to Rs. 125 per patient.

### Problem 3:

In the production shop of a company the breakdown of the machines is found to be Poisson with an average rate of 3 machines per hour. Breakdown time at one machine costs Rs. 40 per hour to the company. There are two choices before the company for hiring the repairmen. One of the repairmen is slow but cheap, the other fast but expensive. The slow-cheap repairman demands Rs. 20 per hour and will repair the broken down machines exponentially at the rate of 4 per hour. The fast-expensive repairman demands Rs. 30 per hour and will repair machines exponentially at an average rate of 6 per hour. Which repairman should be hired?

**Solution.** In this problem, we compare the total expected daily cost for both the repairmen. This would equal the total wages paid plus the downtime cost.

Case 1. Slow-cheap repairman

$\lambda = 3$  machines per hour and  $\mu = 4$  per hour.

1. Average downtime of a machine

$$= \frac{1}{\mu-\lambda}$$

$$= \frac{1}{4-3}$$

$$= 1 \text{ hour.}$$

2. The downtime of 3 machines that arrive in an hour  $= 1 \times 3 = 3$  hours

Downtime cost  $= \text{Rs. } 40 \times 3 = \text{Rs.}120$ ,



Charges paid to the repairman = Rs.  $20 \times 3$  = Rs. 60

Total cost = Rs. 120 + Rs. 60 = Rs. 180.

Case 2. Fast-expensive repairman

$\lambda = 3$  machines per hour and  $\mu = 6$  machines per hour.

∴ Average downtime of a machine

$$\begin{aligned} &= \frac{1}{\mu - \lambda} \\ &= \frac{1}{6 - 3} \\ &= \frac{1}{3} \text{ hours.} \end{aligned}$$

∴ The downtime of 3 machines that arrive in an hour =  $\frac{1}{3} \times 3 = 1$  hour

Downtime cost = Rs.  $40 \times 1$  = Rs. 40,

Charges paid to the repairman = Rs.  $30 \times 1$  = Rs. 30

Total cost = Rs. 40 + Rs. 30 = Rs. 70.

From the above two cases, the decision of the company should be to engage the fast-expensive repairman.

#### **Problem 4**

The arrivals at the counter in a bank occur in accordance with a Poisson process at an average rate of 8 per hour. The duration of service of a customer has an exponential distribution with a mean of 6 min. Find the probability that an arriving customer (i) has to wait (ii) finds 4 customers in the system.

**Solution.**

$$\lambda = 8/hr, \mu = \frac{1}{6}min = 10/hr$$

(i) Probability that a customer has to wait = Probability that the system is busy.

$$= \frac{\lambda}{\mu} = \frac{8}{10} = 0.8$$

(ii) Probability that there are 4 customer in the system

tributions with mean of 10 min and service times are exponentially distributed with mean of 6 min. The space in front of the window can accommodate only 3 vehicles including the serviced one other vehicles have to wait outside this space. Calculate (a) Probability that an arriving customer can drive directly to the space in front of the window .(b) Probability that an arriving customer will have to wait outside the directed space. (c) How long an arriving customer is expected to wait before getting the service?

**Solution.**

From the data of the probability, we have  $\lambda = 6$ , customers per hour  
 $\mu = 10$  customers per hour.

(a) Probability that an arriving customer can drive directly to the space in front of the window.

$$\begin{aligned} P_0 + P_1 + P_2 &= (1 - \frac{\lambda}{\mu}) + \frac{\lambda}{\mu}(1 - \frac{\lambda}{\mu}) + (\frac{\lambda}{\mu})^2(1 - \frac{\lambda}{\mu}) \\ &= (1 - \frac{\lambda}{\mu})[1 + \frac{\lambda}{\mu} + (\frac{\lambda}{\mu})^2] \\ &= (1 - \frac{6}{10})[1 + \frac{6}{10} + (\frac{6}{10})^2] \\ &= \frac{98}{125} \text{ (or) } 0.784 \end{aligned}$$

(b) Probability that an arriving customer will have to wait outside the directed space.

$$\begin{aligned} 1 - (P_0 + P_1 + P_2) &= 1 - 0.784 \\ &= 0.216 \end{aligned}$$

(c) Expected waiting time of a customer before getting the service is

$$\begin{aligned} W_q &= \frac{\lambda}{\mu(\mu - \lambda)} \\ &= \frac{6}{10(10 - 6)} \\ &= \frac{3}{20} \text{ hour (or) } 9 \text{ min} \end{aligned}$$

#### 4. NON-POISSON QUEUEING SYSTEMS

Queues in which arrivals and/or departures may not follow the Poisson axioms, are called *non-Poisson queues*. The development of these queueing systems is more complicated, mainly because the Poisson axioms no longer hold good. Following are the techniques which are usually employed in studying the non-Poisson queues :

(a) Phase technique, (b) The technique of imbedded Markov Chain, and (c) The Supplementary Variable technique.

The *phase technique* is used when an arrival demands phases of service, say  $k$  in number. The technique by which non-Markovian queues are reduced to Markovian is termed as *Imbedded Markov Chain technique*. When one or more random variables are added to convert a non-Markovian process into a Markovian one, the technique involved in this conversation is called the *Supplementary variable technique*. This technique has been applied in studying the queueing systems :  $GI/G/C$ ,  $M/G/I$ ,  $GI/M/S$ ,  $GI/E_k/I$ ,  $D/E_K/I$ . The study of these non-Poisson queues is not presented here initially.

However, we shall introduce the queueing system  $M/E_K/I$  and the steady-state results of  $M/G/I$ .

##### 4.1. ERLANGIAN SERVICE TIME DISTRIBUTION WITH $k$ -PHASES

When a unit has to pass through  $k$  stages for its service, then these  $k$  stages of servicing are called  $k$ -phases. The distribution of servicing time in each of these phases will be an independent variable and the distribution of total servicing time of a unit in the system will be some combined distribution of time in all these phases. In the previous section we have assumed that the servicing time of a unit follows the exponential distribution given by  $\mu e^{-\mu t}$  with  $1/\mu$  as the average servicing time. Since the exponential distribution involves only one parameter (the parameter  $\mu$ ), it is known as one parameter distribution. A two-parameter generalization of the exponential distribution is called the *Erlangian (Gamma) service time distribution*. The p.d.f. of this Erlangian service time distribution is defined by

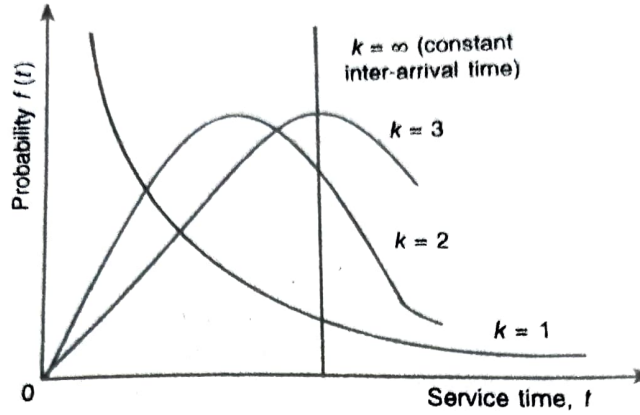


Figure 2:

$$g(t; \mu; k) = C_k t^{k-1} e^{-k\mu t}, k = 1, 2, \dots$$

where  $0 \leq t < \infty$  and  $C_k$  are constants.

The value of constant  $C_k$  is given by

$$\begin{aligned} \int_0^\infty g(t; \mu; k) dt &= 1 \\ \text{or} \\ \int_0^\infty C_k t^{k-1} e^{-k\mu t} dt &= 1 \\ \text{or} \\ \frac{C_k}{(k-\mu)^k} \int_0^\infty y^{k-1} e^{-y} dy &= 1 \text{ for } k\mu t = y \end{aligned}$$

where

$\mu$  = expected number of customers completing service per unit of time, and

$k$  = a positive constant.

The expected value of the total service time and variance of Erlang distribution are  $k(1/k\mu)$ , i.e.,  $1/\mu$  and  $k(1/k\mu^2)$  respectively.

The mode of service time,  $t$ , for the  $k$ th Erlang is  $(k-1)$

The following figure shows how the shape of the Erlang distribution changes for various values of  $k$ . When  $k=1$ , it reduces to the exponential distribution, and for  $1 < k < \infty$ , it reduces to a constant distribution for customer inter-arrival times.

#### 4.2. Model 1 $\{M/E_k/1\} : (\infty/FIFO)$ .

This model consist of a single service channel queueing system in



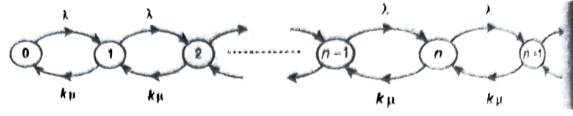


Figure 3:

which there are  $n$  phases in the system (waiting or in service). It has been assumed that a new arrival creates  $k$ -phases of service and departure of one customer reduces  $k$ -phases of service. Let

$n$  = number of customers in the system,

$\lambda_n = \lambda$ , constant arrival rate per unit time,

$\mu_n = k\mu$ ,  $k$  phases of service per unit time.

When  $P_n$  denotes the steady-state probability of  $n$  phases in the system, the transition-rate diagram of model under consideration is :

The balance equations, therefore, are :

$$\lambda P_{n-k} + k\mu P_{n+1} = \lambda P_n + k\mu P_n, \quad n \geq k$$

and

$$k\mu P_1 = \lambda P_0, \quad n = 0$$

Letting  $(\lambda/k\mu) = \rho$ , these equations are:

$$(1 + \rho)P_n = \rho P_{n-k} + P_{n+1}, \quad n \geq k$$

and

$$P_1 = \rho P_0 \quad n = 0.$$

The solution of these difference equations is beyond the scope of this book. However, we discuss characteristics of this model using these difference equations.

#### 4.2.1. Characteristics of Model 1

(i) Average number of phase in the system  $E(n_p)$  is obtained as follows:

Multiplying by  $n^2$  on both sides and then taking summation, the above difference equations gives

$$(1 + \rho) \sum_{n=k}^{\infty} n^2 P_{n-k} + \sum_{n=1}^{\infty} n^2 P_{n+1} = \rho \sum_{x=0}^{\infty} (x+k)^2 P_x + \sum_{y=2}^{\infty} (y-1)^2 P_y,$$

(where  $n-k=x$  and  $n+1=y$ )

$$= \rho \sum_{y=0}^{\infty} (x^2 + 2xk + k^2) P_x + \sum_{y=1}^{\infty} (y^2 - 2y + 1) P_y,$$

$$\text{since } \sum_{y=2}^{\infty} (y-1)^2 P_y = \sum_{y=1}^{\infty} (y-1)^2 P_y$$

or

$$(1 + \rho) \sum_{n=1}^{\infty} n^2 P_n = (1 + \rho) \sum_{n=1}^{\infty} n^2 P_n + \rho \sum_{n=0}^{\infty} (2nk +$$

$$k^2) P_n + \sum_{n=1}^{\infty} (-2n + 1) P_n$$

or

$$0 = \rho [2k \sum_{n=0}^{\infty} n P_n + k^2 \sum_{n=0}^{\infty} P_n] + \sum_{n=1}^{\infty} P_n - 2 \sum_{n=1}^{\infty} n P_n$$

or

$$2(1 - k\rho) \sum_{n=0}^{\infty} n P_n = \rho k^2 - P_0 + 1,$$

$$\text{since, } \sum_{n=0}^{\infty} P_n = 1 \text{ and } \sum_{n=1}^{\infty} n P_n = \sum_{n=0}^{\infty} n P_n.$$

$$\therefore E(n_p) = \frac{\rho k^2 + 1 - 1(1 - k\rho)}{2(1 - k\rho)} = \frac{k(k+1)\rho}{2(1 - k\rho)},$$

where  $P_0 = (1 - k\rho)$ .

$$\text{i.e., } E(n_p) = \frac{k(k+1)}{2} \cdot \frac{\lambda/k\mu}{1 - k\lambda/k\mu} = \frac{k+1}{2} \frac{\lambda}{\mu - \lambda}.$$

(ii) Average waiting time of the phases in the system is given by

$$E(w_p) = \frac{E(n_p)}{\mu} = \frac{k+1}{2\mu} \frac{\lambda}{\mu - \lambda}.$$

(iii) Average waiting time of an arrival is given by

$$E(w) = \frac{E(w_p)}{k} = \frac{k+1}{2k} \frac{\lambda}{\mu(\mu - \lambda)}.$$

(iv) Average time an arrival spends in the system is given by



$$E(v) = E(w) + \frac{1}{\mu} = \frac{k+1}{2k} \frac{\lambda}{\mu(\mu-\lambda)} + \frac{1}{\mu}.$$

(v) Average number of units in the system is given by

$$E(n) = \lambda E(v) = \frac{k+1}{2k} \frac{\lambda^2}{\mu(\mu-\lambda)} + \frac{\lambda}{\mu}.$$

(vi) Average queue length is given by

$$E(m) = E(n) - \frac{\lambda}{\mu} = \frac{k+1}{2k} \frac{\lambda^2}{\mu(\mu-\lambda)}.$$

### Sample Problems

**Problem 1.** A hospital clinic has a doctor examining every patient brought in for a general check-up. The doctor spends 4 minutes on each phase of the check-up although the distribution of time spent on each phase is approximately exponential. If each patient goes through four phases in the check-up and if the arrivals of the patients to the doctors's office are approximately Poisson at the average rate of three per hour, what is the average time spent by a patient waiting in the doctor's office? What is the average time spent in the check-up? What is the most probable time spent in the check-up?

**Solution.**

We are given,

$k=4$  and

Mean arrival rate = 3 patients per hour, i.e.,  $\lambda=3$  per hour.

Service time per phase =  $\frac{1}{4\mu} = 4$  minutes

$$\begin{aligned}\mu &= \frac{1}{4 \times 4} \\ &= \frac{1}{16} \text{ patients per minute.}\end{aligned}$$

and

$$\begin{aligned}E(w) &= \frac{4+1}{4 \times 4} \frac{3}{\frac{1}{4}(\frac{1}{4}-3)} \\ &= 40 \text{ minutes.}\end{aligned}$$

Average time spent in the examination  $1/\mu = 16$  minutes.

Most probable time spent in the examination

$$\begin{aligned}&= \frac{k-1}{k\mu} \\ &= \frac{4-1}{4 \times \frac{1}{16}}\end{aligned}$$

$$\begin{aligned}
&= \frac{3}{1/4} \\
&= 12 \text{ minutes.}
\end{aligned}$$

**Problem 2.** An airline maintenance base has facilities for overhauling only one airplane engine at a time. Hence, to return the airplanes into use at the earliest, the policy is to stagger the overhauling of the 4 engines of each airplane. In other words only one engine is overhauled each time an airplane comes into the base. Under this policy, airplanes have arrivals according to a Poisson process at a mean rate of 1/day. The time required for an engine overhaul has an exponential distribution with a mean of 1/2 day.

A proposal has been made to change the policy so as to overhaul all four engines consecutively each time an aeroplane comes into the shop. It is pointed out that although this will quadruple the expected service time, each plane would need to come into the shop only one-fourth time as often. Compare the two alternatives on a meaningful basis.

### Solution.

The two alternatives will be compared on the basis of the waiting

time cost of the airplanes requiring overhauling.

*First alternative:* {M/M/1}:( $\infty$ /FIFO) queueing system.

Given that  $\lambda = 1$  airplanes per day ;  $\mu = 2$  airplanes per day.

Therefore, average number of airplanes in the system are

$$E(n) = \frac{\lambda}{\mu - \lambda} = \frac{1}{2-1} = 1.$$

*Second alternative:* {M/ $E_K$ /1}:( $\infty$ /FIFO) queueing system.

Given that  $\lambda = \frac{1}{4}$  airplane per day;  $k = 4$ ,

Since service time per airplane is  $4 \times (1/2) = 2$  days, therefore mean service rate,  $\mu_t = 1/2$  airplane per day. Thus

Average number of airplanes in the system are

$$\begin{aligned}
E(n) &= \frac{k+1}{2k} \cdot \frac{\lambda^2}{\mu(\mu-\lambda)} + \frac{\lambda}{\mu} \\
&= \frac{4+1}{2 \cdot (4)} \cdot \frac{(1/4)^2}{\frac{1}{2}(\frac{1}{2}-\frac{1}{4})} + \frac{1/4}{1/2} \\
&= \frac{13}{16} \text{ or } 0.81.
\end{aligned}$$

Since  $E(n) = 0.81$  in the second alternative is less than its value in the first alternative, therefore the waiting cost for requiring overhauling in the overhauling in the second alternative will be less, Hence,

the proposal should be accepted.

**Problem 3.**

At a certain airport it takes exactly 5 minutes to land an aeroplane, once it is given the signal to land. Although incoming planes have scheduled arrival times the wide variability in arrival times produces an effect which makes the incoming planes appear to arrive in a Poisson fashion at an average rate of 6 per hour. This produces occasional stockups at the airport which can be dangerous and costly. Under these circumstances, how much time will a pilot expect to spend circling the field waiting to land?

**Solution.**

From the data of the problem, we have

$\lambda = 6$  per hour (or)  $1/10$  per minute;

$\mu = 1/5$  per minute and

$k = \infty$ , as service time is constant.

Hence, the average time which a pilot expects to spend circling the field waiting to land is given by,

$$\begin{aligned} E(w) &= \lim_{k \rightarrow \infty} \frac{k+1}{2k} \frac{\lambda}{\mu(\mu-\lambda)} \\ &= \lim_{k \rightarrow \infty} \frac{1}{2} \left(1 + \frac{1}{k}\right) \left(\frac{\lambda}{\mu(\mu-\lambda)}\right) \\ &= \frac{1}{2} \left(1 + \frac{1}{\infty}\right) \frac{1/10}{(\frac{1}{5})(\frac{1}{5} - \frac{1}{10})} \\ &= \frac{5}{2} \quad (\text{or}) \quad 2.5 \text{ minutes.} \end{aligned}$$

**Problem 4.** In an airport, it takes exactly 6 minutes to land an aeroplane, once the signal for landing is given. Although the incoming planes have scheduled arrival times, the wide variability of arrival times produces an effect which makes the incoming planes appear to arrive in a Poisson fashion at an average of 5 per hour. This produces occasional stock-ups at the airport, which can be costly and dangerous. Under these circumstances, how much time does a pilot expect to spend circling the field waiting to land?

**Solution**

The mean arrival rate of the aeroplanes,  $\lambda = 5$  per hour.

The mean service rate, i.e., the landing rate,  $\mu = \frac{1}{6}$  per minute.

$$= \frac{1}{6} \times 60 = 10 \text{ per minute.}$$

$$k = \infty$$

( $\therefore$  the service time is constant).

$\therefore$  The average time a pilot expects to spend waiting for landing is  $W_q$ .

Therefore ,

$$\begin{aligned} W_q &= \lim_{k \rightarrow \infty} \left( \frac{k+1}{2k} \right) \frac{\lambda}{\mu(\mu-\lambda)} \\ &= \lim_{k \rightarrow \infty} \frac{1}{2} \left( 1 + \frac{1}{k} \right) \cdot \frac{\lambda}{\mu(\mu-\lambda)} \\ &= \frac{1}{2} \left( 1 + \frac{1}{\infty} \right) \cdot \frac{5}{10(10-5)} \\ &= \frac{1}{2} (1 + 0) \frac{5}{10 \times 5} \\ &= \left( \frac{1}{2} \times \frac{1}{10} \right) \\ &= \frac{1}{20} \text{ hours} \\ &= \frac{1}{20} \times 60 \\ &= 3 \text{ minutes} \\ W_q &= 3 \text{ minutes.} \end{aligned}$$

\*\*\*\*\*



## 5. COST MODELS IN QUEUEING

The *service level* in a queueing facility is a function of the service rate  $\mu$ , that balances the following two conflicting costs:

1. Cost of offering the service, and
2. Cost incurred due to delay in offering the service(customer waiting time).

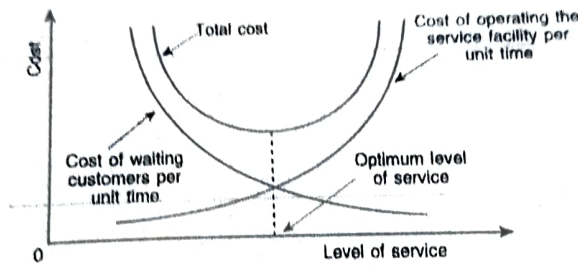


Figure 4:

The two types of costs conflict because an increase in the existing service facility would reduce the customer's waiting time (i.e., cost of waiting). On the other hand, a decrease in the level of service would increase the cost of waiting. "figure-4" illustrates both types of costs as a function of level of service. The optimum service level is one that minimizes the sum of two costs.

The cost model can be expressed as

$$T = K\mu + CE(n).$$

where  $K$  = cost of service rate per unit time,  
 $C$  = cost of waiting customers per unit time, and  
 $E(n)$  = average number of customers in the system.

### 5.1. Optimum service rate.

The procedure for determining the optimum service rate,  $\mu$ , is based on Model I of section Poisson Queuing systems.

Now, in the case of an (M/M/I) : ( $\infty$ /FIFO) model,

$$E(n) = \frac{\lambda}{\mu - \lambda} ; \text{ and therefore } T = K\mu + \frac{C\lambda}{\mu - \lambda}$$

Since the service rate  $\mu$  is continuous, its optimum value can be obtained using the techniques of maxima and minima. Thus, we get

$$\frac{d}{d\mu}(T) = K - \frac{C\lambda}{(\mu - \lambda)^2} \quad \text{and} \quad \frac{d^2}{d\mu^2}(T) = \frac{2\lambda C}{(\mu - \lambda)^3} > 0.$$

Hence, the optimum value of  $\mu$  is  $\mu^0 = \lambda + \sqrt{\lambda C / K}$ .

This result shows that the optimum value of  $\mu$  is not only dependent on K and C, but on the arrival rate  $\lambda$  also.

**Remark.** In the case of (M/M/I):(N/FIFO), the above cost model may be modified to reflect that the larger the value of N, the smaller will be the number of lost customer. The modified cost function is:

$$T = K\mu + CE(n) + D_1N + \lambda P_N D_2$$

where  $D_1$  = cost of servicing each additional customer per unit time,

$D_2$  = cost per lost customer, and

$\lambda P_N$  = number of lost customers per unit time.

## 5.2. Optimum Number of Servers.

For determining the optimum number of service channels, we consider the operational characteristics of queueing model (M/M/S) : ( $\infty$ /FIFO). In this case the cost function can be written as:

$$T = KS + CE(n)$$

where

K = cost per serve per unit time,



$C$  = cost of waiting per customers per unit time,  
 $E(n)$  = average number of customers in the system when there are  $S$  servers.

As the number of serves  $S$  can be measured in discrete units only, differentiation applicable to this case. Instead, one can apply the method of finite differences. The function have a local minima at  $S = S_0$ , if

$$\Delta T(S_0 - 1) < 0 < \Delta T(S_0).$$

Now, considering

$$\begin{aligned} \Delta T(S_0 - 1) &< 0, \\ \text{i.e., } T(S_0) - T(S_0 - 1) &< 0, \end{aligned}$$

we have

$$\begin{aligned} [K \times S_0 + C \times L(S_0)] &< K(S_0 - 1) + C \times L(S_0 - 1) \\ &\text{or} \\ L(S_0 - 1) - L(S_0) &> K/C. \end{aligned}$$

Similarly,  
 considering  $\Delta T(S_0) > 0$ ,  
 we get  $L(S_0) - L(S_0 + 1) < K/C$ .  
 Both these results together yield

$$L(S_0) - L(S_0 + 1) < \frac{K}{C} < L(S_0 - 1) - L(S_0).$$

The value  $K/C$  indicates where the search for optimum  $S$  should start.

### 5.3. OTHER QUEUEING MODELS

We have discussed the models that are rather simple to analyse. However, it is easy to conceive several other models by changing the queue discipline or by considering different patterns of arrival and service rates other than Poisson and exponential. We may have finite or infinite queues, circular queues, queues in series, queues with

balking and reneging. We list some of them:

- (i) Queues in series where input of a queue is derived from the output of another queue, preemptive priority, etc.

### 5.3.QUEUEING CONTROL

The origin of Queue Theory dates back to attempts of determining the, now, well-known queueing characteristics, such as, mean queue length, variance, etc. During the past 30 years or so, there has been a rapidly increasing interest in the study of designing and controlling of the queueing system behaviour (prescriptive models as opposed to descriptive models that make up the majority of queueing literature to date).

Prescriptive models were viewed as static optimization models. The static optimization models are those in which we set up steady-state conditions for the system and determine some long-run average criterion such as cost and/or profit. In static models, the system configuration is set once for all. When the queueing systems are allowed to depend upon time and are controlled, then these systems are known as *dynamic control systems*.

Interestingly, some optimization models are the mixture of these categories. It should be noted that if the state dependent system is controlled, then it is called under *dynamic control*. A lot of work has been done in dynamic control of queueing system. Following are the objectives of dynamic control :

#### **Arrival Process Control**

- (a) To accept or reject control.
- (b) To adjust mean arrival rate.
- (c) Customer exercised control.
- (d) Self versus social optimization.
- (e) Projection times.

#### **Service Process Control**

- (a) Varying a number of servers.
- (b) Varying the service times.

Apart from the two aspects, namely, Static (design) and Dynamic (control) Optimization, there is one more aspect of optimization,

known as "Control of Queue Discipline"-Priority Models, Scheduling Models and allocation of customers to multi-servers fall in this category.

\*\*\*\*\*

## Conclusion

### OTHER QUEUEING MODELS

We have discussed the models that are rather simple to analyse. However, it is easy to conceive several other models by changing the queue discipline or by considering different patterns of arrival and service rates other than Poisson and exponential. We may have finite or infinite queues, circular queues, queues in series, queues with balking and reneging. We list some of them:

- (i) Queues in series where input of a queue is derived from the output of another queue,
- (ii) Circular queues where a customer after being served joins back the queue again,
- (iii) Multi-channel queues with different service rate at each channel,
- (iv) Queues with priority or pre-emptive priority, etc.

### Advantages of Queueing Theory

The following are some of the main advantages of the queueing theory.

- (1) Preparing a plan for transporting fleets.
  - (2) Managing and controlling inventory.
  - (3) Minimizing congestion caused by traffic delays at red light.
  - (4) Preparing a plan for aircraft landing and take-off.
- \*\*\*\*\*

## Reference

- (1) Kanti Swarup, P.K. Gupta, Man Mohan "Operation Research", Published by Sultan Chand and Sons.
- (2) J.K. Sharma "Operation Research Problems and solutions" Third Edition.
- (3) A.K. Malik "Operation Research"
- (4) Dr. D.S. Hira "Operation Research" Revised Edition, Published by S. Chand.
- (5) Hamdy A. Taha "Operation Research An Introduction" 9th Edition



# Game Theory

Project report submitted to

**ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**

Affiliated to

**MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELI**

In partial fulfillment of the requirement for the award of degree of

**Bachelor of science in Mathematics**

Submitted by

**NAMES**

**REG.NO.**

**CHARLET JENISHA. J**

**18AUMT11**

**ELIZABETH SOOSMA. A**

**18AUMT18**

**JASMINE. N**

**18AUMT23**

**JENCY. P**

**18AUMT25**

**MARIA STEPHINO. X**

**18AUMT35**

**SURIYA PRABHA. A**

**18AUMT51**

Under the Guidance of

**Dr. A. PUNITHA THARANI, M.Sc., M.Phil., Ph.D.,**

Head and Associate Professor of Department of Mathematics

**ST. MARY'S COLLEGE(AUTONOMOUS),THOOTHUKUDI.**



Department of Mathematics

**ST.MARY'S COLLEGE (AUTONOMOUS),**

Thoothukudi.

(2020-2021)

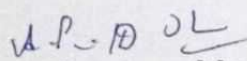


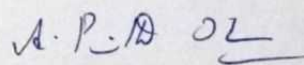
## CERTIFICATE

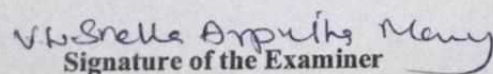
We hereby declare that the project report entitled "GAME THEORY" being submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli in partial fulfillment for the award of degree of Bachelor of science in Mathematics and it is a record of work done during the year 2020-2021 by the following students:

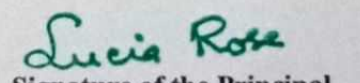
CHARLET JENISHA. J  
ELIZABETH SOOSMA. A  
JASMINE. N  
JENCY. P  
MARIA STEPHINO. X  
SURIYA PRABHA. A

18AUMT11  
18AUMT18  
18AUMT23  
18AUMT25  
18AUMT35  
18AUMT51

  
Signature of the Guide

  
Signature of the HOD

  
Signature of the Examiner

  
Signature of the Principal  
St. Mary's College (Autonomous)  
Thoothukudi - 628 001

## DECLARATION

We hereby declare that the project report entitled "GAME THEORY" is our original work. It has not been submitted to any university for any degree or diploma.

J. Charlet Jenisha  
(CHARLET JENISHA.J)

A. Elizebeth Soosma  
(ELIZABETH SOOSMA.A)

Jasmine. N  
(JASMINE. N)

P. Jency  
(JENCY. P)

X. Maria Stephino  
(MARIA STEPHINO.X)

A. Suriya Prabha  
(SURIYA PRABHA.A)

## **ACKNOWLEDGEMENT**

**First of all, we thank Lord Almighty for showering his blessings to undergo this project.**

**With immense pleasure, we register our deep sense of gratitude to our guide Dr. A. Punitha Tharani M.Sc., M.Phil., Ph.D., Head of the department and Associate Professor of Mathematics, for having imported necessary guidelines throughout the period of our studies.**

**We thank our beloved Principal Rev. Dr. Sr. A.S.J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., for providing us the help to carry out our project work successfully.**

**Finally, we thank all those who extended their helping hands regarding this project.**

## **CONTENT**

<b>Chapter</b>	<b>Topics</b>	<b>Page. No.</b>
	<b>Introduction</b>	
1.	Saddle point and value of the game	3
2.	Games without saddle point – Mixed	
	<b>Strategy</b>	11
3.	Dominance property	17
4.	Graphical method of solution of $2 \times n$ and	
	$m \times n$ matrix	25
5.	Reduction of a game problem to lpp	35
	<b>Conclusion</b>	
	<b>References</b>	

## INTRODUCTION

In real life situation we are to face struggles and competitions. In many practical problems have competitive situations where, there are two or more opponent parties with conflicting interests requires decision making and in these problem the actions to be taken by one depends upon the action taken by the other, called opponent. As for example, candidates for an election countries involved in military battles etc, have their conflicting interests. A competitive situation with a finite number of competitors may termed as a competitive game. The mathematical analysis of competitive problems is fundamentally based on *The Minimax and Maximin criterion*.

Chapter I deals with the basic concepts of game theory, saddle point of payoff matrix and value of the game.

Chapter II deals with the game without saddle point with mixed strategy

Chapter III deals with the dominance property for solving two person zero sum game.

Chapter IV deals with the graphical method of solving  $2 \times n$  or  $m \times n$  games.

Chapter V deals with the reduction of a game problem to an lpp.

# CHAPTER 1

## SADDLE POINT AND VALUE OF THE GAME

### 1.1 INTRODUCTION:

In this chapter we deal about the Strategy of a player, Payoff matrix, The maximum and minimum principles, saddle point of a payoff matrix and the value of the game.

### 1.2 STRATEGY:

In a game, a strategy of a player is a set of rules (or programmes) that will specify which of the available courses of action will be chosen by a player at each play. There are two types of strategies – Pure strategy and Mixed strategy.

(a) Pure strategy: Pure strategy is a decision making rule in which one particular course of action is to be selected by a player. Number of pure strategies may be finite or infinite.

(b) Mixed Strategy: In mixed strategy a player is always kept guessing as to which course of action is to be selected by the other player and hence a mixed strategy is a decision making rule in which a player decides in advance to choose his courses of action with some definite probability distribution. Thus a mixed strategy is a selection among pure strategies with fixed probabilities. For a player the number of mixed strategies is infinite.

### 1.3 PAYOFF MATRIX:

In a game of two players A and B, let A has  $m$  courses and B has  $n$  courses of action. Then the payoff matrix  $(a_{ij})$ ,  $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$  for the player A will be constructed obeying the following rules:



- (i) Row suffixes for each matrix denote the course of action available to the player A.
- (ii) Column suffixes for each matrix denote the course of action available to the player B.
- (iii) Cell entry  $a_{ij}$  denotes the payment to A by B i.e. the gain of the maximization player (row player) when A chooses the  $i$ th course of action and B chooses the  $j$ th courses of action ( $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ ). The payoff matrix of B will simply be constructed replacing  $a_{ij}$  by  $-a_{ij}$ , since the lose to the player A is equivalent to the gain to the player B. Writing in details, the payoff matrix for the player A is given in the following table.

Table 1 (payoff matrix of A)

		Player B				
		1	2	3 ...	n	
Player A	1	$a_{11}$	$a_{12}$	$a_{13}$	...	$a_{1n}$
	2	$a_{21}$	$a_{22}$	$a_{23}$	...	$a_{2n}$
	3	$a_{31}$	$a_{32}$	$a_{33}$	...	$a_{3n}$
	...	...	...	...	...	
	...	...	...	...	...	
	m	$a_{m1}$	$a_{m2}$	$a_{m3}$	...	$a_{mn}$

## 1.4 THE MAXIMIN AND MINIMAX PRINCIPLES:

To explain the Maximin and Minimax principles for the selection of the optimal strategies by the two players, we assume both the players are conservative. i.e. while thinking

of any strategy, say  $A_1$ , by the player A, the player A believes that his opponent knows in advance his strategy and such is the case also for the player B. The principle of Maximin or Minimax is stated as follows:

*If a player lists the worst possible outcomes of all his potential strategies, then he will choose that strategy to be the most suitable for him which corresponds to the best of these worst outcomes.*

## 1.5 TWO-PERSON ZERO SUM (OR RECTANGULAR)

### GAMES:

If a game be restricted to two players, say A and B, only and if the losses of one player are equivalent to the gains to the other player, then this game will be called *Two person Zero-Sum Game*. In such a game the player A can have  $m$  pure strategies and the player B can have  $n$  pure strategies in general and so this game is also called a *Rectangular Game*. If  $m = n$ , this game is called a *Square Game*.

### 1.5 SADDLE POINT AND VALUE OF THE GAME:

A saddle point or an equilibrium point a payoff matrix is that position in the payoff matrix is that position in the payoff matrix where the maximum value (i.e. the maximum of the row minimum). The payoff at the saddle point is known as the value of the game. Thus in a payoff matrix  $(a_{ij})$   $i = 1, 2, \dots, m; j = 1, 2, \dots, n$  the  $(k, l)$ th position will be a saddle point if

$$a_{kl} = \max_i [\min_j \{a_{ij}\}] = \min_j [\max_i \{a_{ij}\}]$$

If we denote the maximin value by  $\underline{v}$  and minimax value by  $\bar{v}$  and the value of the game by  $v$ , then for a determinable game in general

$$\underline{v} \leq v \leq \bar{v}$$

$\underline{v}$  and  $\bar{v}$  are being called the lower and upper value of the game respectively.

If  $\underline{v} = v = \bar{v}$  the game is said to be strictly determinable and if  $\underline{v} \neq \bar{v}$  the game is said to be non-strictly determinable.

If in a strictly determinable game  $v = 0$  i.e. the value of the game is zero, then the game is said to be fair game. But if  $v > 0$  the game is said to be in favour of the player A or biased to maximizing player A on the other hand, if  $v < 0$  the game is said to be biased to the minimizing player B.

## 1.6 ALGORITHM FOR DETERMINING A SADDLE POINT:

STEP 1: Select the minimum element of each row of the payoff matrix and write that minimum value on the right or cover it by the symbol  $\bigcirc$

STEP 2: Select the maximum element of each column of the payoff matrix and put it below the payoff matrix, or cover it by the symbol  $\square$

STEP 3: Now calculate the maximin value and minimax value and check whether they are equal. If they are equal, then the position of that value in the payoff matrix will give the saddle point and that particular value is the value of the game.

Or check whether there is any element of the payoff matrix covered by both the symbols  $\bigcirc$  and  $\square$ . If yes then the position of that element is called the value of the game. If no such element exists, we say that there is no saddle point of this payoff matrix.

### THEOREM 1.1

Let  $(a_{ij})$ ,  $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$  be an  $m \times n$  payoff matrix for a two-person zero sum game. If  $\underline{v}$  and  $\bar{v}$  be the maximum and minimum values of the game, then  $\bar{v} \geq \underline{v}$  i.e.

$$\min_{1 \leq j \leq n} [\max_{1 \leq i \leq m} \{a_{ij}\}] \geq \max_{1 \leq i \leq m} [\min_{1 \leq j \leq n} \{a_{ij}\}]$$

**Proof:** We have

$$\max_{1 \leq i \leq m} \{a_{ij}\} \geq a_{ij} \quad \forall j = 1, 2, \dots, n$$

$$\min_{1 \leq j \leq n} \{a_{ij}\} \leq a_{ij} \quad \forall i = 1, 2, \dots, m$$

Let the above maximum be attained at  $i = k$  and the minimum be attained at  $j = l$ , then

$$\max_{1 \leq i \leq m} \{a_{ij}\} = a_{kj} \text{ and } \min_{1 \leq j \leq n} \{a_{ij}\} = a_{il}$$

Thus we may say  $a_{kj} \geq a_{ij} \geq a_{il} \quad \forall i = 1, 2, \dots, m \text{ and } j = 1, 2, \dots, n$

From this we may write

$$\min_{1 \leq j \leq n} \{a_{kj}\} \geq a_{ij} \geq \max_{1 \leq i \leq m} \{a_{il}\} \quad \forall i = 1, 2, \dots, m; j = 1, 2, \dots, n.$$

Thus

$$\min_{1 \leq j \leq n} [\max_{1 \leq i \leq m} \{a_{ij}\}] \geq \max_{1 \leq i \leq m} [\min_{1 \leq j \leq n} \{a_{ij}\}]$$

Or in symbolic notations  $\bar{v} \geq \underline{v}$

### EXAMPLE 1.1:

For the game with payoff matrix

		Player A		
		I	II	III
Player B	I	-1	2	-2
	II	6	4	-6

7

Determine the best strategies for player A and B and also the value of the game for them. Is this game (i) fair? (ii) strictly determinable?

**SOLUTION:**

We try to solve this problem using maximin and minimax principle as follows

		A			
		I	II	III	Row. Min
B	I	-1	2	-2	-2
	II	6	4	-6	-6
Col. Max.		6	4	-2	

Now  $\underline{v}$  = Maximum of Row min = Maximum value

$$= \max(-2, -6) = -2$$

= value of the Row Min. of the first row

and  $\overline{v}$  = minimum of column max = minimum value

$$= \min(6, 4, -2) = -2$$

= value of the column Max. of the 3<sup>rd</sup> column

Thus the best strategies for the players A and B will be (III, I) (i.e) (A<sub>3</sub>, B<sub>1</sub>)

Now  $v$  = the game value

= the entry at (1,3) cell of the payoff matrix

$$v = -2$$

Hence (1,3) position is a saddle point and the game is strictly determinable.

Again since the game value  $v = -2 \neq 0$  So this game is not fair.

### EXAMPLE 1. 2:

For what value of  $\lambda$  the game with the following payoff matrix is strictly determinable?

		PLAYER B		
		B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>
PLAYER A	A <sub>1</sub>	$\lambda$	6	2
	A <sub>2</sub>	-1	$\lambda$	-7
	A <sub>3</sub>	-5	4	$\lambda$

### SOLUTION:

Let us first ignore the value of  $\lambda$  and determine the maximin and minimax values of the payoff matrix as shown below

		B			
		B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	Row Min.
A	A <sub>1</sub>	$\lambda$	6	2	2
	A <sub>2</sub>	-1	$\lambda$	-7	-7
	A <sub>3</sub>	-2	4	$\lambda$	-2



Col. Max   -1        6        2

Maximin value ( $\underline{v}$ ) = Max. of ( 2, -7, -2 ) = 2

(value corresponding to first row)

Minimax value ( $\overline{v}$ ) = Min. of ( -1, 6, 2 ) = -1

(value corresponding to first column)

Since the best strategies will be ( A, B ) so the value of the game will be the entry at (1,1) cell, so the value of the game (  $v$  ) =  $\lambda$

Now for strictly determinable game  $-1 \leq \lambda \leq 2$ .

## CHAPTER 2

### GAMES WITHOUT A SADDLE POINT- MIXED STRATEGY

#### 2.1 INTRODUCTION:

If any game has no saddle point, it cannot be solved by using Maximin and Minimax principle. To solve such a two – person rectangular game without a saddle point, we see that none of the two players will have single best strategy. But in this situation each player can improve his expected payoff value by considering a mixed strategy. In mixed strategy each player will utilize all of his strategies with certain probabilities such that his expected payoff value may be optimum.

#### THEOREM 2.1.

If for the payoff function  $E(p, q)$  both  $\max_p \min_q E(p, q)$  and  $\min_q \max_p E(p, q)$  exist,

then

$$\min_q \max_p E(p, q) \geq \max_p \min_q E(p, q)$$

**Proof.**

Let  $\hat{p} \in s_m$  and  $\hat{q} \in s_n$  be some arbitrary chosen points. Then we must have

$$\max_p E(p, \hat{q}) \geq E(\hat{p}, \hat{q}) \text{ and } \min_q E(\hat{p}, q) \leq E(\hat{p}, \hat{q})$$

$$\text{Or } \max_p E(p, \hat{q}) \geq E(\hat{p}, \hat{q}) \geq \min_q E(\hat{p}, q)$$

$$\text{Or } \max_p E(p, \hat{q}) \geq \min_q E(\hat{p}, q)$$

But since  $\hat{q}$  is arbitrarily chosen and  $\hat{q} \in S_n$  so the above inequality holds for all values of  $q \in S_n$ . Hence if  $\hat{q}$  be such a point for which  $\max_p E(p, q)$  is minimum, then also

Then also the above inequality holds and therefore

$$\min_q \max_p E(p, q) \geq \max_p \min_q E(\hat{p}, q)$$

Again since  $\hat{p}$  is any point in  $S_m$  the above inequality still holds even if we choose that  $\hat{p}$  which gives the maximum value of  $\min_q E(p, q)$

Hence

$$\min_q \max_p E(p, q) \geq \max_p \min_q E(p, q)$$

Using the symbols we may write  $\bar{v} \geq \underline{v}$ .

## **THEOREM 2.2 (EXISTENCE OF A SADDLE POINT)**

If  $E(p, q)$  be such that both

$$\min_q \max_p E(p, q) \text{ and } \max_p \min_q E(p, q)$$

exist, then a necessary and sufficient condition for the existence of a saddle point  $(\hat{p}, \hat{q})$  of  $E(p, q)$  is

$$E(\hat{p}, \hat{q}) = \min_q \max_p E(p, q) = \max_p \min_q E(p, q)$$

### **Proof .**

Condition is necessary.

If  $(\hat{p}, \hat{q})$  be a saddle point, then by definition

$$E(\hat{p}, q) \geq E(\hat{p}, \hat{q}) \geq E(p, \hat{q}) \quad \forall p \in s_m, q \in s_n$$

So

$$\min_q E(\hat{p}, q) \geq E(\hat{p}, \hat{q}) \geq \max_p E(\hat{p}, q)$$

Again

$$\max_p \min_q E(p, q) \geq \min_q E(\hat{p}, q)$$

$$\text{and } \min_q \max_p E(p, q) \leq \max_p E(p, \hat{q})$$

In the similar way, using the definition of maxima

$$E(\hat{p}, q) \geq E(\hat{p}, \hat{q}) \quad \forall q \in s_n$$

$$E(\hat{p}, q) \geq E(\hat{p}, \hat{q}) \geq E(p, \hat{q}) \quad \forall p \in s_m \text{ and } q \in s_n$$

Hence by the definition of a saddle point  $(\hat{p}, \hat{q})$  is a saddle points

### **NOTE:**

1. Using symbols the necessary and sufficient condition for existence of saddle point is  $\underline{v} = v = \bar{v}$ .
2. Each and every rectangular game can be solved by using mixed strategy. So whether the payoff matrix has saddle point or not the problem can be solved by using mixed strategies.

## EXAMPLE 2.1

Two players A and B match coins. If the coins match, then A wins two units of value, if the coins do not match, then B wins 2 units of value. Determine the optimum strategies for the players of the game.

### Solution:

The payoff matrix for the matching player is,

Player B

Player A	H	T
H	2	-2
T	-2	2

The payoff matrix has no saddle point. The optimum mixed strategies for players A and B, respectively are determined by

$$p_1 = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{2 - (-2)}{2 + 2 - (-2 - 2)} = \frac{1}{2}, p_2 = 1 - p_1 = \frac{1}{2};$$

$$\text{and } q_1 = \frac{a_{22} - a_{12}}{a_{11} + a_{12} - (a_{12} + a_{21})} = \frac{2 - (-2)}{2 + 2 - (-2 - 2)} = \frac{1}{2}; q_2 = 1 - q_1 = \frac{1}{2}.$$

The expected value of game (corresponding to above strategies) is given by

$$V = \frac{a_{11}a_{22} - a_{21}a_{12}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{2 \times 2 - (-2) \times (-2)}{2 + 2 - (-2 - 2)} = 0$$

Hence the optimum strategies for the two players are: H:  $\frac{1}{2}$ , T:  $\frac{1}{2}$  with  $V = 0$ .

## EXAMPLE 2.2

Players A and B each take out one or two matches and guess how many matches opponent has taken. If one of the players guesses correctly, then the loser has to pay him as many rupees as the sum of the number held by both players. Otherwise, the pay-out is zero. Write down the payoff matrix and obtain the optimal strategies of both players.

