Mathematics Journal

Volume XI

2019-2020

PREFACE

Ever since its inception in 2010, Éclat - the annual academic journal of the Department of Mathematics - has been released annually as a print-journal. However, the volume of the academic session 2019-20, that is, the XI th volume marks the foray of Éclat into e-journal form. Even its new form, this journal continues to propel students into quality undergraduate research. The articles published in this volume arise from the mentorship programme that has been initiated from this academic session. The papers aim to provide theoretical and empirical research from interactions between student and faculty.

A new section "Contribution of Women in Mathematics" has been added to the four pre-existing ones, viz. History of Mathematics, Rigour in Mathematics, Extension of Course Content and Interdisciplinary Aspects of Mathematics. This year's volume focuses on the contributions of Maryam Mirzakhani, the first female Fields medallist. Together, these five sections serve to highlight the multiple facets of Mathematics. As can be seen in the contents of the papers, the focus this year has been on the interdisciplinary aspects of Mathematics, which captures the reality of the growing relevance of Mathematics in various other fields.

We would like to express our thanks to the student team comprising of Lipika Parekh, Khuisangmi Konghay (Batch of 2020), Paavani Mangla and Tanvi Vohra (Batch of 2021) for their assistance in the compilation of this journal, as well as all the authors and mentors who have contributed their articles for this volume. We also thank Jassika Kapoor (Batch of 2021) for designing the cover page of this volume.

We hope that this journal continues to be an intellectual stimulus for its readers whilst motivating students to make valuable contributions to their subject.

Éclat Editorial Board

CONTENTS

Éclat, A Mathematics Journal
Volume 11, 2019-20 (1-9)

Escher Arts: Hyperbolic Geometry and Tessellations

Dr. B. Kaur, N. Nair and T. Vohra

Abstract

M.C. Escher created fascinating artwork using simple drawing tools, recognized by mathematicians as extraordinary visualizations of mathematical principles delving into areas of hyperbolic geometry, spherical geometry, tessellations, polyhedra and impossible figures to name a few. In this paper we delve into the areas of hyperbolic geometry and tessellations emphasizing on Escher's artwork in these areas.

Keywords: Euclidean Space, Curvature, Saddle Point, Hyperbolic Plane, Isometric Projection

1 M. C. Escher

Maurits Cornelis Escher (1898-1972)(Figure 1), born in Leeuwarden, Netherlands, was one of the world's most famous graphic artists who made mathematically inspired artwork exploring concepts of infinity, symmetry, perspective, hyperbolic geometry, tessellations, etc. without any formal training in mathematics. He briefly studied architecture at the Haarlem School of Architecture and Decorative Arts, but switched to decorative arts, studying under the tutelage of the graphic artist Samuel Jessurun de Mesquita [5]. After completing his school, he traveled for a long time through Italy, where he made beautiful works inspired by the Italian countryside. The intricate decorative designs of Moorish architecture of the 14^{th} century Alhambra triggered his interest in the mathematics of tessellation which became a powerful influence on his work. Escher created Circle Limits I-IV and Snakes inspired by hyperbolic plane tessellations. He is most famous for his impossible drawings, such as Ascending and Descending and Relativity, and for his metamorphoses, such as Metamorphosis I, II and III. M.C. Escher was the first artist to create patterns in the hyperbolic plane using both the Poincaré disk model and the Poincaré half-plane model of hyperbolic geometry. During his lifetime, Escher made 448 lithographs[1],

[1]Lithography is a printing process on a flat stone, with the use of grease and water.

woodcuts and wood engravings and more than 2000 drawings and sketches.

2 Hyperbolic Geometry

Hyperbolic geometry is a type of non-Euclidean geometry showcasing the independence of the parallel postulate. Independence of the parallel postulate means that it is posible to construct a consistent system of "geometrical" statements dealing with points, lines, etc., by deduction from a set of axioms in which the parallel postulate is replaced by a contrary postulate[1].
Hyperbolic Geometry was independently discovered by C.F. Gauss, János Bolyai, and N.I. Lobatchevsky and satisfies all the axioms of Euclidean geometry except for the parallel postulate which states that If two lines are drawn which intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines inevitably must intersect each other on that side if extended far enough [3]. In other words, through a point not on a given line there is exactly one line parallel to the given line. In hyperbolic geometry there are infinitely many straight lines through a given point that do not intersect a given line. As a consequence, in hyperbolic geometry, the following hold:

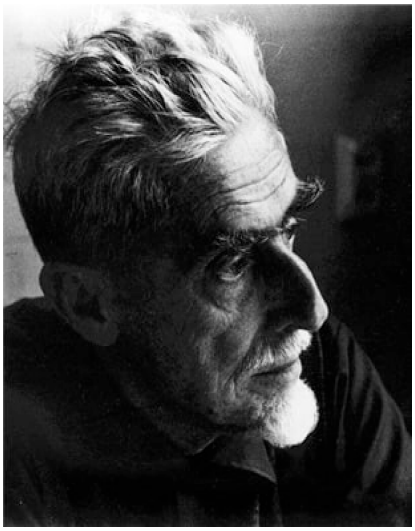


Figure 1: M. C. Escher; (source: https://en.wikipedia.org/wiki/M_C_Escher)

1. All triangles have angle sum less than 180°
2. The interior angles of a quadrilateral sum to less than 360°
3. There are no rectangles
   (In a rectangle, the measure of all the four angles is 90°. Since the sum of the angles of a quadrilateral in hyperbolic plane is less than 360°, therefore not all angles can be 90°.
4. Similar triangles are congruent

Hyperbolic plane geometry is the geometry of surfaces with a constant negative curvature[2] such as saddle surfaces where every point is a saddle point eg. coral reefs.

[2]Curvature is the degree to which something is curved. It is the reciprocal of radius, for each point on the circle.

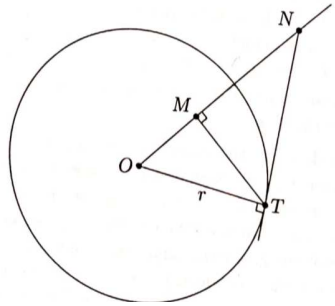


Figure 2: Inversion $|OM|.|ON| = r^2$; (source: Sossinsky,A.B., Geometries, American Mathematical Society (2012)

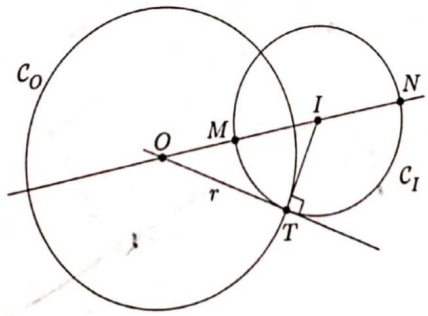


Figure 3: Orthogonal Circles; (source: Sossinsky,A.B., Geometries, American Mathematical Society (2012))

We must rely on non-isometric models of hyperbolic plane as there is no smooth isometric[3] embedding of the hyperbolic plane in Euclidean 3-space, as proved by David Hilbert. The Klein model, the Poincaré disk model, the Poincaré half-plane model, and the Lorentz or hyperboloid model are the four models commonly used for hyperbolic geometry which define a hyperbolic plane satisfying the axioms of a hyperbolic geometry. The Poincaré model for hyperbolic geometry, which is very useful for visualization, is used by Escher for his *Circle Limits I-IV*.

2.1 Inversion

1. Inversions map any circle or straight line orthogonal to the circle of inversion into itself. One such inversion is shown in figure 2.
   In Figure 3, C_0 and C_1 are orthogonal circles with centers O and I respectively. According to the property, inversion takes C_1 to itself and in particular point

[3]Isometric projection means visually representing 3-D objects in 2-D.