### Solution:

The payoff matrix for the two players is given by

		Player B	
Player A		1	2
	1	2	0
	2	0	4

The payoff matrix does not have any saddle point. The optimum mixed strategies for the two players are:

$$p_1 = \frac{4-0}{2+4-(0+0)} = \frac{2}{3}; p_2 = 1-p_1 = \frac{1}{3};$$



$$q_1 = \frac{4-0}{2+4-(0+0)} = \frac{2}{3}; q_2 = 1 - q_1 = \frac{1}{3}.$$

The expected value of the game (corresponding to the above strategies) is given by

$$V = \frac{8-0}{2+4-0} = \frac{4}{3}$$

Hence the optimum strategies for the two players are;

$$\text{Player A: } p_1 = \frac{2}{3}, ; p_2 = \frac{1}{3};$$

$$\text{Player B: } q_1 = \frac{2}{3} ; q_2 = \frac{1}{3} \text{ with } V = \frac{4}{3}.$$

## CHAPTER 3

### DOMINANCE PROPERTY

#### 3.1. INTRODUCTION :

This property is very useful property for solving two-person zero-sum game with  $m \times n$  payoff matrix. Sometimes, it is observed that one of the pure strategies of either player is always inferior to at least one of the remaining ones.

#### 3.2. DOMINANCE PROPERTY

In the case of a row player (Maximizing player) A, if in the payoff matrix each element of  $k$ th row be less than or equal to the corresponding elements of the  $r$ th row, then it is obvious that A will not be interested with the  $k$ th row. This is because A's payoff will always be greater than or at least equal to, if he follows  $r$ th strategy instead of  $k$ th strategy and so he will reject the  $k$ th strategy. In other words the probability of choosing this  $k$ th strategy by A is zero. In this case we say that  *$r$ th row dominates  $k$ th row or  $k$ th row is dominated by  $r$ th row.*

For the column player (Minimizing player) B this dominance property will be applicable on columns but in the reverse order. Finally, we delete that row (or column) which is dominated

from the given payoff matrix and use this new payoff matrix to solve the game and by doing this the optimal strategy will not be changed.

### 3.2. GENERAL RULES FOR DOMINANCE

**Rule 1:** If all the elements of the  $k$ th row of a payoff matrix be less than or equal to the corresponding elements of the  $r$ th row, then we say that  $k$ th row is dominated by  $r$ th row and we discard the  $k$ th row from the payoff matrix.

**Rule 2:** If all the elements of the  $p$ th column of a payoff matrix be greater than or equal to the corresponding elements of the  $q$ th column, then we say that  $p$ th column is dominated by the  $q$ th column and we discard the  $p$ th column from the original payoff matrix.

**Rule 3 : (Modified dominance property):** The dominance property is not only based on the superiority of pure strategies only. The modified dominance property states:

- a) If any convex combination of some rows dominates the  $r$ th row, then  $r$ th row may be deleted from the payoff matrix.
- b) If any convex combination of some columns dominates the  $l$ th column, then the  $l$ th column may be deleted from the payoff matrix.

#### THEOREM 3.1 :

If the  $r$ th row of the payoff matrix of an  $m \times n$  rectangular game be dominated by its  $k$ th row, then the deletion of  $r$ th row from the payoff matrix does not change the set of optimal strategies of the row player (**maximizing** player).

**Proof:**

Let the row and column players of any game with payoff matrix  $(a_{ij})_{m \times n}$  be denoted by **A** and **B** respectively. Now since  $r$ th row is dominated by  $k$ th row, so we have

$$a_{rj} \leq a_{kj}, j = 1, 2, \dots, n; r \neq k$$

and for atleast one  $j$ ,

$$a_{rj} < a_{kj}$$

Let  $\hat{q} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_n)$  be an optimal mixed strategy for the player **B** and therefore,

$$\sum_{j=1}^n a_{rj} \hat{q}_j \leq \sum_{j=1}^n a_{kj} \hat{q}_j \text{ i.e. } E(A_r, \hat{q}) \leq E(A_k, \hat{q})$$

where  $A_r$  and  $A_k$  are the respective  $r$ th and  $k$ th pure strategies of **A**.

If  $v$  be the value of the game, then we may say

$$v \geq E(A_k, \hat{q}) \geq E(A_r, \hat{q})$$

Now let  $\hat{p} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_m)$  be an optimal mixed strategy for the player **A**.

If possible, let  $\hat{p}_r > 0$  from the above equation we can write

$$\hat{p}_r v \geq \hat{p}_r E(A_r, \hat{q})$$

Since by assumption  $(\hat{p}, \hat{q})$  is an optimal solution of the game, so we have

$$\begin{aligned} v &= E(\hat{p}, \hat{q}) = \sum_{i=1}^m \hat{p}_i E(A_i, \hat{q}) \\ &= \hat{p}_r E(A_r, \hat{q}) + \sum_{i \neq r}^m \hat{p}_i E(A_i, \hat{q}) < \hat{p}_r v + v \sum_{i \neq r}^m \hat{p}_i \\ &= v \sum_{i=1}^m \hat{p}_i = v [\text{since } \sum_{i=1}^m \hat{p}_i = 1] \end{aligned}$$

i.e.  $v < v$  which is a contradiction and so our assumption  $\hat{p}_r > 0$  is not true and

hence  $p_r = 0$ . Thus deletion of  $r$ th row is justified and it will not influence the optimal solution

## THEOREM 3.2

If the  $i$ th row of the payoff matrix in an  $m \times n$  rectangular game be strictly dominated by a convex combination of the other rows of the matrix, then the deletion of  $i$ th row from the matrix does not affect the set of optimal strategies for the row player.

### Proof:

Let **A** and **B** denote the row and column players of an  $m \times n$  rectangular game given by the payoff matrix  $(a_{ij})_{m \times n}$ . From the given condition, we have some

scalars (probabilities)  $\lambda_1, \lambda_2, \dots, \lambda_n$  where  $0 \leq \lambda_i \leq 1$ ,  $\lambda_i = 0$  and  $\sum_k \lambda_k = 1$  such that

$$(a_{ij}) \leq \sum_{\substack{k=1 \\ k \neq i}}^m \lambda_k a_{kj}, j = 1, 2, \dots, n$$

$$\text{or} \quad (a_{ij}) \leq \sum_{k=1}^m \lambda_k a_{kj}, (\text{since } \lambda_i = 0)$$

In which for at least once  $j$  the strict inequality holds.

Let  $q^o = (q_1^o, q_2^o, \dots, q_n^o)$  be an optimal strategy for the column Player B. We have,

$$\sum_{j=1}^n a_{ij} p_j^o < \sum_{j=1}^n \sum_{k=1}^m a_{kj} \lambda_k q_j^o$$

$$\text{or} \quad E(A_i, q^o) < \sum_{j=1}^n \sum_{k=1}^m a_{kj} \lambda_k q_j^o \leq v$$

where  $A_i$  denotes  $i$ th strategy of the player A and  $v$  is the value of the game.

Let  $p^o = (p_1^o, p_2^o, \dots, p_n^o)$  be an optimal strategy of the row player A. If possible, let  $p_i^o \neq 0$ .

We get  $E(A_i, q^o) < p_i^o v$  since  $p_i^o \neq 0$ .

Then we get ,  $v = E(p^o, q^o) = \sum_{j=1}^n \sum_{k=1}^m a_{kj} p_k^o q_j^o = p_i^o E(A_i, q^o) + \sum_{\substack{k=1 \\ k \neq i}}^m a_{kj} p_k^o q_j^o$

$$= p_i^o E(A_i, q^o) + \sum_{k \neq i} p_k^o E(A_k, q^o) < p_i^o v + v \sum p_k^o$$

$$= v (\sum_{k=1}^m p_k^o)$$

$$= v$$

This is contradiction to the assumption that  $p_i^o \neq 0$ . Thus  $p_i^o = 0$  which shows that the deletion of the  $i$ th row from the payoff matrix does not affect the optimal solution of the row player.

### EXAMPLE 3.1 :

Use dominance property to solve the game whose payoff matrix is given by

		Player <b>B</b>	
		I	II
Player <b>A</b>	I	9	2
	II	8	6
	III	6	4

**Solution;**

It is clear that the above game has no saddle point .We observe that *player B* is always better off if he plays II, no matter what A plays. (The elements in second column are less than the corresponding elements in the first column) The first column is then dominated by the second



one and thus can be deleted. The resulting payoff matrix now indicates that the second row dominates the first and third rows both and hence these can be deleted, yielding the payoff matrix

	II
II	6

Hence the solution to the game is given by

1. Player A should choose strategy II,
2. Player B should choose strategy II,
3. The value of the game is 6 for Player A and -6 for Player B.

### EXAMPLE 3.2:

Solve the following game

		Player B				
		1	2	3	4	5
Player A	I	3	5	4	9	6
	II	5	6	3	7	8
	III	8	7	9	8	7
	IV	4	2	8	5	3

**Solution:**

It is clear that the fourth column of the payoff matrix is dominated by the first column and the fifth column is dominated by the second one. These can thus be deleted.

Thus we get the payoff matrix (a). Again we observe that the third row dominates all the other rows. Thus the payoff matrix (b) is obtained. Now, the second column dominates both the first and the third column. The reduced payoff matrix (c) is thus obtained

	1	2	3
I	3	5	4
II	5	6	3
III	8	7	9
IV	4	2	8

above table is (a).

	1	2	3
III	8	7	9

(b)

	2
III	7

(c)

Thus the solution to the game is :

1. Player A must choose strategy III;
2. Player B must choose strategy 2;
3. The value of the game is 7 for Player A and -7 for Player B.

### **3.3. THE MODIFIED DOMINANCE PROPERTY:**

The dominance property is not always based on the superiority of pure strategies only. A given strategy can also be said to be dominated if it is inferior to some convex linear combinations of two or more pure strategies. In such case, we delete one of the strategies from amongst those involved in the linear combination.

## CHAPTER 4

### GRAPHICAL METHOD OF SOLUTION OF $(2 \times n)$ AND $(m \times 2)$ MATRIX

#### 4.1 INTRODUCTION :

In previous chapters, you had learnt how to solve a game with pure strategies i.e., how to solve a game whose pay-off matrix was rectangular and of order  $m \times n$ , provided that it had a saddle point but in case there was no saddle point then it could be solved by the method in previous chapter. In other words, a game with mixed strategies could be solved when it was of order  $2 \times 2$ .

But a game with the pay-off matrices of order  $m \times 2$  or  $2 \times n$  could also be solved easily with the help of graphical method. The graphical short-cut enables us to reduce the original size of the game of order  $m \times 2$  or  $2 \times n$  to a much similar size order  $2 \times 2$  which could then be solved by the algebraic or short-cut method.

#### 4.2 OBJECTIVES :

After going through this Chapter, you would be able to

- Reduce the size of the game of order  $2 \times n$  to that of a simple game of order  $2 \times 2$ ,
- Reduce the size of the game of order  $m \times 2$  to that of a simple game of order  $2 \times 2$ ,

- Apply the dominance property to reduce the size of the game,
- Evaluate the value of the game after reducing the size of the game to  $2 \times 2$ .

#### 4.3 GRAPHICAL SOLUTION OF $2 \times N$ AND $M \times 2$ GAMES

The procedure described in the last chapter will generally be applicable for any game with a  $2 \times 2$  payoff matrix, unless it possesses a saddle point. Moreover, the procedure can be extended to any square payoff matrix of any order. But it will not work for the game whose payoff matrix happens to be a *rectangular* one, say  $m \times n$ . In such cases a very simple graphical method is available if either  $m$  or  $n$  is two. The graphical shortcut enables us to reduce the original  $2 \times n$  or  $m \times 2$  game to a much simpler  $2 \times 2$  game. Consider the following  $2 \times n$  game :

	Player B		
	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>
Player A	A <sub>1</sub>	$(a_{11} \quad a_{12} \quad \dots \quad a_{1n})$	
	A <sub>2</sub>	$(a_{21} \quad a_{22} \quad \dots \quad a_{2n})$	

It is assumed that the game does not have a saddle point. Let the optimum mixed strategy for A be given by  $S_A = \begin{bmatrix} A_1 & A_2 \\ p_1 & p_2 \end{bmatrix}$ , where  $p_1 + p_2 = 1$ . The average (*expected*) payoff for A when he plays  $S_A$  against B's pure moves  $B_1, B_2, \dots, B_n$  is given by

B's pure move

A's expected pay off  $E(p)$

$B_1$

$$B_1 E_1(p_1) = a_{11}p_1 + a_{21}p_2$$

$B_2$

$$B_2 E_2(p_1) = a_{12}p_1 + a_{22}p_2$$

$B_n$

$$B_n E_n(p_1) = a_{1n}p_1 + a_{2n}p_2$$

The player B would like to choose that pure strategy  $B_j$  against  $S_A$  for which

$E_j(p_1)$  is a minimum for  $j = 1, 2, \dots, n$ . Let us denote this minimum expected

payoff for A by

$$v = \text{Min. } \{E_j(p_1), j = 1, 2, \dots, n\}.$$

The objective of player A is to select  $p_1$  and (hence)  $p_2$  in such a way

that  $v$  is as large as possible. This may be done by plotting the straight lines

$$E_j(p_1) = a_{1j}p_1 + a_{2j}p_2$$

$$= (a_{1j} - a_{2j})p_1 + a_{2j} \quad (j = 1, 2, \dots, n)$$

as linear functions of  $p_1$ .

Number each plotted line according to its corresponding B's pure strategy.

The lower envelope of these lines (indicate it by thick line segments) should give

the minimum expected payoff to A, a function of  $p_1$ . The value  $E_j(p_1)$  are taken on

a vertical axis and the values of  $p_1$  on a horizontal axis. The highest point on this

lower envelope (indicate it by cross mark) then represents the maximum expected



payoff to A. However, a graphic shortcut is available determine the above maximum

value .Observe that  $p_1$  can be at least 0 and at the most 1.The straight line  $E_j(p_1)$

must pass through the points  $[0, E_j(0)]$ , and  $[1, E_j(1)]$ i.e.  $(0, a_{2j})$  and  $(1, a_{1j})$ .

Thus the lines  $E_j(p_1) = (a_{1j} - a_{2j})p_1 + a_{2j}$  for  $j = 1, 2, \dots, n$  can be plotted by

plotting on two different vertical axis the two payoff corresponding to each of the  $n$

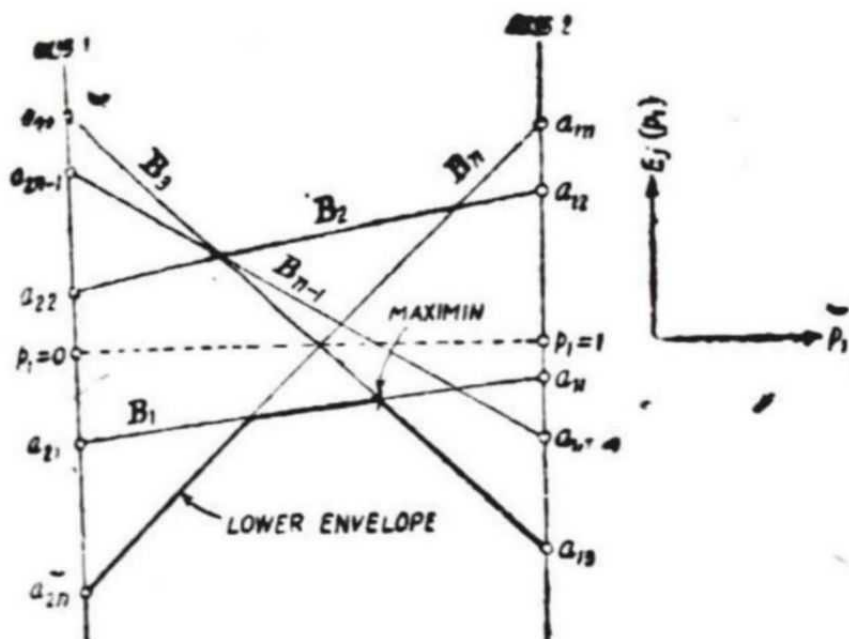
columns. The payoffs in the first row are plotted on axis 2 and those in the second row

on axis 1. The two vertical axes are unit distance apart and are parallel to each other.

The point representing  $a_{1j}$  on axis 2 is then joined to point representing  $a_{2j}$  on axis 1

to yield the straight lines to be numbered  $B_j$ , for  $j = 1, 2, \dots, n$  . A typical example is

illustrated as follows :



The two lines\* passing through the maximum point identify the two critical moves of B which, combined with two of A, yield the  $2 \times 2$  matrix that can be used to determine the optimum strategies of the two players, for the original game, using the results of previous chapter.

### EXAMPLE 4.1:

Solve the following  $2 \times 3$  game graphically:

	Player B		
Player A	1	3	11
	8	5	2

**Solution:** It is obvious that the given payoff matrix has no saddle point. Now let the row player A has the mixes strategies  $(p_1, p_2)$  where  $p_1 + p_2 = 1$  against the column player B.

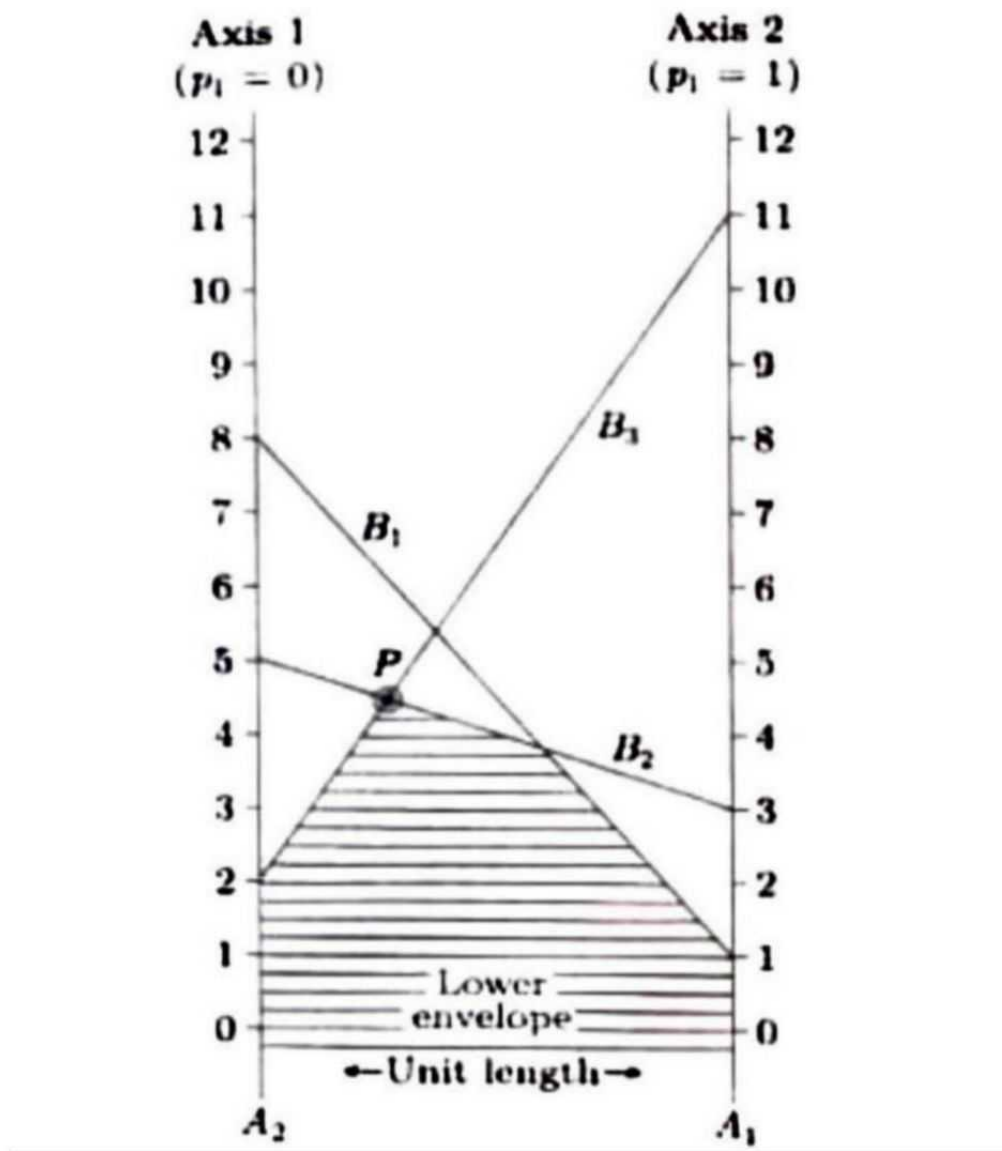
Now we consider the expected payoffs of A against the pure strategies of B as follows:

B's pure strategies	A's expected payoff $E(p)$
$B_1$	$E_1(p) = p_1 + 8p_2 = p_1 8(1 - p_1) = -7p_1 + 8$
$B_2$	$E_2(p) = 3p_1 + 5p_2 = -2p_2 + 5$
$B_3$	$E_3(p) = 11p_1 + 2p_2 = 9p_1 + 2$

Now we plot all the expected payoff equations as functions of  $p_1$

where  $0 \leq p_1 \leq 1$  as shown to the following figure:

Axis 1 ( $p_1 = 0$ ) and Axis 2 ( $p_1 = 1$ ) are two parallel vertical lines the distance between them being one unit. Now to get  $E_1(p)$  we draw the line joining (0,8) and (1,1) i.e., by going the point 8 on the axis 1 and the point 1 on the axis 2 and this line will represent the strategy  $B_1$  for the player  $B$ . In a similar manner the strategies  $B_2$  and  $B_3$  for  $B$  are represented by the lines going  $\{(0,5), (1,3)\}$  and  $\{(0,2), (1,11)\}$  respectively.



From the above figure, it is clear that  $P$  is the highest point of the lower envelope and since,  $P$  is the point of intersection of the lines showing the strategies  $B_2$  and  $B_3$  so, the final reduced  $2 \times 2$  payoff matrix will be

		$B$	
		$B_2$	$B_3$
$A$	$A_1$	3	11
	$A_2$	5	2

Now using formula (13.39), (13.40) and (13.41), we get

$$p_1 = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{2 - 5}{3 + 2 - (11 + 5)} = \frac{3}{11}$$

$$p_2 = 1 - p_1 = 1 - \frac{3}{11} = \frac{8}{11}$$

$$q_1 = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{2 - 11}{3 + 2 - (11 + 5)} = \frac{9}{11}$$

$$\text{Game value } v = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{3 \times 2 - 5 \times 11}{3 + 2 - (11 + 5)} = \frac{49}{11}.$$

Thus for the solution of this game the optimal strategies for the player  $A$  are

$\left(\frac{3}{11}, \frac{8}{11}\right)$  and those for the player  $B$  are  $\left(0, \frac{9}{11}, \frac{2}{11}\right)$ . The value of the game  $v = \frac{49}{11}$ .

### EXAMPLE 4.2:

Use graphical method in solving the following game

	Player B	
Player A	2	4
	2	3
	3	2
	-2	6

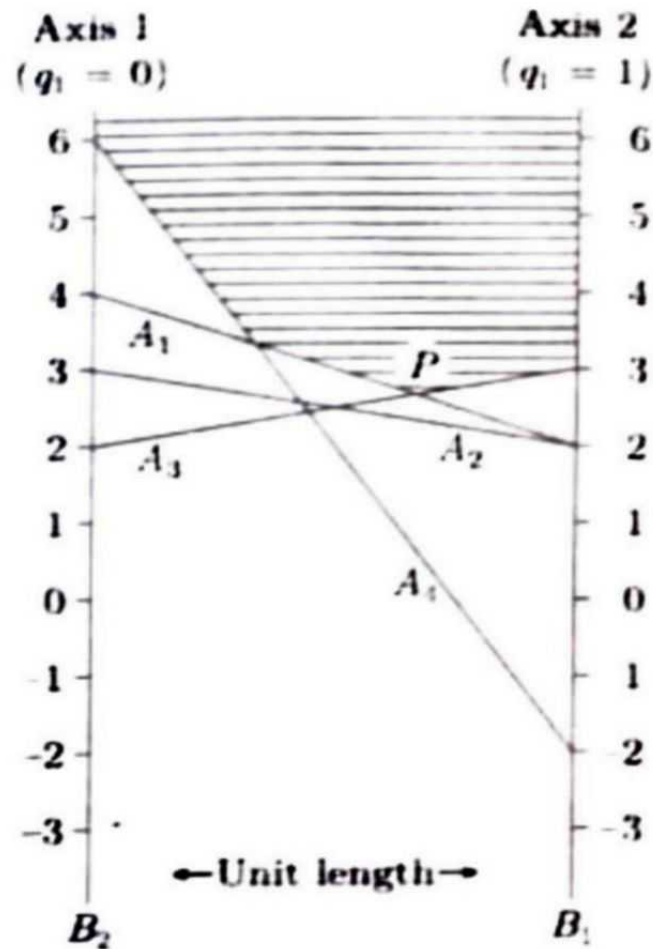
**Solution :**

It is obvious that the given payoff matrix has no saddle point. Let the column player B has the mixed strategies  $(q_1, q_2)$  where  $q_1 + q_2 = 1$  against the row player A.

Now we consider the expected payoff of B against the pure strategies of A as under :

A's pure strategies	B's expected payoff $E(p)$
$A_1$	$E_1(q) = 2q_1 + 4q_2 = 2q_1 + 4(1 - q_1)$ $= -2q_1 + 4$
$A_2$	$E_2(q) = 2q_1 + 3q_2 = -q_1 + 3$
$A_3$	$E_3(q) = 3q_1 + 2q_2 = q_1 + 2$
$A_4$	$E_4(q) = -2q_1 + 6q_2 = -8q_1 + 6$

Now we plot all the expected payoff equations as functions of  $q_1$  where  $0 \leq q_1 \leq 1$  as previous example and they are shown in the figure.



From the figure it is clear that  $P$  is the lowest point of the upper envelope and since,  $P$  is the point of intersection of the lines showing the strategies  $A_1$  and  $A_3$ , so the final reduced  $2 \times 2$  payoff matrix will be

B

		$B_1$	$B_2$
A	$A_1$	2	4
	$A_3$	3	2



Now using formula , we get

$$p_1 = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{2-3}{(2+2) - (4-3)} = \frac{1}{3}$$

$$p_2 = 1 - \frac{1}{3} = \frac{2}{3}$$

$$q_1 = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{2-4}{(2+2) - (4+3)} = \frac{2}{3}$$

$$q_2 = 1 - q_1 = 1 - \frac{2}{3} = \frac{1}{3}$$

$$\text{Game value } v = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - (a_{12} + a_{21})} = \frac{2 \times 2 - 4 \times 3}{(2+2) - (4+3)} = \frac{8}{3}$$

Thus for the solution this optimal strategies for the player A are  $\left(\frac{1}{3}, 0, \frac{2}{3}, 0\right)$  and those for

the player B are  $\left(\frac{2}{3}, \frac{1}{3}\right)$ . The value of the game is  $\frac{8}{3}$ .

# CHAPTER 5

## REDUCTION OF A GAME PROBLEM TO LPP

### 5.1 INTRODUCTION:

Every two – person zero – sum game can always be converted to an lpp .

In this chapter we discuss about the way of converting a game problem to lpp.

### 5.2 REDUTION OF A GAME PROBLEM TO LPP:

Let the payoff matrix of a two-person zero-sum game be an  $m \times n$  rectangular matrix  $(a_{ij})$ ,  $i = 1, 2, 3, \dots, m$  and  $j = 1, 2, 3, \dots, n$ ; and also let  $a_{ij} \geq 0 \forall i$  and  $j$  such that if the value of the game exists , it must be an positive quantity.

Again let

$$P = (p_1, p_2, p_3, \dots, p_m), \sum_{i=1}^m p_i = 1 \text{ and } p_i \geq 0 \forall i$$

and  $Q = (q_1, q_2, q_3, \dots, q_n), \sum_{j=1}^n q_j = 1 \text{ and } q_j \geq 0 \forall j$

be respectively the mixed strategies used by the row player A and the coloumn player B . The net expected gain for A when B plays his  $i$ th strategy  $B_j$  is given by

$$E_j(p) = \sum_{i=1}^m a_{ij} p_i, j = 1, 2, \dots, n$$

Similarly ,the net expected loss for B when A play his  $i$ th strategy  $A_i$  is given by

$$E_i(q) = \sum_{j=1}^n a_{ij} q_j, i = 1, 2, \dots, m$$

Now let  $u$  be the least possible gain for A,

Therefore

$$E_j(p) \geq u \text{ or } \sum_{i=1}^m a_{ij} p_i \geq u, j = 1, 2, \dots, n$$

Also since  $a_{ij} \geq 0$  by assumption, so  $u$  is essentially a positive quantity.

Now the problem may be considered as follows :

The objective of the player A is to select  $(p_1, p_2, p_3, \dots, p_m)$  in such a way that he can maximize his minimum expected gain.

Again since  $u > 0$  so maximization of  $u$  is equivalent to minimization of  $\frac{1}{u}$ .

Writing  $U = \frac{1}{u}$  we see

$$\begin{aligned} U = \frac{1}{u} &= \frac{p_1 + p_2 + p_3 + \dots + p_m}{u} \text{ [ since } \sum_{i=1}^m p_i = 1 \text{ ]} \\ &= p'_1 + p'_2 + p'_3 + \dots + p'_m \text{ where } p'_i = \frac{p_i}{u} \geq 0, i = 1, 2, \dots, m \end{aligned}$$

Therefore,  $\sum_{i=1}^m a_{ij} p_i \geq u$  becomes

$$\sum_{i=1}^m a_{ij} p'_i \geq 1$$

Thus for the player A, the problem can be expressed as an lpp as follows :

$$\text{Minimize } U = p'_1 + p'_2 + \dots + p'_m$$

$$\text{Subject to } \sum_{j=1}^n a_{ij} p'_i \geq 1, j = 1, 2, \dots, n$$

$$\text{and } p'_i \geq 0, i = 1, 2, \dots, m$$

Again since the objective of the player B is to select  $(q_1, q_2, \dots, q_n)$  in such a way that he can minimize his maximum expected loss. If  $v$  denotes the greatest possible loss of B and if  $V = \frac{1}{v}$  then as in the case for A the problem for B can be expressed as an LPP as follows :

$$\text{Maximize } V = q'_1 + q'_2 + \dots + q'_n$$

$$\text{Subject to } \sum_{j=1}^n a_{ij} q'_j \leq 1, i = 1, 2, \dots, m$$

$$\text{And } q'_j \geq 0, j = 1, 2, \dots, n$$

Where

$$q'_j = \frac{q_j}{v}, j = 1, 2, \dots, n$$

## NOTE :

1 ) By assuming initially  $a_{ij} \geq 0 \forall i$  and  $j$  in the payoff matrix  $(a_{ij})$ ,  $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ ; we see that all the variables  $p'_i$  and  $q'_j$  are non negative which are exactly needed for an LPP and also the value of the game is non negative in this case . but if there are some negative elements in the payoff table a suitable constant is to be added to every element in the payoff table so as to make the smallest element greater than or equal to zero . The solution of this new game with the revised payoff matrix will give an optimal mixed strategy for the original problem . The value of the original game will be obtained by subtracting the earlier choosed constant from the value of the new game .

2 ) It is to be noted that the Linear Programming Problem developed for two players is a

primal – dual pair . Therefore , by fundamental theorem of duality ,one can read off the optimal solution of one player just from the optimum simplex table of the other player . So we just need to solve one player's LPP by simplex method .

3 ) If we solve the LPP for the player B by simplex method , then from the optimal simplex table , we get the optimal values of  $(q'_1, q'_2, \dots, q'_n)$  as  $(q^{*'}_1, q^{*'}_2, \dots, q^{*'}_n)$  , and , then

$V^* = V_{max} = \frac{1}{v}$  where v is the value of the game . The corresponding optimal values of

$(q_1, q_2, \dots, q_n)$  say  $(q^*_1, q^*_2, \dots, q^*_n)$  are then given by  $q^*_j = v^* q'_j$

### **THEOREM 5.1: ( Fundamental theorem on rectangular game )**

If mixed strategies be allowed the value of a game exists uniquely .

PROOF :

We know that any rectangular game can be expressed as LPP for both the players A and B

It is interesting to note that the two problems are dual to each other .

Now using a pure strategy say  $A_k$  for A and taking  $0 < u \leq \min ( a_{ij} ) , a_{ij} > 0 , i = 1, 2, 3, \dots, m$  we have

$$p_i = \begin{cases} 1, & \text{for } i = k \\ 0, & \text{for } i \neq k \end{cases}$$

And so

$$p'_i = \frac{p_i}{u} = \begin{cases} \frac{1}{u}, & \text{for } i = k \\ 0, & \text{for } i \neq k \end{cases}$$

In this case

$$\sum_{i=1}^m a_{ij} p'_i = a_{kj} p'_k = \frac{a_{kj}}{u} \geq 1 \text{ [ since } 0 < u < \min a_{ij} \text{ ]}$$

Thus we see that the problem has atleast one feasible solution set  $(0, 0, \dots, p'_k, \dots, 0)$  for A .

Further we see that the objective function  $U = p'_1 + p'_2 + \dots + p'_m \geq 0$  for every such feasible solution, so that the objective function of the minimization problem is never unbounded.

Hence the LPP for A has a finite optimal solution.

Again since B's problem is nothing but the dual of A's problem, so by the important duality theorem If either the primal or the dual problem has a finite optimal solution, then the other problem has a finite optimal solution and the optimal values of the objective functions are equal.

We may say that the problem for B has a finite optimal solution and also

$$\min U = \max V \left( = \frac{1}{v} \right) \text{ say }.$$

Therefore,  $\max u = \min v = v$  or  $v_- = -v = v$  which is the unique value of the game.

## NOTE :

If the payoff matrix has no saddle point, then both the players will use mixed strategies and there will exist optimal mixed strategies for both the players.

## EXAMPLE : 5.1

For the following pay-off matrix, transform the zero-sum game into an equivalent linear programming problem and solve it by using simplex method.

Player B

Player A	$B_1$	$B_2$	$B_3$
$A_1$	1	-1	3
$A_2$	3	5	-3
$A_3$	6	2	-2

SOLUTION :

The first step is to find out the saddle point ( if any ) in the pay-off matrix as shown below :



Player A	Player B			Row minimum
	$B_1$	$B_2$	$B_3$	
$A_1$	1	-1	3	1 ← <i>maximin</i>
$A_2$	3	5	-3	-3
$A_3$	6	2	-2	-2
Column maximum	6	5	3 ← minimax	

The given pay-off matrix does not have a saddle point. Since the maximin value is -1, therefore, it is possible that the value of game ( $v$ ) may be negative or zero because  $-1 < v < 1$ . Thus, a constant which is at least equal to the negative of maximin value, i.e. more than -1 is added to all the elements of the pay-off matrix. Thus, a constant which is at least equal to the negative of maximum value, i.e. more than -1 is added to all the elements of the pay-off matrix. Thus, adding a constant number 4 to all the elements of the pay-off matrix, the pay-off matrix becomes:

Player A	Player B			Probability
	$B_1$	$B_2$	$B_3$	
$A_1$	5	3	7	$p_1$
$A_2$	7	9	1	$p_2$
$A_3$	10	6	2	$p_3$
Probability	$q_1$	$q_2$	$q_3$	

Let  $p_i$  ( $i= 1,2,3$ ) and  $q_j$  ( $j= 1,2,3$ ) be the probabilities of selecting strategies  $A_i$  ( $i= 1,2,3$ ) and

$B_j$  ( $j= 1,2,3$ ) by the players A and B, respectively . The expected gain for player A will be as

follows:

$$\text{For strategy } B_1 : 5P_1 + 7P_2 + 10P_3 \geq V \text{ or } 5\frac{P_1}{V} + 7\frac{P_2}{V} + 10\frac{P_3}{V} \geq 1$$

$$B_2 : 3P_1 + 9P_2 + 6P_3 \geq V \text{ or } \frac{P_1}{V} + 9\frac{P_2}{V} + 6\frac{P_3}{V} \geq 1$$

$$B_3 : 7P_1 + P_2 + 2P_3 \geq V \text{ or } 7\frac{P_1}{V} + \frac{P_2}{V} + 2\frac{P_3}{V} \geq 1$$

$$P_1 + P_2 + P_3 = 1 \text{ or } \frac{P_1}{V} + \frac{P_2}{V} + \frac{P_3}{V} = \frac{1}{V}$$

and  $P_1, P_2, P_3 \geq 0$  .

In order to simplify , we define new variables :  $X_1 = \frac{P_1}{V}, X_2 = \frac{P_2}{V}, X_3 = \frac{P_3}{V}$  . The problem for

player A , therefore becomes ,

$$\text{Minimize } Z_p(=\frac{1}{V}) = X_1 + X_2 + X_3$$

Subject to the constraints

$$5y_1 + 7y_2 + 10y_3 \geq 1 ,$$

$$7y_1 + 9y_2 + y_3 \geq 1 ,$$

$$10y_1 + 6y_2 + 2y_3 \geq 1$$

And  $y_1, y_2, y_3 \geq 0$

Where  $y_1 = q_1/v$ ;  $y_2 = q_2/v$ ;  $y_3 = q_3/v$ .

It may be noted that problem of player A is the dual of the problem of player B .Therefore ,  
solution of the dual problem can be obtained from the optimal simplex table of primal .

To solve the problem of player B , introduce slack variables to convert the three  
inequalities to equalities

The problem becomes

Maximize  $z = y_1 + y_2 + y_3 + 0s_1 + 0s_2 + 0s_3$

Subject to the constraints

$$5y_1 + 3y_2 + 7y_3 + s_1 = 1 ,$$

$$7y_1 + 9y_2 + y_3 + s_2 = 1,$$

$$10 y_1 + 6y_2 + 2y_3 + s_3 = 1$$

And  $y_1, y_2, y_3, s_1, s_2, s_3 \geq 0$

INITIAL SOLUTION TABLE

$$c_j \rightarrow \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0$$

Unit cost	Variables In basis	solution values $y_B (= b)$	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	Min . Ratio $\frac{y_B}{y_1}$
0	$s_1$	1	5	3	7	1	0	0	$1/5$
0	$s_2$	1	7	9	1	0	1	0	$1/7$
0	$s_3$	1	10	6	2	0	0	1	$1/10$
$Z = 0$			0	0	0	0	0	0	
			1	1	1	0	0	0	

Proceeding with usual simplex method ,the optimal solution is shown below ,

OPTIMAL SOLUTION TABLE

$$C_j \rightarrow \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0$$

Unit cost $c_B$	Variables in basis B	Solution Values $y_B = b$	$y_1$	$y_2$	$y_3$	$s_1$	$s_2$	$s_3$
1	$y_3$	$1/10$	$2/5$	0	1	$3/20$	$-1/10$	0
1	$y_2$	$1/10$	$11/15$	1	0	$-1/60$	$7/60$	0
1	$s_3$	$1/5$	$24/5$	0	0	$-1/5$	$-3/5$	1
$Z = \frac{1}{5}$			$17/15$	0	0	$2/15$	$1/15$	0
			$-2/15$	0	0	$-2/15$	$-1/15$	0

The optimal solution (mixed strategies ) for B is :  $y_1 = 0$ ;  $y_2 = \frac{1}{10}$ ;  $y_3 = \frac{1}{10}$  and the expected

value of the game is  $Z = \frac{1}{v} - \text{constraint } (=4) = 5 - 4 = 1$ . These solution values are now converted

back into the original variables : If  $\frac{1}{v} = \frac{1}{5}$  then  $v = 5$

$$y_1 = \frac{q_1}{v}, \text{ then } q_1 = y_1 \times v = 0; y_2 = \frac{q_2}{v}, \text{ then } q_2 = y_2 \times v = \frac{1}{10} \times 5 = \frac{1}{2}$$

$$\text{and } y_3 = \frac{q_3}{v}, \text{ then } q_3 = y_3 \times v = \frac{1}{10} \times 5 = \frac{1}{2}$$

The optimal strategies :  $x_1 = \frac{2}{15}$ ,  $x_2 = \frac{1}{15}$  and  $x_3 = 0$  for player A are obtained from the  $c_j - z_j$  row of the table ( optimal solution ) . Then  $p_1 = x_1 \times v = \left(\frac{2}{15}\right) \times 5 = \frac{2}{3}$  ;  $p_2 = x_2 \times v = \left(\frac{1}{15}\right) \times 5 = \frac{1}{3}$  and  $p_3 = x_3 \times v = 0$ .

Hence the probabilities of using strategies by both the players are : Player A =  $\left(\frac{2}{3}, \frac{1}{3}, 0\right)$  ; Player B =  $\left(0, \frac{1}{2}, \frac{1}{2}\right)$  and , value of the game is ,  $V = 1$ .

## CONCLUSION

The technique of Game Theory provide numerous insights into questions of decision making and strategy. The quantitative predictions are sometimes at odds with intuition or observed behavior indicates that there is still much to say about the nature of rationality, utility and human behaviour. Attempting to reconcile those differences will strengthen both Game Theory and social sciences to which it is applied, and as a relatively new mathematical field, Game Theory almost certainly still has many interesting results to offer.

This project deals about the basic concepts, saddle point and value of the game in the first chapter. In second chapter it deals about the game without saddle point with mixed strategy. In third chapter it deals about the Dominance property for solving two-person zero sum game. In fourth chapter it deals about the graphical method of solving  $2 \times n$  and  $m \times n$  games. In fifth chapter it deals about the reduction of a game problem to an lpp.



## REFERENCES

- [1] P. K. Gupta , Man Mohan, “Linear programming and theory of games, Published by  
. Sultan Chand and Sons, 1979.
- [2] A. Mukherjee, N.K. Bej, “Advanced linear programming and Game theory” Published by  
Book and Allied private ltd, 2013.
- [3] J. K. Sharma, “Operation Research-Problems and solutions” Third Edition.
- [4] Erich Prisner, “ Game theory through examples” Published by the Mathematical  
Association of America”, 2014.
- [5] J.G. Chakravothy, P.R. Ghosh, “Linear programming and game theory” Fourteenth  
Revised Edition.



# **A STUDY ON FUZZY MATRIX THEORY**

Project Report submitted to

**ST. MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**

Affiliated to

**MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELI**

In partial fulfilment of the requirement for the award of degree of

**Bachelor of science in Mathematics**

Submitted by

**NAMES**

**REG.NO.**

**DEILINI SUGEIRTHA D**

**18AUMT14**

**EMIBHA S**

**18AUMT19**

**KEERTHIKA M**

**18AUMT30**

**RAMA DEVI K**

**18AUMT40**

**RISHIBA S**

**18AUMT43**

Under the Guidance of

**Dr. G. PRISCILLA PACIFICA M.Sc., B.Ed., M.Phil., Ph.D., SET**

Assistant Professor of Mathematics

**ST. MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**



Department of Mathematics

**ST.MARY'S COLLEGE (AUTONOMOUS)**

Thoothukudi.

(2020-2021)



## CERTIFICATE

We hereby declare that the project report entitled "**A STUDY ON FUZZY MATRIX THEORY**" being submitted to **St. Mary's College (Autonomous), Thoothukudi** affiliated to **Manonmaniam Sundaranar University, Tirunelveli** in partial fulfilment for the award of degree of **Bachelor of Science in Mathematics** and it is a record of work done during the year 2020-2021 by the following students:

DEILINI SUGEIRTHA D

EMIBHA S

KEERTHIKA M

RAMA DEVI K

RISHIBA S

18AUMT14

18AUMT19

18AUMT30

18AUMT40

18AUMT43

Signature of the Guide

Signature of the HOD

V. S. Stella Apuleio Mary  
Signature of the Examiner

Lucia Rose

Signature of the Principal  
**St. Mary's College (Autonomous)**  
**Theethukudi - 628 001.**



## DECLARATION

We hereby declare that the project reported entitled "A STUDY ON FUZZY MATRIX THEORY" is our original work. It has not been submitted to any university for any degree or diploma.

D. Deilini Sugeirtha.  
(DEILINI SUGEIRTHA.D)

S. Emibha  
(EMIBHA.S)

M. Keerthika  
(KEERTHIKA.M)

K. Rama Devi  
(RAMA DEVI.K)

S. Rishiba.  
(RISHIBA.S)



## ACKNOWLEDGEMENT

First of all, we thank lord Almighty for showering his blessings to undergo this project.

With immense pleasure, we register our deep sense of gratitude to our guide **Dr. G. Priscilla Pacifica M.Sc., B.Ed., M.Phil., Ph.D., SET** and the Head of the Department **Dr. A. Punitha Tharani M.Sc., M.Phil., Ph.D.**, for having imported necessary guidelines throughout the period of our studies.

We thank our beloved Principal **Rev. Dr. Sr. A.S.J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D.**, for providing us the help to carry out our project work successfully.

Finally, we thank all those who extended their helping hands regarding this project.



# CONTENTS

<b>ACKNOWLEDGEMENT</b>	<b>I</b>
<b>PREFACE</b>	<b>II</b>
<b>CHAPTER 1: PRELIMINARIES</b>	<b>1-3</b>
1.1 Fuzzy Matrix Theory	1
1.2 The Determinant Theory of a Square Fuzzy Matrix	1
1.3 The Adjoint Theory of a Square Fuzzy Matrix	2
<b>CHAPTER 2: FUZZY MATRIX THEORY</b>	<b>4-12</b>
2.1 Introduction	4
2.2 Fuzzy Matrix Theory	4
2.3 Operations on Fuzzy Matrices	6
<b>CHAPTER 3: THE DETERMINANT THEORY OF A SQAURE FUZZY</b>	
<b>MATRIX</b>	<b>13-24</b>
3.1 Introduction	13
3.2 The Determinant Theory of a Square Fuzzy Matrix	13
3.3 Properties of Adjoints of Square Fuzzy Matrices	17
<b>CHAPTER 4: THE ADJOINT THEORY OF A SQUARE FUZZY</b>	
<b>MATRIX</b>	<b>25-43</b>
4.1 Introduction	25



4.2 The Adjoint Theory of A Square Fuzzy Matrix	25
4.3 Properties of Adjoints of Square Fuzzy Matrices	27
<b>REFERENCES</b>	<b>45-46</b>

## INTRODUCTION

The concept of the determinant theory and the adjoint of a square fuzzy matrix will be studied. Also, we define the circular fuzzy matrices and show that some properties of a square fuzzy matrix (such as reflexivity, transitivity, and circularity) are carried out to the adjoint of the matrix.

Finally, we show how to construct a transitive fuzzy matrix from a given one through the adjoint matrix of it. Fuzzy matrices for which it is necessary to define Maximum operation, Minimum operation, Max-Min operations and Min-Max operation on fuzzy matrices.