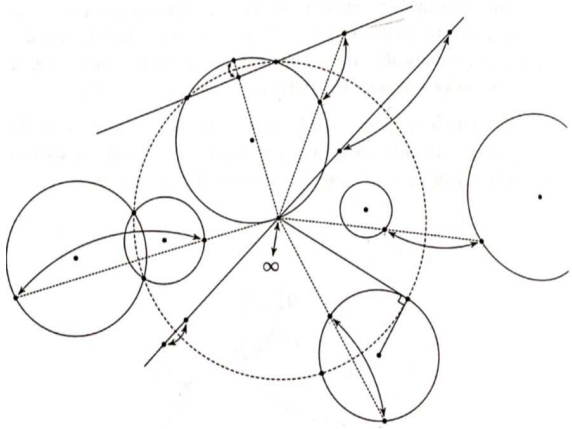


Figure 4: Images of circles and lines under inversion; (source: Sossinsky,A.B., Geometries, American Mathematical Society (2012))

M is mapped to N and the intersection points of the two circles are mapped to itself. The arcs of C_1 cut out by C_0 are mapped to each other. Similar rules hold when C_1 is considered as the circle of inversion.

2. Inversions map any circle or straight line into a circle or straight line
   Consider Figure 4, here the center of inversion is mapped to ∞ and vice versa, hence lines passing through the center of inversion are mapped to themselves but turned inside out. Circles that pass through the center of inversion are mapped to straight lines which do not pass through the center of inversion and vice versa.

3. Inversions preserve the measure of angles
   Angle measurements formed by two intersecting curves are the Euclidean angles formed by their tangents at point of intersection.

2.2 Points, lines and triangles in the hyperbolic plane

Consider the open disk:
$\mathbf{H}^2 := \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 < 1\}$ The points of a hyperbolic plane are simply defined as the points of the open disk $\mathbf{H}^2$ [1].(See figure 5)
The boundary circle of the open disk is referred to as the absolute. Thus, the lines in a hyperbolic plane are defined as the intersections with $\mathbf{H}^2$ of the Euclidean circles orthogonal to the absolute as well as the diameters of the absolute [2].

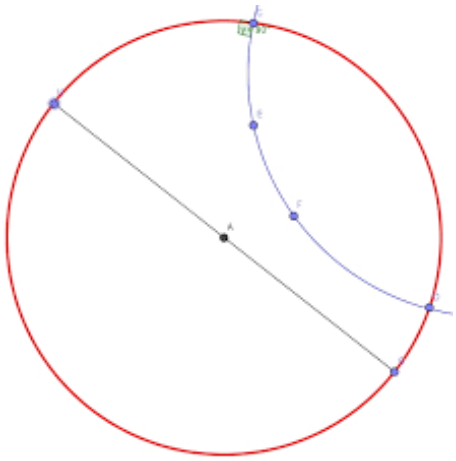


Figure 5: Points and Lines in Hyperbolic Plane;(source: https://ibmathsresources.com/tag/poincare/)

2.3 Poincaré disk model

Poincaré imagined world composed of interior of a circle, such that points form circular arcs perpendicular at their extremities to the circumference. As explained in Section 2.2, these are known as straight lines in Hyperbolic Plane [1].
The Poincaré disk model, proposed by Eugenio Beltrami, is a well-known conformal model of the hyperbolic plane, i.e angles for this model are same as Euclidean angles. The underlying space of this model is the open unit disk $\mathbf{D} = \{z \in \mathbf{C_1} : |x| < 1\}$ [4]. In this model, the plane is the unit disk and points are Euclidean points. The shortest distance between two points, also called a geodesic (lines in the context of Euclidean geometry) are either diameters of the disk or the intersection of a circle C_1 with the disk, where C_1 is perpendicular to the unit circle at its two points of intersection. The distance formula for two complex numbers z and w inside the disk becomes

$$d(z,w) = \operatorname{arccosh}\left(1 + 2\frac{|z-w|^2}{\left(1-|z|^2\right)\left(1-|w|^2\right)}\right),$$

where $\operatorname{arccosh}(x) = \ln\left(x+\sqrt{x^2-1}\right)$ is the inverse function of the hyperbolic cosine. Note that as $|z|$ approaches 1, the distance between z and another point approaches ∞. Thus, the distance to the boundary of the disk is infinite, and postulate 2[4] of Euclidean geometry holds.

[4]Postulate 2 states that any straight line segment can be extended indefinitely in a straight line

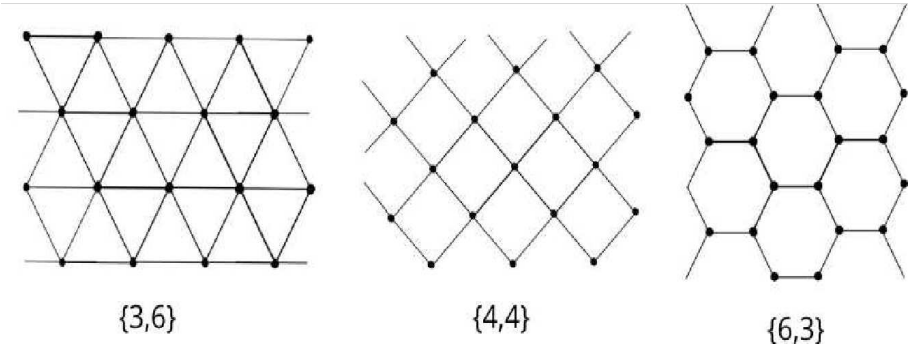


Figure 6: Tessellations;
(source: http://pi.math.cornell.edu/ mec/Winter2009/Mihai/section9.html)

3 Tessellations

A plane tessellation is an infinite set of polygons fitting together to cover the whole plane just once, so that every side of the polygon belongs also to another polygon. No two of the polygons have common interior points [6].

A tessellation is said to be regular if all the faces are regular and equal and the same number of polygons meet at each of the edges. A regular tesselation containing m-sided regular polygons and in which exactly l edges meet at each vertex is denoted by $\{m, l\}$. This is known as the Schläfli symbol[5]. In Euclidean plane, only 3 regular tessellations are possible: $\{3, 6\}$, where 6 equilateral triangles meet at each vertex, $\{6, 3\}$, where 3 hexagons meet at each vertex and $\{4, 4\}$, where 4 squares meet at each vertex, as shown in Figure 6.

This is true because each internal angle of a regular polygon is given by $\pi\left(1 - \frac{2}{m}\right)$. Since l edges meet at each vertex, there are l internal angles that add up to 2π. Hence, we have:

$$l\pi\left(1 - \frac{2}{m}\right) = 2\pi$$

$$\implies \frac{1}{l} + \frac{1}{m} = \frac{1}{2}$$

The above equation has 3 solutions: $m = 3$ and $l = 6$; $m = 4$ and $l = 4$; $m = 6$ and

[5]It is the symbol in geometry that is of the form {p,q,r.....}, that is used to denote regular polytops and tessellations.