A fuzzy matrix arises in many applications one of which is adjacency matrices of fuzzy relations. A fuzzy matrix is a matrix which has its elements from  $[0,1]$  called fuzzy unit interval. The foundation for the determinant theory of a square fuzzy matrix where the elements take its values from the unit interval  $[0,1]$ .

## CHAPTER 1

### PRELIMINARIES

#### 1.1 FUZZY MATRIX THEORY

The purpose of present chapter is to introduce the basic concept of fuzzy matrix theory. Consider the operations defined on these matrices for further treatment of determinant and adjoint theory of square fuzzy matrices. We shall give fuzzy matrix theory and some operations defined on fuzzy matrices.

A fuzzy matrix is a matrix which has its elements from  $[0,1]$ , called fuzzy unit interval.

**Definition 1.1.1:** Consider a matrix  $A = [a_{ij}]_{m \times n}$  where  $a_{ij} \in [0,1]$ ,  $1 \leq i \leq m$  and  $1 \leq j \leq n$ . Then A is fuzzy matrix.

##### 1.1.1 Types of Fuzzy Matrix:

- (i) Fuzzy Rectangular matrix
- (ii) Fuzzy Square matrix
- (iii) Fuzzy Row matrix
- (iv) Fuzzy Column matrix
- (v) Fuzzy Diagonal matrix
- (vi) Fuzzy Scalar matrix

#### 1.2 THE DETERMINANT THEORY OF A SQUARE FUZZY MATRIX

In this chapter we have to investigate some properties of the determinant of a square matrix. The properties of a square fuzzy matrix are somewhere analogous to

the crisp case for determinant of square matrix. We shall introduce the determinant theory of a square matrix and their properties and investigate the elementary properties of determinant theory for fuzzy matrices in matrix semiring  $M_n(F)$ .

The foundation for the determinant theory of a square fuzzy matrix where the elements takes its values from the unit interval  $[0,1]$ .

**Definition 1.2.1:** The determinant  $|A|$  of an  $n \times n$  fuzzy matrix  $A$  is defined as follows:

$$|A| = \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

$$\sum_{\sigma \in S_n} \prod_{i=1}^n a_{i\sigma(i)}$$

Where  $S_n$  denotes the symmetric group of all permutations of the indices  $(1,2,\dots,n)$  may use  $\det(A)$  instead of  $|A|$ . We may call  $\det(A)$  the permanent of  $A$ .

### 1.3 THE ADJOINT THEORY OF A SQUARE FUZZY MATRIX

In this chapter the adjoint theory of a square fuzzy matrix will be studied .We state a formula for the adjoint matrix of a square fuzzy matrix and this formula shall be used anywhere in this chapter .Then we shall establish the relationship between the adjoints of two fuzzy matrices. Also, we shall find the relationship between the adjoints of two matrices corresponding the relationship between the fuzzy matrices. For a square fuzzy matrix satisfying some property, we shall verify the same property for its adjoint. In this chapter we define the symmetric, reflexive, transitive, circular and idempotent fuzzy matrices and show that some properties of a square fuzzy matrix such as symmetry, reflexivity, transitivity, circularity, and idempotence are carried over to the adjoint matrix and shall understand the same with the help of illustrations. For a given square fuzzy matrix  $A$ , through the adjoint matrix  $\text{adj}(A)$ , we



shall construct a transitive fuzzy matrix  $A$  ( $adj A$ ). Along with its illustration, we shall prove it before. We establish some results including that  $A(adjA) \geq |A|I$  where  $|A|$  denotes the determinant of a square fuzzy matrix  $A$  and  $adjA$  denotes the adjoint matrix of a square fuzzy matrix  $A$ .

**Definition 1.3.1:** The adjoint matrix  $B = [b_{ij}]$  of a square matrix  $A = [a_{ij}]$  of order  $n$ , is a square fuzzy matrix of same order  $n$ , denoted by  $adjA$ , is defined as  $b_{ij} = |A_{ji}|$ ; where  $|A_{ji}|$  is the determinant of the square fuzzy matrix of order  $(n-1)$  obtained from a square fuzzy matrix  $A$  of order  $n$  by deleting row  $j$  and column  $i$  and  $B = [b_{ij}] = adjA$ .

## CHAPTER 2

### FUZZY MATRIX THEORY

#### 2.1 Introduction

The purpose of present chapter is to introduce the basic concept of fuzzy matrix theory. Moreover, we shall consider the operations defined on these matrices for further treatment of determinant and adjoint theory of square fuzzy matrices. First of all, we shall give fuzzy matrix theory and some operations defined on fuzzy matrices.

#### 2.2 Fuzzy Matrix Theory

A fuzzy matrix is a matrix which has its elements from  $[0,1]$ , called fuzzy unit interval.

**Definition 2.2.1:** Consider a matrix  $A=[a_{ij}]_{m \times n}$  where  $a_{ij} \in [0,1]$ ,

$1 \leq i \leq m$  and  $1 \leq j \leq n$ . Then A is fuzzy matrix.

##### 2.2.1 Types of Fuzzy Matrices

###### (a) (i) Fuzzy Rectangular Matrix:

Let  $A=[a_{ij}]_{m \times n}$  ( $m \neq n$ ) where  $a_{ij} \in [0,1]$ ,  $1 \leq i \leq m$  and  $1 \leq j \leq n$

Then A is a fuzzy rectangular matrix.

For example:  $\begin{bmatrix} 0 & 1 & 0.2 \\ 0.1 & 0.5 & 0.3 \end{bmatrix}$  is a 2 x 3 fuzzy rectangular matrix.

###### (ii) Fuzzy Square Matrix:



$$\text{Let } A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}$$

where  $a_{ij} \in [0,1]$ ,  $1 \leq i, j \leq n$ .

Then A is a fuzzy square matrix.

**(iii) Fuzzy Row Matrix:**

Let  $A = [a_1 \ a_2 \ \dots \ a_n]$ ,  $a_j \in [0,1]; j = 1, 2, \dots, n$ .

Then A is called a  $1 \times n$  fuzzy row matrix or fuzzy row vector.

For example:  $[0.3 \ 0.7 \ 0.05 \ 1]$  is  $1 \times 4$  fuzzy row matrix.

**(iv) Fuzzy Column Matrix:**

$$\text{Let } A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

where  $a_i \in [0,1]; i = 1, 2, \dots, m$ .

Then A is called a  $m \times 1$  fuzzy column matrix.

For example:  $\begin{bmatrix} 0 \\ 0.4 \\ 0.5 \end{bmatrix}$  is a  $3 \times 1$  fuzzy column matrix or fuzzy column vector.

**(v) Fuzzy Diagonal Matrix:**

A fuzzy square matrix  $A = [a_{ij}]_{m \times n}$  is said to be fuzzy diagonal matrix if

$a_{ij} = 0$  when  $i \neq j$  where  $[a_{ij}] \in [0,1], 1 \leq i, j \leq n$ .

For example:  $\begin{bmatrix} 0.4 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.9 \end{bmatrix}$  is a fuzzy diagonal matrix of order 3.

This diagonal matrix is also denoted by  $[0.4, 0.3, 0.9]$ .

#### (vi) Fuzzy Scalar Matrix:

A fuzzy diagonal matrix is said to be fuzzy scalar matrix, if all its diagonal entries are equal.

Thus, a fuzzy square matrix  $A = [a_{ij}]_{n \times n}$  is said to be a fuzzy scalar matrix if

$$\begin{cases} a_{ij} = 0 & \text{when } i \neq j \\ a_{ij} = \alpha & \text{when } i = j \end{cases} \text{ where } \alpha \in [0, 1], 1 \leq i, j \leq n.$$

For example:  $[0.3]$  and  $\begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix}$  are fuzzy matrices scalar matrices of order 1 and 2 respectively.

- (b) Usual identity matrix and zero matrix are fuzzy matrices as their entries are from the fuzzy crisp set  $\{0, 1\}$ .
- (c) If the entries in upper triangular matrix and lower triangular matrix are from fuzzy unit interval  $[0, 1]$ , then these matrices are said to be fuzzy upper triangular and fuzzy lower triangular matrices respectively.

### 2.2.2 Equality of Fuzzy Matrices

Two Fuzzy matrices of the same type are said to be equal iff their elements in

The corresponding positions are equal.

## 2.3 Operations on Fuzzy Matrices

### 2.3.1 Operations of Maximum and Minimum

We shall define the following three operations on fuzzy matrices:

- (a) maximum of matrices
- (b) minimum of a matrix by a scalar
- (c) max min of matrices

#### (a) Operations I Maximum of Matrices

If two fuzzy matrices are of the same type, then they are said to be comparable for addition. But the question arises that when we add two fuzzy matrices, then the resultant matrix is not a fuzzy matrix. So in case of fuzzy matrices of same type, max operation is defined.

**Definition 2.3.1:** Let  $A = [a_{ij}]_{m \times n}$  and  $B = [b_{ij}]_{m \times n}$  be two fuzzy matrices.

Then their sum, denoted by  $A+B$ , is defined as

$$A+B = \max\{A, B\}$$

$$\text{i.e., } [a_{ij} + b_{ij}]_{m \times n} = [\max(a_{ij}, b_{ij})]_{m \times n}; \text{ for } 1 \leq i \leq m, 1 \leq j \leq n$$

$$\text{For example: Let } A = \begin{bmatrix} 0 & 0.3 & 0.9 \\ 0.4 & 0.3 & 0.1 \\ 1 & 0.8 & 0.4 \\ 0.5 & 0.2 & 0.6 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0.6 & 0.7 & 0 \\ 0.9 & 0.5 & 0.5 \\ 0.7 & 1 & 0 \\ 0.6 & 0.8 & 0.5 \end{bmatrix}$$

Then

$$A+B = \max\{A, B\} = \begin{bmatrix} \max(0, 0.6) & \max(0.3, 0.7) & \max(0.9, 0) \\ \max(0.4, 0.9) & \max(0.3, 0.5) & \max(0.1, 0.5) \\ \max(1, 0.7) & \max(0.8, 1) & \max(0.4, 0) \\ \max(0.5, 0.6) & \max(0.2, 0.8) & \max(0.6, 0.5) \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & 0.7 & 0.9 \\ 0.9 & 0.5 & 0.5 \\ 1 & 1 & 0.4 \\ 0.6 & 0.8 & 0.6 \end{bmatrix}$$



In a similar way, we define the difference of two fuzzy matrices of same type as the max operation.

Thus, in case of fuzzy matrices of same type,  $A-B = \max \{A, B\} = A+B$ .

### (b) Operation II minimum of a matrix by a scalar

**Definition 2.3.2:** Let  $A = [a_{ij}]_{m \times n}$  be any fuzzy matrix and  $k \in F$ , where  $F = [0,1]$  is a fuzzy unit interval. Then scalar multiple of  $A$  by  $k$ , denoted by  $kA$  or  $Ak$  is given by

$$kA = Ak = [ka_{ij}]_{m \times n} = [\min(k, a_{ij})]_{m \times n}; a_{ij} \in [0,1], 1 \leq i \leq m, 1 \leq j \leq n.$$

Thus  $kA$  or  $Ak$  is the matrix obtained when each entry of  $A$  is multiplied by  $k$ .

For example:  $0.3 \begin{bmatrix} 0.4 & 0.5 & 1 \\ 0.2 & 0.8 & 0.6 \end{bmatrix}$

$$= \begin{bmatrix} \min(0.3, 0.4) & \min(0.3, 0.5) & \min(0.3, 1) \\ \min(0.3, 0.2) & \min(0.3, 0.8) & \min(0.3, 0.6) \end{bmatrix}$$

$$= \begin{bmatrix} 0.3 & 0.3 & 0.3 \\ 0.2 & 0.3 & 0.3 \end{bmatrix}$$

### (c) Operation III max min of matrices

If we wish to find the product  $AB$  of two fuzzy matrices  $A$  and  $B$  where  $A$  and  $B$  are compatible under multiplication,

i.e., number of columns of  $A$  = number of rows of  $B$ ; still we may not have the product  $AB$  to be a fuzzy matrix. So in case of fuzzy matrices compatible under multiplication, max min operation is defined.

**Definition 2.3.3:** Let  $A = [a_{ij}]_{m \times n}$  and  $B = [b_{jk}]_{n \times p}$  be two fuzzy matrices.

Then their product, denoted by  $AB$ , is defined to be the fuzzy matrix  $[c_{ik}]_{m \times p}$ ,

where  $c_{ik} = \sum_{j=1}^n a_{ij} b_{jk} = \max\{\min(a_{ij}, b_{jk}); 1 \leq i \leq m, 1 \leq k \leq p\}$  for

$j = 1, 2, \dots, n$ .

**Remark:** If the fuzzy product AB is defined then BA may not be defined.

For example: Let  $A = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.4 & 0 & 0.6 \end{bmatrix}_{2 \times 3}$  and  $B = \begin{bmatrix} 0 & 0.3 & 0.7 & 0.5 \\ 0.4 & 0.7 & 0.4 & 1 \\ 0.6 & 0.3 & 1 & 0.1 \end{bmatrix}_{3 \times 4}$

Since no. of columns of A = no. of rows in B

Then the product AB is defined and is given by

$$AB = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \end{bmatrix}_{2 \times 4}$$

where  $c_{11} = \max\{\min(0.3, 0), \min(0.2, 0.4), \min(0.5, 0.6)\}$

$$= \max\{0, 0.2, 0.5\}$$

$$= 0.5$$

$c_{12} = \max\{\min(0.3, 0.3), \min(0.2, 0.7), \min(0.5, 0.3)\}$

$$= \max\{0.3, 0.2, 0.3\}$$

$$= 0.3$$

$c_{13} = \max\{\min(0.3, 0.7), \min(0.2, 0.4), \min(0.5, 1)\}$

$$= \max\{0.3, 0.2, 0.5\}$$

$$= 0.5$$

$c_{14} = \max\{\min(0.3, 0.5), \min(0.2, 1), \min(0.5, 0.1)\}$

$$= \max \{0.3, 0.2, 0.1\}$$

$$= 0.3$$

$$c_{21} = \max \{ \min (0.4, 0), \min (0, 0.4), \min (0.6, 0.6) \}$$

$$= \max \{0, 0, 0.6\}$$

$$= 0.6$$

$$c_{22} = \max \{ \min (0.4, 0.3), \min (0, 0.7), \min (0.6, 0.3) \}$$

$$= \max \{0.3, 0, 0.3\}$$

$$= 0.3$$

$$c_{23} = \max \{ \min (0.4, 0.7), \min (0, 0.4), \min (0.6, 1) \}$$

$$= \max \{0.4, 0, 0.6\}$$

$$= 0.6$$

$$c_{24} = \max \{ \min (0.4, 0.5), \min (0, 1), \min (0.6, 0.1) \}$$

$$= \max \{0.4, 0, 0.1\}$$

$$= 0.4$$

$$\text{Thus } AB = \begin{bmatrix} 0.5 & 0.3 & 0.5 & 0.3 \\ 0.6 & 0.3 & 0.6 & 0.4 \end{bmatrix}$$

But since, no. of columns of B  $\neq$  no. of rows of A.

Thus the product BA is not defined.

Using the max-min function, we can find the positive integral powers of a square fuzzy matrix.



### 2.3.2 Transpose of fuzzy matrix

**Definition 2.3.4:** Let  $A = [a_{ij}]_{m \times n}$  be any fuzzy matrix. Then the transpose of  $A$ , denoted by  $A'$  or  $A^t$  or  $A^T$ , is  $n \times m$  fuzzy matrix obtained from  $A$  by interchanging its rows and columns.

i.e.,  $A' = [b_{ij}]_{n \times m}$  where  $b_{ij} = a_{ji} \in [0,1]$ ; for  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ .

**Remarks:** 1. The transpose of a fuzzy row matrix is a fuzzy column matrix and vice-versa.

2. The product  $AA'$  and  $A'A$  of two fuzzy matrices are always defined.

**For example:** Let  $A = \begin{bmatrix} 0.2 \\ 0.6 \\ 0.7 \end{bmatrix}_{3 \times 1}$  be a fuzzy row matrix.

Then  $A' = [0.2 \quad 0.6 \quad 0.7]_{1 \times 3}$  is a fuzzy row matrix.

Observe that the products  $AA'$  and  $A'A$  of two fuzzy matrices  $A$  and  $A'$  are defined.

$$\begin{aligned} \text{Now } A'.A &= [0.2 \quad 0.6 \quad 0.7] \begin{bmatrix} 0.2 \\ 0.6 \\ 0.7 \end{bmatrix} \\ &= \max\{0.2, 0.6, 0.7\} \\ &= 0.7 \end{aligned}$$

Then  $A'.A$  is a singleton fuzzy matrix.

$$\text{and } A'.A = \begin{bmatrix} 0.2 \\ 0.6 \\ 0.7 \end{bmatrix} [0.2 \quad 0.6 \quad 0.7]$$

$$= \begin{bmatrix} \min(0.2, 0.2) & \min(0.2, 0.6) & \min(0.2, 0.7) \\ \min(0.6, 0.2) & \min(0.6, 0.6) & \min(0.6, 0.7) \\ \min(0.7, 0.2) & \min(0.7, 0.6) & \min(0.7, 0.7) \end{bmatrix}$$

$$= \begin{bmatrix} 0.2 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.6 \\ 0.2 & 0.6 & 0.7 \end{bmatrix}$$

Then  $A.A'$  is a **symmetric fuzzy matrix** as the matrix obtained by interchanging its rows and columns is the matrix itself and the elements of the matrix belong to fuzzy unit interval  $[0,1]$ .

## CHAPTER 3

### THE DETERMINANT THEORY OF A SQUARE FUZZY MATRIX

#### 3.1 Introduction

The purpose of this chapter is to introduce and investigate some properties of the determinant of a square fuzzy matrix. The properties of a square fuzzy matrix are somewhere analogous to the crisp case for determinant of a square matrix. We shall introduce the determinant theory of a square fuzzy matrix and their properties and investigate the elementary properties of determinant theory for fuzzy matrices in fuzzy matrix semiring  $M_n(F)$

#### 3.2 The Determinant Theory of a square Fuzzy Matrix

This section lays down the foundation of the determinant theory of a square fuzzy matrix where the elements takes its values from the unit interval  $[0,1]$ .

**Definition 3.2.1:** The determinant  $|A|$  of an  $n \times n$  fuzzy matrix  $A$  is defined as follows:

$$\begin{aligned}|A| &= \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \prod_{i=1}^n a_{i\sigma(i)}\end{aligned}$$

where  $S_n$  denotes the symmetric group of all permutations of the indices  $(1,2,\dots,n)$ , we may use  $\det(A)$  instead of  $|A|$ . We may call  $\det(A)$  the permanent of  $A$ .

#### Preliminaries:

1. It may be noted with care that the non-square fuzzy matrices do not have determinants.

2. The elements of the determinant of a fuzzy matrix takes its values from the unit interval  $[0,1]$ .

3. Here multiplication and addition takes respectively the meanings of min. and max. operations as defined usually in fuzzy matrices.

(i) The determinant of a  $1 \times 1$  fuzzy matrix  $[a]$  is denoted by  $|a|$  and is defined as  $a$ .

(ii) The determinant of a  $2 \times 2$  fuzzy matrix  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$  is denoted by  $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$  and is defined as  $ad+bc$  or  $\max \{ \min\{a,d\}, \min\{b,c\} \}$ .

(iii) The determinant of a  $3 \times 3$  fuzzy matrix

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \text{ is denoted by } \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

and is defined as

$$a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} + a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}$$

$$\text{i.e., } a_1(b_2c_3 + b_3c_2) + a_2(b_1c_3 + b_3c_1) + a_3(b_1c_2 + b_2c_1)$$

$$\text{or } a_1 \max \{ \min(b_2, c_3), \min(b_3, c_2) \} + a_2 \max \{ \min(b_1, c_3), \min(b_3, c_1) \}$$

$$+ a_3 \max \{ \min(b_1, c_2), \min(b_2, c_1) \}$$

$$= a_1\lambda + a_2\mu + a_3\nu, \text{ where } \lambda = \max \{ \min(b_2, c_3), \min(b_3, c_2) \}$$

$$\mu = \max \{ \min(b_1, c_3), \min(b_3, c_1) \}$$

$$\nu = \max \{ \min(b_1, c_2), \min(b_2, c_1) \}$$

$$= \min(a_1, \lambda) + \min(a_2, \mu) + \min(a_3, \nu)$$

$$= \max \{ \min(a_1, \lambda), \min(a_2, \mu), \min(a_3, \nu) \}$$



We can expand the determinant along any other row or column as the same value of the determinant can be obtained by expanding along any row or column. It can be easily verified that expanding along any row or column. It can be easily verified that expanding along any row or column gives the value of the determinant as

$$a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 + a_1b_3c_2 + a_2b_1c_3 + a_3b_2c_1$$

**Notation 3.2.1:** The determinant of a square fuzzy matrix  $A$  of order  $n$  is defined as follows:

$$\begin{aligned} \det(A) = |A| &= \sum_{j=1}^n a_{ij} A_{ij}, \quad i \in \{1, 2, \dots, n\} \quad [\text{Expanding along } i\text{th row}] \\ &= \sum_{i=1}^n a_{ij} A_{ij}, \quad j \in \{1, 2, \dots, m\} \quad [\text{Expanding along } j\text{th column}] \end{aligned}$$

where  $A_{ij}$  is the determinant of the fuzzy matrix of order  $(n - 1)$  obtained from a square fuzzy matrix  $A$  of order  $n$  by deleting (striking out) row  $i$  and column  $j$ .

**Explanation:** Consider a  $3 \times 3$  matrix

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

$$A_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad A_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, \quad A_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Then the value of the  $\det(A)$  is given by

$$|A| = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}.$$

We can expand the determinant along any other row or column.

The value of the determinant does not depend on along which row or column it is expanded.

Thus we have:

$$\begin{aligned}
\Delta &= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} \text{ [Expanding by Ist row]} \\
&= a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23} \text{ [Expanding by IInd row]} \\
&= a_{31}A_{31} + a_{32}A_{32} + a_{33}A_{33} \text{ [Expanding by IIIrd row]} \\
&= a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31} \text{ [Expanding by Ist column]} \\
&= a_{12}A_{12} + a_{22}A_{22} + a_{32}A_{32} \text{ [Expanding by IInd column]} \\
&= a_{13}A_{13} + a_{23}A_{23} + a_{33}A_{33} \text{ [Expanding by IIIrd column]}
\end{aligned}$$

**Remark:** It may be noted with care that the value of the determinant of a square fuzzy matrix A is one among the elements of A.

Thus  $0 \leq \det(A) \leq 1$ .

**Example 3.2.1:** For a square fuzzy matrix

$$A = \begin{bmatrix} 0.5 & 0.3 & 0.8 \\ 0.6 & 0.2 & 0.9 \\ 0 & 0.7 & 0.4 \end{bmatrix}$$

We calculate the determinant  $|A|$  as follows:

$$\begin{aligned}
|A| &= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} \\
&= 0.5 \begin{vmatrix} 0.2 & 0.9 \\ 0.7 & 0.4 \end{vmatrix} + 0.3 \begin{vmatrix} 0.6 & 0.9 \\ 0 & 0.4 \end{vmatrix} + 0.8 \begin{vmatrix} 0.6 & 0.2 \\ 0 & 0.7 \end{vmatrix} \\
&= 0.5(0.2+0.7) + 0.3(0.4+0) + 0.8(0.6+0) \\
&= 0.5+0.3+0.6 \\
&= 0.6
\end{aligned}$$

It can be easily verified that the same value of the determinant can be obtained by expanding along any other row or column.



### 3.3 Properties of Determinants of Square Fuzzy Matrices

The following properties of determinant are true for determinants of any order.

However we shall show these for determinants of order 3. These properties are used in order to simplify the determinant before expanding it. Some properties of determinants of square fuzzy matrices are analogous to the properties of determinants of square matrices while some other properties differ a lot and a few properties when considered in fuzzy matrix semiring  $M_n(F)$  give rise to some new problems and sorting out these problems establishes some new results in fuzzy matrix semiring  $M_n(F)$ . before we proceed, it is worth to note that whatever operation or result is true for rows, is also true for columns.

**Property 3.3.1** The value of the determinant remains unaltered by interchanging

its rows and columns.

**Proof:** Let

$$\Delta = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

Expanding along first row, we get

$$\begin{aligned} \Delta &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} + a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \\ &= a_1(b_2c_3 + b_3c_2) + a_2(b_1c_3 + b_3c_1) + a_3(b_1c_2 + b_2c_1) \end{aligned}$$

Interchanging the rows and columns of  $\Delta$ , then the new determinant is

$$\Delta' = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

Expanding along the first column, we get:

$$\begin{aligned}\Delta' &= a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} + a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} \\ &= a_1(b_2c_3 + b_3c_2) + a_2(b_1c_3 + b_3c_1) + a_3(b_1c_2 + b_2c_1) \\ &= \Delta.\end{aligned}$$

Thus we have  $\Delta' = \Delta$

**Remark:** Interchange of the rows and columns does not change the value of the determinant, i.e., if A is a square fuzzy matrix, then  $\det(A) = \det(A')$  where A' denotes the transpose of the square fuzzy matrix A.

**Property 3.3.2** The value of the determinant remains unaltered interchanging its any two rows (or columns).

**Proof:** Let  $\Delta = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$

Expanding along the first row, we get:

$$\begin{aligned}\Delta &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} + a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \\ &= a_1(b_2c_3 + b_3c_2) + a_2(b_1c_3 + b_3c_1) + a_3(b_1c_2 + b_2c_1)\end{aligned}$$

Interchanging the first and second rows, then the new determinant is

$$\Delta' = \begin{vmatrix} b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Expanding along second row, we get:

$$\Delta' = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} + a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_1 \\ c_1 & c_2 \end{vmatrix}$$

$$= a_1(b_2c_3 + b_3c_2) + a_2(b_1c_3 + b_3c_1) + a_3(b_1c_2 + b_2c_1)$$

$$= \Delta.$$

Thus, we have  $\Delta' = \Delta$ .

**Property 3.3.3** If all rows (or columns) of a determinant are identical then its value

is minimum element among all elements of the determinant.

**Proof:** Let  $\Delta = \begin{vmatrix} a & b & c \\ a & b & c \\ a & b & c \end{vmatrix}$

$$= a(bc + bc) + b(ac + ac) + c(ab + ab)$$

$$= a(bc) + b(ac) + c(ab)$$

$$= abc + abc + abc$$

$$= abc$$

$$= \min \{a, b, c\}$$

$$= \text{minimum value among all elements of the determinant}$$

**Definition 3.3.1 :** An  $m \times n$  fuzzy matrix A is said to be constant if  $a_{ik} = a_{jk}$  (or

$a_{ki} = a_{kj}$ ) for all  $i, j, k$  i.e., its rows (or columns) are equal to each other.

**Note:** Thus property 3.3.3 can be stated as:

The determinant of a constant square fuzzy matrix is its minimum element.

**Property 3.3.4** If all the elements of a row (or column) are equal to  $a$  (say) and all

other rows (or columns) have elements  $\geq a$  then the value of the

determinant  $= a$  (which is the minimum element among all elements).



**Proof:** We know that each term in the determinant contains a factor of each row (or column) and hence each term contains a factor of that (or column) in which all elements are equal to  $a$  (say) and all other elements in that term of other rows (or columns)  $\geq a$  so that each term in the determinant of a square fuzzy matrix is equal to  $a$  and consequently  $\Delta = a$ .

**For example:** Let  $\Delta = \begin{vmatrix} a & a & a \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$

$$= a(b_2c_3 + b_3c_2) + a(b_1c_3 + b_3c_1) + a(b_1c_2 + b_2c_1)$$

$$= a(b_2c_3 + b_3c_2 + b_1c_3 + b_3c_1 + b_1c_2 + b_2c_1)$$

$$= a \quad \left[ \begin{array}{l} \because \text{each element} \geq a \\ \therefore \text{min element} = a \end{array} \right]$$

#### Another statement of Property 3.3.4

Let  $A_k = (a_{k1}, a_{k2}, \dots, a_{kn})$  be the  $k$ th row of A. if  $a_{ki} = a \forall i \in \{1, 2, \dots, n\}$

and  $a_{pq} \geq a \forall p, q \in \{1, 2, \dots, n\}$

Then  $\det(A) = a$ .

**Corollary:** The value of the determinant containing a zero row (or column) is zero.

**Proof:** We know that each term in the determinant contains a factor of each row (or column) and hence contains a factor of zero row (or column) so that each term in the determinant of a square fuzzy matrix is equal to zero (min. element) and consequently  $\Delta = 0$ .

**Example:** Let  $\Delta = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ 0 & 0 & 0 \end{vmatrix}$

$$= a_1(0 + 0) + a_2(0 + 0) + a_3(0 + 0)$$

$$= a_1(0) + a_2(0) + a_3(0)$$

$$= 0+0+0$$

$$= 0$$

**Note:** In a square fuzzy matrix, as all the elements of a zero row (or column) are equal to 0 and all other elements  $\geq 0$ . Hence the property 3.3.4, the value of the determinant=0.

**Property 3.3.5** The determinant of a diagonal elements.

**Proof:** Let  $A=[a_{ij}]$  be a diagonal matrix i.e.,  $a_{ij} = 0$  for  $i \neq j$

Take a term  $t$  of  $|A|$ ,

$$t = a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

Let  $\sigma(1) \neq 1$ , then  $a_{1\sigma(1)} = 0$  and thus  $t=0$ ,

This means that each term is zero if  $\sigma(1) \neq 1$ .

Let now  $\sigma(1) = 1$  but  $\sigma(2) \neq 2$ , then  $a_{2\sigma(2)} = 0$  and thus  $t = 0$ .

This mean each term is zero if  $\sigma(1) \neq 1$  or  $\sigma(2) \neq 2$ .

However in the similar way, we can see that each term for which  $\sigma(1) \neq 1$  or

$\sigma(2) \neq 2 \dots$  or  $\sigma(n) \neq n$  must be zero

Consequently  $|A| = a_{11} a_{22} \dots a_{nn} = \prod_{i=1}^n a_{ii}$



= Product of its diagonal elements.

**Corollary 1:** The determinant of a scalar matrix with each of its non-zero (diagonal) element.

**Proof:** Let  $\Delta$  be the determinant of a scalar matrix with each of its non-zero diagonal element equal to  $k$  (say).

Since every scalar matrix is a diagonal matrix.

$\therefore$  By property 3.3.5,  $\Delta$  = product of its diagonal elements.

$$= k \cdot k \dots k$$

$$= k.$$

**Corollary 2** The determinant of an identity matrix is unity.

**Proof:** Since an identity is a diagonal matrix.

$\therefore$  By property 3.3.5,  $|I|$  = products of its diagonal elements.

$$= 1 \cdot 1 \dots \dots 1$$

$$= 1$$

**Corollary 3** The determinant of the matrix obtained from an identity matrix by interchanging its any two rows (or columns) is unity.

**Proof:** By property 3.3.2, the value of determinant remains unaltered by interchanging its any two rows (or columns).

$\therefore |E_{ij}| = |I_n|$  where  $E_{ij}$  is the matrix obtained from identity matrix  $I_n$  by interchanging its rows (or columns).

i.e.,  $|E_{ij}| = 1$ . (by corollary 2)

**Remarks:** 1.  $\det(I_n(A)) = \det(A) = \det(AI_n)$

2.  $\det(E_{ij}A) = \det(A) = \det(AE_{ij})$

**Property 3.3.6** The determinant of a triangular matrix is given by the product of its Diagonal elements.

**Proof:** Suppose  $\Delta = |a_{ij}|$  is triangular from below.

i.e.,  $a_{ij} = 0$  for  $i < j$

Take a term  $t$  of  $\Delta$ ,

$$t = a_{1\sigma(1)}a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

Let  $\sigma(1) \neq 1$ , so that  $1 < \sigma(1)$  and so  $a_{1\sigma(1)} = 0$  and thus  $t=0$ .

This means that each term is zero if  $\sigma(1) \neq 1$ .

Let now  $\sigma(1) = 1$  but  $\sigma(2) \neq 2$ , then  $2 < \sigma(2)$  and

$$a_{2\sigma(2)} = 0 \text{ and thus } t=0.$$

This means that each term is zero if  $\sigma(1) \neq 1$  or  $\sigma(2) \neq 2$ .

However, in a similar manner we can see that each term must be zero if

$$\sigma(1) \neq 1, \text{ or } \sigma(2) \neq 2 \dots \dots \text{ or } \sigma(n) \neq n.$$

$$\text{Consequently } \Delta = a_{11}a_{22} \dots a_{nn} = \prod_{i=1}^n a_{ii}$$

=Product of its diagonal elements.

Similar is the case with the determinant triangular from above

**Property 3.3.7:** If each element of a row (or column) of a determinant is

Multiplied by constant  $k \in F$  where  $F = [0,1]$  is the fuzzy unit

Interval of the real line i.e.,  $k \in [0,1]$  then its value gets multiplied by  $k$ .

**Proof:** By definition of determinant,

$$\begin{aligned}\Delta &= \sum_{\sigma \in S_n} \prod_{i=1}^n a_{i\sigma(i)} \\ &= \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{i\sigma(i)} \dots a_{n\sigma(n)}\end{aligned}$$

Multiply by  $k$ , the elements of the  $i$ th row (say) then the new Determinant is

$$\begin{aligned}\Delta &= \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \dots k a_{i\sigma(i)} \dots a_{n\sigma(n)} \\ &= k \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{i\sigma(i)} \dots a_{n\sigma(n)} \\ &= k \sum_{\sigma \in S_n} \prod_{i=1}^n a_{i\sigma(i)} \\ &= k\Delta\end{aligned}$$

**Remark:** In case of matrices  $kA$  is the matrix obtained when each entry of  $A$  is

Multiplied by  $k \in F = [0,1]$ .

Thus, if  $A$  is a square matrix of order  $n$ , then

$$|kA| = k^n |A| \text{ i.e., } \det(kA) = k^n \det(A) = k \det(A).$$



## CHAPTER 4

### THE ADJOINT THEORY OF A SQUARE FUZZY MATRIX

#### 4.1 Introduction

In this chapter the adjoint theory of a square fuzzy matrix will be studied. The adjoint of a square fuzzy matrix is defined by Thomason [13] and Kim[11]. We state a formula for the adjoint matrix of a square fuzzy matrix and this formula shall be used anywhere in this chapter. Then we shall establish the relationship between the adjoints of two fuzzy matrices. Also, we shall find the relationship between the adjoints of two matrices corresponding the relationship between the fuzzy matrices. For a square fuzzy matrix satisfying some property, we shall verify the same property for its adjoint. In this chapter we define the symmetric, reflexive, transitive, circular and idempotent fuzzy matrices and show that some properties of a square fuzzy matrix such as symmetry, reflexivity, transitivity, circularity, and idempotence are carried over to the adjoint matrix and shall understand the same with the help of illustrations. For a given square fuzzy matrix  $A$ , through the adjoint matrix  $\text{adj}(A)$ , we shall construct a transitive fuzzy matrix  $A(\text{adj } A)$ . Along with its illustration, we shall prove it before. We establish some results including that  $A(\text{adj } A) \geq |A|I$  where  $|A|$  denotes the determinant of a square fuzzy matrix  $A$  and  $\text{adj } A$  denotes the adjoint matrix of a square fuzzy matrix  $A$ .

#### 4.2 The Adjoint Theory of a Square Fuzzy Matrix

Let us first define the notion of adjoint of square fuzzy matrix.

**Definition 4.2.1:** The adjoint matrix  $B = [b_{ij}]$  of a square matrix  $A = [a_{ij}]$  of order  $n$ , is a square fuzzy matrix of same order  $n$ , denoted by  $\text{adj } A$ , is defined as  $b_{ij} = |A_{ji}|$ ;

where  $|A_{ji}|$  is the determinant of the square fuzzy matrix of order (n-1) obtained from a square fuzzy matrix A of order n by deleting row  $j$  and column  $i$  and

$$B=[b_{ij}]=adjA.$$

**Remark:** Note that  $|A_{ji}|$  can be obtained from  $|A|$  by replacing the element  $a_{ji}$  of A by 1 and all other row  $j$  factors  $a_{jk}$ ,  $k \neq i$  by 0.

**Example 4.2.1:** For a square fuzzy matrix

$$A=\begin{bmatrix} 0 & 0.3 & 0.4 \\ 0.2 & 0.4 & 0.5 \\ 1 & 0.3 & 0.7 \end{bmatrix}$$

We find  $adjA$  as follows:

$$b_{11}=|A_{11}|=\begin{vmatrix} 0.4 & 0.5 \\ 0.3 & 0.7 \end{vmatrix} = 0.4+0.3 = 0.4$$

$$b_{12}=|A_{21}|=\begin{vmatrix} 0.3 & 0.4 \\ 0.3 & 0.7 \end{vmatrix} = 0.3+0.3 = 0.3$$

$$b_{13}=|A_{31}|=\begin{vmatrix} 0.3 & 0.4 \\ 0.4 & 0.5 \end{vmatrix} = 0.3+0.4 = 0.4$$

$$b_{21}=|A_{12}|=\begin{vmatrix} 0.2 & 0.5 \\ 1 & 0.7 \end{vmatrix} = 0.2+0.5 = 0.5$$

$$b_{22}=|A_{22}|=\begin{vmatrix} 0 & 0.4 \\ 1 & 0.7 \end{vmatrix} = 0 + 0.4 = 0.4$$

$$b_{23}=|A_{32}|=\begin{vmatrix} 0 & 0.4 \\ 0.2 & 0.5 \end{vmatrix} = 0 + 0.2 = 0.2$$

$$b_{31}=|A_{13}|=\begin{vmatrix} 0.2 & 0.4 \\ 1 & 0.3 \end{vmatrix} = 0.2 + 0.4 = 0.4$$

$$b_{32}=|A_{23}|=\begin{vmatrix} 0 & 0.4 \\ 1 & 0.7 \end{vmatrix} = 0 + 0.4 = 0.4$$

$$b_{33}=|A_{33}|=\begin{vmatrix} 0 & 0.3 \\ 0.2 & 0.4 \end{vmatrix} = 0 + 0.2 = 0.2$$



$$adjA = [b_{ij}] = \begin{bmatrix} 0.4 & 0.3 & 0.4 \\ 0.5 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \end{bmatrix}$$

### 4.3 Properties of Adjoints of Square Fuzzy Matrices

**Notation 4.3.1 :** We also can rewrite the element  $b_{ij}$  of  $adj A$  as

$$b_{ij} = \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} a_{t\sigma(t)}$$

Where  $n_j = \{1, 2, 3, \dots, n\} \setminus \{j\}$

$$n_i = \{1, 2, 3, \dots, n\} \setminus \{i\}$$

and  $S_{n_j n_i}$  is the set of all permutations of set  $n_j$  over the set  $n_i$ .

#### Proposition 4.3.1: Comparison of the adjoints of two fuzzy matrices

For  $n \times n$  fuzzy matrices  $A$  and  $B$ ,

- (i)  $A \leq B \Rightarrow adj A \leq adj B$
- (ii)  $adj A + adj B \leq adj(A + B)$

**Proof:** Let  $A = [a_{ij}]$  and  $B = [b_{ij}]$  where  $i, j \in \{1, 2, \dots, n\}$ .

- (i) since  $A \leq B$

$$\Rightarrow a_{ij} \leq b_{ij} \forall i, j \in \{1, 2, 3, \dots, n\}$$

$$\Rightarrow a_{t\sigma(t)} \leq b_{t\sigma(t)} \text{ for every } t \neq j, \sigma(t) \neq i$$

$$\Rightarrow \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} a_{t\sigma(t)} \leq \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} b_{t\sigma(t)}$$

$$\Rightarrow adj A \leq adj B$$

- (ii) since  $A, B \leq A + B$

$$[\because A+B = \max \{A, B\}]$$

$$\begin{aligned} &\Rightarrow adj A, adj B \leq adj(A+B) & [\because A \leq B \Rightarrow adj A \leq adj B] \\ &\Rightarrow adj A + adj B \leq adj(A+B) \end{aligned}$$

**Proposition 4.3.2 :** The adjoint of the transpose of a matrix is the transpose of the adjoint of the matrix. i.e., for a square fuzzy matrix  $A$  of order  $n$ ,  $adj A' = (adj A)'$ .

**Proof :** Let  $A = [a_{ij}]$  be a  $n \times n$  fuzzy matrix.

$$\text{Let } B = [b_{ij}] = adj A \text{ and } C = [c_{ij}] = adj A'.$$

$$\text{Then } b_{ij} = \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} a_{t\sigma(t)}$$

$$\text{and } c_{ij} = \sum_{\sigma \in S_{n_j n_i}} \prod_{\sigma(t) \in n_i} a_{t\sigma(t)}, \text{ which is element } b_{ji}.$$

Hence  $adj A' = (adj A)'$ , which proves the assertion.

**Proposition 4.3.3 :** Let  $A$  be a  $n \times n$  fuzzy matrix. Then

- (i)  $A (adj A) \geq |A| I_n$
- (ii)  $(adj A) A \geq |A| I_n$

Where  $I_n$  is a unit matrix of order  $n$ .

**Proof (i):** Let  $A = [a_{ij}]_{n \times n}$ , then  $i$ th row of  $A$  is given by  $(a_{i1}, a_{i2}, \dots, a_{in})$ .

Suppose  $B = [b_{ij}]_{n \times n} = adj A$

Then by definition of  $adj A$ ,

The  $j$ th column of  $B = [b_{ij}]_{n \times n} = adj A$  is given by

$$(b_{1j}, b_{2j}, \dots, b_{nj}) = (|A_{j1}|, |A_{j2}|, \dots, |A_{jn}|).$$

Let  $C = [c_{ij}]_{n \times n} = A(adj A)$

Then  $(i,j)th$  element of  $C = [c_{ij}]_{n \times n} = A(adjA)$  is given by

$$c_{ij} = \sum_{k=1}^n a_{ik} |A_{jk}| \geq 0$$

And where  $c_{ii} = \sum_{k=1}^n a_{ik} |A_{ik}| = |A|$

$$|A|I_n = \begin{bmatrix} |A| & 0 & 0 & \dots & 0 \\ 0 & |A| & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & |A| \end{bmatrix}$$

(ii) Let  $A = [a_{ij}]_{n \times n}$

Then  $jth$  column of  $A$  is  $(a_{1j}, a_{2j}, \dots, \dots, a_{nj})$ .

Then by definition of  $adjA$ ,

The  $ith$  row of  $B = [b_{ij}]_{n \times n} = adjA$  is  $(b_{i1}, b_{i2}, \dots, \dots, b_{in})$

$= (|A_{1i}|, |A_{2i}|, \dots, \dots, |A_{ni}|)$

Let  $C = [c_{ij}]_{n \times n} = (adjA)A$

Then  $(i,j)th$  element of  $C = [c_{ij}]_{n \times n} = (adjA)A$  is

$$c_{ij} = \sum_{k=1}^n |A_{ki}| a_{kj} \geq 0$$

And where  $c_{ii} = \sum_{k=1}^n |A_{ki}| a_{ki} = |A|$ .

Thus  $C = (adjA)A \geq |A|I_n$  where  $I_n$  is a unit matrix of order  $n$ .

**Remark:** Any diagonal element of the fuzzy matrix  $A(adjA)$  is  $|A|$  and non-diagonal element  $\geq 0$ .

**Proposition 4.3.4:** Let A be a square fuzzy matrix, then the following properties hold:

- (i) If A has a zero row then  $(adj A)A = O$  (the zero matrix)
- (ii) If A has a zero column then  $A(adj A) = O$  (the zero matrix)

**Proof:** Let  $C = [c_{ij}] = (adj A)A$ , then  $c_{ij} = \sum_k |A_{ki}| a_{kj}$

If the  $i$ th row of A is zero, then  $a_{kj} = 0$  for every  $k=i$

And for  $k \neq i$ ,  $A_{ki}$  contains a zero row and so  $|A_{ki}| = 0$  for  $k \neq i$

(by corollary of property 3.3.4 of determinants)

So that  $\sum_k |A_{ki}| a_{kj} = 0 \forall i, j$ .

Hence  $A(adj A) = O$  (the zero matrix).

**Proposition 4.3.5:** Let A be an  $n \times n$  constant fuzzy matrix, then

- (i)  $(adj A)'$  is constant.
- (ii)  $C = A(adj A)$  is constant and  $c_{ij} = |A|$ , which is the least element in |

**Proof:** Let A be an  $n \times n$  constant fuzzy matrix where its all rows are equal to each-other. i.e.,  $a_{ik} = a_{jk} \forall i, j$

- (i) Let  $B = (adj A)$ , then

$$b_{ij} = \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} a_{t\sigma(t)} \text{ and } b_{ik} = \sum_{\sigma \in S_{n_k n_i}} \prod_{t \in n_k} a_{t\sigma(t)}$$

Since the numbers  $\sigma(t)$  of columns cannot be changed in the two expansions of  $b_{ij}$  and  $b_{ik}$  as A is constant and so  $b_{ij} = b_{ik} \forall i, j, k$ .

In order that,  $b_{ji} = b_{ki} \forall i, j, k$ , we must have  $(adj A)'$  is constant.



(ii) Since A is constant i.e.,  $a_{ik} = a_{jk} \forall i, j, k$ .

Then  $A_{ik} = A_{jk} \forall i, j, k$  and so  $|A_{ik}| = |A_{jk}| \forall i, j, k$ .

Let  $C = [c_{ij}] = (\text{adj}A)A$ , then

$$\begin{aligned} c_{ij} &= \sum_{k=1}^n a_{ik} |A_{jk}| = \sum_{k=1}^n a_{ik} |A_{ik}| \\ &= |A| \forall i, j \end{aligned}$$

Thus  $C = A(\text{adj}A)$  is constant.

Now  $|A| = \sum_{\sigma \in S_n} \prod_{i=1}^n a_{i\sigma(i)}$

$$= \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

$$= a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)} \text{ for any } \sigma \in S_n$$

$$(\because A \text{ is constant i.e., } a_{ik} = a_{jk} \forall i, j, k)$$

Taking  $\sigma$ , the identity permutation i.e.,  $\sigma(i) = i \forall i$ , we get

$|A| = a_{11} a_{22} \dots a_{nn}$ ; which is the least element in a constant fuzzy matrix A.

**Definition 4.3.1:** Let A be a square fuzzy matrix of order n, then following hold:

- (i) A is said to be **reflexive** fuzzy matrix iff  $A \geq I_n$  i.e., iff all diagonal elements in fuzzy matrix A are unity i.e., iff  $a_{ii} = 1 \forall i$ .
- (ii) A is said to be **symmetric** iff  $A' = A$  i.e., iff the square fuzzy matrix A remains unaltered by interchanging its row and columns i.e., iff  $a_{ij} = a_{ji} \forall i, j \in \{1, 2, \dots, n\}$ .
- (iii) A is said to be **transitive** iff  $A^2 \leq A$  i.e., iff the square fuzzy matrix A multiplied by itself gives the elements less than or equal to the



corresponding elements of the square fuzzy matrix A. i.e., iff  $a_{ik}a_{kj} \leq a_{ij}$

for every  $k=1,2,\dots,n$

A square fuzzy matrix is **similarity (equivalence relation)** iff it is reflexive, symmetric and transitive.

Let us understand the same with help of an example.

**Example 4.3.1:** Let A be a square fuzzy matrix of order 3.

(i) Consider a square fuzzy matrix

$$A = \begin{bmatrix} 1 & 0.2 & 0 \\ 0.3 & 1 & 0.4 \\ 0.9 & 0.3 & 1 \end{bmatrix}$$

Since all the diagonal elements in square fuzzy matrix are unity, then A is a reflexive fuzzy matrix.

(ii) Consider a square fuzzy matrix

$$A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.4 & 0.6 & 0.1 \\ 0.5 & 0.1 & 1 \end{bmatrix}$$

Then

$$A' = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.4 & 0.6 & 0.1 \\ 0.5 & 0.1 & 1 \end{bmatrix} = A$$

Thus A is a symmetric fuzzy matrix

(iii) Consider a square fuzzy matrix

$$A = \begin{bmatrix} 0.6 & 0.7 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.7 & 0.6 \end{bmatrix}$$

Then

$$A^2 = \begin{bmatrix} 0.6 & 0.7 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.7 & 0.6 \end{bmatrix} \begin{bmatrix} 0.6 & 0.7 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.7 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & 0.6 & 0.6 \\ 0.5 & 0.5 & 0.5 \\ 0.6 & 0.6 & 0.6 \end{bmatrix} \leq A$$

As the element in  $A^2$  are  $\leq$  the corresponding elements in A. Thus, A is a transitive fuzzy matrix. Next, consider the square fuzzy matrix.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

All diagonal elements in A equal to 1 implies that A is reflexive. Further, observe that  $A' = A$  so A is symmetric. Also  $A^2 = A$  leads to the conclusion that A is similarity.

Let us see how the properties of a square fuzzy matrix are carried over to its adjoint.

**Theorem 4.3.1:** Let A be a square fuzzy matrix of order  $n$ , Then we have the following properties:

- (i) If A is reflexive, then  $adj A$  is reflexive.
- (ii) If A is symmetric, then  $adj A$  is symmetric.
- (iii) If A is transitive, then  $adj A$  is transitive.

**Proof:** (i) Since  $A = [a_{ij}]$  is reflexive, then  $a_{ii} = 1 \forall i$

$$\text{Let } B = [b_{ij}] = adj A$$

$$\text{Then } b_{ij} = \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} a_{t\sigma(t)}$$

$$\text{And so } b_{ii} = \sum_{\sigma \in S_{n_i}} \prod_{t \in n_j} a_{t\sigma(t)}$$

Taking only the identity permutation  $\sigma(t) = t$ ; we get

$$b_{ii} = a_{11}a_{22} \dots a_{(i-1)(i-1)}a_{(i+1)(i+1)} \dots a_{nn}$$

$$\text{i.e; } b_{ii} \geq 1 \forall i \quad a_{ii} = 1 \forall i$$

and so  $b_{ii} \geq 1 \forall i$

Hence  $\text{adj } A$  is reflexive.

(ii) Since  $A$  is symmetric, then  $a_{ij} = a_{ji} \forall i, j$ .

$$\text{Let } B = [b_{ij}] = \text{adj } A$$

$$\begin{aligned} \text{Then } b_{ij} &= \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} a_{t\sigma(t)} = \sum_{\sigma \in S_{n_i n_j}} \prod_{t \in n_i} a_{t\sigma(t)} \quad [\because a_{ij} = a_{ji} \forall i, j] \\ &= b_{ji}. \end{aligned}$$

Hence  $\text{adj } A$  is symmetric.

(iii) Since  $A$  is transitive, then  $a_{ik}a_{kj} \leq a_{ij} \forall i, j$ .

$$\text{Let } B = [b_{ij}] = \text{adj } A$$

Let  $D = A_{ij}$ , we can determine the elements of  $D$  in terms of the elements of  $A$  as follows:

$$d_{hk} = \begin{cases} a_{hk} & \text{if } h < i, k < j, \\ a_{(h+1)k} & \text{if } h \geq i, k < j, \\ a_{h(k+1)} & \text{if } h < i, k \geq j, \\ a_{(h+1)(k+1)} & \text{if } h \geq i, k \geq j. \end{cases}$$

Where  $A_{ij}$  denotes the  $(n-1) \times (n-1)$  fuzzy matrix obtained from  $A$  by deleting  $i$ th and  $j$ th column.

Now we show that  $A_{st}A_{tu} \leq A_{su}$  for every  $t \in \{1, 2, \dots, n\}$

Let  $R = A_{st}, C = A_{tu}, F = A_{su}$  and  $W = A_{st}A_{tu}$ .

Now  $W_{ij} = \sum_{k=1}^{n-1} r_{ik}c_{kj}$

$$= \sum_{k=1}^{n-1} a_{ik}a_{kj} \leq a_{ij} = f_{ij} \text{ if } i < s, k < t, j < u,$$

$$= \sum_{k=1}^{n-1} a_{ik}a_{k(j+1)} \leq a_{i(j+1)} = f_{ij} \text{ if } i < s, k < t, j \geq u,$$

$$= \sum_{k=1}^{n-1} a_{i(k+1)}a_{(k+1)j} \leq a_{ij} = f_{ij} \text{ if } i < s, k \geq t, j < u,$$

$$= \sum_{k=1}^{n-1} a_{i(k+1)}a_{(k+1)(j+1)} \leq a_{i(j+1)} = f_{ij} \text{ if } i < s, k \geq t, j \geq u,$$

$$= \sum_{k=1}^{n-1} a_{(i+1)k}a_{kj} \leq a_{(i+1)j} = f_{ij} \text{ if } i \geq s, k < t, j < u,$$

$$= \sum_{k=1}^{n-1} a_{(i+1)(k+1)}a_{(k+1)j} \leq a_{(i+1)j} = f_{ij} \text{ if } i \geq s, k \geq t, j < u,$$

$$= \sum_{k=1}^{n-1} a_{(i+1)(k+1)}a_{(k+1)(j+1)} \leq a_{(i+1)(j+1)} = f_{ij} \text{ if } i \geq s, k \geq t, j \geq u,$$

$$= \sum_{k=1}^{n-1} a_{(i+1)k}a_{k(j+1)} \leq a_{(i+1)(j+1)} = f_{ij} \text{ if } i \geq s, k < t, j \geq u.$$

Thus  $w_{ij} \leq f_{ij}$  in every case and therefore  $A_{st}A_{tu} \leq A_{su}$  for every  $t \in \{1, 2, \dots, n\}$

Since we know that  $|AB| \geq |A||B|$ ; we have  $|A_{st}||A_{tu}| \leq |A_{st}A_{tu}| \leq |A_{su}|$ .

This means  $b_{ts}b_{ut} \leq b_{us}$  i.e.,  $b_{ut}b_{ts} \leq b_{us}$  for every  $t \in \{1, 2, \dots, n\}$

Hence  $B = \text{adj}A$  is transitive.

$$\sum_{k=1}^{n-1} a_{ik}a_{kj} \leq a_{ij} = f_{ij} \text{ if } i < s, k < t, j < u,$$

**Corollary:** If a square fuzzy matrix is similarity then  $\text{adj} A$  is also similarity.

**Example 4.3.2:** Consider a square fuzzy matrix

$$(i) \quad \text{Let } A = \begin{bmatrix} 1 & 0 & 0.3 \\ 0.1 & 1 & 0 \\ 0.4 & 0.5 & 1 \end{bmatrix} \text{ be a reflexive fuzzy matrix, then}$$



$$adj A = \begin{bmatrix} 1 & 0 & 0.3 \\ 0.1 & 1 & 0 \\ 0.4 & 0.5 & 1 \end{bmatrix}$$

Since all the diagonal elements in  $adj A$  are unity, then  $adj A$  is a reflexive fuzzy matrix.

(ii) Let  $A = \begin{bmatrix} 0.2 & 0 & 0.6 \\ 0 & 1 & 0.1 \\ 0.6 & 0.1 & 0.9 \end{bmatrix}$  be a symmetric fuzzy matrix, then

$$adj A = \begin{bmatrix} 0.9 & 0.1 & 0.6 \\ 0.1 & 0.6 & 0.1 \\ 0.6 & 0.1 & 0.2 \end{bmatrix}$$

Since  $(adj A)' = adj A$  is a symmetric fuzzy matrix.

(iii) Let  $A = \begin{bmatrix} 0.6 & 0.7 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.6 & 0.6 \end{bmatrix}$  be transitive fuzzy matrix, then

$$adj A = \begin{bmatrix} 0.6 & 0.6 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.7 & 0.6 \end{bmatrix}$$

Now

$$(adj A)^2 = \begin{bmatrix} 0.6 & 0.6 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.7 & 0.6 \end{bmatrix} \begin{bmatrix} 0.6 & 0.6 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.7 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & 0.6 & 0.6 \\ 0.5 & 0.6 & 0.5 \\ 0.6 & 0.6 & 0.6 \end{bmatrix}$$

$\leq adj A$ , then  $adj A$  is a transitive fuzzy matrix.

Next, consider the similarity fuzzy matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then  $A$  is reflexive, symmetric and transitive.

Now



$$adj A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then  $adj A$  is reflexive, symmetric and transitive.

Hence  $adj A$  is similarity fuzzy matrix.

**Definition 4.3.2:** A square fuzzy matrix  $A$  of order  $n$  is called **circular** fuzzy matrix

iff  $(A^2)' \leq A$  or more explicitly,  $a_{jk}a_{ki} \leq a_{ij}$  for every  $k = 1, 2, \dots, n$ .

**Theorem 4.3.2:** If a square fuzzy Matrix  $A$  of order  $n$  is circular, then  $adj A$  is circular.

**Proof:** Since  $A = [a_{ij}]$  is circular then  $a_{jk}a_{ki} \leq a_{ij} \forall i, j$

Let  $B = [b_{ij}] = adj A$

Let  $D = A_{ij}$ , we can determine the elements of  $D$  in terms of the elements of  $A$  as follows:

$$d_{hk} = \begin{cases} a_{hk} & \text{if } h < i, k < j, \\ a_{(h+1)k} & \text{if } h \geq i, k < j \\ a_{h(k+1)} & \text{if } h < i, k \geq j \\ a_{(h+1)(k+1)} & \text{if } h \geq i, k \geq j. \end{cases}$$

Where  $A_{ij}$  denotes the  $(n-1) \times (n-1)$  fuzzy matrix of order  $(n-1)$  obtained from  $A$  by deleting  $i$ th row and  $j$ th column.

Now we show that  $A_{st} A_{tu} \leq A_{us}$  for every  $t \in \{1, 2, \dots, n\}$

Let  $R = A_{st}$ ,  $C = A_{tu}$ ,  $F = A_{us}$  and  $W = A_{st} A_{tu}$ .

Now  $W_{ij} = \sum_{k=1}^{n-1} r_{jk} c_{ki}$

$$= \sum_{k=1}^{n-1} a_{jk} a_{ki} \leq a_{ij} = f_{ij} \text{ if } i < s, k < t, j < u,$$

$$\begin{aligned}
&= \sum_{k=1}^{n-1} a_{(j+1)k} a_{ki} \leq a_{i(j+1)} = f_{ij} \text{ if } i < s, k < t, j \geq u, \\
&= \sum_{k=1}^{n-1} a_{j(k+1)} a_{(k+1)i} \leq a_{ij} = f_{ij} \text{ if } i < s, k \geq t, j < u, \\
&= \sum_{k=1}^{n-1} a_{(j+1)(k+1)} a_{(k+1)i} \leq a_{i(j+1)} = f_{ij} \text{ if } i < s, k \geq t, j < u, \\
&= \sum_{k=1}^{n-1} a_{jk} a_{k(i+1)} \leq a_{(i+1)j} = f_{ij} \text{ if } i \geq s, k < t, j < u, \\
&= \sum_{k=1}^{n-1} a_{j(k+1)} a_{(k+1)(j+1)} \leq a_{(i+1)j} = f_{ij} \text{ if } i \geq s, k \geq t, j < u, \\
&= \sum_{k=1}^{n-1} a_{(j+1)(k+1)} a_{(k+1)(j+1)} \leq a_{(i+1)(j+1)} = f_{ij} \text{ if } i \geq s, k \geq t, j \geq u, \\
&= \sum_{k=1}^{n-1} a_{(j+1)k} a_{k(i+1)} \leq a_{(i+1)(j+1)} = f_{ij} \text{ if } i \geq s, k < t, j \geq u.
\end{aligned}$$

Thus  $w_{ij} \leq f_{ij}$  in every case and therefore  $A_{st}A_{tu} \leq A_{us}$  for every  $t \in \{1, 2, \dots, n\}$

Since we know that  $|AB| \geq |A||B|$ ; we have  $|A_{st}| |A_{tu}| \leq |A_{st}A_{tu}| \leq |A_{us}|$

This means  $b_{ts}b_{ut} \leq b_{su}$  i.e.,  $b_{ut}b_{ts} \leq b_{su}$  for every  $t \in \{1, 2, \dots, n\}$

Hence  $B = \text{adj } A$  is circular.

**Theorem 4.3.3:** To construct a transitive fuzzy matrix from a given fuzzy matrix through adjoint matrix.

For any  $n \times n$  fuzzy matrix  $A$ , the fuzzy matrix  $A(\text{adj } A)$  is transitive.

**Proof:** Let  $C = [c_{ij}] = A(\text{adj } A)$

Then  $c_{ij} = \sum_{k=1}^n a_{ik} |A_{jk}| = a_{if} |A_{jf}|$  for some  $f \in \{1, 2, \dots, n\}$

and  $c_{ij}^2 = \sum_{s=1}^n c_{is} c_{sj}$

$$\begin{aligned}
&= \sum_{s=1}^n \left[ \left( \sum_{l=1}^n a_{il} |A_{sl}| \right) \left( \sum_{t=1}^n a_{st} |A_{jt}| \right) \right] \\
&= \sum_{s=1}^n (a_{ih} |A_{sh}|) (a_{su} |A_{ju}|), \text{ for some } h, u \in \{1, 2, \dots, n\}
\end{aligned}$$

$$= a_{ih} |A_{gh}| a_{gu} |A_{ju}|$$

$$\leq a_{ih} |A_{ju}|$$

$$\leq a_{if} |A_{if}| = c_{ij}$$

Thus  $c_{ij}^2 \leq c_{ij}$

And so  $(A \text{ adj} A)^2 \leq A (\text{adj} A)$ .

Hence  $A (\text{adj} A)$  is transitive.

**Example 4.3.3:** For a square fuzzy matrix

$$A = \begin{bmatrix} 0.5 & 0.7 & 0.8 \\ 0.3 & 0.6 & 0.4 \\ 0.9 & 0.2 & 1 \end{bmatrix}$$

$$\text{Then } \text{adj} A = \begin{bmatrix} 0.6 & 0.7 & 0.6 \\ 0.4 & 0.8 & 0.4 \\ 0.6 & 0.7 & 0.5 \end{bmatrix}$$

$$\text{Now } A (\text{adj} A) = \begin{bmatrix} 0.5 & 0.7 & 0.8 \\ 0.3 & 0.6 & 0.4 \\ 0.9 & 0.2 & 1 \end{bmatrix} \begin{bmatrix} 0.6 & 0.7 & 0.6 \\ 0.4 & 0.8 & 0.4 \\ 0.6 & 0.7 & 0.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & 0.7 & 0.5 \\ 0.4 & 0.6 & 0.4 \\ 0.6 & 0.7 & 0.6 \end{bmatrix}$$

Which is transitive fuzzy matrix as

$$\text{Now } (A (\text{adj} A))^2 = \begin{bmatrix} 0.6 & 0.7 & 0.5 \\ 0.4 & 0.6 & 0.4 \\ 0.6 & 0.7 & 0.6 \end{bmatrix} \begin{bmatrix} 0.6 & 0.7 & 0.5 \\ 0.4 & 0.6 & 0.4 \\ 0.6 & 0.7 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & 0.6 & 0.5 \\ 0.4 & 0.6 & 0.4 \\ 0.6 & 0.6 & 0.6 \end{bmatrix}$$

$$\leq A (\text{adj} A)$$

Hence for any square fuzzy matrix  $A$ ,  $A (adj A)$  is transitive.

**Definition 4.3.3:** An  $n \times n$  fuzzy matrix  $A$  is called **idempotent** fuzzy matrix iff

$$A^2 = A.$$

Let us understand by an example to convert reflexive fuzzy matrix into idempotent fuzzy matrix by taking its adjoint matrix:-

**Example 4.3.4:** for a reflexive fuzzy matrix

$$A = \begin{bmatrix} 1 & 0.1 & 0.2 \\ 0.3 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix}$$

Then

$$adj A = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.4 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix}$$

which is idempotent fuzzy matrix as

$$(adj A)^2 = \begin{bmatrix} 1 & 0.1 & 0.2 \\ 0.4 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.4 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.4 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix} = adj A$$

Hence idempotent fuzzy matrix is formed by taking adjoint of reflexive fuzzy matrix.

**Remark:** For a reflexive fuzzy matrix  $A$  of order  $n$ ,  $adj A = A^c$  where  $A^c$  is

idempotent and  $c \leq n-1$ .

**Example 4.3.5:** For a  $3 \times 3$  reflexive fuzzy matrix

$$A = \begin{bmatrix} 1 & 0.1 & 0.2 \\ 0.3 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix}$$

Then



$$\begin{aligned}
adj A &= \begin{bmatrix} 1 & 0.1 & 0.2 \\ 0.3 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.1 & 0.2 \\ 0.3 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.4 & 1 & 0.4 \\ 0.5 & 0.6 & 1 \end{bmatrix} \\
&= adj A
\end{aligned}$$

Then  $A^2$  is idempotent fuzzy matrix.

**Theorem 4.3.4:** Let  $A$  be a  $n \times n$  reflexive fuzzy matrix. If  $A$  is idempotent then  $adj A = A$  is idempotent.

**Proof:** We know that for a reflexive fuzzy matrix  $A$  of order  $n$ ,

$$adj A = A^c (c \leq n - 1) \text{ where } A^c \text{ is idempotent.}$$

But we have also that  $A$  is idempotent and so  $A^c = A$ .

Thus  $adj A = A$ .

Since  $A$  is idempotent and hence  $adj A$  is idempotent.

**Theorem 4.3.5:** Let  $A$  be an  $n \times n$  reflexive fuzzy matrix. Then we have the following properties:

- (i)  $adj A^2 = (adj A)^2 = adj A$
- (ii)  $adj(adj A) = adj A$
- (iii)  $adj A \geq A$
- (iv)  $A(adj A) = (adj A)A = adj A$ .

**Proof:**

- (i) Since  $A$  is reflexive and thus  $A^2$  is also reflexive.



Then  $adjA^2 = (A^2)^c = (A^c)^2$  where  $A^c = adjA$  is idempotent.

$$\therefore adjA^2 = (adjA)^2$$

Since  $adjA$  is idempotent and so  $(adjA)^2 = adjA$

$$\text{Hence } adjA^2 = (adjA)^2 = adjA$$

(ii) Since  $A$  is reflexive and then  $adjA$  is reflexive.

Also, for a reflexive fuzzy matrix  $A$ ,  $adjA$  is idempotent.

$$\text{Hence } adj(adjA) = adjA. \quad (\text{by theorem 4.3.4})$$

(iii) Since  $A$  is reflexive, then  $a_{ii} = 1 \forall i$

$$\text{Let } B = [b_{ij}] = adjA$$

$$\text{Then } b_{ij} = \sum_{\sigma \in S_{n_j n_i}} \prod_{t \in n_j} a_{t\sigma(t)}$$

Taking the permutation of set  $n_j$  over the set  $n_i$  such that

$$\sigma(h) = h, \sigma(i) = j, h \neq i$$

i.e., the permutation

$$\begin{pmatrix} 1 & 2 & 3 & \dots & i & \dots & j-1 & j+1 & \dots & n \\ 1 & 2 & 3 & \dots & j & \dots & j-1 & j+1 & \dots & n \end{pmatrix}$$

Thus  $a_{11}a_{22}a_{33} \dots a_{(j-1)(j-1)}a_{(j+1)(j+1)} \dots a_{nn}$  is a term of  $b_{ij}$  so that

$$b_{ij} \geq a_{11}a_{22}a_{33} \dots a_{(j-1)(j-1)}a_{(j+1)(j+1)} \dots a_{nn} = a_{ij} \quad (\because a_{ii} = 1 \forall i)$$

Hence  $adjA \geq A$ .

(iv) Since  $A = [a_{ij}]$  is reflexive, then  $a_{ii} = 1 \forall i$

$$\text{Let } B = [b_{ij}] = adjA$$

$$\text{Let } C = [c_{ij}] = A(adjA) \text{ and } D = [d_{ij}] = (adjA)A.$$

$$\text{Then } [c_{ij}] = \sum_{k=1}^n a_{ik} |A_{jk}| \geq a_{ii} |A_{ji}| = |A_{ij}| = b_{ij} \quad (\because a_{ii} = 1 \forall i)$$

$$\text{and } d_{ij} = \sum_{k=1}^n |A_{ki}| a_{kj} \geq |A_{ji}| a_{jj} = |A_{ji}| = b_{ij} \quad (\because a_{jj} = 1 \forall j)$$

Thus we have  $A(adjA) \geq adjA$  and  $(adjA)A \geq adjA$ .

$$\text{But } adjA = (adjA)(adjA) \quad [\text{by (i)}]$$

$$\geq A(adjA) \quad [\text{by (iii) and using the result } A \leq B \Rightarrow AC \leq BC]$$

$$\text{So that } A(adjA) = adjA$$

$$\text{Also } adjA = (adjA)(adjA) \geq (adjA)A$$

$$[\text{by (i),(iii) and using the result } A \leq B \Rightarrow CA \leq BA]$$

$$\text{So that } (adjA)A = adjA$$

$$\text{Hence } A(adjA) = adjA = (adjA)A.$$

**Example 4.3.6:** For a reflexive fuzzy matrix

$$A = \begin{bmatrix} 1 & 0.5 & 0.8 \\ 0.3 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix}$$

$$\text{We have } A^2 = \begin{bmatrix} 1 & 0.5 & 0.8 \\ 0.3 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & 0.8 \\ 0.3 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.4 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix}$$

$$adjA^2 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.4 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix}$$

$$adjA = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.4 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} \geq A, (adjA)^2 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.4 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} = adj(adjA)$$

$$A(adjA) = \begin{bmatrix} 1 & 0.5 & 0.8 \\ 0.3 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & 0.8 \\ 0.3 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.4 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix}$$

$$(adj)A = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.4 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & 0.8 \\ 0.3 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.4 & 1 & 0.4 \\ 0.7 & 0.6 & 1 \end{bmatrix}$$

It is clear that this example satisfies all the statements of the above proposition.

$$\text{Hence } adjA^2 = (adj)^2 = adjA = adj(adjA) = A(adjA) = (adjA)A$$

$$\text{and } (adjA) \geq A.$$

## CONCLUSION

The work being presented in the matrix is devoted to “The study of Fuzzy Matrix”. Moreover, operations defined on fuzzy matrices are involved which show the comparative working of these matrix theories. Thus, comparative study also includes the determinant and adjoint theory of square fuzzy matrix. The determinant theory of a square fuzzy matrix contains the determinant theory of a square fuzzy matrix along their properties including another statement, illustration, corollary parts and remarks, wherever observed. Moreover, some theorems for fuzzy matrices are considered. In the adjoint theory of a square fuzzy matrix with their properties including corollary parts and remarks, wherever observed. Some special properties of square fuzzy matrices such as symmetry, reflexive, transitivity, circularity and idempotence are dealt with and carried over to adjoint matrix, which can be easily understood with the help of given examples.



## **REFERENCES**

1. Dubois D. and Prade H., "Fuzzy Sets and Systems: Theory and Applications", *Academic Press, New York* (1986).
2. Horst P., "Matrix Algebra for Social Scientist", *Holt, Rinehart and Winston, inc.* (1963)
3. Kandasamy W.B.V., Smarandache F. and Ilanthendral K., "Elementary Fuzzy Matrix Theory and Fuzzy Model for Social Scientists", *Los Angeles* (2007)
4. Kaufmann A., "Introduction to the theory of Fuzzy subsets", *Vol. 1, Academic press, New York* (1975).
5. Kaufmann A. and Guota M.M., "Introduction to Fuzzy Arithmetic Theory and Applications", *Van Nostrand Reinhold, New York* (1988).
6. Kaufmann A. and Gupta M.M., "Fuzzy Mathematical Models in Engineering and Management Science", *North Holland* (1988).
7. Kim J.B., "Idempotents and Inverses in Fuzzy Matrices", *Malaysian Math6(2)* (1983) 57-61.
8. Kim J.B., "Inverses of Boolean Matrices", *Bull. Inst. Math. Acad. Sinica* 12(2) (1984) 125-128.
9. Kim J.B., "Determinant Theory for Fuzzy and Boolean Matrices". *Congressus Numerantium, Utilitas Mathematica Pub.* (1978) 273-276.
10. Kim J.B., Baartmans A. and Sahadin N.S., "Determinant Theory for Fuzzy Matrices", *Fuzzy Sets and Systems* 29 (1989) 273-276.
11. Ragab M.Z. and Emam E.G., "The Determinant and Adjoint of a Square Fuzzy Matrix", *Fuzzy Sets and Systems* 61 (1994) 297-307.

12. Thomason M.G., "Convergence of Power of Fuzzy matrix", *J. of Math. Anal. Appl.* 57(1977) 476-486
13. Xin L.J., "Controllable Fuzzy Matrices", *Fuzzy Sets and Systems*, 45(1992), 313-319.
14. Zadeh L.A., "Fuzzy Sets as a Basis for Theory of Possibility", *Fuzzy Sets and Systems*, Vol. 1, No.1 (1978) 3-28.
15. Zimmermann H.J., "Fuzzy Set Theory and its Applications", *Kluwer Nijhoff Publishing, Dordrecht* (1985).



# **SOCIAL NETWORKING ANALYTICS**

Project Report submitted to

**ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI**

Affiliated to

**MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELI**

In partial fulfilment of the requirement for the award of degree of

**Bachelor of science in Mathematics**

Submitted by

**NAMES**

**REG.NO.**

**DHARSHANA MISHAL. S**

**18AUMT16**

**DONI LIMRA. M**

**18AUMT17**

**JESSICA VAZ. R**

**18AUMT27**

**SEFFRINA. C**

**18AUMT45**

**SNOWTHISHA. S**

**18AUMT50**

Under the Guidance of

**Dr. Sr. S. KULANDAI THERESE M.Sc., B.Ed., M.Phil., Ph.D.,**

Assistant Professor of Mathematics

St.Mary's College (Autonomous), Thoothukudi



**Department of Mathematics**

St. Mary's College (Autonomous)

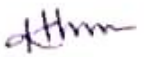
Thoothukudi

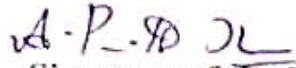
(2020-2021)

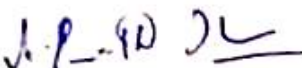
## CERTIFICATE

We hereby declare that the project report entitled "SOCIAL NETWORKING ANALYTICS" being submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli in partial fulfillment for the award of degree of Bachelor of Science in Mathematics and it is a record of work done during the year 2020-2021 by the following students:

NAMES	REG.NO.
DHARSHANA MISHAL. S	18AUMT16
DONI LIMRA. M	18AUMT17
JESSICA VAZ. R	18AUMT27
SEFFRINA. C	18AUMT45
SNOWTHISHA. S	18AUMT50

 09/04/21  
Signature of the Guide

  
Signature of the HOD

  
Signature of the Examiner

  
Signature of the Principal  
St. Mary's College (Autonomous)  
Thoothukudi - 628 001

## DECLARATION

We hereby declare that the project reported entitled "SOCIAL NETWORKING ANALYTICS" is our original work. It has not been submitted to any university for any degree or diploma.

P. Dharshana Mishal

(DHARSHANA MISHAL. S)

M. Doni Limra

(DONI LIMRA. M)

R. Jessica Vaz

(JESSICA VAZ. R)

C. Seffrina

(SEFFRINA. C)

S. Snowthisha

(SNOWTHISHA. S)

# **SOCIAL NETWORKING ANALYTICS**

Project Report submitted to

**ST.MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI**

Affiliated to

**MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELI**

In partial fulfilment of the requirement for the award of degree of

**Bachelor of science in Mathematics**

Submitted by

**NAMES**

**REG.NO.**

**DHARSHANA MISHAL. S**

**18AUMT16**

**DONI LIMRA. M**

**18AUMT17**

**JESSICA VAZ. R**

**18AUMT27**

**SEFFRINA. C**

**18AUMT45**

**SNOWTHISHA. S**

**18AUMT50**

Under the Guidance of

**Dr. Sr. S. KULANDAI THERESE M.Sc., B.Ed., M.Phil., Ph.D.,**

Assistant Professor of Mathematics

St.Mary's College (Autonomous), Thoothukudi



**Department of Mathematics**

St. Mary's College (Autonomous)

Thoothukudi

(2020-2021)

## ACKNOWLEDGEMENT

First of all, we thank lord Almighty for showering his blessings to undergo this project.

With immense pleasure, we register our deep sense of gratitude to our guide **Dr. Sr. S. Kulandai Therese M.Sc., B.Ed., M.Phil., Ph.D.**, and the Head of the Department **Dr. A. Punitha Tharani M.Sc., M.Phil., Ph.D.**, for having imported necessary guidelines throughout the period of our studies.

We thank our beloved Principal **Rev. Dr. Sr. A.S.J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D.**, for providing us the help to carry out our project work successfully.

Finally, we thank all those who extended their helping hands regarding this project.



# **SOCIAL NETWORKING ANALYTICS**

April 2021

## Preface

The topic of this project is *Social Networking Analytics*, with focus on underlying concepts of the discipline, behaviour aspects in social networks and link prediction modeling. As the interest of individual in virtual social networking grows, more scientific attention is given to them. Systems are being developed for understanding how and who acts in such social networks. These are tracking every possible social networking activity: usage, topics, who interact with who, for how long, user specific interests etc. Social Networking Analytics(SNA) is the discipline incorporating such scientific interests, arose from a long standing practice called Social Network Analysis. After the introduction of virtual social networks, it was a natural progression to apply the learned concepts and practices in the internet world. As networks continue to increase in numbers and technology becomes more advanced, even more tools for social networking analytics will come on the market, each delving deeper into the system and offering more and more insight. If used correctly, social networking analytics may be a key tool in helping an organization to find and connect to the right markets and audiences, on a personal level. The paper is structure in five chapters.

**Chapter 1** presents briefly the evolution of Social Network Analytics.

**Chapter 2** introduces the fundamental concepts and metrics in Social Networking Analytics.

**Chapter 3** deals with stuctural balance and transitivity of Social Networking.

**Chapter 4** focuses on the blockmodels in Social Networking namely Perfect Fit, Zeroblock Criterion, Oneblock Criterion and  $\alpha$  Density Criterion.

**Chapter 5** deals with the behavioral aspects of Social Networking, introducing a set of established link prediction models.

Finally it gives the conclusion of the project in Social Networking Analytics.

# Contents

<b>List of Notations Used</b>	<b>4</b>
<b>1 Evolution of Social Network Analytics</b>	<b>6</b>
<b>2 Basic Concepts In Social Networking</b>	<b>8</b>
2.1 Graph Theory . . . . .	9
2.2 Sociomatrices . . . . .	12
2.3 Measures in Social Networking . . . . .	15
<b>3 Structural Balance and Transitivity</b>	<b>23</b>
3.1 Structural Balance of Social Networks . . . . .	23
3.2 Transitivity . . . . .	25
<b>4 Blockmodels</b>	<b>28</b>
4.1 Definition . . . . .	28
4.2 Building Blocks . . . . .	29
4.2.1 Perfect Fit (Fat Fit) . . . . .	30
4.2.2 Zeroblock (Lean Fit) Criterion . . . . .	31
4.2.3 Oneblock Criterion . . . . .	31
4.2.4 $\alpha$ Density Criterion . . . . .	32
<b>5 Link Prediction Models</b>	<b>33</b>
5.1 Mathematical framework . . . . .	34

5.1.1	Models based on node similarity . . . . .	34
5.1.2	Models based on topological patterns . . . . .	36
5.1.3	Models based on a probabilistic model . . . . .	39
<b>Conclusion</b>		<b>45</b>
<b>References</b>		<b>46</b>

# List of Notations

$f$	the total number of actors (nodes)
$L$	the number of connections of the network
$\vartheta_1, \vartheta_2$	the sets of properties characterizing the two nodes
$\mathfrak{B}_k$	an equivalence class or position
$g_k$	the number of actors in equivalence class $k$
$\mathbf{B}=\{b_{klr}\}$	an image matrix for relation $x_r$
$B$	the number of positions (equivalence classes)
$\Delta_{klr}$	the density of block $b_{klr}$
$\mathbf{X}=\{x_{ijr}\}$	super-sociomatrix of size $g \times g \times r$



# List of Figures

2.1	Social network . . . . .	8
2.2	Example of Dyad . . . . .	10
2.3	Example of Triad . . . . .	10
2.4	Directional connection in graphs . . . . .	11
2.5	Maximum number of connections in non-directional graphs . . . .	12
2.6	Network of non-directional business relations . . . . .	13
2.7	Directional graph with sales connections between companies . . . .	14
2.8	Density of different non-directional graphs . . . . .	18
2.9	Types of connectivity in directional graphs . . . . .	19
2.10	Example of geodesic in an undirected and a directed network . . .	20
2.11	Example of centrality . . . . .	21
3.1	Structural balance for sets of three nodes . . . . .	24
3.2	The sixteen possible triads for transitivity in a digraph . . . . .	26
3.3	The type 16 triad, and all six triples of actors . . . . .	27

# Chapter 1

## Evolution of Social Network Analytics

Due to the recent globalization of the commercial environment and the impact of the new technologies, the analysis of social networks represents a major interest. This rather new area of research grew out of social and exact sciences, computers supporting today modeling and complex mathematical calculations, previously impossible. The analysis of social networks is driven by business and social interests, combining various academic fields.

The term social networks was used for the first time in 1950 in sociometrics, the science that seeks to obtain data on social behavior and to analyze it. The latter incorporation of mathematical tools and computing triggered the evolution of Social Network Analysis and Analytics.

The mathematical basis of SNA arose out of the fields of graph theory, statistical and probability theory, game theory as well as algebraic models. In fact, it was from these theories, especially graphs, that the Internet and various virtual networking concepts were derived.

Networks are generally studied based on the participants and their actions in the network, with little or no emphasis on the relationships. Particularly, in Social Networking and SNA the type and the forms of relationships between the network members are fundamental.

Social networking data comes today in many forms: blogs (Blogger, LiveJournal), micro-blogs (Twitter), social networking (Facebook, LinkedIn), wiki sites (Wikipedia, Wetpaint) and multimedia sharing (Flickr, Youtube).

Online social networking represents a fundamental shift of how information is being produced, transferred and consumed. User generated content, in any data form, establishes a connection between producers and consumers of information. For consumers, the abundance of share data and opinions is a support in making more informed decisions.

SNA is applicable in various domains and fields: organizational behavior, terrorist networking, political and economic systems, inter-relationships between banks and companies, social influence, educational systems and many others. Some of the current interests and challenges in the discipline of SNA are:

- Collecting massive amounts of data and preventing information overload for the users.
- Extracting and modeling temporal patterns of information growth and fade over time.
- Correcting effects and biases generated by incomplete or missing data.
- Handling unreliable or conflicting information
- Classification and tracking of topics.
- Predicting and identifying emerging or popular topics.
- Detecting, quantifying and maximizing the individuals influence.
- Identification of topic relevance.
- Determining implicit links between users.
- Understanding of sentiment flow through networks and polarization.

## Chapter 2

# Basic Concepts In Social Networking

A *social network* can be defined as a finite set of actors and their *relationships*. This is a simple and direct concept, allowing everyone to understand the social network according to the complete data and the connectivity of a considered network. This definitions does not say much though over the types of relationships of certain groups (i. e. the number of times they take part in the same programs or activities).

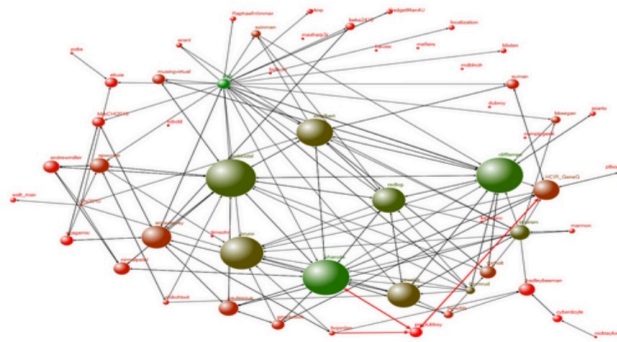


Figure 2.1: Social network

An *actor* is the social entity who participates in a certain network and who is able to act and form connections with other actors. It could be an individual, a corporation or a social body. Examples of actors could be the students in a classroom, the departments in a company, etc. When all the actors of a network are of the same type, the network is called *monomodal*. But there are cases in which there are different actors in a network. In a multi-agent system, the actor is called an *agent*.

A link between two actors in a social network is called a *connection*. It is defined by some type of *relationship* between these actors, depending on the type

of society. Between companies, the connection could be a business contract of supply, between people in a company, it could be the hierarchic relationship, if considering the organizational structure, or it could be the sending of e-mails in a network of relationships between friends. Other examples include the relationships of friendship or respect between students in a classroom, the biological relationships (in a family), the associations of members to clubs, the diplomatic relationships between countries etc. In the graph theory section presented later in the paper it will be shown that connections may have a value as well as a direction.

To study networks of various relationships in an objective way, models need to be created to represent them. There are three *notations* currently in use in the social network analysis:

- Graph Theory – the most common model for visual representation, it is graph base.
- Sociometrics – proposes matrices representation, also called sociomatrices.
- Algebraic – proposes algebraic notations for specific cases, especially for multiple relationships.

Each notation scheme has different applications and will enable different developments and analyses. Further, this chapter presents concepts and notations used for representations with graphs and *sociomatrices*. The combination of these two techniques has helped significantly the evolution of social network analysis

## 2.1 Graph Theory

The Graph theory has been widely used in analyses of social networks due to its representational capacity and simplicity. Basically, the graph consists of **nodes** ( $n$ ) and of **connections** ( $l$ ) which connect the nodes. In social networks the representation by graphs is also called **sociogram**, where the nodes are the actors or events and the lines of connection establish the set of relationships in a two dimensional drawing.

**Dyad** is the simplest network, composed of only two nodes, that may be connected or not. If connected, this represents a property of the **pair**.





Figure 2.2: Example of Dyad

**Triad** is a network formed by three nodes and the possible connections between them. The triad brings some important concepts into question, such as the equilibrium and the transitivity which are presented later on. There are maximum three dyads in a triad. In business relationships, this can be an important factor because if Node 1 has a relationship with Node 2, and they in turn with Node 3, there is a possible path through Node 2 and on to Node 1 to make transactions with Node 3.

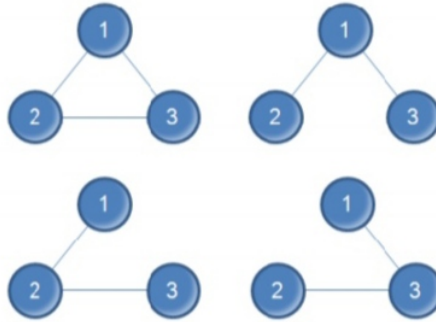


Figure 2.3: Example of Triad

**Relationship.** The set of connections of a given type defines the relationship found in the social network under analysis. Whereas a connection is only between two actors or nodes, the relationship is defined for the whole set of connections. Thus, we can talk about social relationships, business relationships, educational relationships etc. In the social network, there may be a connection between two actors (a situation where often the variable is set to “1” in a table or matrix), or there is none (represented with a “0”).

There are also relationships which imply values, when there is a connection and this connection can be attributed with a value (i.e. the financial worth of the business relationships between companies). The social networks where values are also involved, have a greater degree of complexity. This also due to the possibility of direction within a graph (i.e. a given company buys from another, but sells nothing to it).

The actors of a network will be noted  $n$ , and the set of actors as  $N$ . The connections of a network will have notation  $l$ , and the set of connections will be  $L$ .

Thus, a network of “f” actors and of “h” connections will have the sets of actors and of connections defined respectively by:  $N = n_1, n_2, \dots, n_f$  and  $L = l_1, l_2, \dots, l_h$ .

As the connection is always between two actors, then the connection defines a pair of actors (or dyad). If saying that a connection  $l_1$  refers to the connection between actors  $n_2$  and  $n_5$ , then we can write:  $l_1 = \langle n_2, n_5 \rangle$

Up to this point it has been defined a connection between two actors without being concerned about the type of relationship. Many of these connections are **non-directional**, meaning that a connection between two actors is established and that the relationship is not in any specific direction. For example, marriage establishes a relationship which is non-directional as it is not possible for a member to be married to another and that the inverse is not also true. If considering that the type of connection between companies to be the existence or otherwise of a contract, such a connection is non-directional.

A **directional** connection is that which represents a connection which goes from an actor (origin) and ends at another (destination). For example, if making an analysis which considers purchases and sales between companies of a network, there will be a direction in the connections. The image below (Figure 2.4) exemplifies the concept. In the first case, the direction of the arrow shows that actor 1 sells to actor 2; in the second, actor 2 sells to 1, and in the last case, the graph represents that actor 1 sells to actor 2 and also that actor 2 sells to actor 1.

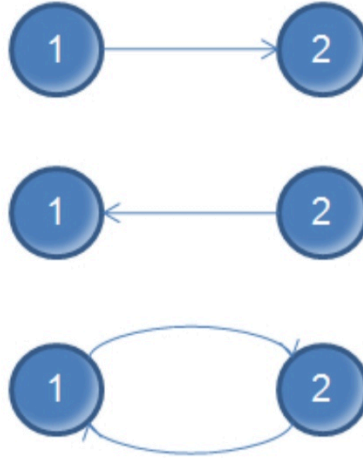


Figure 2.4: Directional connection in graphs

So if a connection  $l_1$  refers to the directional connection of actor  $n_2$  to actor  $n_5$  :  $l_1 = \langle n_2 \rightarrow n_5 \rangle$ .

For a network with the number of actors equal to “f”, the maximum number  $l_{max}$  of connections in a non-directional graph can be written using the expression:

$$l_{max} = \frac{f(f-1)}{2}$$

In other words, for two actors the maximum is one connection, for three the maximum is three, for four, it's six, and so on, as shown in Figure 2.5 below:

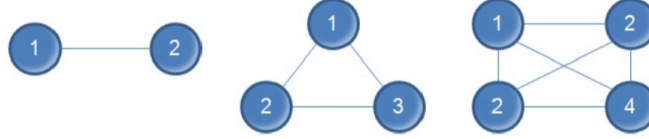


Figure 2.5: Maximum number of connections in non-directional graphs

In directional graphs, the maximum number of connections (arrows) between two actors is two arrows (one in each direction), for three actors the maximum is six, and so on. The expression which defines the maximum number of directional connections is:  $l_{maxdr} = f(f - 1)$ . One example of directional graph which has the maximum number of connections is the Brazilian soccer championship. There are twenty teams playing for the championship, each team plays against all the other teams, once at home and once away (outward game and return match, two directions). The total of the connections (games) in this network (championship) will be 380.

Graphs enable many interesting analyses to be made and have visual appeal which help us to understand the structure and behavior of social networks. However, for networks with many actors and connections, this becomes impossible. Similarly, some important information, such as the frequency of occurrence and specific values, are difficult to apply in a graph.

## 2.2 Sociomatrices

For making possible the analysis of networks with many actors and connection, the matrices developed by sociometrics, *sociomatrices*, are being used. Thus, sociometrics and its sociomatrice complement the Graph theory, establishing a mathematical basis for analyses of social networks.

Figure: M1 presents a **matrix** which shows the existence of the connections between the various actors of the network proposed in Figure 2.6, represented by a non-directional graph. In being **non-directional**, a matrix is symmetrical.

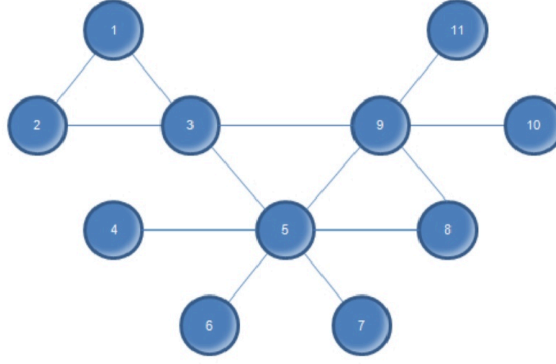


Figure 2.6: Network of non-directional business relations

	1	2	3	4	5	6	7	8	9	10	11
1	0	1	1	0	0	0	0	0	0	0	0
2	1	0	1	0	0	0	0	0	0	0	0
3	1	1	0	0	1	0	0	0	1	0	0
4	0	0	0	0	1	0	0	0	0	0	0
5	0	0	1	1	0	1	1	1	1	0	0
6	0	0	0	0	1	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	1	0	0
9	0	0	1	0	1	0	0	1	0	1	1
10	0	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	1	0	0

Figure M1: Symmetrical matrix for the non-directional graph in Fig. 2.6

Each element of the matrix shows a connection, or the lack of it, between two actors and is notated " $x_{line,column}$ ", with the sub-indices indicating the actor of a given line and the actor of a given column. If considering the values of "i" and "j" as these indices, each element will be identified by  $x_{ij}$  or algebraically:

- $x_{ij}$  - when there is a connection between  $n_i$  and  $n_j$
- $x_{ij}=0$ - when there is no connection
- $x_{ii} = x_{jj} = 0$  - when the connection does not exist

and in the symmetrical matrix:  $x_{ij} = x_{ji}$  Therefore, if the connections are directional, the graph is directional, and in this case the notation will be:

- $x_{ij}=1$  - when there is a connection from  $n_i$  to  $n_j$

- $x_{ji}=1$  - when there is a connection from  $n_j$  to  $n_i$
- $x_{ij}$  - when there is no connection

and here the matrix is rarely symmetrical. In Figure 2.7 is presented a directional graph where the companies have selling relationships between each other. The arrows point in the direction of the sale.