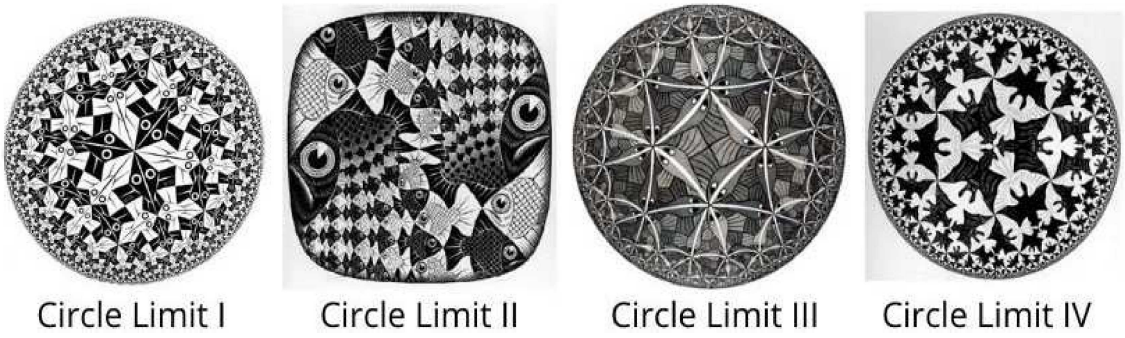


Figure 7: Circle Limit I-IV;
(source: https://mathstat.slu.edu/escher/index.php/Circle_Limit_Exploration)

$l = 3$.
However, there are infinitely many such pairs in the hyperbolic plane since the sum of the angles of a triangle is less than 180°, the equation becomes:

$$\frac{1}{l} + \frac{1}{m} < \frac{1}{2}$$

which has infinitely many solutions.

4 Escher's Artwork

Escher was highly inspired by the hyperbolic plane tessellations and utilized the Poincaré disk model for his artwork. As a result, he created four different pieces known as *Circle Limits I-IV* (Shown in figure 7), which were all wood cuts. His aim was to depict infinity in a finite space. Escher's ideas about structure, pattern, and infinity were enhanced when he came across the work of the geometer H. S. M. Coxeter (1907-2003). Coxeter and Escher struck up a correspondence when Coxeter hoped to use Escher's unique depictions of symmetry in a presentation for the Royal Society of Canada. Coxeter sent Escher a copy of the talk, which included an illustration depicting a tessellation of the hyperbolic plane. This image sparked a new area of Escher's exploration of infinity [7].

4.1 Circle Limit I

This was Escher's first attempt in the execution of his idea of creating infinity in a finite space. In this work both quadrilateral and hexagonal tessellations can be seen. The Euclidean polygons and lines tend to diminish as we reach the boundary

of the disk. The question was how much do they scale. The distance is calculated in the similar way as in Euclidean geometry. In the hyperbolic plane, the lengths are inversely proportional to the distance from the center. For example, a segment half way between the center and the boundary has twice the Euclidean length of one that is one quarter of the way from the boundary to the center. The area of polygons share the same relationship.
Escher was displeased with his first work because the fish didn't face the same direction, the colours did not alternate well and the fish didn't look realistic. Although, the basic structure was to his satisfaction upon which he based his further paintings.

4.2 Circle Limit III

Circle limit III is where Escher resolved his self- criticisms. In this painting, the fish are properly organised, the colouring alternates and the fish look real. This was done by alternating triangles and quadrilaterals. It was a 4- colour symmetry based on $\{8, 3\}$ tessellation, which implies that 3 octagons meet at each vertex. It can be observed that the nose tips and fin tips form the vertices of an octagon. Although this time the painting was aesthetically pleasing, it was not approved by Coxeter. In a hyperbolic tessellation, the polygons are made from perpendicular circle segments, unlike in this piece of work, where the arcs meet the boundary at 80° angles.

4.3 Circle Limit IV

Circle limit IV is the last work of the series. This is based on a hexagonal tiling. Here, the toe tips and wing tips form the vertices of the hexagon. This piece is commonly known as *Heaven and Hell* and alternates angles and devils.

5 Conclusion

M. C. Escher's work is a beautiful amalgamation of the wonderful realities of the world around us and an expression of his own mind-boggling ideas. His work includes the application of Hyperbolic Plane models such as the Poincaré Disk Model. One can clearly observe the use of patterns and tessellations, and such other concepts that also have mathematical significance.

References

[1] Courant R. and Robbins H., *What is Mathematics*, Oxford University Press 2007.

[2] Sossinsky A. B., *Geometries*, American Mathematical Society 2012.

[3] Brannan D. A., *Geometry*, Cambridge University Press 2012.

[4] Dunham D., *M.C. Escher's Use of the Poincaré Models of Hyperbolic Geometry*, University of Minnesota, 2012.

[5] https://mcescher.com/about/biography/

[6] http://pi.math.cornell.edu/ mec/Winter2009/Mihai/section9.html

[7] http://web.colby.edu/thegeometricviewpoint/2016/12/21/tessellations-of-the-hyperbolic-plane-and-m-c-escher/

(Mentor) Dr. Bhavneet Kaur, Assistant Professor, Department of Mathematics, Lady Shri Ram College for Women, New Delhi
bhavneet.lsr@gmail.com

Nayana Nair, B.Sc.(H) Mathematics, 4th Semester, Lady Shri Ram College for Women, New Delhi
nayana.nair391@gmail.com

Tanvi Vohra, B.Sc.(H) Mathematics, 4th Semester, Lady Shri Ram College for Women, New Delhi
vohratanvi5@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (10-17)

Sundial: The Mathematical Angle

Dr. B. Kaur, D. Natu and D. Bansal

Abstract

This paper aims to discuss the working of different types of sundial and extend the research to study the structure of Samrat Yantra which is based on the concept of equatorial sundials. Built in 18th century, Samrat Yantra monitors and maps out the movement of the sun, moon and the stars.

Keywords: Gnomon, Equinoctial, Hour Angle, Latitude, Elliptical Dial

1 Introduction

Architectural wonders have always been closely integrated to the field of mathematics and most architects implemented mathematics to create marvels all over the world. This paper corroborates on the geometrical interpretation of intricate patterns and designs used in one specific architectural marvel- *the sundial.* The earliest form of time keeping, known through archaeological findings, was done with the use of shadow clocks also known as sundials. One such sundial was the Samrat Yantra, built in the 18th century in India by Maharaja Sawai Jai Singh. This magnificent structure intrigued us and inspired us to dwell deeper into its mathematical aspect. To think of something so great that involves some of the most complex mathematical and astronomical principles, that too in an age when there was no internet or concrete existence of any such knowledge is in itself a great achievement.

2 Sundial

Centuries ago, shadow of sticks was used as an archaic instrument for calculating the time throughout the day. The stick then came to be known as gnomon which is parallel to the Earth's axis. The Greek mathematicians developed the trigonometry in such a way that hour lines could be determined arithmetically rather than by geometry, and that lead to more sophisticated sundials. There are three main kinds of sundial- equatorial, horizontal and vertical.

2.1 Equatorial sundial

Consider the Earth as a giant sundial as shown in figure 1. The axis of the Earth is tilted at an angle of 23.5° to the plane of its orbit around the Sun. A vertical stick is placed at the pole and so when the earth rotates on its axis, the shadow of the stick would form a circle on the surface of the earth such that it is parallel to the equator. If this circle is divided into 24 equal hour marks, the position of the shadow around the circle would give the time. Sundials that work on this principle are called equatorial sundial. The gnomon makes an angle of L with the horizontal plane which is equal to the latitude of the location so that the gnomon is parallel to the axis of the earth and the dial is inclined at an angle of 90°-L making it perpendicular to the gnomon and parallel to the equator. Thus achieving accuracy in time prediction.

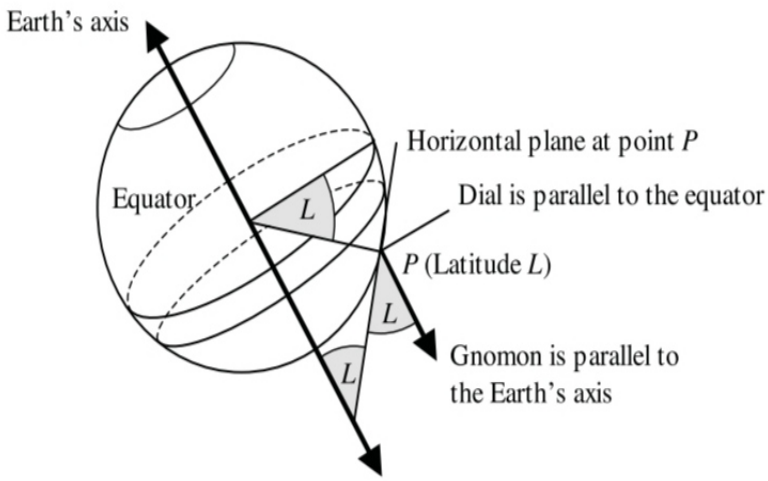


Figure 1: Inclination of the dial and the gnomon of an equatorial sundial; (source: Vincent,J.,(2008) The mathematics of sundials, Australian Senior Mathematics Journal

2.2 Horizontal sundial

In a horizontal sundial, the shadow is tracked on a horizontal plane. The circular equatorial dial which is parallel to the plane of the equator is projected onto a horizontal plane as an ellipse. The gnomon makes an angle (L) equal to the local latitude with the horizontal plane. The motion of the shadow of the gnomon on the plane of the dial surface is not uniform and so the hour lines marked on the dial surface are not equidistant. In figure 2, the radius of the equatorial dial and the semi-minor (east-west) axis is a whereas the semi-major (north-south) axis of the ellipse is b. Using trigonometry, we get

$$\sin L = \frac{a}{b} \quad (1)$$

$$\text{therefore,} \quad b = \frac{a}{\sin L} \quad (2)$$

Equation of ellipse is:

$$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1 \tag{3}$$

substituting (2) in (3), we get

$$\begin{aligned}
\frac{x^2 \sin^2 L}{a^2} + \frac{y^2}{a^2} &= 1 \\
x^2 \sin^2 L + y^2 &= a^2 \\
\text{Also,} \qquad y &= a \sin T \qquad \text{(from } \Delta\text{OCA)} \\
\text{therefore,} \qquad x^2 &= \frac{a^2 - a^2 \sin^2 T}{\sin^2 L} \\
x^2 &= \frac{a^2 \cos^2 T}{\sin^2 L} \\
x &= \frac{a \cos T}{\sin L}
\end{aligned}$$

The values of x and y will vary according to the quadrant in which the shadow of the gnomon lies. In figure 2, the shadow lies in the first quadrant and hence positive values of x and y are chosen.T is the hour angle of the equatorial sundial and H is the hour angle of the horizontal sundial.Thus, the parametric equations for the ellipse of the horizontal sundial are:

$$x = \frac{a \cos T}{\sin L}$$

$$y = a \sin T$$

The hour angle on a horizontal plane can be calculated by:
From triangle OBC,

$$\begin{aligned}
\tan H &= \frac{y}{x} \\
\tan H &= \frac{a \sin T \sin L}{a \cos T} \\
H &= \tan^{-1}(\tan T \sin L)
\end{aligned}$$

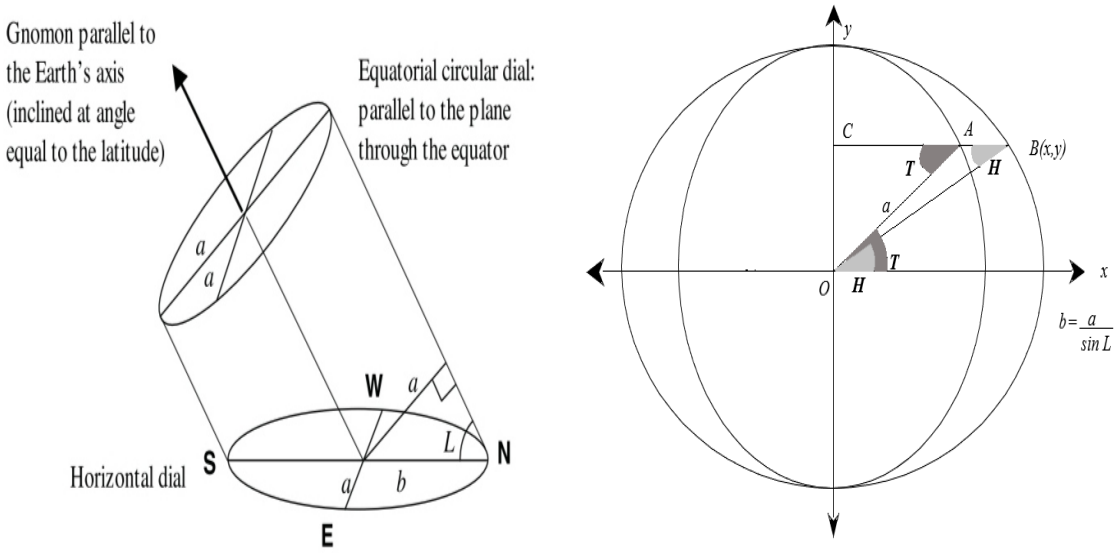


Figure 2: Relationship between the hour angle T of the equatorial dial and the projected hour angle H of the horizontal dial [2];(source: Vincent,J.,(2008) The mathematics of sundials, Australian Senior Mathematics Journal)

2.3 Vertical sundial

Vertical sundials are those sundials whose shadow receiving plane is aligned vertically. The gnomon's style (hypotenuse) is the shadow casting element and is aligned with the Earth's axis of rotation. An equatorial sundial has equal angles between each hour line however a vertical sundial is not equiangular. An equatorial sundial is projected onto a vertical plane as an elliptical dial, as in figure 3, where the radius of the equatorial dial, a, is equal to the semi minor axis of this dial and b is the semi major axis with L being the latitude of the place where the sundial is situated. Using trigonometry, we get:

$$\begin{aligned} \cos L &= \frac{a}{b} \\ b &= \frac{a}{\cos L} \end{aligned}$$

Thereby, the equation of ellipse and the above trigonometric formula gives us the hour angle of the vertical ellipse. The parametric equations for the ellipse are

$$\begin{aligned} x &= a \sin T \\ y &= \frac{a \cos T}{\cos L} \end{aligned}$$

From triangle OBC,

$$\begin{aligned} \tan H &= \frac{x}{y} \\ \tan H &= \frac{a \sin T \cos L}{a \cos T} \end{aligned}$$

$$H = \tan^{-1}(\tan T \cos L)$$

gives the projected hour angle of the vertical sundial.

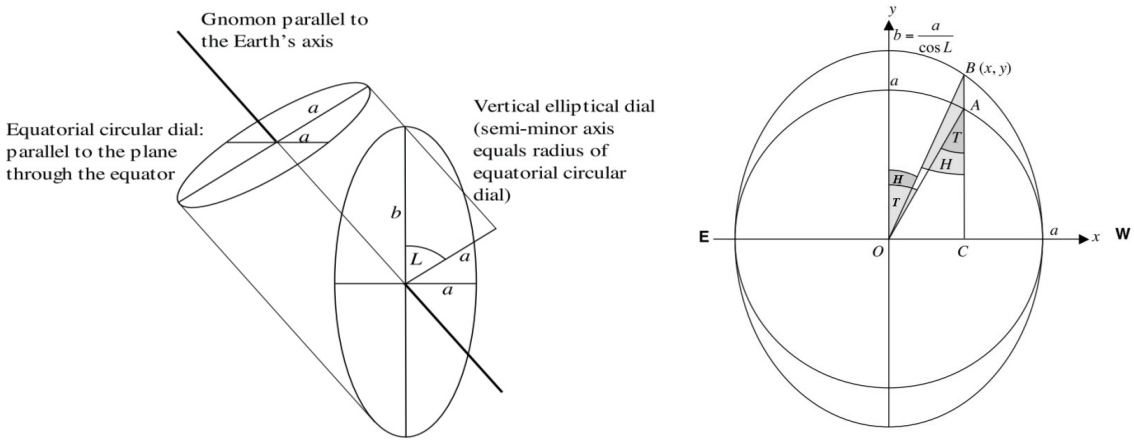


Figure 3: Relationship between the hour angle T and the projected hour angle H; (source: Vincent,J.,(2008) The mathematics of sundials, Australian Senior Mathematics Journal)

3 Samrat Yantra

Jantar Mantar is known to be one of the greatest designed monuments in the history of art, mathematics and astronomy. The architecture of the place is like a rich tapestry woven with threads of astronomy and mathematics. It is a gigantic astronomical observatory which was built in 1730s by Maharaja Sawai Jai Singh II of Jaipur. He constructed five such observatories in five different cities. The founding concept of Jantar Mantar lies in the conquest of science and cosmic energy that governs the movement of various celestial bodies. Samrat Yantra is an equinoctial [3] sundial that is present in all five observatories. The Samrat Yantra translated as Supreme Instrument is the largest sundial in the world. It was designed with mathematical precision to serve various functions in monitoring and scaling out the movements of the sun, moon and stars.