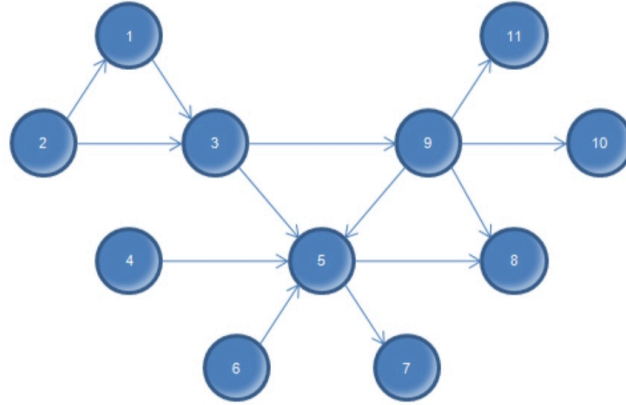


Figure 2.7: Directional graph with sales connections between companies

Figure:M2 presents the corresponding sociomatrix, where can be seen the asymmetry and that the main diagonal is empty.

	1	2	3	4	5	6	7	8	9	10	11
1	-	0	1	0	0	0	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0	0	0
3	0	0	-	0	1	0	0	0	1	0	0
4	0	0	0	-	1	0	0	0	0	0	0
5	0	0	0	0	-	0	1	1	0	0	0
6	0	0	0	0	1	-	0	0	0	0	0
7	0	0	0	0	0	0	-	0	0	0	0
8	0	0	0	0	1	0	0	-	0	0	0
9	0	0	0	0	1	0	0	1	-	1	1
10	0	0	0	0	0	0	0	0	0	-	0
11	0	0	0	0	0	0	0	0	0	0	-

Fig.M2 Sociomatrix corresponding to the directional graph in Fig.2.7

In the next section, using the basic knowledge of graphs and sociomatrixes, various characteristics of the networks of business relationships, such as prestige, social role of the actors and other definitions which are useful in the practical analyses in business and social environments are being defined.



## 2.3 Measures in Social Networking

The use of *graphs* and *sociomatrices* is necessary in order to create **models**, or simplified representation systems of networks of relationship. However, with graphs and sociomatrices it is not possible to represent the whole of the characteristics and attributes of a network, nor all of its limits and variations. In order to make analyses therefore, the model is simplified and the analysis is based on various measures. The main measures used for social network analysis are presented in this section.

### Nodal degree

In a non-directional network, it is measure the number of connections at a node and this number is called the nodal degree. The degree of a node can vary from zero, when there is no connection at this node to any other node of the network, through to the value  $f-1$ , when there is a connection at this node with all the other nodes on the network. The measure of the degree of a node can define its importance, for example, in a network where there are various connections, this is something of interest to the members of the network.

To obtain a graph of the degree of a given node,  $g(ni)$ , count the number of lines which are connected to this node. Considering the example shown in Figure 2.6 and then checking the degree of each node, in decreasing order, as follows:

- $g(n_5) = 6$
- $g(n_9) = 5$
- $g(n_3) = 4$
- $g(n_1) = g(n_2) = g(n_8) = 2$
- $g(n_4) = g(n_6) = g(n_7) = g(n_{10}) = g(n_{11}) = 1$

An important piece of data in business networks is the average number of relationships between the members of the network. This can be measured by obtaining the average degree of the network. The average degree is defined by the sum of all the degrees divided by the number of actors in the network or algebraically:

$$\bar{g} = \frac{\sum_{i=1}^f g(n_i)}{f} = \frac{2L}{f}$$

where  $L$  is the number of connections of the network and  $f$  is the total number of actors (nodes). For the network from the previous example, the value of  $\bar{g} = 2.36$ .

## Nodal degree (directional graph)

In directional graphs, the measure of the degree is slightly different, as it is interesting to know how many connections the origin node has and how many connections it has as destination.

The number of connections this node has as destination is called **nodal-in degree**. For the nodal-in degree of node  $n_i$ , obtained by counting the number of arrows pointing towards it. The used notation is  $gi(n_i)$ .

The number of connections this node has as origin is called **nodal-out degree**. For the nodal-out degree of node  $n_i$ , obtained by counting the number of arrows pointing from it. The used notation is  $go(n_i)$ . These measures are very important in a network, as the **nodal-out degree** can indicate the capacity of expansion of a given actor, whilst the nodal-in degree can represent their popularity. The measure of the nodal-in degree, for example, is one of the factors which determines the status of a given web site when making a search using Google. The position in the ranking of a page shown in the search results is determined by the number of sites which link to that page on the network, in other words, the nodal-in degree of the page.

For the business network considered in Figure 2.7, showing the directed connections for sales from one actor to another, the next nodal-out degree and the nodal-in degree are calculated for each node:

Nodal-out degree	Nodal-in degree
$go(n_1) = 1$	$gi(n_1) = 1$
$go(n_2) = 2$	$gi(n_2) = 0$
$go(n_3) = 2$	$gi(n_3) = 2$
$go(n_4) = 1$	$gi(n_4) = 0$
$go(n_5) = 2$	$gi(n_5) = 4$
$go(n_6) = 1$	$gi(n_6) = 0$
$go(n_7) = 0$	$gi(n_7) = 1$
$go(n_8) = 1$	$gi(n_8) = 2$
$go(n_9) = 4$	$gi(n_9) = 1$
$go(n_{10}) = 0$	$gi(n_{10}) = 1$
$go(n_{11}) = 0$	$gi(n_{11}) = 1$

Table 1 Nodal-out and Nodal-in degrees corresponding to the directional graph in Fig. 2.7

In the table above can be seen that for the same node the nodal-out degree and the nodal-in degree may be either equal or not. Based on the differences of in and out degrees, the theoreticians of directional graphs have created different names for the roles of the nodes. This is of special interest in business networks, as they define the behavior of the actor in the network of relationships.

Furthermore, depending on the number and type of connection, different types of node are defined:

- **Isolated**      if  $gi(n_i) = go(n_i) = 0$  - neither the origin nor destination of connections
- **Transmitter**    if  $gi(n_i) = 0$  and  $go(n_i) \geq 1$  - not the destination of connection, but the origin
- **Receptor**        if  $go(n_i) = 0$  and  $gi(n_i) \geq 1$  - not the origin of connection, but the destination
- **Carrier**            if  $gi(n_i) \geq 1$  and  $go(n_i) \geq 1$  - the origin and destination of connection

For the considered example, the company node 5 is a carrier and acts as intermediary as a seller in this network, but also concentrates most of the buying (its nodal-in degree is by far the highest).

As for the non-directional graph, it is important to find the average nodal-in degree and the average nodal-out degree of the members of such a network. The average nodal-in degree, denoted by  $\bar{ge}$ , is defined as the sum of all the nodal-in degrees divided by the number of actors of the network, that is:

$$\bar{ge} = \frac{\sum_{i=1}^f ge(n_i)}{f}$$

where  $f$  is the total number of actors (nodes). Similarly, the average nodal-out degree, denoted by  $\bar{gs}$  is defined as the sum of all the nodal-out degrees divided by the number of actors of the network, that is

$$\bar{gs} = \frac{\sum_{i=1}^f gs(n_i)}{f}$$

The total number of “ins” have necessarily to be equal to the total of the “outs” (the sum of all the origins should be equal to the sum of all the destinations). The next formulation is possible:

$$\bar{gs} = \bar{ge} = \frac{L}{f}$$

where  $L$  is the number of connections of the network. For the network in the above example, the value of  $\bar{ge} = \bar{gs} = 1,27$ , which represents a directional network with low connectivity.

**Density of the network.** Whilst the degree of the node is important to define the number of relationships of a given actor, another important piece of data of a network is its density, in other words, the measurement of the number of existent connections. Dense networks are those in which there are many connections and sparse networks are those where there are few connections. Environments where there are intense business relationships, such as between the countries of the European Union form dense networks.

The measurement of the density of a non-directional network is denoted by  $\Delta$  and it is defined by the number of connections  $L$  of this network divided by the maximum number  $l_{max}$  of connections. The expression for the density for the non-directional graph is:

$$\Delta = \frac{L}{\frac{f(f-1)}{2}} = \frac{2L}{f(f-1)}$$

If the graph has no connections, it is said to be empty and the density is equal to 0. If it has the maximum number of connections, then it is said to be full and the density is equal to 1. Figure 2.8 exemplifies the empty, the full and the intermediate graph, for a network with four nodes.

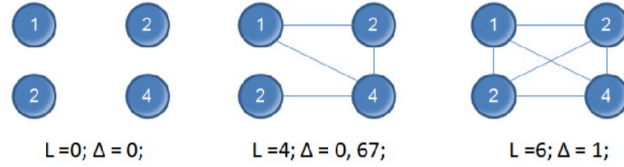


Figure 2.8: Density of different non-directional graphs

For a directional network, the measurement of the density is denoted by  $\Delta$  and is defined by the number of  $L$  connections (arrows) of this network divided by the maximum number  $l_{max.dir}$ . The expression for the density for the directional graph is:

$$\Delta = \frac{L}{f(f-1)}$$

### Searchability and directional connectivity

In a network, if there is a path between two nodes, this means that these two nodes can establish some type of relationship along this path formed by the path, that is, a node can find the other node along the path. This possibility of relationship is called **searchability**.

In a directional graph, *searchability* can be established at different levels, depending on the direction of the arrows along the path. For a node to be able to *find* the other node in a directional network, there are four types of connectivity,

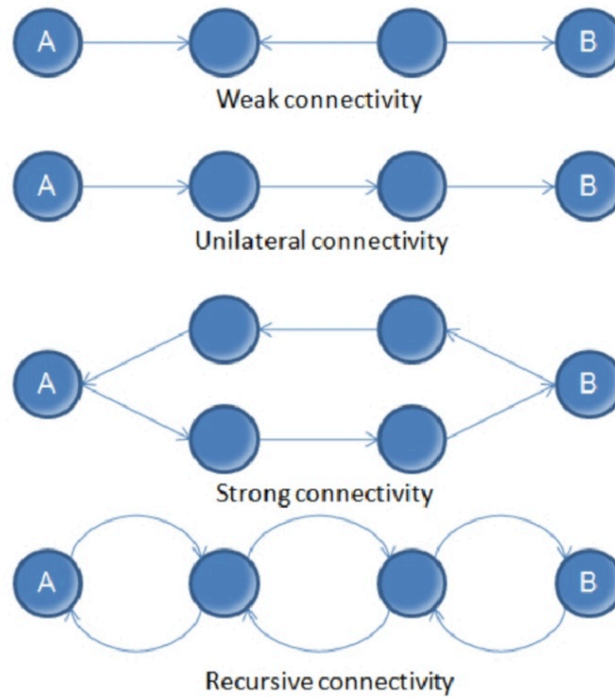


Figure 2.9: Types of connectivity in directional graphs

as shown in the example of types of paths between nodes A and B in Figure 2.9. These are the four types of connectivity:

- The nodes A and B have **weak connectivity** between them when there is a semi-path between them (at least one arrow in the opposite direction)
- The nodes A and B have **unilateral connectivity** between them when there is a directional path from A to B or from B to A between them (all arrows point in the same direction)
- The nodes A and B have **strong connectivity** between them when there is a directional path from A to B and another directional path from B to A (passing through different nodes and connections)
- The nodes A and B have **recursive connectivity** between them when there is a directional path from A to B and from B to A passing through the same nodes and connections.

Every directional graph comes within one of these types of connectivity. Their interpretation is:



- The directional graph has **weak connectivity** if **all** the pairs of nodes have weak connectivity
- The directional graph has **unilateral connectivity** if **all** the pairs of nodes are connected unilaterally
- The directional graph has **strong connectivity** if **all** the pairs of nodes have strong connectivity
- The directional graph has **recursive connectivity** if **all** the pairs of nodes have recursive connectivity

**Note:** These ideas are important for the analysis of cohesion between the members of a given network. If there is weak connectivity between A and B in a business network of sales, the possibility of A selling to B is less than if the connectivity were strong.

## Geodesic

The shortest path between two nodes is called **geodesic**, and the length of this path, in number of intermediate connections, is called **geodesic distance**. This minimum distance is very interesting because it allows the analyst to see how many connections and how many nodes are intermediaries in a relationship between two actors of a network. The geodesic distance between any two nodes  $n_i$  and  $n_j$ , is noted  $d(n_i, n_j)$ .

If there is no geodesic for any two nodes, that is, if there is no possibility of any path between them, their distance is considered infinite and the network will be disconnected.

For a directional network, the geodesic is considered as the shortest directed path between two nodes. Considering that in a directed path all the arrows have to be in the same direction, the geodesic from  $n_i$  to  $n_j$  will not always be the same geodesic from  $n_j$  to  $n_i$ . See an example of this type in figure 2.10. The sequence which defines the geodesic from  $n_1$  to  $n_3$  is  $\{n_1, l_2, n_2, l_3, n_4, l_4, n_3\}$ , with the geodesic distance  $d(n_1, n_3)=3$ . Whereas for the geodesic from  $n_3$  to  $n_1$ , the sequence is  $\{n_3, l_5, n_1\}$ , with the geodesic distance  $d(n_3, n_1)=1$ .

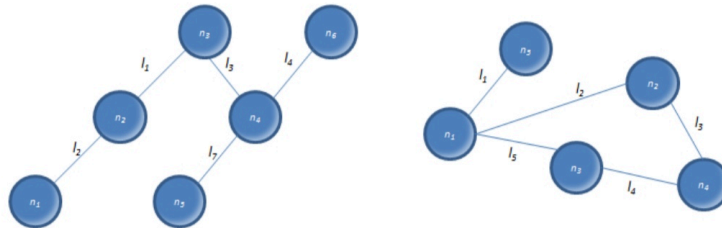


Figure 2.10: Example of geodesic in an undirected and a directed network

## Graphs with sign and with value

For each relationship established by a connection in a graph, two further pieces of additional information can be included: a **sign** and a **value**. The inclusion of a positive or negative sign for a connection can show us that a relationship is good or bad. An example of this type of network is a graph showing the relationships of affinity between students in a classroom. Usually (+) indicates that there is friendship and (−) indicates enmity.

The inclusion of *value* can add a number to a connection. An is indicating on the graph the business relationship between companies, the value of the connection representing the amount in millions of dollars in a sale transaction.

## Centrality and prestige

Two important concepts in a network are the ideas of **centrality** and **prestige** of an actor. There are various definitions and forms of calculating centrality. For a given actor  $n_i$ , the **centrality** is denoted as  $C(n_i)$  and the measure will be given by the degree of the node, that is, by the number of connections of this node in the network. *Centrality* can be also considered the measure that gives the indication of power and influence of the individual nodes of the network based on how well they are connected. The fundamental measures of centrality are: *Betweenness*, *Closeness*, and *Degree*.

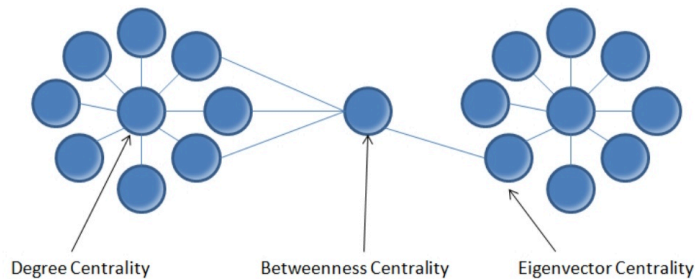


Figure 2.11: Example of centrality

**Betweenness** measures the number of subjects whom an individual is connecting indirectly, through their direct links.

**Closeness** indicates how near is a subject to all other individuals in a network, directly or indirectly.

Closeness centrality is the inverse measure of the sum of the shortest distances between each individual and everyone else in the network.

**Centralization** is the difference between the numbers of links of each node in the network divided by maximum possible sum of differences. A centralized network will have many of the links dispersed around a certain node(s) while

a decentralized will have nodes with comparable number of links.

The concept of **prestige** of an individual  $n_i$  is related to the concept of directional networks. The centrality of an individual  $n_i$ , considering the arrows directed towards them (i.e. their nodal-in degree) defines his prestige,  $P(n_i)$ , in the network.

### Other metrics in Social Networking

- **Clustering coefficient** is the measure representing the probability of a future link between two unconnected neighbors of a considered node.
- **Cohesion** represents the degree in which nodes are connected directly among each other by cohesive bonds.
- **Radiality** represents the degree with which the network of a certain individual reaches out into the global network providing content and inducing influence.
- **Reach** represents the degree in which any node of a network can reach the other nodes.
- **Structural cohesion** measures the minimum number of nodes that would disconnect the network or the group if removed.
- **Structural equivalence** represents the degree in which nodes share a common set of links connecting them to other nodes in the network.

# Chapter 3

## Structural Balance and Transitivity

### 3.1 Structural Balance of Social Networks

Social relationships have a profound impact on human development, in all life stages. Such relationships are of positive nature (i.e. friendship, collaboration, trust, support etc.) or of negative nature (i.e. oppression, dislike, harassment, intimidation etc.). A social network captures all such types of relations defined between a finite set of members. Individual characteristics and shared relationships change in time and continuously impact the entire community (social network).

Clearly, the tension executed between every two network entities, be it positive or negative is a fundamental aspect in social networking. The framework of this type of analysis is the structural balance, which aims to extract and store the relationship information in a clear and structured way. The structural balance concept is based on social psychology theories being helped by graphical and mathematical representations. The structural balance theory is based in fact on pure mathematical analysis.

The structural balance theory is based on identifying the nature of relationship between two individuals by initially isolating them. If these individuals share some level of friendship, support or collaboration, their link is marked positive: “+”, else the link is marked negative: “-“. The theory looks at subgroups of three individuals sharing a particular configuration of positive and negative values. In fact, there are possible four distinctive configuration cases between three individuals A, B, C. These are presented in Figure 3.1 below.

In such reduced systems, clear conclusion of structural balance can be drawn:

**Case 1:** A, B, C are mutual friends. This is a natural situation of three persons that are mutually friends. There are no instability sources in such system,

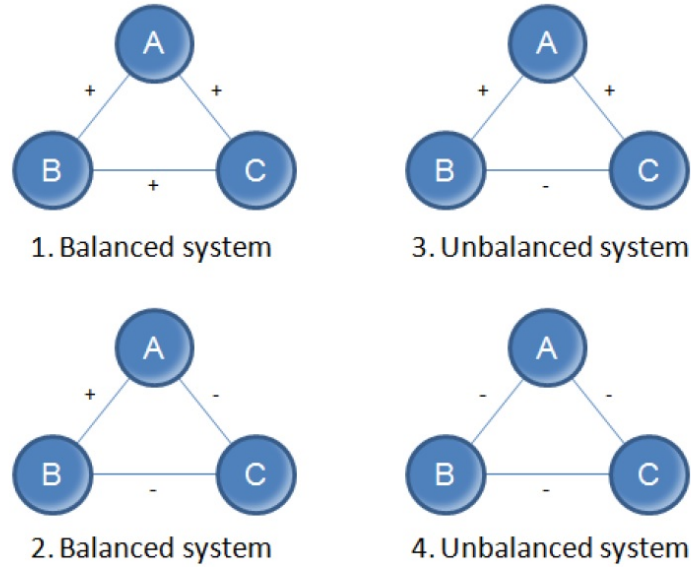


Figure 3.1: Structural balance for sets of three nodes

therefore the system is balanced.

**Case 2:** A, B are friends and C is a mutual enemy. This is also a natural situation between three individuals, two of the three are in a relationship of friendship and both dislike the third individual. As the system has clear friendship and enemy bindings and therefore no instability sources, this system is balanced.

**Case 3:** A is friend with B and C, but B and C are enemies. In such system there is present, in some degree, a psychological stress or instability into the formed relationships: one individual is in a friendship relation with two other individuals that dislike each other. The instability source comes from the fact that individual A might try changing the negative relation between B and C in positive one or might take side and become enemies with one of the individuals B or C. Based on this instability reasoning, this system is unbalanced.

**Case 4:** A, B, C are mutual enemies. In this type of system there are also present instability aspects. The reasoning is based on the fact that two individuals might start collaborating against the third individual in the system. In this case a negative link might transform in a positive one. This is why this system too is considered unbalanced.

In conclusion, the structural balance of a sub-system of three individuals connected by three links is achieved if: either all three links are positive or else, only exactly one of the links is positive. This consideration is known as the structural balanced property and is at the basis of the global structural balance of the network.

The global structural balance of the network is expressed as the problem of eliminating the unbalanced triangles. This expression is not convenient due to the involved computation, but it represents the basic start point in the concept of structural balance of social networks.



*“If a labeled complete graph is balanced, then either all pairs of nodes are friends, or else the nodes can be divided into two groups,  $X$  and  $Y$ , such that every pair of nodes in  $X$  like each other, every pair of nodes in  $Y$  like each other, and everyone in  $X$  is the enemy of everyone in  $Y$ .”*

Today structural balance is highly relevant in the on-line social media where individual opinion is intensively expressed, often in a context of influence. Another example is the international relations, representing the relationship between various countries.

Understanding the mechanism of positive and negative relationships helps the studies of behavior, structure and influence in the social field. These are important aspects in managing social or business contexts. Research is only starting exploring these fundamental questions, aiming to understand how, out of large scale datasets, balance and related theories can bring out knowledge.

## 3.2 Transitivity

### Definition

*The triad involving actors  $i$ ,  $j$ , and  $k$  is transitive if whenever  $i \rightarrow j$  and  $j \rightarrow k$  then  $i \rightarrow k$ .*

If either of the two conditions of this statement is not met (if  $i \not\rightarrow j$  and/or  $j \not\rightarrow k$ ), then the triple is termed vacuously transitive. Vacuously transitive triples are neither transitive nor intransitive. Note how the focus has shifted from cycles in signed graphs to semicycles in signed digraphs to transitive triads in ordinary digraphs.

From this definition we have the following theorem:

**Theorem** *A digraph is transitive if every triad it contains is transitive.*

We note that if a transitive digraph has no asymmetric dyads -that is, if all choices are reciprocated -then it is clusterable. Clusterable digraphs require mutual dyads to be within and null dyads to be between clusters. Thus, clusterability is a special case of transitivity. Ranked clusterable digraphs are also transitive. In fact, transitivity is the most general idea of this type for graphs and digraphs.

Refer again to Figure 3.2. The following triads are transitive: 6, 7, 8, 9. Triads 1, 2, 3, 4, 5 are vacuously transitive. They do not contain enough arcs to meet the conditions of the theorem, so cannot be transitive or intransitive. Triads 10, 11, 12, 13, 14, 15, 16 are intransitive. Vacuously transitive triads can occur and the digraph itself can still be transitive. Now, rather than eight "miserable" triples from ranked clusterability, there are only seven intransitive triads.

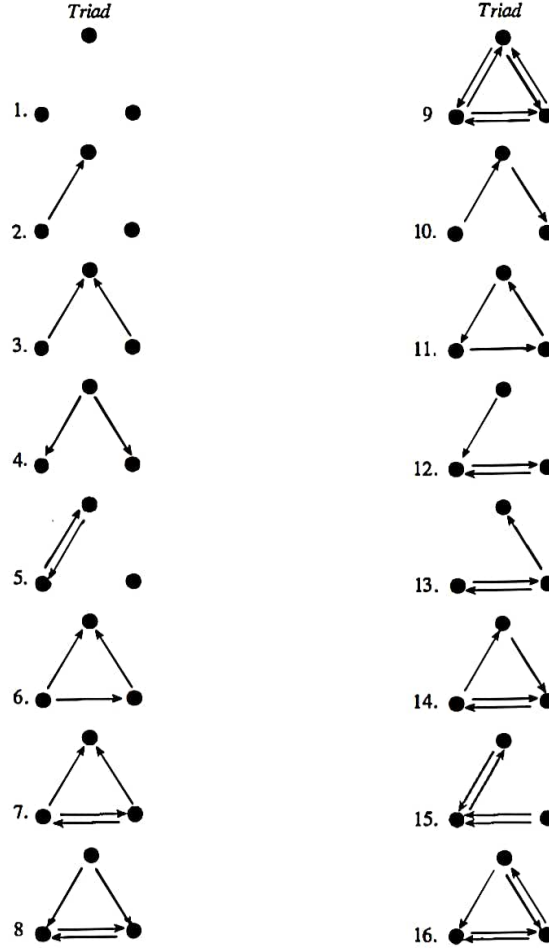


Figure 3.2: The sixteen possible triads for transitivity in a digraph

Thus, we must look at ordered triples rather than triads. Note also that each threesome of actors consists of six distinct ordered triples of actors. Some of these triples may have transitive choices while others may be intransitive. Still others may be vacuously transitive. A triple must be of one of these types. For the triad itself to be labeled transitive, all ordered triples of actors present in a triad must be either transitive or vacuously transitive. If any one of the triples is intransitive, so is the triad.

For example, look at triad 16 in Figure 3.2. As is the case with all triads, triad 16 has six triples. This triad, along with its triples and their statuses, are listed in Figure 3.3. Three of the triples are transitive, while one of them (the second) is not. The other two triples are vacuously transitive (for example, the first triple,  $n_i n_j n_k$  is neither transitive nor intransitive since actor  $i$  does not have a tie to actor  $j$ ). The second triple,  $n_i n_j n_k$ , is clearly intransitive, since  $n_i \rightarrow n_k, n_k \rightarrow n_j$ , but  $n_i \not\rightarrow n_j$ . Thus, this triad is considered intransitive because of this single intransitive triple. The number of transitive and/or intransitive triples within a particular type of triad is very important when quantitatively and statistically assessing the amount of transitivity in a digraph.

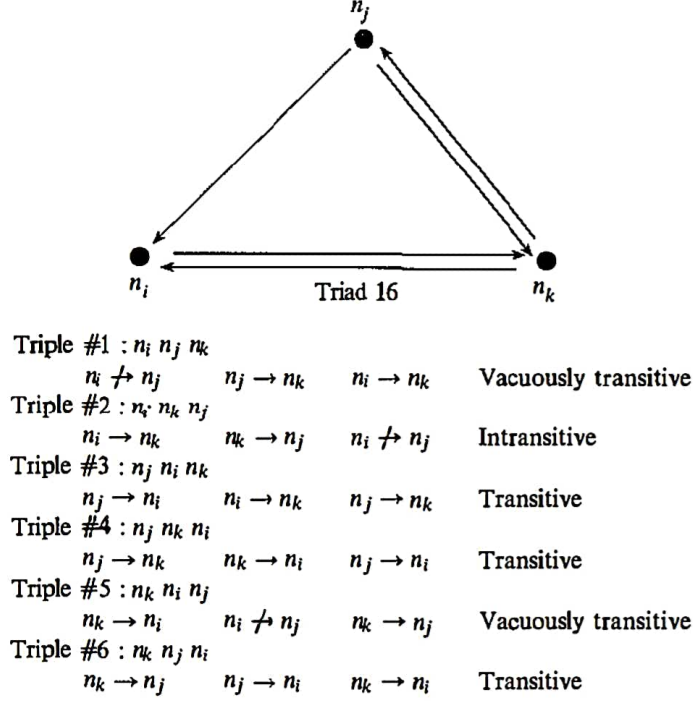


Figure 3.3: The type 16 triad, and all six triples of actors

The generality of transitivity can be seen, for example, by looking at triad 2 from Figure 3.2. This triad, which is not allowed under ranked clusterability, has just a single asymmetric dyad, so it is vacuously transitive. Vacuously transitive triples are allowed under transitivity, so type 2 triads can arise, without invalidating the idea.

The other triad that was problematic for ranked clusterability was triad 16. But this triad is almost transitive. Only one of its six triples is intransitive. So, the presence of this 5/6th's transitive/vacuously transitive triad in a data set is not such a big deal (assuming transitivity is operating).

# Chapter 4

## Blockmodels

### 4.1 Definition

We begin with a set of  $R$  dichotomous relations defined on a one-mode network of  $g$  actors. A *blockmodel* consists of two things:

1. A partition of actors in the network into discrete subsets called positions.
2. For each pair of positions a statement of the presence or absence of a tie within or between the positions on each of the relations.

A blockmodel is thus a *model*, or a *hypothesis* about a multirelational network. It presents general features of the network, such as the ties between positions, rather than information about individual actors.

We can define a blockmodel more precisely in terms of a mapping of the actors in the network onto the positions in the blockmodel. A *blockmodel* is a partition of the actors in  $\mathcal{N}$  into  $B$  positions,  $\mathfrak{B}_1, \mathfrak{B}_2, \dots, \mathfrak{B}_B$ , and onto mapping  $\phi$  from  $\mathcal{N}$  onto the collection of positions, where  $\phi(i) = \mathfrak{B}_k$  if actor  $i$  is in position  $\mathfrak{B}_k$ . A blockmodel also specifies the ties between and within the  $B$  positions. We let  $b_{klr}$ , indicate the presence or absence of a tie from position  $\mathfrak{B}_k$  to position  $\mathfrak{B}_l$  on relation  $\mathfrak{X}_r$ , where  $b_{klr} = 1$  if there is a tie from position  $\mathfrak{B}_k$  to position  $\mathfrak{B}_l$  on relation  $\mathfrak{X}_r$ , and  $b_{klr}=0$  otherwise.

A blockmodel is also represented by an image matrix,  $\mathbf{B}=\{b_{klr}\}$ . The image matrix is a  $B \times B \times R$  array, with entries  $b_{klr}$  indicating the presence or absence of a tie from position  $\mathfrak{B}_k$  to position  $\mathfrak{B}_l$  on relation  $\mathfrak{X}_r$ . Each layer of  $\mathbf{B}$  describes the hypothesized ties between and within positions on the specific relation. The matrix  $\mathbf{B}$  has also been referred to as a blockmodel, since it specifies the presence or absence of ties between positions. Whereas the original relational data are presented in the usual  $g \times g \times R$  multirelational sociomatrix, a blockmodel is a simplification in that it consists of a smaller  $B \times B \times R$  array,  $\mathbf{B}$ , that presents ties between positions.

A blockmodel thus has two components: the mapping,  $\phi$ , that describes the assignment of actors to positions, and the matrix,  $\mathbf{B}$ , that specifies the presence or absence of ties between and within positions on each relation. Each actor is assigned to one and only one of the positions, and If the assignment is the same across relations.

Each of the entries in the  $B \times B \times R$  matrix  $\mathbf{B}$  is called a *block*. Each block,  $b_{klr}$ , in the blockmodel corresponds to a submatrix of the original sociomatrix that contains the relevant interposition or intraposition ties. A block containing a 1 is called a *oneblock*, and indicates the presence of a tie from the row position to the column position. A oneblock may also be referred to as a *bond*. A block containing a 0 is called a *zeroblock*, and indicates the absence of a tie from the row position to the column position. More formally, if there is a hypothesized tie from position  $\mathfrak{B}_k$  to position  $\mathfrak{B}_l$  on relation  $\mathfrak{X}_r$ , then  $b_{klr} = 1$  in the blockmodel;  $b_{klr}$  is a oneblock. If there is no hypothesized tie from position  $\mathfrak{B}_k$  to position  $\mathfrak{B}_l$  then  $b_{klr} = 0$  in the blockmodel;  $b_{klr}$  is a zeroblock.

A blockmodel is a simplified representation of multirelational network that captures some of the general features of a network's structure. Specifically, positions in a blockmodel contain actors who are approximately structurally equivalent. Actors in the same position have identical or similar ties to and from all actors in other positions. For example all actors in position  $\mathfrak{B}_k$  have similar ties to actors in positions  $\mathfrak{B}_l, \mathfrak{B}_m$ , and so on. Thus, the blockmodel is stated at the level of the positions, not individual actors.

## 4.2 Building Blocks

Suppose that we start with a partition of actors into  $B$  positions, and have permuted the rows and the columns of the sociomatrix for each relation so that actors who are assigned to the same position occupy adjacent rows and columns in the permuted sociomatrix. In the permuted sociomatrix, all entries,  $x_{ij}$ , are the observed values of the ties between actors in the positions and all ties pertaining to ties between or within positions will be contained in submatrices of the sociomatrix. If all actors within each position are perfectly structurally equivalent, then all submatrices corresponding to ties within and between positions, for all relations, will be filled either completely with 0's or completely with 1's. However, in real network data, pairs (or collections) of actors are seldom structurally equivalent. In the permuted sociomatrix the submatrices corresponding to inter- and intraposition ties will usually contain both 1's and 0's. Therefore, determining whether a block in a blockmodel is a oneblock or a zeroblock is not straightforward. Constructing a blockmodel requires a rule which governs the assignment of a 0 or 1 to the tie between positions in the model.

There are several criteria which have proved useful for deciding whether a block should be coded as a zeroblock or a oneblock. These include:



- Perfect fit (fat fit)
- Zeroblock (lean fit)
- Oneblock
- $\alpha$  density criterion
- Maximum value -for valued data
- Mean value - for valued data

We first define each of these rules and then discuss when each one might be appropriate.

In a blockmodel, each of the  $B \times B \times R$  elements of  $\mathbf{B}$  contains the hypothesized value of the tie from the row position to the column position on the layer relation. As described above,  $b_{klr}$  denotes the value of the hypothesized tie from position  $\mathfrak{B}_k$  to position  $\mathfrak{B}_l$  on relation  $r$ . If the block is a oneblock then  $b_{klr} = 1$ , and if the block is a zeroblock then  $b_{klr} = 0$ . The decision about whether a tie exists or not in each block of  $\mathbf{B}$  depends on the observed values of the ties between actors in the positions. That is,  $b_{klr}$  depends on the values of  $x_{ijr}$  for  $i \in \mathfrak{B}_k$  and  $j \in \mathfrak{B}_l$ . We will let  $g_k$  be the number of actors in position  $\mathfrak{B}_k$  and  $g_l$  be the number of actors in position  $\mathfrak{B}_l$ . For distinct  $\mathfrak{B}_k$  and  $\mathfrak{B}_l$ , there will be  $g_k \times g_l$  ties from members of position  $\mathfrak{B}_k$  to members of position  $\mathfrak{B}_l$ . For ties among members of the same position, there will be  $g_k \times (g_k - 1)$  ties among actors in position  $\mathfrak{B}_k$ . Note that in a blockmodel, ties from a position to itself are meaningful, and often quite important theoretically, in contrast to reflexive ties for actors and diagonal entries in a sociomatrix, which are often undefined.

The most common criteria for defining oneblocks and zeroblocks are based on the density of ties within a block. The density of ties in block  $b_{klr}$  will be denoted by  $\Delta_{klr}$  and (for a dichotomous relation) is defined as the proportion of ties that are present. For  $k \neq l$  this proportion is:

$$\Delta_{klr} = \frac{\sum_{i \in \mathfrak{B}_k} \sum_{j \in \mathfrak{B}_l} X_{ijr}}{g_k g_l}$$

The density of ties within a position, for example block  $b_{klr}$ , is equal to:

$$\Delta_{klr} = \frac{\sum_{i \in \mathfrak{B}_k} \sum_{j \in \mathfrak{B}_k} X_{ijr}}{g_k (g_k - 1)}, \text{ for } i \neq j.$$

We can now specify more formally some useful criteria for defining zeroblocks and oneblocks in a blockmodel.

#### 4.2.1 Perfect Fit (Fat Fit)

The perfect fit (or fat fit) blockmodel occurs if all actors in each position are structurally equivalent. This ideal situation results in submatrices in the permuted sociomatrix filled with all 1's or with all 0's. The criterion for a perfect fit blockmodel requires that the tie between two positions on a given relation is equal to 1

only if all actors in the row position have ties to all actors in the column position, and a tie between positions is equal to 0 only if there are no ties from actors in the row position to actors in the column position.

$$b_{klr} = \begin{cases} 0 & \text{if } X_{ijr} = 0 \text{ for all } i \in \mathfrak{B}_k, j \in \mathfrak{B}_l, \text{ and} \\ 1 & \text{if } X_{ijr} = 1, \text{ for all } i \in \mathfrak{B}_k, j \in \mathfrak{B}_l. \end{cases}$$

The only way that this criterion can be met for all blocks is if all actors in all positions are structurally equivalent. Thus, it is quite unlikely that this criterion will be useful in practice. However, as an ideal, the perfect fit criterion can provide a baseline for assessing the goodness-of-fit of a blockmodel.

#### 4.2.2 Zeroblock (Lean Fit) Criterion

The zeroblock criterion states that the tie between two positions on a given relation is 0 only if there are no ties from actors in the row position to actors in the column position on the specified relation, otherwise the block is a oneblock.

$$b_{klr} = \begin{cases} 0 & \text{if } X_{ijr} = 0, \text{ for all } i \in \mathfrak{B}_k, j \in \mathfrak{B}_l \\ 1 & \text{otherwise.} \end{cases}$$

The focus on zeroblocks as structurally important phenomena arises because of the expectation that while one blocks might not be completely filled with 1's, blocks that contain no observed ties indicate important structural patterns. Substantively, if we expect that effort is required to maintain a tie, then a single observed "1" in a submatrix should be taken as an important tie in the blockmodel. For example if we recorded the incidence of military interventions by countries during a given year, these rare events would nevertheless indicate an important political tie, not only between individual countries, but also between positions. The zeroblock criterion is reasonable if ties are scarce and/or if the density of the sociomatrix is small. The fact that although zeroblocks should contain only 0's, oneblocks might contain both 1's and 0's gives rise to the alternative label lean fit. The oneblocks might be "lean" rather than "fat."

#### 4.2.3 Oneblock Criterion

The oneblock criterion focuses on oneblocks rather than on zeroblocks. This criterion requires that the submatrix of the sociomatrix corresponding to the intra- or interposition ties be completely filled with 1's. All possible ties from actors in the row position to actors in the column position need to be present in order to define a oneblock, otherwise it is a zeroblock:

$$b_{klr} = \begin{cases} 1 & \text{if } x_{ijr} = 1, \text{ for all } i \in \mathfrak{B}_k, j \in \mathfrak{B}_l \\ 0 & \text{otherwise.} \end{cases}$$

The oneblock criterion might be most appropriate when the relation is dense, rather than sparse. However, in practice, oneblocks seem to be quite rare.

#### 4.2.4 $\alpha$ Density Criterion

Since real social network data rarely contain (perfectly) structurally equivalent actors, blockmodels that are based on the property of structural equivalence are unlikely to contain blocks all of which are either perfect oneblocks or perfect zeroblocks. For various reasons we expect that oneblocks might contain some 0's and zeroblocks might contain some 1's. Therefore it is reasonable to define a threshold density,  $\alpha$ , such that if the observed block density,  $\Delta_{klr}$  is greater than or equal to  $\alpha$  then the block will be coded as oneblock, and if the observed block density is less than  $\alpha$  then the block is coded as a zeroblock. We define the  $\alpha$  criterion as:

$$b_{klr} = \begin{cases} 0 & \text{if } \Delta_{klr} < \alpha \\ 1 & \text{if } \Delta_{klr} \geq \alpha \end{cases}$$

One guideline for choosing a value of  $\alpha$  is that it should depend on the density of the relations in the analysis. Two commonly used values are the overall (grand) density computed across all relations, or, since all relations are unlikely to have the same density, there could be  $R$  separate  $\alpha$ 's, one for each relation ( $\alpha_r = \Delta_r$ ).

# Chapter 5

## Link Prediction Models

Social networks present high dynamics and a continuous transformation by adding new nodes and edges. This behavior causes changes in the nature of the social interaction and the structure of the network. For various domains it would be a great benefit to be able to understand and therefore control the mechanism of evolution of social networks.

Apart from influence, another fundamental topic in the evolution of social networks is the link prediction problem. This subject has captured the attention of various scientists, especially in the artificial intelligence sector and data mining.

Many studies refer business and professional collaborations generated by informal social interactions in such networks. Other studies focus on the impact of the social hierarchy in the professional network or inferring missing links. It is interesting to notice that most of these studies conclude that effective and concrete link prediction methods can be used to analyze social networks so to predict future interactions that might help organizations, businesses or investigations.

The social network analysis proved a significant role in domains as security, terrorism, biology, sales and many others. In some of the domains, such as security and terrorism, the type of prediction is of a link between groups of individuals that collaborate, but not by an obvious connection. In domains similar to sales, a typical type of link predictions regards the potential collaboration based on observations of business and informal interests and actions.

Today, due to the large amount of available social networking data, studies and simulation of different nature are possible. These contribute significantly to understanding the properties and the behavior of social networks.

## 5.1 Mathematical framework

Consider a social network  $G=(V,E)$ , where  $V$  is the set of network nodes and  $E$  is the set of edges between the network nodes, the problem of link prediction is the task to predict how likely a new link  $e_{i,j} \notin E$  will exist between a pair of existing nodes in the network  $(v_1, v_2)$ .

Often the time dimension is added to the link prediction problem so to measure the growth of the network. In this case the discussed problem should be seen as the task of accurate prediction of the edges that will be added to the network between two deterministic points in time.

The link prediction problem addresses four main aspects: link existence, link type, link weight and link cardinality. Many link prediction studies concentrate on the problem of link existence - whether a new link between two nodes in a given social network will exist in the future or not. The link existence problem is extended by the other two problems of link prediction: link weight – the links between different network nodes are given different weights and link cardinality – two nodes of a given social network are connected to more than one link. The fourth problem, the link type is a more particular problem - it refers to possible different roles of the one relationship between the same two nodes of the given social network.

The link prediction problem can be treated with techniques of various natures: statistics, probability, graph theory, machine learning etc. Depending though on the approach of analysis, the techniques can be classified in three groups:

- **Models based on node similarity** – regards the similarity measurement between two nodes.
- **Models based on topological patterns** - local or global patterns that could define the network
- **Methods based on probabilistic models** – a defined model that could abstract the network

### 5.1.1 Models based on node similarity

The models based on node similarity propose measurements of similarity for pairs of network nodes. In this context, the task of link prediction is the consideration of new edges between network nodes presenting a considerable similarity, usually measured against a threshold. In general, the measurement of similarity is either (pre)defined or learned (using machine learning techniques), depending on the studied domain or the type of network.

The similarity between two network nodes  $(v_1; v_2)$  can be defined by the percentage of the common information in the total set of properties characterizing



the two nodes. The measurement is applicable in case of a probabilistic model for the studied case:

$$sim(v_1; v_2) = \frac{\log P(common(\vartheta_1; \vartheta_2))}{\log P(description(\vartheta_1; \vartheta_2))}$$

where  $\vartheta_1, \vartheta_2$  are the sets of properties characterizing the two nodes  $v_1; v_2$ . Another similarity distance measurement was given by Bennett and Li [2004] and refers to the Kolmogorov complexity measurement between the set of properties of the two nodes ( $v_1; v_2$ ). the Kolmogorov complexity measurement of a binary string  $v$  is defined as the length of the shortest program for an Universal Turing Machine (UTM) to correctly reproduce the considered string,  $v$ . Consider  $v_i, v_j$  the binary strings corresponding to the set of properties of the two nodes, for a given UTM, the Kolmogorov complexity measurement  $K(v_i|v_j)$  is the length of the shortest program for the UTM to output  $v_i$  when given  $v_j$  as input. In this context, the similarity measurement is formulated as:

$$dis(v_1; v_2) = \frac{\max\{K(v_2|v_1), K(v_1|v_2)\}}{\max\{K(v_1), K(v_2)\}}$$

The disadvantage of such predefined similarity measurements is that they do not consider the network context. For this reason, the adaptive similarity functions are frequently learned using supervised learning techniques. Some of the most representative techniques are: Binary classifiers, Kernel methods and Statistical Relational Learning (SRL).

**Binary classifiers** are proposing training a binary classifier to determine the similarity between two network nodes, based on their content information. A mapping feature function is used to extract the content features of the two network nodes in a single vector  $\hat{a}(v_1; v_2)$ . Considering a simple linear regression, the objective of the function is learning a set of parameters  $w$  that can indicate best similarity. For a candidate node pair, the link prediction problem is reduced to:

$$link(v_1; v_2) = \begin{cases} DoesExist, & \text{if } w'\hat{a}(v_1; v_2) > 0 . \\ DoesNotExist, & \text{if } w'\hat{a}(v_1; v_2) < 0 . \end{cases}$$

Within the set of pairs not selected as candidates (negative examples), it is possible and should be considered that new links might exist. Another conclusion is that in networks with few or sparse links, the number of candidates and non-candidates pairs is considerably unbalanced. The binary classifiers are best applicable when nodes of a certain class have many features in common, else finding pairs is very difficult and the consequence is a high recall.

**Kernel matrices** methods are proposing an alternative to the binary classifiers that suit also the case when the set of common features between nodes

of the same class is reduced. One approach is capturing the content information of the network nodes in Cartesian products for pairs of features  $\langle v^\alpha; v^\beta \rangle$

$$\hat{\mathbf{a}}_{Cart}^{<\alpha,\beta>} = (v_1^\alpha v_1^\beta, v_1^\alpha v_2^\beta, \dots, v_1^\alpha v_n^\beta, v_2^\alpha v_1^\beta, v_2^\alpha v_2^\beta, \dots, v_n^\alpha v_1^\beta, v_n^\alpha v_2^\beta, \dots, v_n^\alpha v_n^\beta)$$

The problem with this approach is that the dimension of the feature set is  $n^2$ . Clearly, the involved computation is not practical in the case of networks with a large set of node-features. Also, conducting learning in a high dimensional feature-space is challenging and may lead to over-fitting.