3.1 Working

The gnomon of the Samrat Yantra (figure 4) rises over 74 feet [5] above its base and has marble faced quadrants on both sides of the gnomon that are 9 feet in width and create an arc which reaches 45 feet in height.The instrument has a triangular central wall that points towards the north celestial pole, making an angle equal to the latitude of the location where the yantra has been placed and thus the hypotenuse is parallel to the axis of rotation of the Earth. The quadrant arcs are placed perpendicular to the inclined wall of the instrument and are in plane of the equator. The shadow of the gnomon sweeps the quadrant from west to east as the sun journeys from east to west. At any given moment, the time is indicated by the shadow's edge on a quadrant scale. The edge of each quadrant is graduated in hours, minutes and seconds. The shadow on the gnomon moves approximately by 4 meters in one hour. This translates to 6 cm every minute and, with each minute sub-divided into thirty fractions, the sundial is theoretically able to provide an accuracy of 2 seconds [1]. When the sun rises in the east, the shadow rails on the top of the western quadrant, gradually descending across its curvature to the midpoint of the structure and when noon approaches there is no shadow. In the afternoon, the shadow correspondingly rises up the eastern quadrant until it reaches its most distant point at sunset. The Samrat Yantra measures the time with the help of shadow formed on the quadrant however it measures the local time and not the standard time of the country and so a correction has to be applied to its readings in order to obtain the standard time. The correction required for Indian Standard Time is as follows:

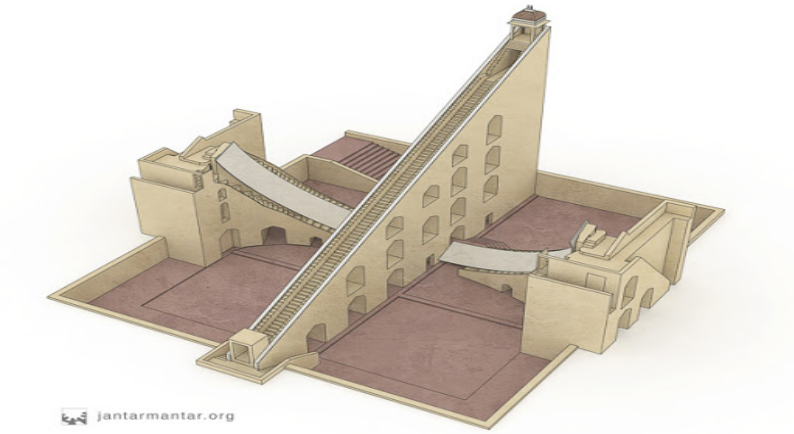


Figure 4: Samrat Yantra; (source: jantarmantar.org.)

Indian Standard Time = Local Time ± Equation of Time [1] ± Longitude difference [2] [4]

The latitude of Jaipur is 27° and the height of the gnomon is 74 ft. Using simple trigonometry we can compute the length of style which will be approximately equal to 164.4 ft.

4 Conclusion

Our article shows the underlying concept of mathematics used in architecture and the combination of the two used in analyzing many astronomical phenomena.

References

[1] Volwahsen A., *Cosmic Architecture in India: The Astronomical Monuments of Maharaja Jai Singh II*, Prestel Publishing 2001

[2] Vincent J., *The mathematics of sundials*, Australian Senior Mathematics Journal 2008

[3] https://www.ias.ac.in/article/fulltext/reso/022/03/0201-0212

[4] https://www.jantarmantar.org/Architecture-Science-web.pdf

[5] https://www.jantarmantar.org/gallery/renderingsSamrat/index.php

(Mentor) Dr.Bhavneet Kaur, Professor, Department of Mathematics, Lady Shri Ram College for Women, New Delhi
bhavneet.lsr@gmail.com

Disha Natu, B.Sc.(H) Mathematics, 2nd semester,Lady Shri Ram College for Women, New Delhi

[1] Equation of time is the difference between mean solar time (as shown by clocks) and apparent solar time (indicated by sundials), which varies with the time of year.

[2] Longitude difference is the difference between the longitudes of the place where sundial is situated and the place which is used for calculation of Indian standard time. The time difference between each longitude (each degree) is 4 minutes.

dishanatu@gmail.com

Diya Bansal,B.Sc.(H) Mathematics, 2nd semester,Lady Shri Ram College for Women, New Delhi
bansal.diya16@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (18-24)

The Magic of Vedic Mathematics

Ms. R. Agarwal, C. Narang and K. Sharma

Abstract

Mathematics is a subject known for endless yet precise assumptions. The Indian historical mathematics profoundly known as "Vedic Mathematics" was discovered by Jagadguru Sankaracharya Sri Bharati Krishna Tirthaji Maharaj through assumptions made on the grounds of Sutras extracted from "The Vedas ". In this paper, we will discuss the successful applications of Vedic mathematics in the field of trigonometry, algebra, calculus, calculating squares and coordinate geometry.

Keywords: Vedic Mathematics, Sutras, Sub-sutras, Pythagorean Triplets, Integration.

1 Introduction

Vedic mathematics is a system of mathematics which was discovered by Indian mathematician Jagadguru Shri Bharati Krishna Tirthaji (1884-1960) known as the Father of Vedic mathematics. After practising metaphysics and studying the Vedas[1], he formulated 16 Sutras[2], 13 Sub-sutras[3] and various tehniques in the period between 1911-1918 and published his findings in a Vedic Mathematics book [1] (Vedic Ganitagya by Bharati Krishna Tirthaji of Govardhana Matha) for solving arithmetic, algebra, geometry, calculus and conics problems, etc. in an easy and faster way.

2 History of Shri Bharati Krishna Tirthaji

Shri Bharati Krishna Tirthaji was born in March 1884 at Tinnivelli, Tamil Nadu to highly learned and pious parents. He grew up to be a brilliant student and invariably won the first positions in all the subjects in all classes throughout his educational career. During his school days, he was a student of National College Trichana-palli, Church Missionary Society College and Hindu College, Tinnivelly. His extraordinary proficiency in Sanskrit earned him the title "Saraswati" from the Madras Sanskrit Association in July 1899. After securing the highest position in the B.A examination, he appeared for the M.A. examination of the

[1]The Vedas are a collection of hymns and other religious texts written in India by scholars between 1500 and 1000 BCE. One of the 4 Vedas written, it is the Atharva Veda which is rich in knowledge about mathematical formulas.

[2]The word Sutra means Formula.

[3]The word Sub-sutra is Upa-sutra which means Corollary.

American College of Sciences, Rochester, New York from the Bombay center in 1903. After his M.A., he worked under Gopal Krishna Gokhale for the National Education Movement, also taught at the college for three years, but his dedication and devotion towards the Hindu studies led him to join the Sringeri Matha in Mysore completely in 1911 and he became a disciple of Sri Satchidananda Sivabhinava Nrisimha Bharati Swami. He spent the next eight years in the profoundest study of the most advanced Vedanta[4] Philosophy and practice of the Brahmasadhana[5]. After several years, in 1921, he was installed on the pontifical throne of Sharada Peetha Sankaracharya and later in 1925, he became the pontifical head of Sri Govardhan Math Puri where he served the remainder of his life spreading the holy spiritual teachings of Sanatana Dharma[6]. The reader can refer to [1] and [3] for more information of his life history.

3 The Sutras

The real beauty of Vedic mathematics can be endured with the close study of the Sutras. The following table depicts the Sutras, along with their corollary and meaning.

Table 1: The Sutras

Sutras	**Corollary**	**Meaning**
Ekadhikina Purvena	Anurupyena	By one more than the previous one
Nikhilam Navatashcaeamam Dashatah	Sisyate Sesasamjnah	All from 9 and last from 10
Urdhva-Tiryagbyham	Adyamadyenantyamantyena	Vertically and crosswise
Paraavartya Yojayet	Kevalaih Saptakam Gunyat	Transpose and Adjust
Shunyam Saamyasamuccaye	Vestanam	When sum is the same that sum is zero
(Anurupye)Shunyamanyat	Yavadunam Tavadunam	If one is in ratio, other is zero
Sankalana-vyavakalanabhyam	Yavadunam tavadunikritya Varga Yojayet	By addition and by subtraction
Puranapuranabyham	Antyayordashake'pi	By the completion or non-completion
Chalana-Kalanabyham	Antyayoreva	Differences and similarities
Yaavadunam	Samuccayagunitah	Whatever the extent of its deficiency
Vyashtisamanstih	lopanasthapanabhyam	Part and whole
Shesanyankena Charamena	Vilokanam	The remainders by the last digit
Sopaantyadvayamantyam	Gunitasamuccayah Samuccayagunitah	The ultimate and twice the penultimate
Ekanyunena Purvena	Dhvajanka	By one less than the previous one
Gunitasamuchyah	Dwandwa Yoga	The product of the sum is equal to the sum of the product
Gunakasamuchyah	Adyam Antyam Madhyam	The factors of the sum is equal to the sum of the factors

4 Applications of Vedic Mathematics

In this section, some applications of Vedic mathematics in various fields of mathematics such as trigonometry, algebra, calculus, calculating squares, and coordinate geometry are

[4]Vedanta, one of the six systems (darshans) of Indian philosophy. The term Vedanta means in Sanskrit the "conclusion" (anta) of the Vedas, the earliest sacred literature of India. It applies to the Upanishads, which were elaborations of the Vedas metaphysics

[5]when spiritual practice is performed with the ideation of Brahma, it is known as Brahma sadhana, which is a perfect mode of meditation

[6]Sanatana Dharma, in Hinduism, term used to denote the "eternal" or absolute set of duties or religiously ordained practices incumbent upon all Hindus, regardless of class, caste, or sect.

given. The reader can refer to [2] and [4] for more examples and details.

4.1 Application of vedic mathematics in trigonometry

Trigonometry binds the ratio of sides of a right angled triangle in specified way, defining the various trigonometry ratios- sine, cosine, tangent, cotangent, cosecent and secant. Trigonometry demands the user to avail the Pythagorean Triplet. Using Vedic Mathematics we can find the triplets in very quick yet precise manner. The method is divided into two cases.

Case 1: When one number is odd.
The following procedure is followed to attain the result:
Step 1: Square the number (square of any odd number is odd)
Step 2: Divide the squared number by 2.
Step 3: Round off the number to its greatest and least integer.

Example: The length of a side of a right angled triangle is 7 units. Find the length of the other two sides.
Step 1. $7 \times 7 = 49$
Step 2. $49/2 = 24.5$
Step 3. The sides are 7, 24 and 25

Case 2: When one number is even.
The following procedure is followed to attain the result:
Step 1: Divide the number by an even number, such that dividend is an odd number.
Step 2: Square the dividend i.e. the odd number (square of any odd number is odd)
Step 3: Divide the squared number by 2.
Step 4: Round off the number to its greatest and least integer.
Step 5: Multiply the answer with the divisor used in step 1.

Example: A side of a right angled triangle is given 10 units. Find the length of the other two sides.
Step 1. $10/2 = 5$
Step 2. $5 \times 5 = 25$
Step 3. $25/2 = 12.5$
Step 4: The triplet with one odd number is 5, 12 and 13
Step 5: The required triplet is $2 \times (5, 12, 13)$ i.e. 10, 24 and 26.

4.2 Application of vedic mathematics in straight lines

Using the Vedic-one-line method, we will find the equation of a line that passes through two given points. The Sutra used here is "Adyam Sutra[7]".

[7] Adyam Sutra is the extended version of Gunitasamuchyah Sutra [See Table 1]

Procedure: The required equation will be of the form $ax - by = c$
where,

a = difference in y coordinates of two points on the line.

b = difference in x coordinates of two points on the line.

c = product of means minus product of extremes.

Example: Find the equation of a line passing through (3,4) and (8,10).
Solution: Given points are (8,10) and (3,4). So here,

$$\begin{aligned} a &= 10 - 4 = 6 \\ b &= 8 - 3 = 5 \\ c &= 10 \times 3 - 8 \times 4 = -2 \end{aligned}$$

Hence, the required equation is $6x - 5y = -2$.

4.3 Application of vedic mathematics in calculus

In Calculus, and more generally in mathematical analysis, integration by parts is a rule that transforms the integral of products of functions into other, hopefully simpler integrals. Integration by parts is normally done using the formula:

$$\int u\frac{dv}{dx}dx = uv - \int v\frac{du}{dx}dx$$

But you will be surprised how Vedic mathematics can be used to solve this problem in just few seconds. The Sutra used here is "Paravartya Adyamadya [8]" in the following procedure. Let us take an example:

$$I = \int \frac{2x}{(x-1)(x-2)}dx \tag{1}$$

Procedure: Step 1: Write the given integral in the form:

$$I = \int \left[\frac{}{(x-1)} + \frac{}{(x-2)}\right] dx$$

Step 2: Take the first term of integral $\left[\frac{}{(x-1)}\right]$. Putting $x - 1 = 0$, we get $x = 1$. Now, use $x = 1$ in equation (1), which gives,

$$\frac{2 \times 1}{(1-1)(1-2)}$$

Since, equations with zero denominator are undefined, eliminate the term $(1 - 1)$, hence we get

$$\frac{2 \times 1}{1-2} = -2$$

[8] Alternate method to reverse it's exchange.

So, -2 is the required numerator of first term.
In the similar way, 4 is the numerator of second term, where second term is $\left[\frac{}{x-2}\right]$.

Step 3: Integrating the simplified integral

$$I = \int \left[\frac{-2}{(x-1)} + \frac{4}{(x-2)}\right] dx$$

$$\Rightarrow I = -2\log|x-1| + 4\log|x-2| + C$$

4.4 Application of vedic mathematics in finding squares

We generally have a hard time calculating squares of numbers. Vedic mathematics helps us compute the square of any length in two really easy and quick steps.
Procedure:
Step 1: Find the duplex of the given number. The duplex of number is denoted by D(number). The duplex can be found in the following manner.

Number of digits	**Formula**
Single digit	$D(a) = a^2$
Two digits	$D(ab) = 2ab$
Three digits	$D(abc) = 2ac + b^2$
Four digits	$D(abcd) = 2(ad + cb)$
Five digits	$D(abcde) = 2ae + 2bd + c^2$

Step 2: Finding square from the duplex of digits of given number.