A better solution is the approach of Support Vector Machine (SVM) learning algorithms, suggesting pairing nodes as inner products  $\langle v_1; v_2 \rangle$  and not considering the nodes individually. In this way, by using kernel functions  $K(v_1; v_2)$  for the defined inner products, the challenge of classification in higher dimensional feature-space can be solved.

$$K(\hat{v}_1, \hat{v}_2) = K((v_1^\alpha, v_1^\beta), (v_2^\alpha, v_2^\beta)) = \langle v_1^\alpha, v_2^\alpha \rangle \langle v_1^\beta, v_2^\beta \rangle$$

when  $(\hat{v}_1) = (v_1^\alpha, v_1^\beta)$  and  $(\hat{v}_2) = (v_2^\alpha, v_2^\beta)$  are instances of feature-pairs of the considered nodes. The proposed kernel is actually a tensor product between two linear kernels representing the inner products.

The link prediction problem considers the space of node-pairs as input space of nodes and the similarity between such pairs is defined by the explicit form of the proposed kernel. A high value of the kernel indicates high node-similarity. This approach has a wide applicability, especially in prediction of rating or collaborations. One specific domain of collaboration is the scientific co-authorship, and link prediction in such a community represents the subject of the second study presented in the paper.

**Statistical Relational Learning (SRL)** incorporates a variety of approaches and techniques. The nature of these methods can be statistical, probabilistic, logic-based algorithms etc. An established approach suggested for link prediction was established by Popescul [2003] and suggests using aggregation of relational features for measuring similarity. Various classification algorithms have been proposed and studied, many known from other disciplines such as data mining and machine learning. A particular approach is translating the link prediction problem in an optimization problem by mapping the network nodes to Euclidean spaces.

### 5.1.2 Models based on topological patterns

This approach is focused on identifying global or local topological patterns in the entire network or partial network. For fundamental concept in this approach is

scoring the weight of the link between the nodes of a pair  $(v_1; v_2)$  , in rapport to the determined topological pattern(s).

Depending on the leading element in determining the topological patterns, there can be distinguished three types of topological patterns approaches: Node based, Path based or Graph based.

**Node based approaches** take into consideration the neighborhood information of a node, for example the set of first neighbors that a node has. One consideration in this area is that two network nodes would more probably establish a link if they have a large number of common neighbors.

In the proposed link prediction study in the co-authorship world, due to the nature of the domain, such information is relevant and important. Scientists and researchers tend to set new collaboration with colleagues in the same area, based on the recommendations received from their collaboration partners, the first neighbors. In other words there is a high probability that a scientist will collaborate with his second neighbors. This is one topological feature considered in the algorithm comparison.

A number of measurements of this nature have been already formulated and standardized. These intend to define a scoring function for a potential link between two nodes  $(v_1; v_2)$ , most often based on structural considerations such as the number of direct neighbors a node has, noted  $\Gamma(v_i)$  and respectively  $\Gamma(v_j)$ . The most common node-based scoring functions are:

- **Common neighbors method** – proposes a scoring function of the link between two nodes  $(v_1; v_2)$  based on the number of common neighbors these nodes share:

$$score(v_1; v_2) = |\Gamma(v_1) \cap \Gamma(v_2)|$$

- **Jaccard coefficient** – proposes a scoring function of the link between two nodes  $(v_1; v_2)$  based on the ratio between their common neighbors and the total number of their neighbors:

$$score(v_1; v_2) = \frac{|\Gamma(v_1) \cap \Gamma(v_2)|}{|\Gamma(v_1) \cup \Gamma(v_2)|}$$

- **Adamic/Adar coefficient** – proposes a scoring function of the link between two nodes  $(v_1; v_2)$  based on the number of their common neighbors, weighting more those neighbors  $x \in \Gamma(v_1) \cap \Gamma(v_2)$  that the two nodes share least with other nodes in the network:

$$score(v_1; v_2) = \sum_{x \in \Gamma(v_1) \cap \Gamma(v_2)} \frac{1}{\log|\Gamma(x)|}$$

- **Preferential attachment method** – proposes a scoring function of the link between two nodes based on the premise that node  $v_1$  will receive a connection from node two  $v_2$  with a probability proportional to the number of neighbors of  $v_2$ ,  $|\log \Gamma(v_2)|$ . And vice versa:

$$score(v_1; v_2) = |\Gamma(v_1)| |\Gamma(v_2)|$$

**Path based approaches** take in consideration the path connectivity information between two network nodes. The main idea of this type of approaches is that the more indirect paths are connecting two nodes the higher the possibility that a link will connect them directly. Many studies contributed to the theory of shortest-path distance based on analysis of the entire set of indirect links connecting two network nodes.

As in the case of the node similarity approach, a number of measures based on the path similarity have been already established. The main ones are:

- **Katz measure** - proposes a scoring function of the link between two nodes based on the sum of the total number of paths weighted according their length. If the  $paths_{v_1;v_2}^{(l)}$  denotes all paths of length  $l$  between two network nodes  $(v_1; v_2)$  then the formulation of the Katz measure is:

$$score(v_1; v_2) = \sum_{l=1}^{\infty} \delta^l |paths_{v_1;v_2}^{(l)}|$$

where  $\delta^l > 0$  is a parameter of the predictor.

- **Hitting time measure** – proposes a scoring function of the link between two nodes based on the required steps to reach one of the nodes,  $v_2$  when starting from a certain node  $v_1$  and when using a random walk to move through the neighborhoods or the considered start node. The required number of steps is also called hitting time and is often notated with  $H_{v_1;v_2}$ . It is important to realize that this measure is not always symmetric. This is also why often an extension of the hitting time measure is used, the commute time:  $C_{v_1;v_2} = H_{v_1;v_2} + H_{v_2;v_1}$ . The scoring function  $score(v_1; v_2)$  is obtained by negating one of the two measures, hitting time or commute time.
- **PageRank measure** – proposes a scoring function of the link between two nodes  $(v_1; v_2)$  that measures the probability with which node  $v_2$  is present in a random walk that is returning to  $v_1$ . The measurement uses a parameter  $\delta \in [0, 1]$  considering that, at every step, the stationary probability of  $v_2$  in the walk is  $\delta$  and the probability of a move to another random neighbor is  $1 - \delta$ .
- **SimRank measure** – proposes a scoring function of the link between two nodes  $(v_1; v_2)$  indicating whether the similarity of the considered two nodes is shared by also with other neighbors of theirs. The measure is in fact a fixed point of the previous recursive formulation defined by the condition that for a parameter  $\delta \in [0, 1]$  the scoring function  $score(v_1; v_2) = 1$ . In this context, the SimRank measure is formulated as:

$$score(v_1; v_2) = \delta \frac{\sum_{x \in \Gamma(v_1)} \sum_{y \in \Gamma(v_2)} score(x, y)}{|\Gamma(v_1)| |\Gamma(v_2)|}$$

### 5.1.3 Models based on a probabilistic model

The fundamental concept of this approach is to learn a model based on a given network based on certain strategies of optimization such as *Maximum Likelihood (ML)* and *Maximum a Posteriori (MAP)*. Consider a network graph  $G = (V, E)$  and  $\partial$  the set of parameters of the learned model, the candidate future links  $l_{ij}$  are defined as variables in probabilistic models and can be defined as:  $p(l_{ij}|\partial)$ .

In this approach, three significant sub-categories of models are: Probabilistic Relational Models (PRM), Bayesian Relational Models (BRM) and Stochastic Relational Models (SRM). The first two categories are based on specific database structure representations: the PRMs are corresponding to the Relational Model and the BRMs, defined on the *Directed Acyclic Probabilistic Entity Relationship (DAPER)* framework, are corresponding to the Entity-Relationship Model.

**Probabilistic Relational Models**, different than the classical graph models, propose a set of three graphical models for representing the network relational data: data graph ( $G_D = (V_D, E_D)$ ), model graph ( $G_M = (V_M, E_M)$ ) and inference graph ( $G_I = (V_I, E_I)$ ). The original application of these models was in the problem of attribute prediction for relational data. The probabilistic relational models reduce the link prediction problem to the task of prediction of the existence of attributes for potential new network links. Therefore, with the PRM framework, the link prediction problem requires setting up an *< exist >* attribute.

The data graph ( $G_D = (V_D, E_D)$ ) contains the network information as the set of nodes,  $v_i \in V_D$  and the set of links defined between these nodes,  $e_j \in E_D$ . Each node and link have associated a type  $t_i \in T : T(v_i) = t_{v_i}$  and  $T(e_j) = t_{e_j}$  and implicitly by a set of attributes corresponding to this type,  $Z^{t_i} = (Z_1^{t_i}, \dots, Z_{m_{t_i}}^{t_i})$ . As the PRMs consider a joint probability distribution over the network data information (attributes), in the given context, this can be formulated then as:

$$z = \{z_{v_i}^{t_{v_i}} : v_i \in V_D, T(v_i) = t_{v_i}\} \cup \{z_{e_j}^{t_{e_j}} : e_j \in E_D, T(e_j) = t_{e_j}\}$$

The model graph ( $G_m = (V_m, E_m)$ ) has the purpose to present the dependencies between the type attributes  $Z$  characterizing the set of network nodes  $V_D$ . There can be probabilistic dependencies between attributes of the same type or different types. The model graph ties together the network entities with the same type as well as the attributes of these entities. In this way, a decomposition of the data graph per type can be achieved, this leading to a joint model of type attributes dependencies. Aside the structure of dependencies between the defined type attributes, a second component of the model graph is the *Conditional Probability Distributions (CPD)* associated with the network nodes.

The inference graph ( $G_I = (V_I, E_I)$ ) is generated based on the prior two models  $G_D$  and  $G_M$  through a process similar to the one used by the Hidden Markov Models (HMM) to instantiate sequence models. In this process, the structure of  $G_I$  is defined based on the  $G_D$  and  $G_M$ , with the particularity that for each node-attribute pair in  $G_D$  a local copy of the correspondent CPD from  $G_M$  is made in  $G_I$ .

The PRMs differ among them mainly in the definition of the model graph  $G_M$ , the learning models and inference procedures. A number of *Probabilistic Relational Models* are next introduced:

- **Relational Bayesian Networks (RBN)** use the object oriented approach for extending the Bayesian networks concept. The model graph ( $G_m^D = (V_m, E_m^D)$ ) in this case is a *Directed Acyclic Graph* (DAG) representing the joint distribution over the network entity type attributes by a set of CPDs. A CPD corresponding to an attribute  $Z$  is specified by the likelihood  $p(Z|pa(Z))$ , where  $pa(Z)$  represents the value of the parents of  $Z$ . In general though, a network object is characterized by a set of attributes  $(Z_1, Z_2, \dots, Z_n)$ , the DAG and CPT specifying the Bayesian network and representing the distribution for the  $n$ - dimensional random attributes as:

$$p(Z_1, Z_2, \dots, Z_n) = \prod_{i=1}^n p(Z_i|pa(Z_i))$$

Corresponding to the dependencies in the DAG structure, the joint probabilistic distribution can be expressed as a factorization of the following form:

$$p(z) = \prod_{t \in T} \prod_{z_i^t \in z^t} \prod_{v: T(v)=t} p(z_{v_i}^t | pa_{z_{v_i}^t}) \prod_{e: T(e)=t} p(z_{e_i}^t | pa_{z_{e_i}^t})$$

where  $v_i \in V_D$  are the network nodes,  $e_i \in E_D$  are the set of links defined between these nodes,  $t_i \in T$  are the set of types associated to the network nodes and links:  $T(v_i) = t_{v_i}$  and  $T(e_i) = t_{e_i}$ . Each  $t_i \in T$  is defined by a set of attributes  $Z^{t_i} = (Z_1^{t_i}, \dots, Z_{m_{t_i}}^{t_i})$ .

The structure learning problem in a Bayesian network is similar to searching the optimum in the space of all DAGs. RBNs use closed-form parameter estimation techniques, helping the structure learning. The learning methods for RBN are similar to the ones used for Bayesian networks, the efficiency of such parameter learning techniques representing the strength of this approach.

For reasons of simplicity, accuracy and efficiency, the *Relational Bayesian Networks* propose a *belief propagation inference*.

**Relational Markov Networks (RMN)** extend the concepts of conditional Markov Networks for relational data. The model graph in this case is an undirected graph ( $G_M^U = (V_M, E_M^U)$ ) and represents the joint distribution over



the attribute  $z$  as a set of potential functions  $\phi = \{\phi_c | C \in \mathbf{C}\}$  where  $C_i \in \mathbf{C}$  is a set of templates of relational cliques specified by a RMN model for defining all cliques. For a graph  $G$ , a clique is a set of nodes  $V_c$  in  $G$ , not necessarily maximal (can be also one single node), such that each  $V_i, V_j \in V_c$  is connected by an edge in  $G$ . The combined probabilistic model for a set of variables  $Z$  is:

$$p(Z) = \frac{1}{N} \prod_{C_i \in \mathbf{C}} \prod_{c_j \in C_j} \phi_{C_i}(z_{c_j})$$

where  $N$  is a normalization constant and  $C_j$  represents all instantiations of the set of clique templates,  $\mathbf{C}$ .

The RMN models extend the learning techniques of the Markov networks with an approach of parameter estimation "*maximum-a-posteriori*" using Gaussian priors. The approach considers predefined clique templates, reducing the prediction problem to optimizing the potential functions  $\phi = \{\phi_c | C \in \mathbf{C}\}$ . With RMN models, the learning efficiency is not as high as in the case of RBN (the structure is not defined nor improved by learning) but this category of models presents flexible and detailed representations.

Similar to RBN models, in this case too, a *belief propagation* approach is used as inference procedure.

• **Relational Dependency Networks (RDN)** propose an extension of the dependency networks for relational data. The model graph in this case is a bi-directed graph ( $G_M^B = V_M, E_M^B$ ) presenting a set of CPDs. RDN models try to maximize the pseudo-likelihood for each variable  $z$  independently. For a considered graph data  $G_D$ , the pseudo-likelihood PL is formulated as the product over network item types  $t \in T$ , the set of type attributes  $Z^t$  and the nodes  $v_i$  and the links  $e_i$  of the considered type:

$$PL(G_D; \partial) = \prod_{t \in T} \prod_{z_i^t \in Z^t} \prod_{v: T(v)=t} p(z_{v_i}^t | pa_{z_{v_i}^t}; \partial) \prod_{e: T(e)=t} p(z_{e_i}^t | pa_{z_{e_i}^t}; \partial)$$

where  $\partial$  is global the set of parameters of the learned model and  $pa_z$  represents the value of the parents of  $z$ .

In this approach, there are used specific queries to define the relational neighborhoods. The learning algorithm used by the RDN models takes in consideration these queries, on one hand for structuring the learning and on the other hand for the parameter estimation. Different than RBN and RMN models, the CPDs of RDN models do not need factoring over the data model, being considered that for an attribute  $z_{v_i}^t$  the parent values are conditioned  $pa_{z_{v_i}^t}$ , independent of the fact that the parent values might have been conditioned by the considered attribute in their CPD estimation. The downside of the approach of independent CPD learning is that it does not lead with certainty to a consistent joint distribution.

In what concerns the inference approach, the RDN models propose the *Gibbs sampling* technique.

**Bayesian Relational Models** are based on the *Directed Acyclic Probabilistic Entity Relationship* (DAPER) framework, a probabilistic framework defined for the Entity-Relationship database model. The framework proposes the modeling of data in specific classes: *entities, relationships, arcs, attributes, constraints and local distribution*. Classes are connected by dashed lines. For link prediction, the entities and relationship classes are given equal importance. In real-world it is often encountered that in the defined relationships, one part is defined with certainty and the other part presents uncertainty. In these cases, uncertainty referencing is used.

A Bayesian approach is applicable to relational modeling as it proposes a clear representation of parameters and hyper parameters, not at global level, but at network component level (nodes and relationships). This approach supports the *Hierarchical Bayesian Framework* (HB), structure that centers the parameterization of the prior distribution on the consideration that the prior distribution should represent both the *prior belief* and *learned prior*. The DAPER framework is most often considered in the context of a *Hierarchical Bayesian Framework*, in either a parametric or a nonparametric form.

The parametric form, *Parametric Hierarchical Bayesian Relational Model*, is applicable in cases when the individual parameterization of network entities can be assumed to derive from a common prior distribution which can be learned and shared globally by the network entities.

Often the parameterization of prior belief and learned prior are different distribution types and therefore a non-parametric prior distribution presents more flexibility. This model is known as the *Non Parametric Hierarchical Bayesian Relational Model* and is based on specifying the prior distribution as a sample from a Dirichlet Process (DP), seen as a generalization of the *Dirichlet* distribution, infinitely dimensioned.

*Dirichlet Enhanced Relational Learning Model* (DERL) is formulated as:  $G_{pc} \sim DP(G_0, \alpha_0)$ , a sample from a DP where the base distribution  $G_0$  presents uncertain prior belief and  $\alpha_0 \geq 0$  represents the parameter reflecting the prior belief certainty. The flexibility of this approach lies in the fact that a multinomial parameter  $\partial_{|pc,pa}$  can be expressed as samples from the  $G_{pc}$  prior, when this is rich:  $\partial_{|pc,pa} \sim G_{pc}$

A relational learning model is expected to predict new entities and relationship attributes based on the already defined relationship attributes.

A generalization of the nonparametric DERL model is the Infinite Hidden Relational Model (IHRM), which combines the Hidden Relational Model with a DP Mixture Model. The DP *Mixture Model* aims to determine in an organized manner the appropriate number of latent states by embedding an infinite number of DP mixture models, which based on the considered data, are limited

automatically to a finite number of mixture components.

A challenge in the relational learning the large number of features that might characterize an attribute. A solution is capturing information in latent variables so that information can be distributed at global level in the network and the need of extensive structural learning is reduced. From this perspective, the *Hidden Relational Model* can be considered as a generalization of *Hidden Markov Models* (HMM) using hidden Markov random fields.

A second particular nonparametric model is the *Infinite Relational Model*, very much alike IHRM, though independently formulated. The main difference between the two models is that IHRM is able to define a CPD for an attribute based on structural consideration (considering its structural parents) and IRM, by modeling attributes as unary predicates, represents the CPD in a logical binary form.

**Stochastic Relational Models** propose a *Gaussian Process* (GP) framework based on the consideration that, for prediction tasks, the training models using a discriminative approach perform better than the generative models. The pioneers of this framework are Yu and Chu. The principal of *Stochastic Relational Models* is defining a GP for each entity type and then using a tensor composed by the set of such defined GPs for modeling the stochastic network link structure. The approach considers that the candidate links are local derivatives of a latent relational function:  $\tau : UXV \rightarrow E$ . A candidate link  $l_{i,j}$  is dependent on its correspondent latent value  $\tau_{i,j}$  and is modeled by the probability  $p(l_{i,j}|\tau_{i,j})$ . The candidate links introduce a set of *Stochastic Relational Processes* (SRP) defined on  $U \times V$ , generating the function  $\tau$  via the tensor interaction of two GP kernel functions, one defined on  $U$  and one defined on  $V$  ( $U, V$  could have infinite number of network entities). The SRPs are described by a set of two hyper parameters  $\sigma = \{\sigma_\epsilon, \sigma_\vartheta\}$ , corresponding to the GP kernel functions on  $U$ , respectively  $V$ .

In this context, the *Stochastic Relational Models* (SRM) define a Bayesian-prior for latent variables  $\tau$  denoted  $p(\tau|\sigma)$ . For a set of candidate links  $C$ , the marginal probability is then formulated as:

$$p(L_I|\sigma) = \int \prod_{(i,j) \in C} p(l_{i,j}|\tau_{i,j}) p(\tau|\sigma) d\tau, \quad \sigma = \{\sigma_\epsilon, \sigma_\vartheta\} \text{ and } L_I = \{l\}_{(i,j) \in C}$$

By estimating the hyper parameters  $\sigma = \{\sigma_\epsilon, \sigma_\vartheta\}$  with the maximum marginal probability, the link prediction problem is realized by marginalization:  $p(\tau|E_I, \sigma)$ . This type of prediction is similar to general GP regressions, with the difference that the GP approach makes use of a set of hyper parameters. With the same constraint, the GP approach can be compared to a classification task.

A challenge of the approach is the scaling of GP inferences. Such attempts, due to the cubic complexity of GP inference, present computational risks even for networks of reduced size. If considering a network graph  $G = (V, E)$ , where  $V$  represents the set of network nodes and  $E$  represents the set of links between these nodes, the size of observations of missing links scales in  $\theta(VE)$ . GP inference has the computational complexity cubic to the missing data size,

$\theta(V^3E^3)$ , an extremely complex computation.

A solution for this problem is given by the *Stochastic Relational Process* (SRP). This approach starts from the probabilistic model which considers that the link candidate solution is generated by the latent function  $\tau : UXV \rightarrow E$  following the GP process  $GP(u, K)$  where  $u$  is the mean function and  $K$  is the kernel function between network links. Considering two network links:  $(v_i, v_j)$  and  $(v'_i, v'_j)$ , the  $K$  covariance function can be expressed depending on the other two kernel functions,  $\epsilon, \vartheta$ , defined on  $U$  and  $V : K((v_i, v_j), (v'_i, v'_j)) = \epsilon(v_i, v'_i)\vartheta(v_j, v'_j)$ .

The link structure dependency can be expressed by node dependency. In this way, based on a similarity notion ensured by the kernel function, if considering two pair of similar nodes:  $v_i$  with  $v'_i$  and  $v_j$  with  $v'_j$ , then also  $\tau(i, j)$  is similar with  $\tau(i', j')$ . The edge descriptive function  $\tau$  can be defined thus by a factorization of two node descriptive functions which are samples of the priors:  $GP(0, \epsilon)$  and  $GP(0, \vartheta)$ . In this way the computational complexity of the GP is of range  $\theta(V^3 + E^3)$ , a significant complexity reduction.

A second approach of improving the GP scaling complexity is based on a link descriptive covariance:

$$K((v_i, v_j), (v'_i, v'_j)) = \frac{1}{\sqrt{2}}(C(v_i, v'_i)C(v_j, v'_j) + C(v_i, v'_j)C(v_j, v'_i)), \text{ where } C(v_i, v_j) = \langle z_i, z_j \rangle$$

This approach is very similar to the previous one presented, based on a node descriptive covariance. With this approach the computational complexity of the GP is of range  $\theta(\rho^3 + \rho^2|\mathbb{O}|)$ , where  $|\mathbb{O}|$  represents the input network links and  $\rho$  represents a small value.

# Conclusion

Social networks are a popular way to model the individual interaction within an organized group or community. Such social structure can be visualized as a network or graph, where an actor represents a group member and a link represents the form of association between two members of the group. Social Network Analytics combines the concept of the sociogram with elements of graph theory to analyze patterns of interaction among the group members, allowing quantitative comparisons between different network structures.

Due to the recent globalization of the commercial environment and the impact of the new technologies, the analysis of social networks represents a major interest. This rather new area of research grew out of social and exact sciences, computers supporting today modeling and complex mathematical calculations, previously impossible. The analysis of social networks is driven by business and social interests, combining various academic fields.

The current paper introduced the fundamental concepts and metrics in Social Network Analytics and proposed a set of mathematical models that can be applied for the problem of link prediction.

Link prediction is a measure of social proximity between two individuals in a community that can be used to optimize an objective function over the entire social network. The link prediction problem implies modeling the way an information, a trend, a piece of knowledge etc. propagates via a social network. Such knowledge supports the development of tools for detection of hidden, missing or potential new links within a group. These type of problems are critical in many domains: security and criminal investigation, biology, marketing and sales, CRM, knowledge management systems and so on. A common weakness in the link prediction studies, is the fact that social structures and their evolutions are studied separately. Therefore, some of the major interests in the domain are the link prediction problem in dynamic social networks and the knowledge exchange between heterogeneous social networks.

Social networking provides clear advancements in communication and self expression. Businesses uses social networking to promote products, concepts and services. But if not understood and managed properly, social networking could cost the reputation of business and individuals.

# References

- [1] Borgatti, S. P., and Everett, M. G. *A Graph-theoretic perspective on centrality*. In Social Networks 28: 466-484, 2006.
- [2] Elena Pupazan *Social Networking Analytics.*, 2011.
- [3] Freeman, L. C. *Centrality in social networks conceptual clarification*. In Social Networks 1: 215-239, 1978.
- [4] Liben-Nowell, D., and Kleinberg, J. M. *The link-prediction problem for social networks*. 1019–1031, 2007.
- [5] Popescul, R., and Ungar, L. H. *Statistical relational learning for link prediction*. 2003.
- [6] Wasserman, S, and Faust, K. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 243-246, 394-406 1994.
- [7] White, D. R., and Borgatti, S. P. *Betweenness centrality measures for directed graphs*. In Social Networks 16: 335-346, 1994.



# ORBITAL MECHANICS

Project Report submitted to

**ST. MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**

Affiliated to

**MANONMANIAM SUNDARANAR UNIVERSITY, TIRUNELVELI**

In partial fulfilment of the requirement for the award of degree of

**Bachelor of science in Mathematics**

Submitted by

NAMES

V. VEERALAKSHMI  
A. RATHNA ABISHA  
P. EMIMA  
M. SEYAD ALI FATHIMA  
N. MARIA LURDU ABIYA  
P.MUTHURAMU

REG.NO.

18AUMT53  
18AUMT42  
18AUMT20  
18AUMT46  
18AUMT36  
18AUMT38

Under the Guidance of

**Dr. G. PRISCILLA PACIFICA M.Sc., B.Ed., M.Phil., Ph.D., SET**

Assistant Professor of Mathematics

**ST. MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.**



Department of Mathematics

**ST. MARY'S COLLEGE (AUTONOMOUS)**  
Thoothukudi. (2020-2021)

# CERTIFICATE

We hereby declare that the studies on **“ORBITAL MECHANICS”** being submitted to **St. Mary’s College (Autonomous), Thoothukudi** affiliated to **Manonmaniam Sundaranar University, Tirunelveli** in partial fulfilment for the award of degree of **Bachelor of science in Mathematics** and it is a record of work done during the year 2019-2020 by the following Students:

NAMES	REG.NO.
V. VEERALAKSHMI	18AUMT53
A. RATHNA ABISHA	18AUMT42
P. EMIMA	18AUMT20
M. SEYAD ALI FATHIMA	18AUMT46
N. MARIA LURDU ABIYA	18AUMT36
P. MUTHURAMU	18AUMT38

Jeffrey Pacific  
09/04/21  
Signature of the Guide

Signature of the HOD

V.L. Snelle Appelhe. May  
Signature of the Examiner

Lucia Rose  
Signature of the Principal  
St. Mary's College (Autonomous)  
Thoothukudi - 628 001

## DECLARATION

We hereby declare that the project report entitled "ORBITAL MECHANICS" is our original work. It has not been submitted to any university for any degree or diploma.

V. Veeralakshmi  
(VEERALAKSHMI.V)

M. Seyad Ali Fathima  
(SEYAD ALI FATHIMA.M)

A. Rathna Abisha  
(RATHNA ABISHA.A)

P. Muthuramu.  
(MUTHURAMU.P)

P. Emima  
(EMIMA.P)

N. Mariya Lurdu Abiya  
(MARIA LURDU ABIYA.N)

## ACKNOWLEDGEMENT

First of all, we thank lord Almighty for showering his blessings to undergo this project.

With immense pleasure, we register our deep sense of gratitude to our guide **Dr. G. Priscilla Pacifica M.Sc., B.Ed., M.Phil., Ph.D., SET** and the Head of the Department **Dr. A. Punitha Tharani M.Sc., M.Phil., Ph.D.**, for having imported necessary guidelines throughout the period of our studies.

We thank our beloved Principal **Rev. Dr. Sr. A.S.J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D.**, for providing us the help to carry out our project work successfully.

Finally, we thank all those who extended their helping hands regarding this project.

# Content

## Certificate

## Acknowledgement

pg.no

## Introduction

1

## Chapter

### 1 Dynamics of point masses

2 - 17

1.1 Introduction

2

1.2 Kinematics

2

1.3 Mass, force and Newton's law of gravitation

6

1.4 Newton's law of motion

8

1.5 Time derivatives of moving vectors

11

## Chapter

### 2 The two-body problem

14 - 22

2.1 Introduction

14

2.2 Equations of motion in an inertial frame

14

2.3 Equations of relative motion

16

2.4 Angular momentum and the orbit formulas

17

## Chapter

### 3 Orbits in three dimensions

23- 30

3.1 Introduction

23

3.2 Geocentric right ascension–declination frame

24

3.3 State vector and the geocentric equatorial frame

26

3.4 Orbital elements and the state vector

## Chapter

### 4 Preliminary orbit determination

31 - 36

4.1 Gibbs' method of orbit determination from three position vectors

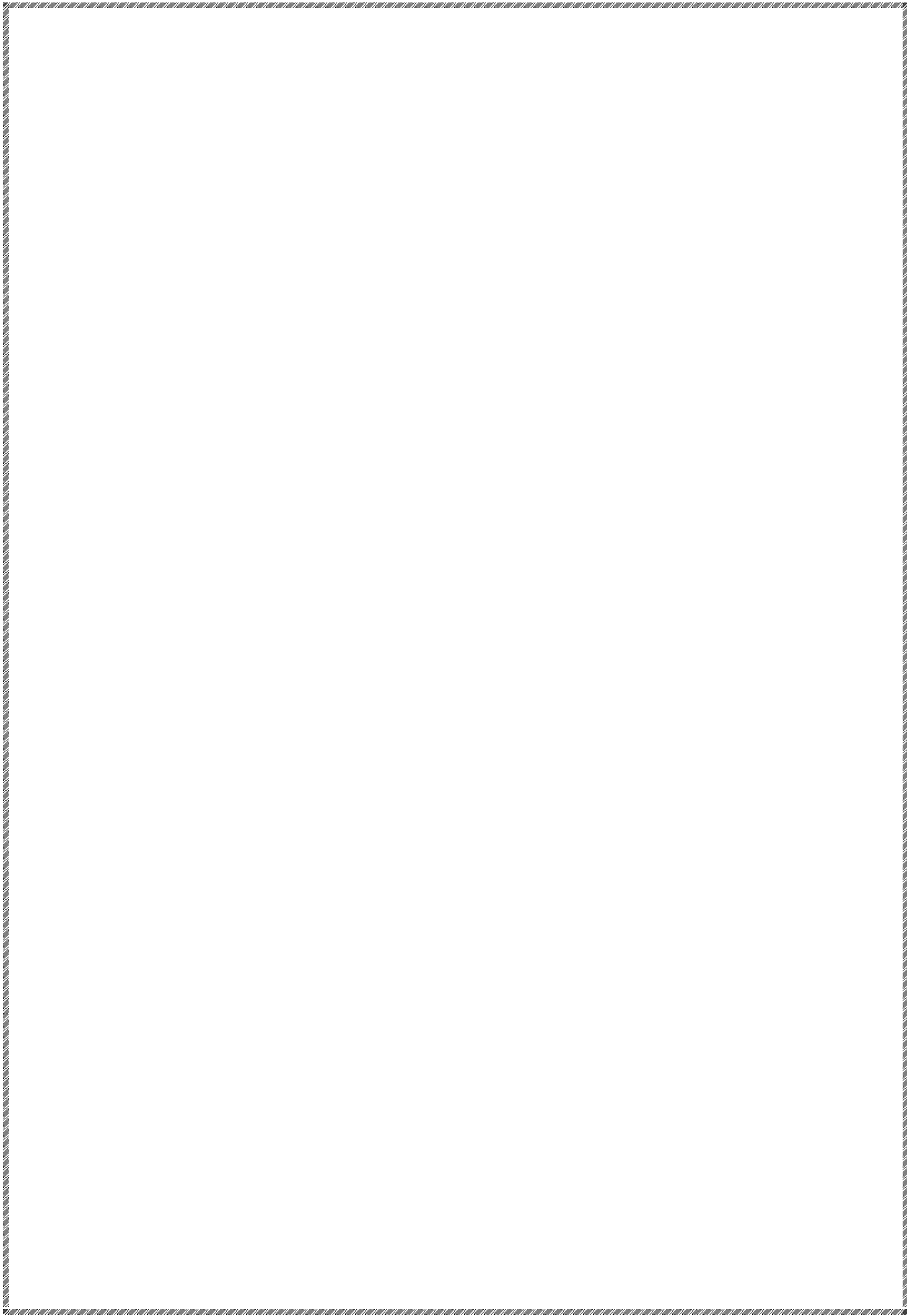
31

## Conclusion

37

## References

38





# Introduction

A orbit is a regular, repeating path that one object in space takes around another one. Orbits are the result of a perfect balance between the forward motion of a body in space such as a planet, moon, and the pull of gravity on it from another body in space. Without gravity, an Earth-orbiting satellite would go off into space along a straight line. Escape velocity depends on the mass of the plane. Each planet has a different escape velocity. The object's distance from the planet's center is also important. There are six orbital elements, they are, Epoch time, orbital inclination, Right Ascension of Ascending Node, Eccentricity, Argument of Perigee, Mean motion, Mean Anomaly.

**Johannes Kepler** was the first to successfully model planetary orbits to a high degree of accuracy, publishing his laws in 1605. **Issac Newton** published more general laws in 1687. **Newton's** method of successive approximation was formalised into an analytic method by **Euler** in 1744, whose work was in turn generalised to elliptical and hyperbolic orbits by **Lambert** in 1761 – 1777. Another milestone in orbit determination was **Carl Friedrich Gauss's** assistance in the “recovery” of the *dwarf planet ceres* in 1801.

A brief perusal of the Contents shows that there are more than enough topics to cover in a single semester or term. Chapter 1 is a review of vector kinematics in three dimensions and of Newton's laws of motion and gravitation. It also focuses on the issue of relative motion, crucial to the topics of rendezvous and satellite attitude dynamics. Chapter 2 presents the vector-based solution of the classical two-body problem, coming up with a host of practical formulas for orbit and trajectory analysis. The restricted three-body problem is covered in order to introduce the notion of Lagrange points. Chapter 3 is devoted to describing orbits in three dimensions and accounting for the major effects of the earth's oblate, non-spherical shape. Chapter 4 is an introduction to preliminary orbit determination, including Gibbs' and Gauss's methods and the solution of Lambert's problem. Auxiliary topics include topocentric coordinate systems, Julian day numbering and sidereal time.

# CHAPTER 1

## 1. Dynamics of point masses

### 1.1 Introduction

This chapter serves as a self-contained reference on the kinematics and dynamics of point masses as well as some basic vector operations. The notation and concepts summarized here will be used in the following chapters. Those familiar with the vector-based dynamics of particles can simply page through the chapter and then refer back to it later as necessary. Those who need a bit more in the way of review will find the chapter contains all of the material they need in order to follow the development of orbital mechanics topics in the upcoming chapters. We begin with the problem of describing the curvilinear motion of particles in three dimensions. The concepts of force and mass are considered next, along with Newton's inverse-square law of gravitation. This is followed by a presentation of Newton's second law of motion ('force equals mass times acceleration') and the important concept of angular momentum.

### 1.2 Kinematics

To track the motion of a particle  $P$  through Euclidean space we need a frame of reference, consisting of a clock and a cartesian coordinate system. The clock keeps track of time  $t$  and the  $x y z$  axes of the cartesian coordinate system are used to locate the spatial position of the particle. In non-relativistic mechanics, a single 'universal' clock serves for all possible cartesian coordinate systems. So when we refer to a frame of reference we need think only of the mutually orthogonal axes themselves.

The unit of time used throughout this book is the second (s). The unit of length is the meter (m), but the kilometre (km) will be the length unit of choice when large distances and

velocities are involved. Conversion factors between kilometres, miles and nautical miles are listed in Table A.3.

Given a frame of reference, the position of the particle  $P$  at a time  $t$  is defined by the position vector  $\mathbf{r}(t)$  extending from the origin  $O$  of the frame out to  $P$  itself as illustrated in Figure 1.1. (Vectors will always be indicated by boldface type.)

The components of  $\mathbf{r}(t)$  are just the  $x$ ,  $y$  and  $z$  coordinates,  
 $\mathbf{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}$

$\hat{i}$ ,  $\hat{j}$  and  $\hat{k}$  are the unit vectors which point in the positive direction of the  $x$ ,  $y$  and  $z$  axes, respectively. Any vector written with the overhead hat (e.g.  $\hat{\mathbf{a}}$ ) is to be considered a vector of unit dimensionless magnitude. The distance of  $P$  from the origin is the magnitude or length of  $\mathbf{r}$ , denoted  $\|\mathbf{r}\|$  or just  $r$

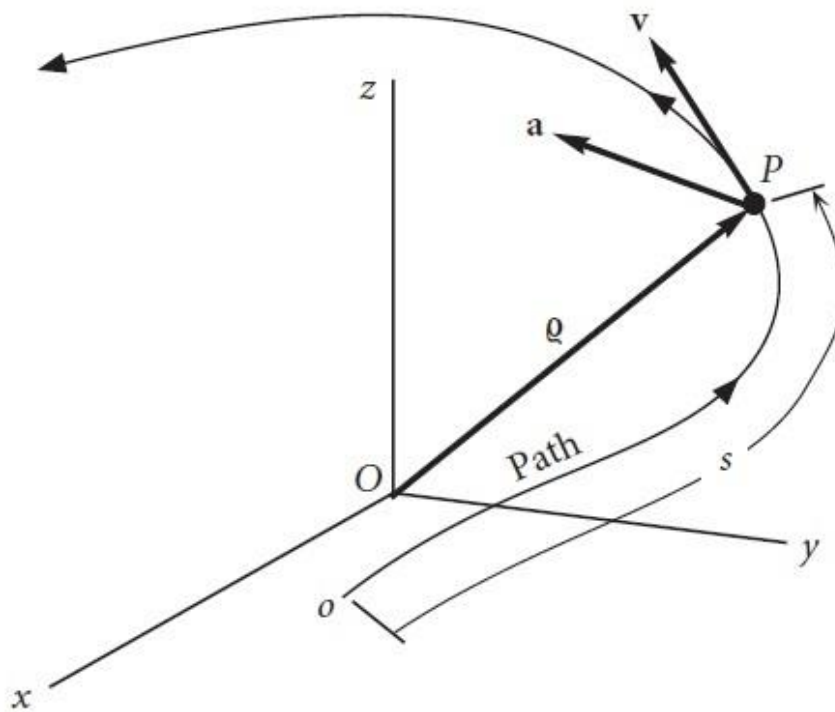


figure 1.1

$$\|\mathbf{r}\| = r = \sqrt{x^2 + y^2 + z^2}$$

The magnitude of  $\mathbf{r}$ , or any vector  $\mathbf{A}$  for that matter, can also be computed by means of the dot product operation

$$r = \sqrt{\mathbf{r} \cdot \mathbf{r}} \quad \|\mathbf{A}\| = \sqrt{\mathbf{A} \cdot \mathbf{A}}$$

The velocity  $\mathbf{v}$  and acceleration  $\mathbf{a}$  of the particle are the first and second time derivatives of the position vector,

$$\mathbf{v}(t) = \frac{dx(t)}{dt} \hat{i} + \frac{dy(t)}{dt} \hat{j} + \frac{dz(t)}{dt} \hat{k} = v_x(t)\hat{i} + v_y(t)\hat{j} + v_z(t)\hat{k}$$

$$\mathbf{a}(t) = \frac{dv_x}{dt}(t) \hat{i} + \frac{dv_y}{dt}(t) \hat{j} + \frac{dv_z}{dt}(t) \hat{k} = a_x(t) \hat{i} + a_y(t) \hat{j} + a_z(t) \hat{k}$$

It is convenient to represent the time derivative by means of an overhead dot. In this shorthand notation, if  $()$  is any quantity, then

$$() = \frac{d()}{dt} \quad () = \frac{d()}{dt} \quad () = \frac{d()}{dt}$$

Thus, for example,

$$\mathbf{v} = \dot{\mathbf{r}}$$

$$\mathbf{a} = \dot{\mathbf{v}} = \ddot{\mathbf{r}}$$

$$v_x = \dot{x} \quad v_y = \dot{y} \quad v_z = \dot{z}$$

$$a_x = \dot{v}_x = \ddot{x} \quad a_y = \dot{v}_y = \ddot{y} \quad a_z = \dot{v}_z = \ddot{z}$$

The locus of points that a particle occupies as it moves through space is called its path or trajectory. If the path is a straight line, then the motion is rectilinear. Otherwise, the path is curved, and the motion is called curvilinear. The velocity vector  $\mathbf{v}$  is tangent to the path. If  $\hat{\mathbf{u}}_t$  is the unit vector tangent to the trajectory, then

$$\mathbf{v} = v \hat{\mathbf{u}}_t$$

where  $v$ , the speed, is the magnitude of the velocity  $\mathbf{v}$ . The distance  $ds$  that  $P$  travels along its path in the time interval  $dt$  is obtained from the speed by

$$ds = v dt$$

In other words,

$$v = \dot{s}$$

The distance  $s$ , measured along the path from some starting point, is what the odometers in our automobiles record. Of course,  $\dot{s}$  our speed along the road, is indicated by the dial of the speedometer.

Note carefully that  $v \neq \dot{s}$ , i.e., the magnitude of the derivative of  $\mathbf{r}$  does not equal the derivative of the magnitude of  $\mathbf{r}$ .

## EXAMPLE 1

Relative to a Cartesian coordinate system, the position, velocity and acceleration of a particle relative at a given instant are

$$\mathbf{r} = 250\hat{i} + 630\hat{j} + 430\hat{k} \text{ (m)}$$

$$\mathbf{v} = 90\hat{i} + 125\hat{j} + 170\hat{k} \text{ (m/s)}$$

$$\mathbf{a} = 16\hat{i} + 125\hat{j} + 30\hat{k} \text{ (m/s}^2\text{)}$$

Find the coordinates of the center of curvature at that instant.

First, we calculate the speed  $v$ ,

$$v = \|v\| = \sqrt{90^2 + 125^2 + 170^2} = 229.4 \text{ m/s}$$

The unit tangent is, therefore,

$$\widehat{\mathbf{u}}_t = \frac{\mathbf{v}}{v} = \frac{90\hat{i} + 125\hat{j} + 170\hat{k}}{229.4} = 0.3923\hat{i} + 0.5449\hat{j} + 0.7411\hat{k}$$

We project the acceleration vector onto the direction of the tangent to get its tangential component  $a_t$ ,

$$a_t = \mathbf{a} \cdot \widehat{\mathbf{u}}_t = (16\hat{i} + 125\hat{j} + 30\hat{k}) \cdot (0.3923\hat{i} + 0.5449\hat{j} + 0.7411\hat{k}) = 96.62 \text{ m/s}^2$$

The magnitude of  $\mathbf{a}$  is

$$a = \sqrt{16^2 + 125^2 + 30^2} = 129.5 \text{ m/s}^2$$

since  $\mathbf{a} = a_t \widehat{\mathbf{u}}_t + a_n \widehat{\mathbf{u}}_n$  and  $\widehat{\mathbf{u}}_t$  and  $\widehat{\mathbf{u}}_n$  are perpendicular to each other it follows that  $a^2 = a_t^2 + a_n^2$ , which means

$$a_n = \sqrt{a^2 - a_t^2} = \sqrt{129.5^2 - 96.62^2} = 86.29 \text{ m/s}^2$$

Hence,

$$\widehat{\mathbf{u}}_n = \frac{1}{a_n} (\mathbf{a} - a_t \widehat{\mathbf{u}}_t)$$

$$\begin{aligned} \widehat{\mathbf{u}}_n &= \frac{1}{86.29} [(16\hat{i} + 125\hat{j} + 30\hat{k}) - 96.62(0.3923\hat{i} + 0.5449\hat{j} + 0.7411\hat{k})] \\ &= -0.2539\hat{i} + 0.8385\hat{j} - 0.4821\hat{k} \end{aligned}$$

The equation  $a_n = \frac{v^2}{\rho}$  can now be solved for  $\rho$  to yield

$$\rho = \frac{v^2}{a_n} = \frac{229.4^2}{86.29} = 609.9 \text{ m}$$

Let  $\mathbf{r}_c$  be the position vector of the center of curvature  $C$ . Then

$$\begin{aligned} \mathbf{r}_c &= \mathbf{r} + \mathbf{r}_{C/P} \\ r_c &= \mathbf{r} + \rho \widehat{\mathbf{u}}_n = 250\hat{i} + 630\hat{j} + 430\hat{k} + 609.9(-0.2539\hat{i} + 0.8385\hat{j} - 0.4821\hat{k}) \end{aligned}$$

$$\mathbf{r}_c = 95.16\hat{i} + 1141\hat{j} + 136.0\hat{k} \text{ (m)}$$

That is the coordinates of  $C$  are

$$X=95.16 \text{ m} \qquad y=1141 \text{ m} \qquad z=136.0 \text{ m}$$

## 1.3 Mass, force and Newton's law of gravitation

Mass, like length and time, is a primitive physical concept: it cannot be defined in terms of any other physical concept. Mass is simply the quantity of matter. More practically, mass is a measure of the inertia of a body. Inertia is an object's resistance to changing its state of motion. The larger its inertia (the greater its mass), the more difficult it is to set a body into motion or bring it to rest. The unit of mass is the kilogram (kg).

Force is the action of one physical body on another, either through direct contact or through a distance. Gravity is an example of force acting through a distance, as are magnetism and the force between charged particles. The gravitational force between two masses  $m_1$  and  $m_2$  having a distance  $r$  between their center is

$$F_g = G \frac{m_1 m_2}{r^2} \quad (1.1)$$

This is Newton's law of gravity, in which  $G$ , the universal gravitational constant, has the value  $6.6742 \times 10^{-11} \text{ m}^3/\text{kg} \cdot \text{s}^2$ . Due to the inverse-square dependence on distance, the force of gravity rapidly diminishes with the amount of separation between the two masses. In any case, the force of gravity is minuscule unless at least one of the masses is extremely big.