Number of digits	**Formula**
Single Digit	$D(a) = a^2$
Two digits	$D(ab) = D(a)/D(ab)/D(b)$
Three digits	$D(abc) = D(a)/D(ab)/D(abc)/D(bc)/D(c)$

4.5 Application of vedic mathematics in algebra

We will be solving the following equation using the Sutra "Sinyam[9]".

$$\frac{2x+5}{2x+11} = \frac{2x+11}{2x+5}$$

The Sub-sutra "Samuchchaya" used in this method means combination. If the sum of numerators is equal to the sum of denominators, then equate that sum to zero.

[9]Sinyam is a synonym for *Tulyashodhana* which means simplifying algebraic equations for eliminating same terms.

Check whether the sum of numerators is equal to the sum of denominators.
Sum of numerators $= (2x + 5) + (2x + 11) = (4x + 16)$
Sum of denominators $= (2x + 11) + (2x + 5) = (4x + 16)$
Since sum is equal, equate the sum to zero

$$4x + 16 = 0 \Rightarrow x = -4$$

4.6 Application of vedic mathematics in factorizations

Here we will use the method of factorization by "Alternate Elimination and Retention". The Sutra used in this method is "Lopan Sthapanabhyam" and the Sub-sutra used is "Adyamadyenantyamantya". Let us take an example:
Factorise $2x^2 + 2y^2 + z^2 + 5xy + 3xz + 3yz$.
Solution: Step 1: Assume either of the three variables is zero. Let $z = 0$. Then

$$2x^2 + 2y^2 + 5xy = 0$$

$$\Rightarrow (x + 2y)(2x + y) = 0 \tag{2}$$

Step 2: Again assume either of the variables zero, but note that this variable should be other than the one used in step 1. Let $y = 0$. Then

$$2x^2 + z^2 + 3xz = 0$$

$$\Rightarrow (x + z)(2x + z) = 0 \tag{3}$$

Step 3: In step 1 and step 2, we acquired equations (2) and (3). From the equations we can note that x and $2x$ have different factors. Now we extract the factors of x and $2x$ and contract them into two factors. Hence, factorized form of the given example is $(x+2y+z)(2x+y+z)$

5 Conclusion

In this paper we have compiled some interesting and infamous Vedic Sutras and their applications which provides us with an easy route for problem solving and develops strategic thinking. In addition to this, these Sutras encourage an intelligent and holistic approach to the subject. However, there has also been serious criticism of its "Vedicness", as it is sometimes suggested that Tirthaji himself might have invented the tricks given the fact that the 16 Sutras expounded by Thirthaji do not appear in any known edition of the Atharva Veda. Nevertheless, Vedic mathematics induces creativity and develops a flexible and logical approach to the subject.

References

[1] Jagadguru Shankaracharya Shri Bharati Krishna Tirthaji Maharaja, *Vedic Mathematics*, 1^{st} edition, Motilal banarsidass publications, 1990.

[2] Thakur R. K., *The essentials of Vedic Mathematics*, 2^{nd} edition, Rupa publications India Pvt. Ltd., 2013.

[3] Kandasamy W. B. Vasantha, *Neutrosophic analysis of Vedic mathematics*, volume 1, 2006, 27.

[4] Satapathy H. K., *Ancient Indian Mathematics*, 1^{st} edition, Rashtriya Sanskrit Vidyapeetha publishers, 2011.

(Mentor) Ms. Reema Agarwal, Assistant Professor, Lady Shri Ram College For Women, New Delhi
reemaagarwal001@gmail.com

Chhavi Narang, B.Sc.(H) Mathematics, 2nd Semester, Lady Shri Ram College For Women, New Delhi
chhaviprth@gmail.com

Kaaya Sharma, B.Sc.(H) Mathematics, 2nd Semester, Lady Shri Ram College For Women, New Delhi
kaayasharma001@icloud.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (25-33)

Bayesian Analysis: A Tool for Geographic Profiling

Dr. S. Nayak, O. Khare and S. Ahmed

Abstract

This paper provides a review of the application of Bayesian statistics in Geographic Profiling and how the inclusion of this statistical technique has modified the existing models to attain higher accuracy.

Keywords: Likelihood Function, Marginal Distribution, Prior Distribution, Posterior Distribution, Normal Distribution, Geographic Profiling

1 Introduction

The Bayesian approach in statistical sciences and data analysis has emerged as an increasingly practical and efficient technique. It requires a sampling model and a prior distribution on all the unknown parameters. The Bayesian inferences arise from the conditional distribution of the unknown parameters given the observed data. It means that the Bayesian approach conditions on the data and integrates over the parameters.
It is used in a wide range of fields including engineering, economics, sports and medicine. N. Thompson Hobbs and Mevin B. Hooten, in their book 'Bayesian Models: A Statistical Primer for Ecologists' [1], describe how Bayesian modeling is an indispensable tool for ecological research as it deals with the complexity in a statistically comprehensible way. The recent advances in Bayesian methods make it extremely useful for the analysis of epidemiological data. 'Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology' [2] by Andrew B. Lawson proffers a coherent account of the full range of Bayesian disease mapping methods and applications. Applications of Bayesian statistics are pervasive and include clinical decision making, bio-monitoring, econometric, etc.
This paper elaborates upon the application of Bayesian Analysis in Geographic Profiling. The paper comprises of 5 sections. Section 2 discusses the basic Bayesian model, providing an overview and understanding of the Bayesian approach. Section 3 discusses the Criminal Geographic Targeting Model given by D.K. Rossmo [4] and how the drawbacks observed for this model led to the application of Bayesian Analysis by researchers. It also explains the Mike O' Leary's Model[5]. Section 4 compares the two models and lastly, section 5 provides a summary of our review research paper, highlighting the recent advancements in the field of Geographic Profiling.

2 Basic Bayesian Model

In Bayesian statistics, some of the following notions are predominantly used:
If $X = (x_1, x_2, \ldots, x_n)$ and θ are two random variables having joint probability distribution $f_{X,\theta}(x, \theta)$

which can be factored as

$$f_{\mathrm{X},\theta}(x,\theta) = f(x|\theta)f(\theta)$$

The function of θ defined by

$$f\left(x_1, x_2, \ldots, x_{\mathrm{n}}|\theta\right)$$

is the **likelihood function**. It represents the support provided by the data for each possible value of the parameter θ.
Given, X and θ have the joint probability distribution function $f_{\mathrm{X},\theta}(x,\theta)$, then the **marginal distribution** of X is given by

$$m(x) = \int f_{\mathrm{X},\theta}(x,\theta)d\theta$$

$$m(x) = \int f(x|\theta)f(\theta)d\theta$$

In Bayesian statistics, before the data is analysed the unknown parameter i.e. θ is modeled as a random variable having distribution $f(\theta)$, called the **prior distribution**.
The conditional distribution of θ given $X = x$ is

$$f(\theta|x) = \frac{f(x,\theta)}{m(x)} = \frac{f(x|\theta)f(\theta)}{\int f(\theta)d\theta}$$

This is known as the **posterior distribution** of θ. It represents the information about the parameter after having observed the data X.
Basic Bayesian model has two stages, with a likelihood specification $Y|\theta \sim p(Y|\theta)$ and a prior specification $\theta \sim p(\theta)$. In Bayesian approach, instead of assuming that θ is a fixed parameter, it is also assumed as a random quantity. The posterior distribution of θ is given by (2)

$$p(\theta|Y) = \frac{p(Y,\theta)}{p(Y)}$$

$$p(\theta|Y) = \frac{p(Y,\theta)}{\int p(Y,\theta)d\theta}$$

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta} \tag{1}$$

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \tag{2}$$

where,
$p(\theta|Y)$ is the posterior distribution, $p(Y|\theta)$ is the likelihood distribution, $p(\theta)$ is the prior distribution and $p(Y)$ is the marginal distribution.
Bayes' Theorem can be explained in a simpler formulation, given an event A and collection of events $B_i, i = 1, 2, 3 \ldots i$, that are mutually exclusive and exhaustive. Given $P\left(B_i\right)$ and $P\left(A|B_i\right)$, from fundamental probability, we have

$$P\left(B_i|A\right) = \frac{P\left(A, B_i\right)}{P(A)} = \frac{P\left(A, B_i\right)}{\sum_{i=1}^{i} P\left(A|B_i\right) P\left(B_i\right)} \tag{3}$$

Equation(3) is just a discrete finite version of equation(2) with B_i playing the role of parameter θ and A playing the role of the data Y.