The force of a large mass (such as the earth) on a mass many orders of magnitude smaller (such as a person) is called weight,  $W$ . If the mass of the large object is  $M$  and that of the relatively tiny one is  $m$ , then the weight of the small body is

$$W = G \frac{Mm}{r^2} = m \left( \frac{GM}{r^2} \right)$$

or

$$W = mg \quad (1.2)$$

where

$$g = \frac{GM}{r^2} \quad (1.3)$$

$g$  has units of acceleration ( $\text{m/s}^2$ ) and is called the acceleration of gravity. If planetary gravity is the only force acting on a body, then the body is said to be in free fall. The force of gravity draws a freely falling object towards the center of attraction (e.g., center of the earth) with an acceleration  $g$ . Under ordinary conditions, we sense our own weight by feeling contact forces acting on us in opposition to the force of gravity.

In free fall there are, by definition, no contact forces, so there can be no sense of weight. Even though the weight is not zero, a person in free fall experiences weightlessness, or the absence of gravity.

Let us evaluate Equation 1.3 at the surface of the earth, whose radius according to Table A.1 is 6378 km. Letting  $g_0$  represent the standard sea-level value of  $g$ , we get

$$g_0 = \frac{GM}{R_E^2} \quad (1.4)$$



In SI units,

$$g_0 = 9.807 \text{ m/s}^2 \quad (1.5)$$

Substituting Equation 1.4 into Equation 1.3 and letting  $z$  represent the distance above the earth's surface, so that  $r = R_E + z$ , we obtain

$$g = g_0 \frac{R_E^2}{(R_E + z)^2} = \frac{g_0}{(1 + \frac{z}{R_E})^2} \quad (1.6)$$

Commercial air liners cruise at altitudes on the order of 10 kilometers (six miles). At that height, Equation 1.6 reveals that  $g$  (and hence weight) is only three-tenths of a percent less than its sea-level value. Thus, under ordinary conditions, we ignore the variation of  $g$  with altitude. A plot of Equation 1.6 out to a height of 1000 km (the upper limit of low-earth orbit operations) is shown in Figure 1.1. The variation of  $g$  over that range is significant. Even so, at space station altitude (300 km), weight is only about 10 percent less than it is on the earth's surface. The astronauts experience weightlessness, but they clearly are not weightless.

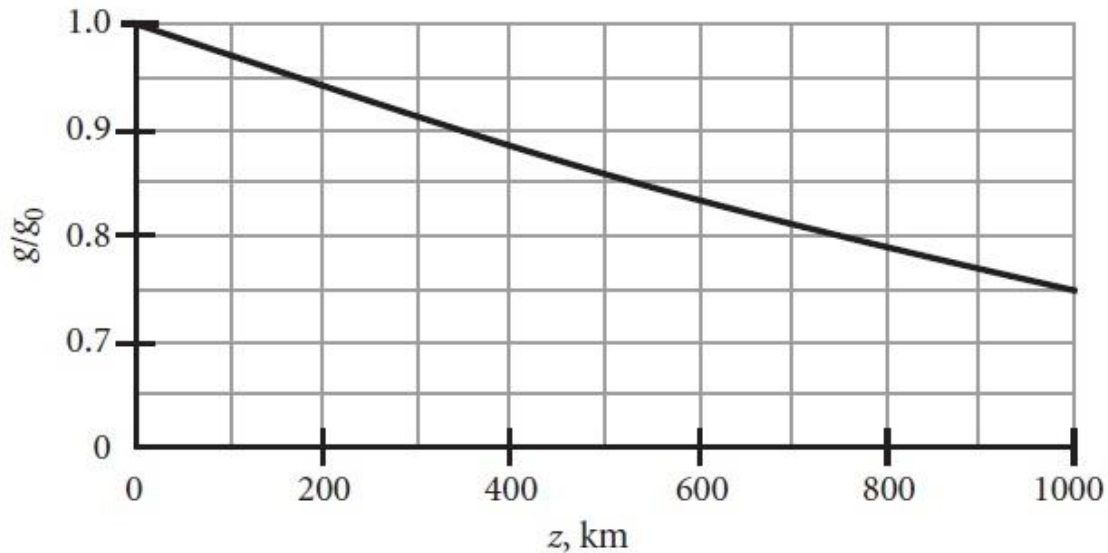


Figure 1.2

## EXAMPLE 2

Show that in the absence of an atmosphere, the shape of a low altitude ballistic trajectory is a parabola. Assume the acceleration of gravity  $g$  is constant and neglect the earth's curvature.

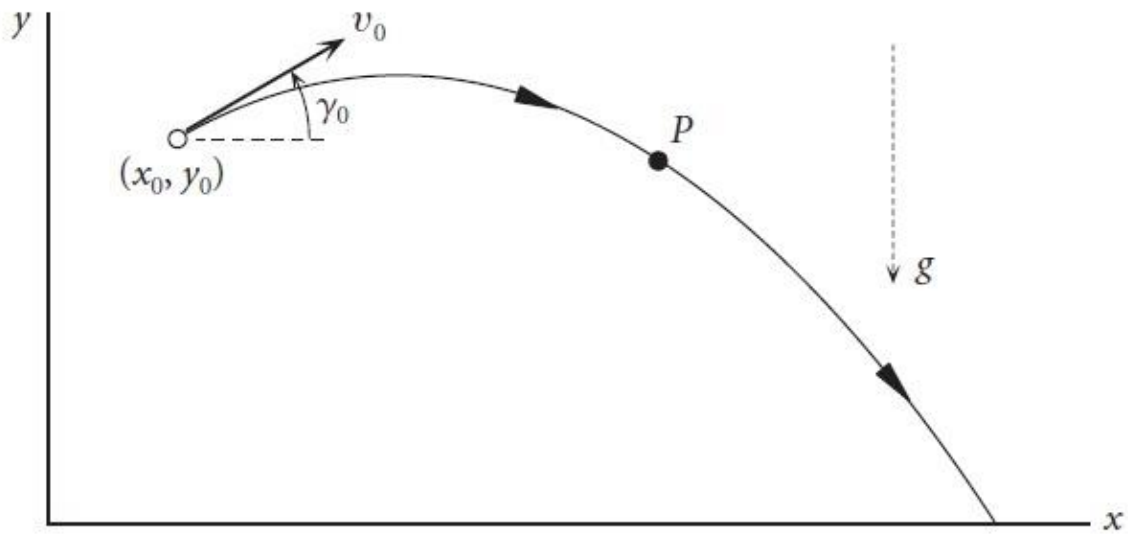


Figure 1.3

Figure 1.3 shows a projectile launched at  $t=0$  with a speed  $v_0$  at a flight path angle  $\gamma_0$  from the point with coordinates  $(x_0, y_0)$ . Since the projectile is in free fall after launch, its only acceleration is that of gravity in the negative  $y$ -direction:

$$\ddot{x} = 0 \qquad \ddot{y} = -g$$

Integrating with respect to time and applying the initial conditions leads to

$$x = x_0 + (v_0 \cos \gamma_0)t \qquad (a)$$

$$y = y_0 + (v_0 \sin \gamma_0)t - \frac{1}{2}gt^2 \qquad (b)$$

Solving (a) for  $t$  and substituting the result into (b) yields

$$y = y_0 + (x - x_0) \tan \gamma_0 - \frac{1}{2} \frac{g}{v_0^2 \cos^2 \gamma_0} (x - x_0)^2 \qquad (c)$$

This is the equation of a second-degree curve, a parabola, as sketched in Figure 1.3.

## 1.4 Newton's law of motion

Force is not a primitive concept like mass because it is intimately connected with the concepts of motion and inertia. In fact, the only way to alter the motion of a body is to exert a force on it. The degree to which the motion is altered is a measure of the force. This is quantified by Newton's second law of motion. If the resultant or net force on a body of mass  $m$  is  $\mathbf{F}_{net}$ , then

$$\mathbf{F}_{net} = m\mathbf{a} \qquad (1.7)$$

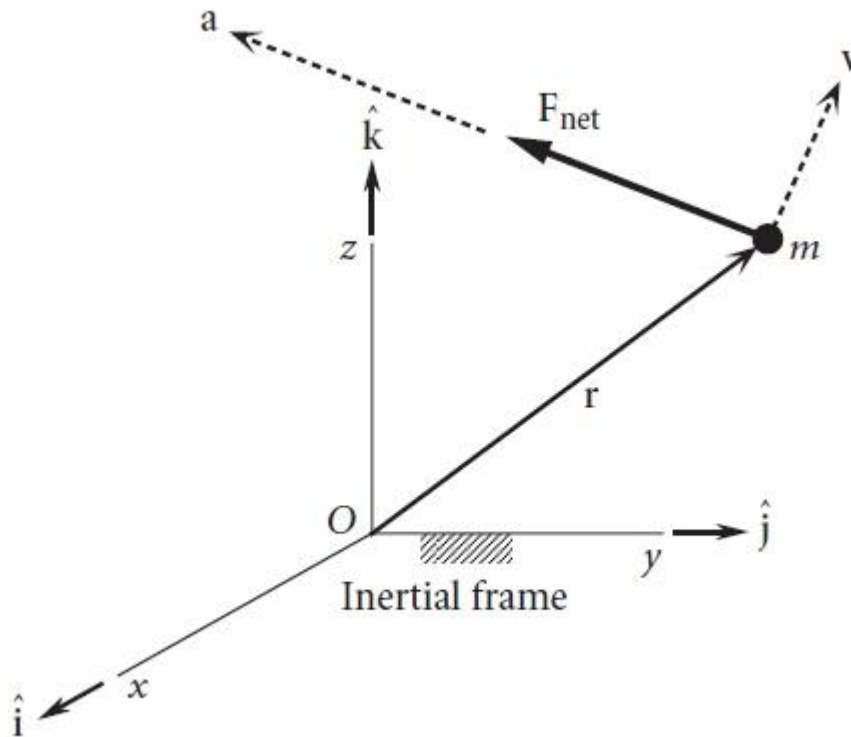


Figure 1.4

In this equation,  $\mathbf{a}$  is the absolute acceleration of the center of mass. The absolute acceleration is measured in a frame of reference which itself has neither translational nor rotational acceleration relative to the fixed stars. Such a reference is called an absolute or inertial frame of reference.

Force, then, is related to the primitive concepts of mass, length and time by Newton's second law. The unit of force, appropriately, is the Newton, which is the force required to impart an acceleration of  $1 \text{ m/s}^2$  to a mass of  $1 \text{ kg}$ . A mass of one kilogram therefore weighs  $9.81$  Newtons at the earth's surface. The kilogram is not a unit of force.

Confusion can arise when mass is expressed in units of force, as frequently occurs in US engineering practice. In common parlance either the pound or the ton (2000 pounds) is more likely to be used to express the mass. The pound of mass is officially defined precisely in terms of the kilogram as shown in Table A.3. Since one pound of mass weighs one pound of force where the standard sea-level acceleration of gravity ( $g_0 = 9.80665 \text{ m/s}^2$ ) exists, we can use Newton's second law to relate the pound of force to the Newton:

$$\begin{aligned} 1 \text{ lb (force)} &= 0.4536 \text{ kg} \times 9.807 \text{ m/s}^2 \\ &= 4.448 \text{ N} \end{aligned}$$

The slug is the quantity of matter accelerated at one foot per *second*<sup>2</sup> by a force of one pound. We can again use Newton's second law to relate the slug to the kilogram. Noting the relationship between feet and meters in Table A.3, we find

$$1 \text{ slug} = \frac{1 \text{ lb}}{1 \text{ ft/s}^2} = \frac{4.448 \text{ N}}{0.3048 \text{ m/s}^2} = 14.59 \frac{\text{kg} \cdot \text{m/s}^2}{\text{m/s}^2}$$

$$= 14.59 \text{ kg}$$

### Example 1.3

On a NASA mission the space shuttle Atlantis orbiter was reported to weigh 239 255 lb just prior to lift-off. On orbit 18 at an altitude of about 350 km, the orbiter's weight was reported to be 236 900 lb. (a) What was the mass, in kilograms, of Atlantis on the launch pad and in orbit? (b) If no mass were lost between launch and orbit 18, what would have been the weight of Atlantis in pounds?

(a) The given data illustrates the common use of weight in pounds as a measure of mass. The 'weights' given are actually the mass in pounds of mass. Therefore, prior to launch

$$m_{\text{launch pad}} = 239\,255 \text{ lb (mass)} \times \frac{0.4536 \text{ kg}}{1 \text{ lb (mass)}} = 108\,500 \text{ kg}$$

In orbit,

$$m_{\text{orbit 18}} = 236\,900 \text{ lb (mass)} \times \frac{0.4536 \text{ kg}}{1 \text{ lb (mass)}} = 107\,500 \text{ kg}$$

The decrease in mass is the propellant expended by the orbital maneuvering and reaction control rockets on the orbiter

(b) Since the space shuttle launch pad at Kennedy Space Center is essentially at sea level, the launch-pad weight of Atlantis in lb (force) is numerically equal to its mass in lb (mass). With no change in mass, the force of gravity at 350 km would be, according to Equation 1.5

$$W = 239\,255 \text{ lb (force)} \times \left( \frac{1}{1 + \frac{350}{6378}} \right)^2 = 215\,000 \text{ lb (force)}$$

The integral of a force  $\mathbf{F}$  over a time interval is called the impulse  $I$  of the force

$$I = \int_{t_1}^{t_2} \mathbf{F} \, dt \quad (1.8)$$

From Equation 1.8 it is apparent that if the mass is constant, then

$$I_{\text{net}} = \int_{t_1}^{t_2} m \frac{dv}{dt} dt = mv_2 - mv_1 \quad (1.9)$$

That is the net impulse on a body yields a change  $m\Delta v$  in its linear momentum, so that

$$\Delta v = \frac{I_{\text{net}}}{m} \quad (1.10)$$

If  $F_{\text{net}}$  is constant then  $I_{\text{net}} = F_{\text{net}} \Delta t$ , in which case equation (1.10) becomes

$$\Delta v = \frac{F_{\text{net}}}{m} \Delta t \quad (\text{if } F_{\text{net}} \text{ is constant}) \quad (1.11)$$

Let us conclude this section by introducing the concept of angular momentum. The moment of the net force about  $O$  in Figure 1.6 is

$$\mathbf{M}_{\text{Onet}} = \mathbf{r} \times \mathbf{F}_{\text{net}}$$

Substituting Equation 1.10 yields

$$\mathbf{M}_{\text{Onet}} = \mathbf{r} \times m\mathbf{a} = \mathbf{r} \times m \frac{d\mathbf{v}}{dt} \quad (1.12)$$

But Keeping in mind that the mass is constant

$$\mathbf{r} \times m \frac{d\mathbf{v}}{dt} = \frac{d}{dt} (\mathbf{r} \times m\mathbf{v}) - \left( \frac{d}{dt} \times m\mathbf{v} \right) = \frac{d}{dt} (\mathbf{r} \times m\mathbf{v}) - (\mathbf{v} \times m\mathbf{v})$$

since  $\mathbf{v} \times m\mathbf{v} = m(\mathbf{v} \times \mathbf{v}) = 0$  it follows that equation 1.12 can be written

$$\mathbf{M}_{\text{Onet}} = \frac{d\mathbf{H}_O}{dt} \quad (1.13)$$

Where  $\mathbf{H}_O$  is the angular momentum about O,

$$\mathbf{H}_O = \mathbf{r} \times m\mathbf{v} \quad (1.14)$$

Thus, just as the net force on a particle changes its linear momentum  $m\mathbf{v}$ , the moment of that force about a fixed point changes the moment of its linear momentum about that point. Integrating Equation 1.16 with respect to time yields

$$\int_{t_1}^{t_2} \mathbf{M}_{\text{Onet}} dt = \mathbf{H}_{O_1} - \mathbf{H}_{O_2} \quad (1.15)$$

The integral on the left is the net angular impulse. This angular impulse–momentum equation is the rotational analog of the linear impulse–momentum relation given above in Equation 1.9

## 1.5 Time derivatives of moving vectors

Figure 1.5(a) shows a vector  $\mathbf{A}$  inscribed in a rigid body  $B$  that is in motion relative to an inertial frame of reference (a rigid, cartesian coordinate system which is fixed relative to the fixed stars). The magnitude of  $\mathbf{A}$  is fixed. The body  $B$  is shown at two times, separated by the differential time interval  $dt$ . At time  $t + dt$  the orientation of vector  $\mathbf{A}$  differs slightly from that at time  $t$ , but its magnitude is the same. According to one of the many theorems of the prolific eighteenth century Swiss mathematician Leonhard Euler (1707–1783), there is a unique axis of rotation about which  $B$  and, therefore,  $\mathbf{A}$  rotates during the differential time interval. If we shift the two vectors  $\mathbf{A}(t)$  and  $\mathbf{A}(t + dt)$  to the same point on the axis of rotation, so that they are tail-to-tail

as shown in Figure 1.8(b), we can assess the difference  $d\mathbf{A}$  between them caused by the infinitesimal rotation. Remember that shifting a vector to a parallel line does not change the vector. The rotation of the body  $B$  is measured in the plane perpendicular to the instantaneous axis of rotation. The amount of rotation is the angle  $d\theta$  through which a line element normal to the rotation axis turns in the time interval  $dt$ . In

Figure 1.8(b) that line element is the component of  $\mathbf{A}$  normal to the axis of rotation. We can express the difference  $d\mathbf{A}$  between  $\mathbf{A}(t)$  and  $\mathbf{A}(t + dt)$  as

$$d\mathbf{A} = \overbrace{[\|\mathbf{A}\| \sin\phi] d\theta}^{\text{magnitude of } d\mathbf{A}} \hat{\mathbf{n}} \quad (1.16)$$

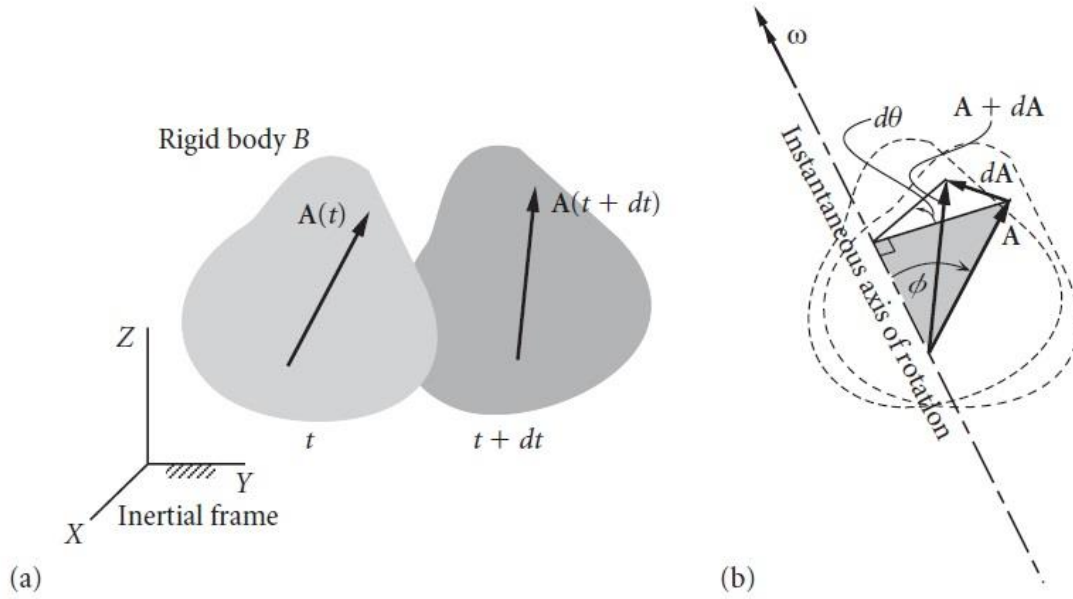


Figure 1.5

where  $\hat{\mathbf{n}}$  is the unit normal to the plane defined by  $\mathbf{A}$  and the axis of rotation, and it points in the direction of the rotation. The angle  $\phi$  is the inclination of  $\mathbf{A}$  to the rotation axis. By definition,

$$d\theta = \|\boldsymbol{\omega}\| dt \quad (1.17)$$

where  $\boldsymbol{\omega}$  is the angular velocity vector, which points along the instantaneous axis of rotation and its direction is given by the right-hand rule. That is, wrapping the right hand around the axis of rotation, with the fingers pointing in the direction of  $d\theta$ , results in the thumb's defining the direction of  $\boldsymbol{\omega}$ . This is evident in Figure 1.8(b). It should be pointed out that the time derivative of  $\boldsymbol{\omega}$  is the angular acceleration, usually given the symbol  $\boldsymbol{\alpha}$ . Thus,

$$\boldsymbol{\alpha} = \frac{d\boldsymbol{\omega}}{dt} \quad (1.18)$$

Substituting Equation 1.20 into Equation 1.19, we get

$$d\mathbf{A} = \|\mathbf{A}\| \sin\phi \|\boldsymbol{\omega}\| dt \cdot \hat{\mathbf{n}} = (\|\boldsymbol{\omega}\| \cdot \|\mathbf{A}\| \sin\phi) \hat{\mathbf{n}} dt \quad (1.19)$$

By definition of the cross product,  $\boldsymbol{\omega} \times \mathbf{A}$  is the product of the magnitude of  $\boldsymbol{\omega}$ , the magnitude of  $\mathbf{A}$ , the sine of the angle between  $\boldsymbol{\omega}$  and  $\mathbf{A}$  and the unit vector normal to the plane of  $\boldsymbol{\omega}$  and  $\mathbf{A}$ , in the rotation direction. That is

$$\boldsymbol{\omega} \times \mathbf{A} = \|\boldsymbol{\omega}\| \cdot \|\mathbf{A}\| \sin\phi \cdot \hat{\mathbf{n}} \quad (1.20)$$

Substituting Equation 1.16 into Equation 1.17 yields



$$d \mathbf{A} = \boldsymbol{\omega} \times \mathbf{A} dt \quad (1.21)$$

Dividing through by  $dt$ , we finally obtain

$$\frac{d\mathbf{A}}{dt} = \boldsymbol{\omega} \times \mathbf{A} \quad (1.22)$$

Equation 1.22 is a formula we can use to compute the time derivative of any vector of constant magnitude.

# Chapter 2

## THE TWO –BODY PROBLEM

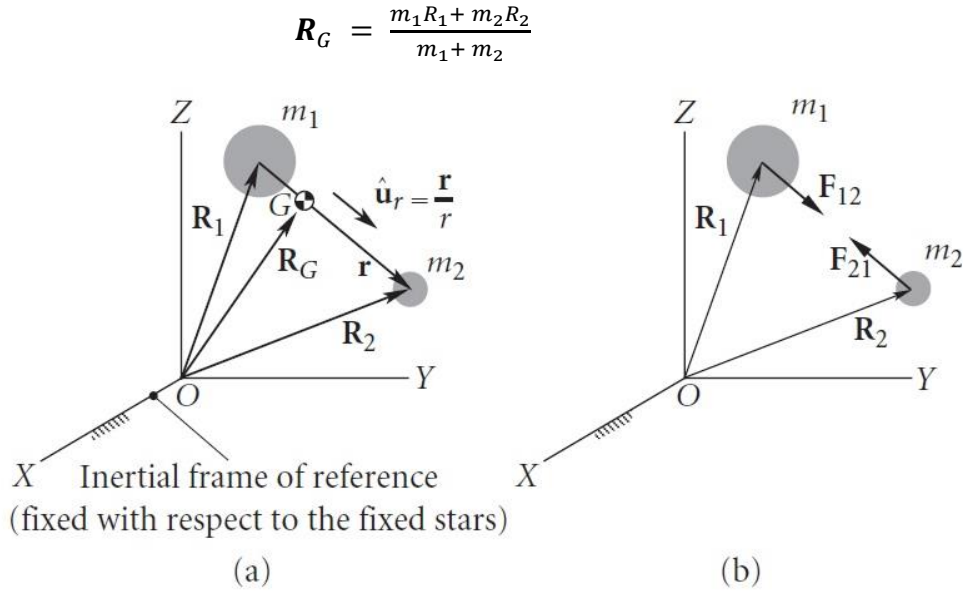
### 2.1 INTRODUCTION

This chapter presents the vector- based approach to the classical problem of determining the motion of two bodies due solely to their own mutual gravitational attraction. We show that the path of one of the masses relative to the other is a conic section (circle, ellipse, parabola or hyperbola) whose shape is determined by the eccentricity. Several fundamental properties of the different types of orbits are developed with the aid of the laws of conservation of angular momentum and energy.

### 2.2 EQUATIONS OF MOTION IN AN INERTIAL FRAME

Figure 2.1 shows two point masses acted upon only by the mutual force of gravity between them. The positions of their center of mass are shown relative to an inertial frame of reference XYZ. The origin O of the frame may move with constant velocity (relative to the fixed stars), but the axes do not rotate. Each of the two bodies is acted upon by the gravitational attraction of the other.  $F_{12}$  is the force exerted on  $m_1$  by  $m_2$ , and  $F_{21}$  is the force exerted on  $m_2$  by  $m_1$ .

The position vector  $\mathbf{R}_G$  of the center of mass G of the system in Figure 2.1 (a) is, defined by the formula



(2.1)

Figure 2.1 (a) Two masses located in an inertial frame. (b) Free-body diagrams

$$\mathbf{V}_G = \dot{\mathbf{R}}_G = \frac{m_1 \dot{\mathbf{R}}_1 + m_2 \dot{\mathbf{R}}_2}{m_1 + m_2}$$

(2.2)

$$\mathbf{a}_G = \ddot{\mathbf{R}}_G = \frac{m_1 \ddot{\mathbf{R}}_1 + m_2 \ddot{\mathbf{R}}_2}{m_1 + m_2}$$

(2.3)

The adjective ‘absolute’ means that the quantities are measured relative to an inertial frame of reference.

Let  $\mathbf{r}$  be the position vector of  $m_2$  relative to  $m_1$ . Then

$$\mathbf{r} = \mathbf{R}_2 - \mathbf{R}_1$$

(2.4)

Furthermore, let  $\mathbf{u}_r$  be the unit vector pointing from  $m_1$  towards  $m_2$ , so that

$$\hat{\mathbf{u}}_r = \frac{\mathbf{r}}{r}$$

(2.5)

Where  $r = \|\mathbf{r}\|$ , the magnitude of  $\mathbf{r}$ . The body  $m_1$  is acted upon only by the force of gravitational attraction towards  $m_2$ . The force of gravitational attraction,  $F_g$ , which acts along the line joining the centers of mass of  $m_1$  and  $m_2$ , is given by Equation 1.3. The force exerted on  $m_2$  by  $m_1$  is

$$F_{21} = \frac{Gm_1m_2}{r^2}(-\widehat{u}_r) = -\frac{Gm_1m_2}{r^2}(\widehat{u}_r)$$

Where  $-\widehat{u}_r$  accounts for the fact that the force vector  $F_{21}$  is directed from  $m_2$  towards  $m_1$ . (Do not confuse the symbol  $G$ , used in this context to represent the universal gravitational constant, with its use elsewhere in the book to denote the center of mass.) Newton's second law of motion as applied to body  $m_2$  is  $F_{21} = m_2\ddot{\mathbf{R}}_2$ , where  $\ddot{\mathbf{R}}_2$  is the absolute acceleration of  $m_2$ . Thus

$$-\frac{Gm_1m_2}{r^2}(\widehat{u}_r) = m_2\ddot{\mathbf{R}}_2 \quad (2.7)$$

By Newton's third law (the action-reaction principle),  $F_{12} = -F_{21}$ , so that for  $m_1$  we have

$$\frac{Gm_1m_2}{r^2}(\widehat{u}_r) = m_1\ddot{\mathbf{R}}_1 \quad (2.8)$$

Equation 2.7 and 2.8 are the equation of the two bodies in inertial space. By adding each side of these equations together, we find  $m_1\ddot{\mathbf{R}}_1 + m_2\ddot{\mathbf{R}}_2 = 0$ . According to Equation 2.3, that means the acceleration of the center of mass  $G$  of the system of two bodies  $m_1$  and  $m_2$  is zero.  $G$  moves with a constant velocity  $\mathbf{V}_G$  in a straight line, so that its position vector to XYZ given by

$$\mathbf{R}_G = \mathbf{R}_{Go} + \mathbf{v}_G t \quad (2.9)$$

Where  $\mathbf{R}_{Go}$  is the position of  $G$  at time  $t=0$ . The center of mass of a two-body system may therefore serve as the origin of an inertial frame.

## 2.3 EQUATION OF RELATIVE MOTION

Let us now multiply Equation 2.7 by  $m_1$  and Equation 2.8 by  $m_2$  to obtain

$$-\frac{Gm_1^2m_2}{r^2}\widehat{u}_r = m_1m_2\ddot{\mathbf{R}}_2$$

$$\frac{Gm_1m_2^2}{r^2}\widehat{u}_r = m_1m_2\ddot{\mathbf{R}}_1$$

Subtracting the second of these two equations from the first yields

$$m_1m_2(\ddot{\mathbf{R}}_2 - \ddot{\mathbf{R}}_1) = -\frac{Gm_1m_2}{r^2}(m_1 + m_2)\widehat{u}_r$$

Canceling the common factor and using Equation 2.4 yields

$$\ddot{\mathbf{r}} = -\frac{G(m_1+m_2)}{r^2}\widehat{u}_r \quad (2.13)$$

Let the gravitational  $\mu$  parameter be defined as

$$(2.14) \quad \mu = G(m_1 + m_2)$$

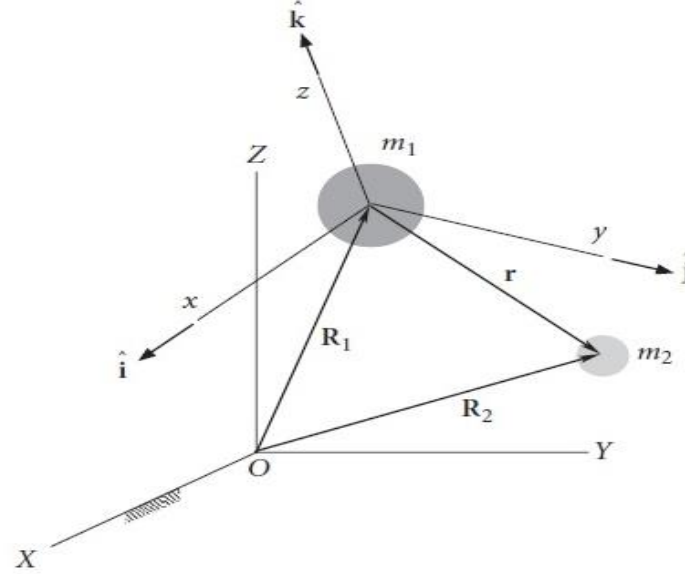
The units of  $\mu$  are  $km^3s^{-2}$ . Using Equation 2.14 together with Equation 2.5, we can write Equation 2.13 as

$$(2.15) \quad \ddot{\mathbf{r}} = -\frac{\mu}{r^3} \mathbf{r}$$

This is the second order differential equation that governs the motion of  $m_2$  relative to  $m_1$ . It has two vector constants of integration, each having three scalar components. Therefore, Equation 2.15 has six constants of integration. Note that interchanging the roles of  $m_1$  and  $m_2$  in all of the above amounts to simply multiplying Equation 2.15 through by  $-1$ , which, of course, changes nothing. Thus, the motion of  $m_2$  as seen from  $m_1$  is precisely the same as the motion of  $m_1$  as seen from  $m_2$ .

The relative position vector  $\mathbf{r}$  in Equation 2.15 was defined in the inertial frame (Equation 2.4). It is convenient, however, to measure the components of  $\mathbf{r}$  in a frame of reference attached to and moving with  $m_1$ . In a co-moving reference frame, such as the  $xyz$  system illustrated in Figure 2.2,  $\mathbf{r}$  has the expression

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$$



**Figure 2.2** Moving reference frame  $xyz$  attached to the center of mass of  $m_1$ .

## 2.4 ANGULAR MOMENTUM AND THE ORBIT FORMULAS

The angular momentum of body  $m_2$  relative to  $m_1$  is the moment of  $m_2$ 's relative linear momentum  $m_2\dot{\mathbf{r}}$  (cf. Equation 1.17),

$$H_{1/2} = \mathbf{r} \times m_2\dot{\mathbf{r}}$$

where  $\dot{\mathbf{r}} = \mathbf{v}$  is the velocity of  $m_2$  relative to  $m_1$ . Let us divide this equation through by  $m_2$  and let  $\mathbf{h} = H_{1/2}/m_2$ , so that

$$(2.18) \quad \mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}}$$

$\mathbf{h}$  is the relative angular momentum of  $m_2$  per unit mass, that is, the specific relative angular momentum. The units of  $\mathbf{h}$  are  $km^3s^{-1}$ .

Taking the time derivative of  $\mathbf{h}$  yields

$$\frac{d\mathbf{h}}{dt} = \dot{\mathbf{r}} \times \dot{\mathbf{r}} + \mathbf{r} \times \ddot{\mathbf{r}}$$

But  $\dot{\mathbf{r}} \times \dot{\mathbf{r}} = 0$ . Furthermore,  $\ddot{\mathbf{r}} = -(\mu/r^3)\mathbf{r}$ , according to Equation 2.15, so that

$$\mathbf{r} \times \ddot{\mathbf{r}} = \mathbf{r} \times \left(-\frac{\mu}{r^3}\mathbf{r}\right) = -\frac{\mu}{r^3}(\mathbf{r} \times \mathbf{r}) = 0$$

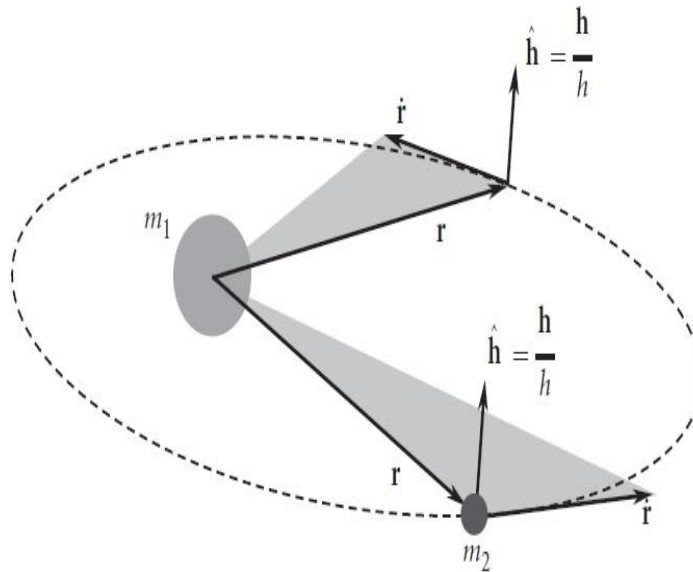


Figure 2.7 The path of  $m_2$  around  $m_1$  lies in a plane whose normal is defined by  $\mathbf{h}$ .

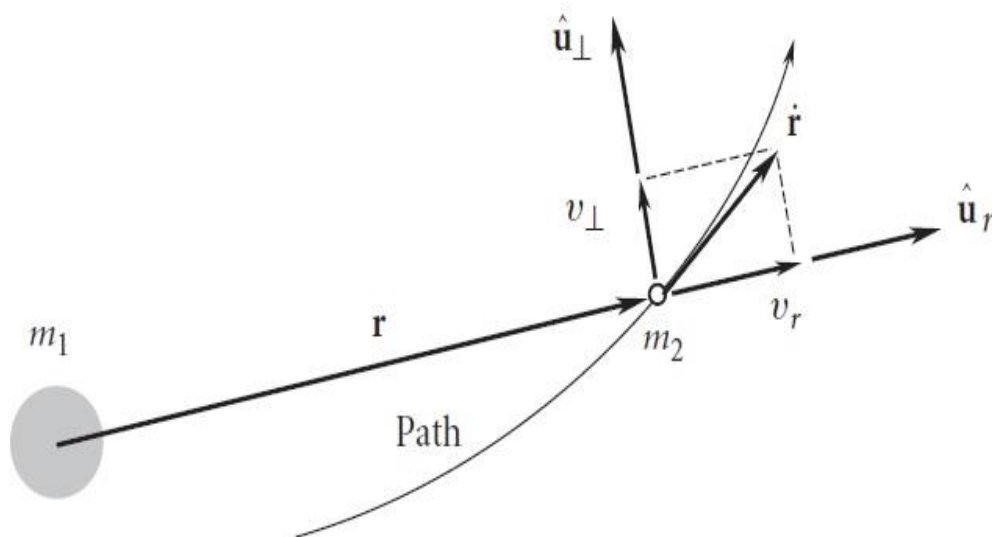


Figure 2.8 Components of the velocity of  $m_2$ , viewed above the plane of the orbit.



Therefore,

$$\frac{d\mathbf{h}}{dt} = 0 \quad (\text{or } \mathbf{r} \times \dot{\mathbf{r}} = \text{constant})$$

At any given time, the position vector  $\mathbf{r}$  and the velocity vector  $\dot{\mathbf{r}}$  lie in the same plane, as illustrated in Figure 2.7. Their cross product  $\mathbf{r} \times \dot{\mathbf{r}}$  is perpendicular to that plane.

Since  $\mathbf{r} \times \dot{\mathbf{r}} = \mathbf{h}$ , the unit vector normal to the plane is

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{h}$$

(2.20)

But, according to Equation 2.19, this unit vector is constant. Thus, the path of  $m_2$ , around  $m_1$ , lies in a single plane.

Since the orbit of  $m_2$ , around  $m_1$ , forms a plane, it is convenient to orient oneself above that plane and look down upon the path, as shown in Figure 2.8. Let us resolve the relative velocity vector  $\dot{\mathbf{r}}$  into components  $v_r = v_r \hat{\mathbf{u}}_r$  and  $\mathbf{v}_\perp = v_\perp \hat{\mathbf{u}}_\perp$  along the outward radial from  $m_1$  and perpendicular to it, respectively, where  $\hat{\mathbf{u}}_r$  and  $\hat{\mathbf{u}}_\perp$  are the radial and perpendicular (azimuthal) unit vectors. Then we can write Equation 2.18

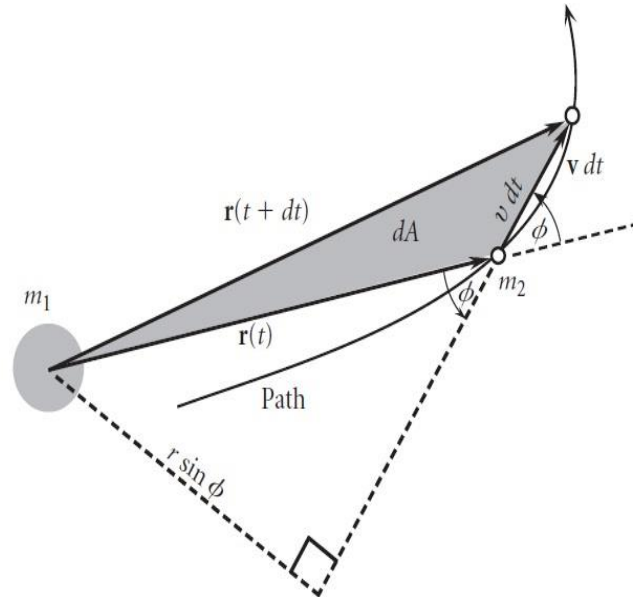


Figure 2.9 Differential area  $dA$  swept out by the relative position vector  $\mathbf{r}$  during time interval  $dt$ .

as

$$\mathbf{h} = \mathbf{r} \hat{\mathbf{u}}_r \times (v_r \hat{\mathbf{u}}_r + v_\perp \hat{\mathbf{u}}_\perp) = \mathbf{r} v_\perp \hat{\mathbf{h}}$$

That is,

$$\mathbf{h} = \mathbf{r} \mathbf{v}_\perp$$

Clearly, the angular momentum depends only on the azimuth component of the relative velocity.

During the differential time interval  $dt$  the position vector  $\mathbf{r}$  sweeps out an area  $dA$ , as shown in Figure 2.9. From the figure it is clear that the triangular area  $dA$  is given by

$$dA = \frac{1}{2} \times \text{base} \times \text{altitude} = \frac{1}{2} \times v dt \times r \sin \phi = \frac{1}{2} r (v \sin \phi) dt = \frac{1}{2} r v_\perp dt$$

Therefore, using Equation 2.21 we have

$$\frac{dh}{dt} = \frac{h}{2}$$

(2.23)

$dA/dt$  is called the areal velocity, and according to Equation 2.22 it is constant. Named after the German astronomer Johannes Kepler (1571–1630), this result is known as Kepler's second law: equal areas are swept out in equal times. Before proceeding with an effort to integrate Equation 2.15, recall the vector identity known as the *bac–cab* rule:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B} (\mathbf{A} \cdot \mathbf{C}) - \mathbf{C} (\mathbf{A} \cdot \mathbf{B})$$

(2.23)

Recall as well that

$$\mathbf{r} \cdot \mathbf{r} = r^2 \quad (2.24)$$

So that

$$\frac{d}{dt}(\mathbf{r} \cdot \mathbf{r}) = 2r \frac{dr}{dt}$$

But

$$\frac{d}{dt}(\mathbf{r} \cdot \mathbf{r}) = \mathbf{r} \cdot \frac{d\mathbf{r}}{dt} + \frac{d\mathbf{r}}{dt} \cdot \mathbf{r} = 2\mathbf{r} \cdot \frac{d\mathbf{r}}{dt}$$

Thus, we obtain the important identity

$$\mathbf{r} \cdot \dot{\mathbf{r}} = r\dot{r}$$

(2.25a)

since  $\dot{r} = v$  and  $r = ||r||$ , this can be written alternatively as

$$\mathbf{r} \cdot \mathbf{v} = ||r|| \frac{d||r||}{dt}$$

(2.25b)

Now let us take the cross product of both sides of Equation 2.15 [ $\ddot{\mathbf{r}} = -(\frac{\mu}{r^3}) \mathbf{r}$ ] with the specific angular momentum  $\mathbf{h}$ :

$$\ddot{\mathbf{r}} \times \mathbf{h} = -\frac{\mu}{r^3} \mathbf{r} \times \mathbf{h}$$

(2.26)

Since  $\frac{d}{dt}(\dot{\mathbf{r}} \times \mathbf{h}) = \ddot{\mathbf{r}} \times \mathbf{h} + \dot{\mathbf{r}} \times \dot{\mathbf{h}}$ , the left-hand side can be written

$$\ddot{\mathbf{r}} \times \mathbf{h} = \frac{d}{dt}(\dot{\mathbf{r}} \times \mathbf{h}) - \dot{\mathbf{r}} \times \dot{\mathbf{h}}$$

But according to Equation 2.19, the angular momentum is constant ( $\dot{h} = 0$ ), so this reduces to

$$\dot{\mathbf{r}} \times \mathbf{h} = \frac{d}{dt}(\dot{\mathbf{r}} \times \mathbf{h}) \quad (2.27)$$

$$\frac{1}{r^3} \mathbf{r} \times \mathbf{h} = -\frac{d}{dt}\left(\frac{\mathbf{r}}{r}\right) \quad (2.28)$$

Substituting Equation 2.27 and 2.28 into Equation 2.26, we get, The right – hand side Equation 2.26 can be transformed by the following sequence of substitutions:

$$\begin{aligned} \frac{1}{r^3} \mathbf{r} \times \mathbf{h} &= \frac{1}{r^3} [\mathbf{r} \times (\mathbf{r} \times \dot{\mathbf{r}})] \quad (\text{Equation 2.18 } [\mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}}]) \\ &= \frac{1}{r^3} [r(\mathbf{r} \cdot \dot{\mathbf{r}}) - \dot{\mathbf{r}}(\mathbf{r} \cdot \mathbf{r})] \quad (\text{Equation 2.23 [bac - cab rule]}) \\ &= \frac{1}{r^3} [r(r\dot{r}) - \dot{r}r^2] \quad (\text{Equation 2.24 and 2.25}) \\ &= \frac{r\dot{r} - \dot{r}r}{r^2} \end{aligned}$$

But

$$\frac{d}{dt}\left(\frac{\mathbf{r}}{r}\right) = \frac{r\dot{\mathbf{r}} - \dot{r}\mathbf{r}}{r^2}$$

Therefore

$$= -\frac{r\dot{r} - \dot{r}r}{r^2}$$

$$\frac{d}{dt}(\dot{\mathbf{r}} \times \mathbf{h}) = \frac{d}{dt}\left(\mu \frac{\mathbf{r}}{r}\right) \quad \text{or}$$

$$\frac{d}{dt}\left(\dot{\mathbf{r}} \times \mathbf{h} - \mu \frac{\mathbf{r}}{r}\right) = 0$$

That is,

$$\dot{\mathbf{r}} \times \mathbf{h} - \mu \frac{\mathbf{r}}{r} = \mathbf{c} \quad (2.29)$$

Where the vector  $\mathbf{c}$  is an arbitrary constant of integration having the dimensions of  $\mu$ . Equation 2.29 is the first integral of the equation of motion,  $\ddot{\mathbf{r}} = -\left(\frac{\mu}{r^3}\right) \mathbf{r}$ . Taking the dot product of both sides of Equation 2.29 with the vector  $\mathbf{h}$  yields

$$(\dot{\mathbf{r}} \times \mathbf{h}) \cdot \mathbf{h} - \mu \frac{\mathbf{r} \cdot \mathbf{h}}{r} = \mathbf{c} \cdot \mathbf{h}$$

Since  $\dot{\mathbf{r}} \times \mathbf{h}$  is perpendicular to both  $\dot{\mathbf{r}}$  and  $\mathbf{h}$ , it follows that  $(\dot{\mathbf{r}} \times \mathbf{h}) \cdot \mathbf{h} = 0$ . Likewise, since  $\mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}}$  is perpendicular to both  $\mathbf{r}$  and  $\dot{\mathbf{r}}$ , it is true that  $\mathbf{r} \cdot \mathbf{h} = 0$ . Therefore, we have  $\mathbf{c} \cdot \mathbf{h} = 0$ , i.e.,  $\mathbf{c}$  is perpendicular to  $\mathbf{h}$  which is normal to the orbital plane. That of course means  $\mathbf{c}$  must lie in the orbital plane.

$$\frac{\mathbf{r}}{r} + \mathbf{e} = \frac{\dot{\mathbf{r}} \times \mathbf{h}}{\mu} \quad (2.30)$$

Where  $\mathbf{e} = \mathbf{c}/\mu$ . The dimensionless vector  $\mathbf{e}$  is called the eccentricity vector. The line defined by the vector  $\mathbf{e}$  is commonly called the apse line. In order to obtain a scalar equation, let us take the dot product of both sides of Equation 2.30 with  $\mathbf{r}$ :

$$\frac{\mathbf{r} \cdot \mathbf{r}}{r} + \mathbf{r} \cdot \mathbf{e} = \frac{\mathbf{r} \cdot (\dot{\mathbf{r}} \times \mathbf{h})}{\mu} \quad (2.31)$$

In order to simplify the right-hand side, we can employ the useful vector identity, known as the interchange of the dot and the cross,

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} \quad (2.32)$$

To obtain

$$\mathbf{r} \cdot (\dot{\mathbf{r}} \times \mathbf{h}) = (\mathbf{r} \times \dot{\mathbf{r}}) \cdot \mathbf{h} = \mathbf{h} \cdot \mathbf{h} = h^2 \quad (2.33)$$

Substituting this expression into the right-hand side of Equation 2.31, and substituting  $\mathbf{r} \cdot \mathbf{r} = r^2$  on the left yields

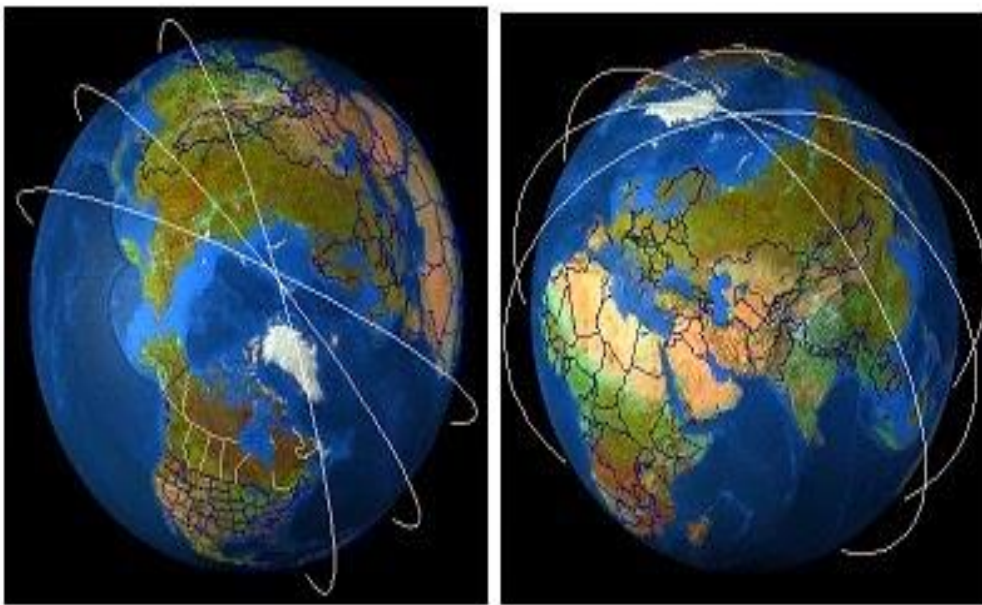
$$r + \mathbf{r} \cdot \mathbf{e} = \frac{h^2}{\mu} \quad (2.34)$$

observe that by following the steps leading from Equation 2.30 to 2.34 we have lost track of the variable time. This occurred at Equation 2.33, because his constant Finally, from the definition of the dot product we have

$$\mathbf{r} \cdot \mathbf{e} = r e \cos \theta$$

# Chapter 3

## ORBITS IN THREE DIMENSION



### 3.1. Introduction:

*The discussion of orbital mechanics up to now has been confined to two dimensions to the plane of the orbits themselves.*

*This chapter explores the means of describing orbits in three-dimensional space, which, of course, is the setting for real missions and orbital maneuvers. Our focus will be on the orbits of earth satellites.*

*We begin with a discussion of the ancient concept of the celestial sphere and the use of right ascension and declination to define the locations of stars, planets and other celestial objects on the sphere.*

*This leads to the establishment of the inertial geocentric equatorial frame of reference and the concept of state vector.*

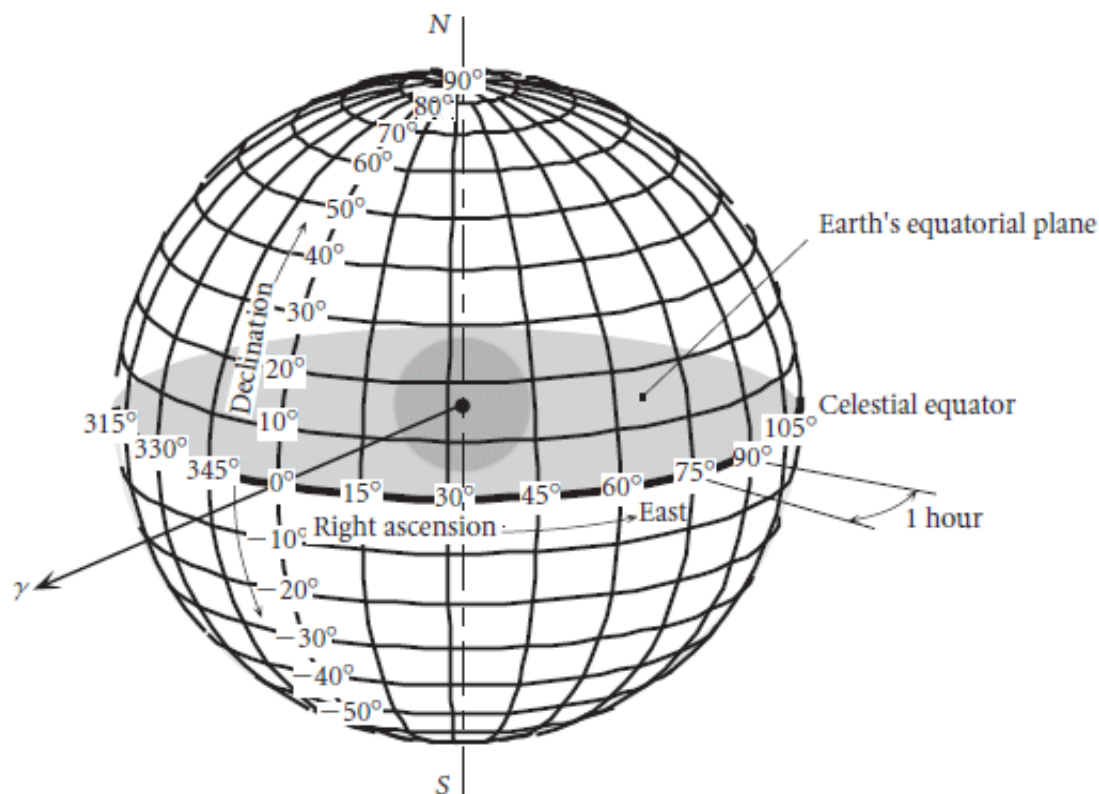
*The six components of this vector give the instantaneous position and velocity of an object relative to the inertial frame and define the characteristics of the orbit.*

*Following that discussion is a presentation of the six classical orbital elements, which also uniquely define the shape and orientation of an orbit and the location of a body on it.*

*We then show how to transform the state vector into orbital elements.*

## 3.2. Geocentric Right Ascension-Declination frame:

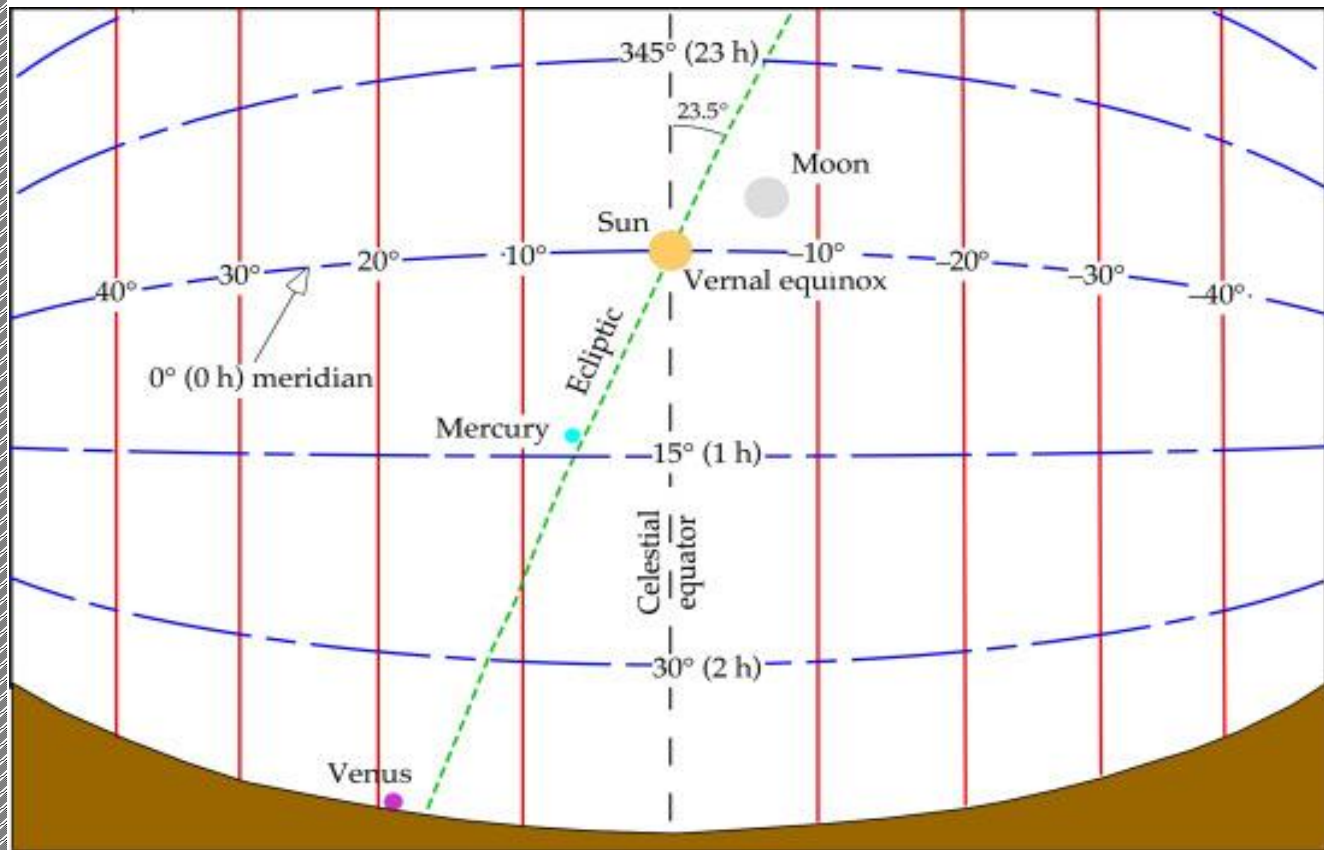
In the human eye, objects in the night sky appears as points on a celestial sphere surrounding the earth, as given in below diagram. The north and south poles of this fixed sphere correspond to those of the earth rotating within it. Coordinates of latitude and longitude are used to locate points on the celestial sphere in much the same way as on the surface of the earth. The projection of the earth's equatorial plane outward onto the celestial sphere defines the celestial equator. The vernal equinox  $\gamma$ , which lies on the celestial sphere, is the origin for measurement of longitude, which in astronomical parlance is called right ascension. Right ascension (RA or  $\alpha$ ) is measured along the celestial equator in degrees east from the vernal equinox. (Astronomers measure right ascension in hours instead of degrees, where 24 hours equals  $360^\circ$ ).



[The celestial sphere, with grid lines of right ascension and declination.]



Latitude on the celestial sphere is called declination. Declination (Dec or  $\delta$ ) is measured along a meridian in degrees, positive to the north of the equator and negative to the south.



[A view from the sky above the eastern horizon from 0° longitude at the equator 9cm local time, 20 March 2004.]

Above figure is a sky chart showing how the heavenly grid appears from a given point on the earth. Notice that the sun is located at the intersection of the equatorial and ecliptic planes, so this must be the first day of spring. Stars are so far away from the earth that their positions relative to each other appear stationary on the celestial sphere. Planets, comets, satellites, etc., move upon the fixed backdrop of the stars. The coordinates of celestial bodies as a function of time is called an ephemeris, for example, the *Astronomical Almanac* (US Naval Observatory, 2004).

## Venus and moon ephemeris for 0 hours universal time (Precession epoch: AD 2000)

Date	Venus		Moon	
	RA	Dec	RA	Dec
1 Jan 2004	21hr 05.0 min	-18°36_	1hr 44.9 min	+8°47_
1 Feb 2004	23hr 28.0 min	-04°30_	4hr 37.0 min	+24°11_
1 Mar 2004	01hr 30.0 min	+10°26_	6hr 04.0 min	+08°32_
1 Apr 2004	03hr 37.6 min	+22°51_	9hr 18.7 min	+21°08_
1 May 2004	05hr 20.3 min	+27°44_	11hr 28.8 min	+07°53_
1 Jun 2004	05hr 25.9 min	+24°43_	14hr 31.3 min	-14°48_
1 Jul 2004	04hr 34.5 min	+17°48_	17hr 09.0 min	-26°08_
1 Aug 2004	05hr 37.4 min	+19°04_	21hr 05.9 min	-21°49_
1 Sep 2004	07hr 40.9 min	+19°16_	00hr 17.0 min	-00°56_
1 Oct 2004	09hr 56.5 min	+12°42_	02hr 20.9 min	+14°35_
1 Nov 2004	12hr 15.8 min	+00°01_	05hr 26.7 min	+27°18_
1 Dec 2004	14hr 34.3 min	-13°21_	07hr 50.3 min	+26°14_
1 Jan 2005	17hr 12.9 min	-22°15_	10hr 49.4 min	+11°39_

or epoch, for we know that even the positions of the stars relative to the equinox change slowly with time.

Above table is an abbreviated ephemeris for the moon and for Venus. An ephemeris depends on the location of the vernal equinox at a given time.

#### **Variation of the coordinates of the star Regulus due to precession of the equinox.**

<i>Precession epoch</i>	<i>RA</i>	<i>Dec</i>
AD 1700	9hr 52.2 min (148.05°)	+13°25_AD
AD1800	9hr 57.6 min (149.40°)	+12°56_
AD1900	10hr 3.0 min (150.75°)	+12°27_
AD 1950	10hr 5.7 min (151.42°)	+12°13_
AD 2000	10hr 8.4 min (152.10°)	+11°58_

For example, Above table shows the celestial coordinates of the star Regulus at five epochs since AD 1700. Currently, the position of the vernal equinox in the year 2000 is used to define the standard grid of the celestial sphere.

In 2025, the position will be updated to that of the year 2050; in 2075 to that of the year 2100; and so on at 50year intervals. Since observations are made relative to the actual orientation of the earth, these measurements must be transformed into the standardized celestial frame of reference. As above table suggests, the adjustments will be small if the current epoch is within 25 years of the standard precession epoch.

### **3.3. State Vector and the geocentric equatorial frame:**

At any given time, the state vector of a satellite comprises its velocity  $\mathbf{v}$  and acceleration.

- a. Orbital mechanics is concerned with specifying or predicting state vectors over intervals of time. From Chapter 2, we know that the equation governing the state vector of a satellite traveling around the earth is, under the familiar assumptions,

$$\mathbf{r}'' = -\frac{\mu}{r^3} \mathbf{r} \quad (3.1)$$

$\mathbf{r}$  is the position vector of the satellite relative to the center of the earth. The components of  $\mathbf{r}$  and, especially, those of its time derivatives  $\dot{\mathbf{r}} = \mathbf{v}$  and  $\ddot{\mathbf{r}} = \mathbf{a}$ , must be measured in a non-rotating frame attached to the earth. A commonly used nonrotating right-handed cartesian coordinate system is the geocentric equatorial frame shown in below diagram. The  $X$  axis points in the vernal equinox direction. The  $XY$  plane is the earth's equatorial plane, and the  $Z$  axis coincides with the earth's axis of rotation and points northward.

The unit vectors  $\hat{\mathbf{i}}$ ,  $\hat{\mathbf{j}}$  and  $\hat{\mathbf{k}}$  form a right-handed triad. The non-rotating geocentric equatorial frame serves as an inertial frame for the two-body earth satellite problem, as embodied in Equation 3.1. It is not truly an inertial frame, however, since the center of the earth is always accelerating towards a third body, the sun (to say nothing of the moon), a fact which we ignore in the two-body formulation.

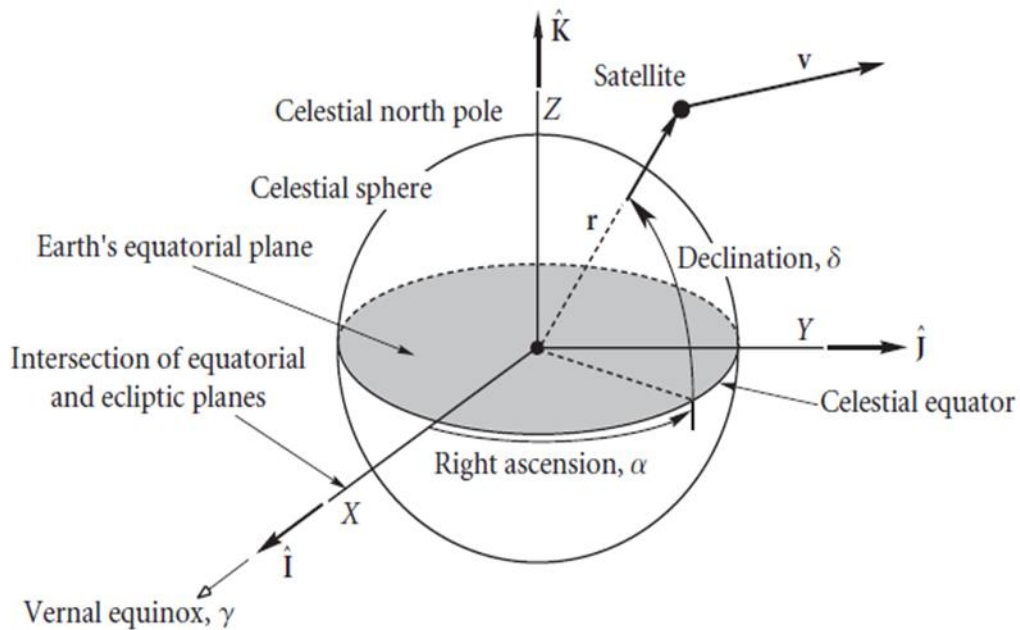
In the geocentric equatorial frame the state vector is given in component form by,

$$\mathbf{r} = X \hat{\mathbf{i}} + Y \hat{\mathbf{j}} + Z \hat{\mathbf{k}} \quad (3.2)$$

$$\mathbf{v} = v_X \hat{\mathbf{i}} + v_Y \hat{\mathbf{j}} + v_Z \hat{\mathbf{k}} \quad (3.3)$$

If  $r$  is the magnitude of the position vector then,

$$\mathbf{r} = r \hat{\mathbf{u}}_r \quad (3.4)$$



[Geocentric equatorial frame.]sssss

From above figure we see that the components of  $\hat{\mathbf{u}}_r$  (the direction cosines of  $\mathbf{r}$ ) are found in terms of the right ascension  $\alpha$  and declination  $\delta$  as follows,

$$\hat{\mathbf{u}}_r = \cos \delta \cos \alpha \hat{\mathbf{i}} + \cos \delta \sin \alpha \hat{\mathbf{j}} + \sin \delta \hat{\mathbf{k}} \quad (3.5)$$

Therefore, given the state vector, we can then compute the right ascension and declination. However, the right ascension and declination alone do not furnish  $\mathbf{r}$ . For that we need the distance  $r$  to obtain  $\mathbf{r}$  from equation 3.4.

### Example: 3.1

If the position vector of the International Space Station is

$$\mathbf{r} = -5368 \hat{\mathbf{i}} - 1784 \hat{\mathbf{j}} + 3691 \hat{\mathbf{k}} \text{ (km)}$$

what are its right ascension and declination

The magnitude of  $\mathbf{r}$  is

$$r = (-5368)^2 + (-1784)^2 + 3691^2 = 6754 \text{ km}$$

Hence,

$$\begin{aligned} \hat{\mathbf{u}}_r &= \mathbf{r}/r \\ &= -0.7947 \hat{\mathbf{i}} - 0.2642 \hat{\mathbf{j}} + 0.5464 \hat{\mathbf{k}} \end{aligned}$$

From this and equation 4.5 we see that  $\sin \delta = 0.5464$  which means,

$$\delta = \sin^{-1} 0.5464 = 33.12^\circ$$

There is no quadrant ambiguity since, by definition, the declination lies between  $-90^\circ$  and  $+90^\circ$ , which is precisely the range of the principal values of the arc sin function. It also follows that  $\cos \delta$  cannot be negative. From equation 4.5 and equation (a) just above we have,

$$\cos \delta \cos \alpha = -0.7947 \quad (b)$$

$$\cos \delta \sin \alpha = -0.2642 \quad (c)$$

Therefore,

$$\begin{aligned} \cos \alpha &= -0.7947 / \cos 33.12^\circ \\ &= -0.9489 \end{aligned}$$

which implies,

$$\begin{aligned} \alpha &= \cos^{-1} (-0.9489) \\ &= 161.6^\circ \text{ (second quadrant) or } 198.4^\circ \text{ (third quadrant)} \end{aligned}$$

From (c) we observe that  $\sin \alpha$  is negative, which means  $\alpha$  lies in the third quadrant,

$$\alpha = 198.4^\circ$$

If we are provided with the state vector  $\mathbf{r}_0, \mathbf{v}_0$  at a given instant, then we can determine the state vector at any other time in terms of the initial vector by means of the expressions.

$$\begin{aligned} \mathbf{r} &= f \mathbf{r}_0 + g \mathbf{v}_0 \\ \mathbf{v} &= \dot{f} \mathbf{r}_0 + \dot{g} \mathbf{v}_0 \end{aligned} \quad (36)$$

where the Lagrange coefficients  $f$  and  $g$  and their time derivatives are given in Equation 3.66. Specifying the total of six components of  $\mathbf{r}_0$  and  $\mathbf{v}_0$  therefore completely determines the size, shape and orientation of the orbit.

## 3.4. Orbital Elements and State Vector

### Algorithm 3.1:

Obtain orbital elements from the state vector. A MATLAB version of this procedure appears in Appendix D.8. Applying this algorithm to orbits around other planets or the sun amounts to defining the frame of reference and substituting the appropriate gravitational parameter  $\mu$ .

1. Calculate the distance,

$$r = \sqrt{\mathbf{r} \cdot \mathbf{r}} = \sqrt{X^2 + Y^2 + Z^2}$$

2. Calculate the speed,

$$v = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_X^2 + v_Y^2 + v_Z^2}$$

3. Calculate the radial velocity,

$$V_r = \mathbf{r} \cdot \mathbf{v} / r = (Xv_X + Yv_Y + Zv_Z) / r$$

Note that if  $v_r > 0$ , the satellite is flying away from perigee. If  $v_r < 0$ , it is flying towards perigee.

4. Calculate the specific angular momentum,

$$\mathbf{h} = \mathbf{r} \times \mathbf{v} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ X & Y & Z \\ v_X & v_Y & v_Z \end{vmatrix}$$

5. Calculate the magnitude of the specific angular momentum,

$$h = \sqrt{\mathbf{h} \cdot \mathbf{h}} \text{ the first orbital element.}$$

6. Calculate the inclination,

$$i = \cos^{-1} (h_Z / h) \quad (3.7)$$

This is the second orbital element. Recall that  $i$  must lie between  $0^\circ$  and  $180^\circ$ , so there is no quadrant ambiguity. If  $90^\circ < i \leq 180^\circ$ , the orbit is retrograde.

7. Calculate

$$\mathbf{N} = \hat{\mathbf{K}} \times \mathbf{h} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 0 & 0 & 1 \\ h_X & h_Y & h_Z \end{vmatrix} \quad (3.8)$$

This vector defines the node line.

8. Calculate the magnitude of  $\mathbf{N}$ ,

$$N = \sqrt{\mathbf{N} \cdot \mathbf{N}}$$

9. Calculate the RA of the ascending node,

$$\Omega = \cos^{-1} (N_X / N)$$

the third orbital element. If  $(N_X / N) > 0$ , then  $\Omega$  lies in either the first or fourth quadrant. If  $(N_X / N) < 0$ , then  $\Omega$  lies in either the second or third quadrant. To place  $\Omega$  in the proper quadrant, observe that the ascending node lies on the positive side of the vertical XZ plane ( $0^\circ \leq \Omega < 180^\circ$ ) if  $N_Y > 0$ . On the other hand, the ascending node lies on the negative side of the XZ plane ( $180^\circ \leq \Omega < 360^\circ$ ) if  $N_Y < 0$ . Therefore,  $N_Y > 0$  implies that  $0^\circ \leq \Omega < 180^\circ$ , whereas  $N_Y < 0$  implies that  $180^\circ \leq \Omega < 360^\circ$ . In summary,

$$\begin{aligned} \Omega &= \cos^{-1} (N_X / N) & (N_Y \geq 0) \\ \Omega &= 360^\circ - \cos^{-1} (N_X / N) & (N_Y < 0) \end{aligned} \quad (3.9)$$

10. Calculate the eccentricity vector. Starting with Equation 2.30,

$$\mathbf{e} = 1/\mu [\mathbf{v} \times \mathbf{h} - \mu_r / r]$$

$$\begin{aligned}
&= 1/\mu [\mathbf{v} \times (\mathbf{r} \times \mathbf{v}) - \mu \frac{\mathbf{r}}{r}] \\
&\quad \text{bac-cab rule} \\
&= 1/\mu [\mathbf{r}v^2 - \mathbf{v}(\mathbf{r} \cdot \mathbf{v}) - \mu \frac{\mathbf{r}}{r}]
\end{aligned}$$

so that

$$\mathbf{e} = 1/\mu [(v^2 - \mu/r) \mathbf{r} - v_r \mathbf{v}] \quad (3.10)$$

11. Calculate the eccentricity,

$$e = \sqrt{\mathbf{e} \cdot \mathbf{e}}$$

the fourth orbital element. substituting equation 3.10 leads to a form depending only on the scalars obtained thus far,

$$e = 1/\mu \sqrt{(2\mu - rv^2)rv^2r + (\mu - rv^2)^2}$$

12. Calculate the argument of perigee,

$$\omega = \cos^{-1} (\mathbf{N} \cdot \mathbf{e}/Ne)$$

the fifth orbital element. If  $\mathbf{N} \cdot \mathbf{e} > 0$ , then  $\omega$  lies in either the first or fourth quadrant. If  $\mathbf{N} \cdot \mathbf{e} < 0$ , then  $\omega$  lies in either the second or third quadrant. To place  $\omega$  in the proper quadrant, observe that perigee lies above the equatorial plane ( $0 \leq \omega < 180^\circ$ ) if  $\mathbf{e}$  points up (in the positive Z direction), and perigee lies below the plane ( $180^\circ \leq \omega < 360^\circ$ ) if  $\mathbf{e}$  points down. Therefore,  $e_Z \geq 0$  implies that  $0 < \omega < 180^\circ$ , whereas  $e_Z < 0$  implies that  $180^\circ < \omega < 360^\circ$ . To summarize,

$$\omega = \begin{cases} \cos^{-1} \left( \frac{\mathbf{N} \cdot \mathbf{e}}{Ne} \right) & (e_Z \geq 0) \\ 360^\circ - \cos^{-1} \left( \frac{\mathbf{N} \cdot \mathbf{e}}{Ne} \right) & (e_Z < 0) \end{cases}$$

13. Calculate the true anomaly,

$$\theta = \cos^{-1} (\mathbf{e} \cdot \mathbf{r}/er)$$

the sixth and final orbital element. If  $\mathbf{e} \cdot \mathbf{r} > 0$ , then  $\theta$  lies in the first or fourth quadrant. If  $\mathbf{e} \cdot \mathbf{r} < 0$ , then  $\theta$  lies in the second or third quadrant. To place  $\theta$  in the proper quadrant, note that if the satellite is flying away from perigee ( $\mathbf{r} \cdot \mathbf{v} \geq 0$ ), then  $0 \leq \theta < 180^\circ$ , whereas if the satellite is flying towards perigee ( $\mathbf{r} \cdot \mathbf{v} < 0$ ), then  $180^\circ \leq \theta < 360^\circ$ . Therefore, using the results of step 3 above

$$\theta = \cos^{-1} (\mathbf{e} \cdot \mathbf{r}/er) \quad (e_r \geq 0)$$

$$360^\circ - \cos^{-1} (\mathbf{e} \cdot \mathbf{r}/er) \quad (e_r < 0)$$

Substituting Equation 4.10 yields an alternative form of this expression,

$$\theta = \cos^{-1} \left[ \frac{1}{e} \frac{h^2}{\mu r} - 1 \right] \quad (v_r \geq 0)$$

$360^\circ - \cos^{-1} \left[ \frac{1}{e} \frac{h^2}{\mu r} - 1 \right] \quad (v_r < 0)$  The procedure described above for calculating the orbital elements is not unique.



# Chapter 4

## Preliminary Orbits

## Determination

### 4.1 Gibb's method of orbits

#### determination with three positions vectors

let observations of a space object at the three successive times  $t_1, t_2$  and  $t_3$  ( $t_1 < t_2 < t_3$ ) the geocentric position vectors  $\mathbf{r}_1, \mathbf{r}_2$  and  $\mathbf{r}_3$ . The problem is to determine the velocities  $v_1, v_2$  and  $v_3$  at  $t_1, t_2$  and  $t_3$  assuming that the object is in a two-body orbit.

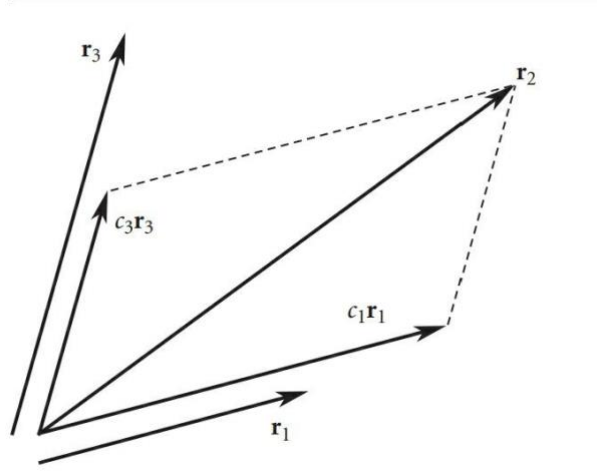
The unit vector normal to the plane of  $\mathbf{r}_2$  and  $\mathbf{r}_3$  must be perpendicular to the unit vector in the direction of  $\mathbf{r}_1$ . Thus, if  $\widehat{\mathbf{u}}_{r1} = \mathbf{r}_1/r_1$  and  $\widehat{\mathbf{C}}_{23} = (\mathbf{r}_2 \times \mathbf{r}_3) / \|\mathbf{r}_2 \times \mathbf{r}_3\|$ , then the dot product of these two unit vectors must vanish,

$$\widehat{\mathbf{u}}_{r1} \cdot \widehat{\mathbf{C}}_{23} = 0$$

$\mathbf{r}_1, \mathbf{r}_2$  and  $\mathbf{r}_3$  lie in the same plane. Apply scalar factors  $c_1$  and  $c_3$  to  $\mathbf{r}_1$  and  $\mathbf{r}_3$  so that  $\mathbf{r}_2$  is the vector sum of  $\mathbf{r}_1 c_1$  and  $\mathbf{r}_3 c_3$

$$\mathbf{r}_2 = c_1 \mathbf{r}_1 + c_3 \mathbf{r}_3 \quad (4.1)$$

The coefficients  $c_1$  and  $c_3$  are readily obtained from  $\mathbf{r}_1, \mathbf{r}_2$  and  $\mathbf{r}_3$



**Figure 4.1** Any one of a set of three coplanar vectors ( $r_1, r_2, r_3$ ) can be expressed as the vector sum of the other two.

To find the velocity  $v$  corresponding to any of the three given position vectors  $r$ , in Equation 2.30, which may be written

$$v \times h = \mu \cdot \left( \frac{r}{r} + e \right)$$

where  $h$  is the angular momentum and  $e$  is the eccentricity vector. The cross product of this equation with the angular momentum,

$$h \times (v \times h) = \mu \cdot \left( \frac{h \times r}{r} + h \times e \right) \quad (4.2)$$

By means of the bac – cab rule (Equation 2.23), the left side becomes

$$h \times (v \times h) = v(h \cdot h) - h(h \cdot v)$$

But  $h \cdot h = h^2$  and  $v \times h = 0$ , since  $v$  is perpendicular to  $h$ . Therefore

$$h \times (v \times h) = h^2 v$$

which means Equation 4.2 may be written

$$v = \mu/h^2 \cdot h \times r/r + h \times e \quad (4.3)$$

the unit vector  $\hat{p}$  lies in the direction of the eccentricity vector  $e$  and  $\hat{w}$  is the unit vector normal to the orbital plane, in the direction of the angular momentum vector  $h$ . Thus, we can write

$$e = e\hat{p} \quad (4.4a)$$

$$h = h\hat{w} \quad (4.4b)$$

Equation 4.3 becomes

$$v = \frac{\mu}{h^2} \cdot (h\hat{w} \times r/r + h\hat{w} \times e\hat{p}) = \frac{\mu}{h} [\hat{w} \times r/r + e(\hat{w} \times \hat{p})] \quad (4.5)$$

Since  $\hat{p}$ ,  $\hat{q}$  and  $\hat{w}$  form a right-handed triad of unit vectors,  $\hat{p} \times \hat{q} = \hat{w}$ ,  $\hat{q} \times \hat{w} = \hat{p}$  and

$$\hat{w} \times \hat{p} = \hat{q} \quad (4.6)$$

Equation 4.5 reduces to

$$v = \mu/h \cdot \hat{w} \times r/r + e\hat{q} \quad (4.7)$$

Use the position vectors  $r_1, r_2$  and  $r_3$  to calculate  $q, w, h$  and  $e$ , let us take the dot product of Equation 4.1 with the eccentricity vector  $e$  to obtain the scalar equation

$$r_2 \cdot e = c_1 r_1 \cdot e + c_3 r_3 \cdot e \quad (4.8)$$

According to Equation 2.34 – the orbit equation – we have the following relation among  $h, e$  and each of the position vectors,

$$r_1 \cdot e = \frac{h^2}{\mu} - r_1 \quad r_2 \cdot e = \frac{h^2}{\mu} - r_2 \quad r_3 \cdot e = \frac{h^2}{\mu} - r_3 \quad (4.9)$$

Substituting these relations into Equation 5.8 yields

$$\left(\frac{h^2}{\mu} - r_2\right) = c_1\left(\frac{h^2}{\mu} - r_1\right) + c_3\left(\frac{h^2}{\mu} - r_3\right) \quad (4.10)$$

To eliminate  $c_1$  and  $c_2$  from this expression, let us take the cross product of Equation 4.1 first with  $r_1$  and then  $r_3$ , both having  $r_3 \times r_1$  on the right,

$$\mathbf{r}_2 \times \mathbf{r}_1 = c_3(\mathbf{r}_3 \times \mathbf{r}_1) \quad \mathbf{r}_1 \times \mathbf{r}_3 = -c_1(\mathbf{r}_3 \times \mathbf{r}_1) \quad (4.11)$$

multiply Equation 4.10 by the vector  $\mathbf{r}_3 \times \mathbf{r}_1$

$$\frac{h^2}{\mu} (\mathbf{r}_3 \times \mathbf{r}_1) - \mathbf{r}_2(\mathbf{r}_3 \times \mathbf{r}_1) = c_1(\mathbf{r}_3 \times \mathbf{r}_1) \left(\frac{h^2}{\mu} - r_1\right) + c_3(\mathbf{r}_3 \times \mathbf{r}_1) \left(\frac{h^2}{\mu} - r_3\right)$$

Using Equations 5.11, this becomes

$$\frac{h^2}{\mu} (\mathbf{r}_3 \times \mathbf{r}_1) - \mathbf{r}_2(\mathbf{r}_3 \times \mathbf{r}_1) = -(\mathbf{r}_2 \times \mathbf{r}_3) \left(\frac{h^2}{\mu} - r_1\right) + (\mathbf{r}_2 \times \mathbf{r}_1) \left(\frac{h^2}{\mu} - r_3\right)$$

$$\frac{h^2}{\mu} (\mathbf{r}_1 \times \mathbf{r}_2 + \mathbf{r}_2 \times \mathbf{r}_3 + \mathbf{r}_3 \times \mathbf{r}_1) = \mathbf{r}_1(\mathbf{r}_2 \times \mathbf{r}_3) + \mathbf{r}_2(\mathbf{r}_3 \times \mathbf{r}_1) + \mathbf{r}_3(\mathbf{r}_1 \times \mathbf{r}_2) \quad (4.12)$$

This is an equation involving the given position vectors and the unknown angular momentum

h. Let us introduce

$$\mathbf{N} = \mathbf{r}_1(\mathbf{r}_2 \times \mathbf{r}_3) + \mathbf{r}_2(\mathbf{r}_3 \times \mathbf{r}_1) + \mathbf{r}_3(\mathbf{r}_1 \times \mathbf{r}_2) \quad (5.13)$$

$$\mathbf{D} = \mathbf{r}_1 \times \mathbf{r}_2 + \mathbf{r}_2 \times \mathbf{r}_3 + \mathbf{r}_3 \times \mathbf{r}_1 \quad (4.14)$$

Then Equation 5.12 may be written

$$\mathbf{N} = \frac{h^2}{\mu} \mathbf{D}$$

$$N = \frac{h^2}{\mu} D$$

where  $N = \|\mathbf{N}\|$  and  $D = \|\mathbf{D}\|$ . It follows from Equation 4.15 that the angular momentum  $h$  is determined from  $\mathbf{r}_1$ ,  $\mathbf{r}_2$  and  $\mathbf{r}_3$  by the formula

$$h = \sqrt{\mu \frac{N}{D}} \quad (4.16)$$

Since  $\mathbf{r}_1$ ,  $\mathbf{r}_2$  and  $\mathbf{r}_3$  are coplanar, all of the cross products  $\mathbf{r}_1 \times \mathbf{r}_2$ ,  $\mathbf{r}_2 \times \mathbf{r}_3$  and  $\mathbf{r}_3 \times \mathbf{r}_1$  lie in the same direction, namely, normal to the orbital plane.  $\mathbf{D}$  must be normal to the orbital plane.  $\hat{\mathbf{w}}$  to denote the orbit unit normal.

$$\hat{\mathbf{w}} = \frac{\mathbf{D}}{D} \quad (4.17)$$

Find  $h$  and  $\hat{\mathbf{w}}$  in terms of  $\mathbf{r}_1$ ,  $\mathbf{r}_2$  and  $\mathbf{r}_3$ . Find an expression for  $q$  to use in Equation 4.7. From Equations 4.4a, 4.6, and 4.17 it follows that

$$\hat{\mathbf{q}} = \hat{\mathbf{w}} \times \hat{\mathbf{p}} = \frac{1}{De} (\mathbf{D} \times \mathbf{e}) \quad (4.18)$$

Substituting Equation 4.14 we get

$$\hat{\mathbf{q}} = \frac{1}{De} [(\mathbf{r}_1 \times \mathbf{r}_2) \times \mathbf{e} + (\mathbf{r}_2 \times \mathbf{r}_3) \times \mathbf{e} + (\mathbf{r}_3 \times \mathbf{r}_1) \times \mathbf{e}] \quad (5.19)$$

Apply the bac – cab rule to the right side by noting

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = -\mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{A}(\mathbf{B} \cdot \mathbf{C})$$

Using this vector identity we obtain

$$(\mathbf{r}_2 \times \mathbf{r}_3) \times \mathbf{e} = \mathbf{r}_3(\mathbf{r}_2 \cdot \mathbf{e}) - \mathbf{r}_2(\mathbf{r}_3 \cdot \mathbf{e})$$

$$(\mathbf{r}_3 \times \mathbf{r}_1) \times \mathbf{e} = \mathbf{r}_1(\mathbf{r}_3 \cdot \mathbf{e}) - \mathbf{r}_3(\mathbf{r}_1 \cdot \mathbf{e})$$

$$(\mathbf{r}_1 \times \mathbf{r}_2) \times \mathbf{e} = \mathbf{r}_2(\mathbf{r}_1 \cdot \mathbf{e}) - \mathbf{r}_1(\mathbf{r}_2 \cdot \mathbf{e})$$

employing Equations 4.9,

$$(\mathbf{r}_2 \times \mathbf{r}_3) \times \mathbf{e} = \mathbf{r}_3 \left( \frac{h^2}{\mu} - r_2 \right) - \mathbf{r}_2 \left( \frac{h^2}{\mu} - r_3 \right) = \frac{h^2}{\mu} (\mathbf{r}_3 - \mathbf{r}_2) + r_3 \mathbf{r}_2 - r_2 \mathbf{r}_3$$

$$(\mathbf{r}_3 \times \mathbf{r}_1) \times \mathbf{e} = \mathbf{r}_1 \left( \frac{h^2}{\mu} - r_3 \right) - \mathbf{r}_3 \left( \frac{h^2}{\mu} - r_1 \right) = \frac{h^2}{\mu} (\mathbf{r}_1 - \mathbf{r}_3) + r_1 \mathbf{r}_3 - r_3 \mathbf{r}_1$$

$$(\mathbf{r}_1 \times \mathbf{r}_2) \times \mathbf{e} = \mathbf{r}_2 \left( \frac{h^2}{\mu} - r_1 \right) - \mathbf{r}_1 \left( \frac{h^2}{\mu} - r_2 \right) = \frac{h^2}{\mu} (\mathbf{r}_2 - \mathbf{r}_1) + r_2 \mathbf{r}_1 - r_1 \mathbf{r}_2$$

Summing these three equations, collecting terms and substituting the result into Equation 4.19 yields

$$\hat{\mathbf{q}} = \frac{1}{De} \mathbf{S} \quad (4.20)$$

where

$$\mathbf{S} = \mathbf{r}_1(r_2 - r_3) + \mathbf{r}_2(r_3 - r_1) + \mathbf{r}_3(r_1 - r_2) \quad (4.21)$$

substitute Equations 4.16, 4.17 and 4.20 into Equation 4.7

$$\mathbf{v} = \frac{\mu}{h} (\hat{\mathbf{w}} \times \mathbf{r} / r + \mathbf{e} \hat{\mathbf{q}})$$

$$= \frac{\mu}{\sqrt{\mu_D^N}} \left[ \frac{\frac{D}{D} \times \mathbf{r}}{r} + \mathbf{e} \left( \frac{1}{De} \mathbf{S} \right) \right]$$

$$\mathbf{v} = \sqrt{\frac{\mu}{ND}} \left( \frac{D \times \mathbf{r}}{r} + \mathbf{S} \right) \quad (4.22)$$

All of the terms on the right depend only on the given position vectors  $r_1$ ,  $r_2$  and  $r_3$



# Conclusion:

Humanity requires more efficient, more sustainable, and much less costly access to space, if it wants to dramatically expand its use of Earth orbit and make interplanetary space part of its economical sphere. We need ways to get into orbit to reach other planets that do not leave large amount of debris, require enormous amounts of propellant, or take incredibly long period of time. The space tether systems described in this book offer various solutions. Space elevators could provide an easy and regular way to get into Earth orbit, and electrodynamic momentum exchange tethers could send space-craft from low orbits up into higher ones and vice versa. Tethers could even de-orbit return capsules or send space craft on their way to other planets.

The earth's gravity field provides a record of the mass distribution within the system and can be used to understand the evolution and dynamics needed to maintain that distribution. For the fluid portions of the earth, gravity measurements can be used to sense directly the motions of mass within the system. The inversion of the gravity signals to obtain the mass distribution and the dynamics that cause them is not a straightforward problem, but through a combination of spatial and temporal analyses, knowledge of the gravity field and its temporal variations can provide insights into the process that can control these dynamics.

# References

- Bate, R. R., Mueller, D., and White, J. E. (1971). *Fundamentals of Astrodynamics*, Dover Publications.
- Battin, R. H. (1999). *An Introduction to the Mathematics and Methods of Astrodynamics*, Revised Edition, AIAA Education Series.
- Beyer, W. H., ed. (1991). *Standard Mathematical Tables and Formulae*, 29th Edition, CRC Press.
- Bond, V. R. and Allman, M. C. (1996). *Modern Astrodynamics: Fundamentals and Perturbation Methods*, Princeton University Press.
- Boulet, D. L. (1991). *Methods of Orbit Determination for the Microcomputer*, Willmann-Bell.
- Chobotov, V. A., ed. (2002). *Orbital Mechanics*, Third Edition, AIAA Education Series, AIAA.
- Coriolis, G. (1835). ‘On the Equations of Relative Motion of a System of Bodies’, *J. École Polytechnique*, Vol. 15, No. 24, 142–54.
- Hahn, B. D. (2002). *Essential MATLAB® for Scientists and Engineers*, Second Edition, Butterworth-Heinemann.
- Hale, F. J. (1994). *Introduction to Space Flight*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Hohmann, W. (1925). *The Attainability of Celestial Bodies* (in German), R. Oldenbourg.
- Kaplan, M. H. (1976). *Modern Spacecraft Dynamics and Control*, Wiley.
- Kermit, S. and Davis, T. A. (2002). *MATLAB Primer*, Sixth Edition, Chapman & Hall/CRC.
- Likins, P.W. (1967). ‘Attitude Stability Criteria for Dual Spin Spacecraft’, *Journal of Spacecraft and Rockets*, Vol. 4, No. 12, 1638–43.
- Magrab, E. B., ed. (2000). *An Engineer’s Guide to MATLAB®*, Prentice-Hall.
- NASA Goddard Space Flight Center (2003). *National Space Science Data Center*, <http://nssdc.gsfc.nasa.gov>.
- Nise, N. S. (2003). *Control Systems Engineering*, Fourth Edition, Wiley.
- Ogata, K. (2001). *Modern Control Engineering*, Fourth Edition, Prentice-Hall.
- Palm W. J. (1983). *Modeling, Analysis and Control of Dynamic Systems*, Wiley.
- Prussing, J. E. and Conway, B. A. (1993). *Orbital Mechanics*, Oxford University Press.