2.1 Sequential use of Bayes' theorem

It is observed that the equation(2) can be expressed as the following,

$$p(\theta|Y) \propto p(Y|\theta)p(\theta)$$

That is, "posterior is proportional to likelihood times prior". Bayesian approach can be used sequentially as well. Suppose we have two sets of independently collected samples of data Y_1, and Y_2. Then,

$$\begin{aligned} p\left(\theta|Y_1, Y_2\right) &\propto p\left(Y_1, Y_2|\theta\right) p(\theta) \\ p\left(\theta|Y_1, Y_2\right) &= p_2\left(Y_2|\theta\right) p_1\left(Y_1|\theta\right) p(\theta) \\ p\left(\theta|Y_1, Y_2\right) &\propto p_2\left(Y_2|\theta\right) p\left(\theta|Y_1\right) \end{aligned}$$

That is, the posterior distribution of the full data set (Y_1, Y_2) can be obtained by first finding out $p\left(\theta|Y_1\right)$ and then using it as prior for the second section of data Y_2

2.2 The normal distribution

The most important example of continuous distribution is the normal distribution which is met at almost every turn of statistics. The general formula for the probability density function of the standard normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2}\right)$$

Now, for $X \sim N\left(\mu, \sigma^2\right), x \in R, -\infty < \mu < \infty, \sigma^2 > 0$, normal distribution is given as [3]

$$p\left(x|\mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

where, $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$
Since the distribution is symmetric around the mean, the median and mode both are equal to the mean. The area under the bell curve produced by normal distribution is 1.0, that is equal to the total probability of all of the values that the variable can take.
It is observed that 68% of the area of a normal distribution is within one standard deviation of the mean and approximately 95% of the area is within two standard deviations of the mean. These features of normal distribution give a more accurate result and hence it is so widely used.
Carlin and Louis [3] explain that the likelihood, prior and posterior distributions can be expressed using normal distribution. The prior probability density function can be given as $N\left(\mu, \tau^2\right)$, where τ^2 is the variance. The likelihood probability density function can be given as $N\left(\theta, \sigma^2\right)$, where σ^2 is the variance.
Then, given the prior and likelihood distributions as above, the posterior probability density function is given with mean and variance [3]

$$\begin{aligned} E(\theta|Y) &= B\mu + (1-B)Y \\ \mathrm{Var}(\theta|Y) &= (1-B)\sigma^2 \end{aligned}$$

where $B = \frac{\sigma^2}{(\sigma^2+\tau^2)}$ and $0 < \mathrm{B} < 1$
If $\sigma^2 > \tau^2$, i.e. the prior information is more precise than the data information, then B is close to

1 and the posterior mean is close to the prior mean. If $\sigma^2 < \tau^2$, i.e., the prior information is less precise than the data information, then B is close to 0 and the posterior mean is close to the data value Y.

3 Geographic Profiling

Geographical profiling is one of the most important criminal investigative methodologies that analyzes data related to spatial locations of crime scenes to predict the most probable area of criminal operation or residence.

3.1 Criminal Geographic targeting model: Rossmo's formula

Rossmo's formula [4] helps in assigning a certain score to various places which helps the crime agencies to decide whether a particular place is a criminal's anchor point or not. Anchor points refer to locations where a serial criminal or offender is operating.
This model uses 'Manhattan Metric'. In the Manhattan metric, the distance is calculated by adding vertical and horizontal changes in distance between two points. For example, the coordinates of point A are (x_a, y_a) and the coordinates of point B are (x_b, y_b). The distance between the two points denoted by $D(a, b)$ is calculated calculated as:

$$D(a, b) = |x_a - x_b| + |y_a - y_b|$$

The model follows the given procedure:

1. Rossmo gave the following formula to assign scores to each grid to determine whether that place/ grid is the offender's anchor point[4].

$$p_{i,j} = k \sum_{n-1}^{(\text{total crimes})} \left[\underbrace{\frac{\phi_{ij}}{(|X_i - x_n| + |Y_j - y_n|)^f}}_{1^{st}\text{ term}} + \underbrace{\frac{(1 - \phi_{ij})\left(B^{g-f}\right)}{(2B - |X_i - x_n| - |Y_j - y_n|)^g}}_{2^{nd}\text{ term}} \right] \tag{4}$$

$$\text{where: } \phi_{ij} = \begin{cases} 1, & \text{if } (|X_i - x_n| + |Y_j - y_n|) > B \\ 0, & \text{otherwise} \end{cases}$$

where,
p_{ij} represents the resultant probability for point ij,
ϕ_{ij} is the weighing factor,
k is an empirically determined constant,
B is the radius of buffer zone,
C is the number of crime sites,
f, g are empirically determined exponents,
X_i, Y_j are the coordinates of point ij,
x_n, y_n are the coordinates of the n^{th} crime site.

The formula consists of two terms. The first term explains the idea of decreasing probability with increasing distance since the criminal would find it risky to commit crimes far from

his/her known places. The second term indicates the concept of buffer zone. The criminal isn't likely to commit crimes too close to his or her residence or the sites where he/she has already committed crimes, as there is a fear of being recognized. This gives us a distance decay function.

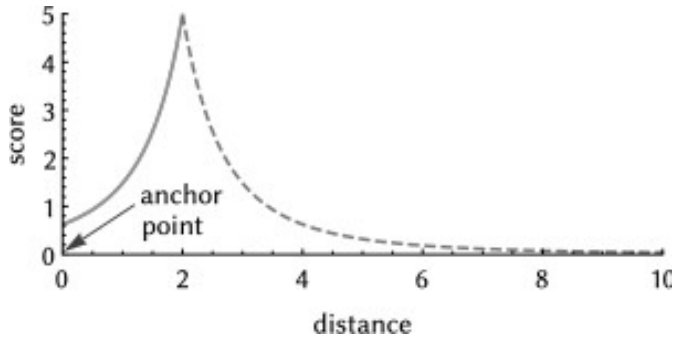


Figure 1: Buffer/distance decay function; (source: Faulkner, S., Stevens, M., Geographic profiling - Murder, maths, malaria and mammals, Chalkdust Magazine, 2017

2. A three-dimensional surface is produced when scores for each point are calculated and represented on the z axis, while coordinates of the same are represented on x-y axis. The performance of the model can be evaluated by determining the proportion of the total hunting area covered before the criminal's residence or area of operation is encountered.

$$\text{hit score} = \frac{\text{number of grid cells searched before finding the criminal}}{\text{total number of grid cells}}$$

3.1.1 Shortcomings of criminal geographic targeting model

1. This "delta function-like" behaviour indicates that the criminal's anchor point is either right next to the crime scene or on the boundary defined by the model. Hence, the B-value becomes exceptionally important and needs its own procedure to ensure its accuracy [7].

2. The model uses the Manhattan metric which might prove to be inappropriate if the cities where crimes are taking place are characterized by concentric layouts and streets.

Keeping these drawbacks aside, the most prominent issue which rose while employing this model was that it provided a score and not an explicit probability of a place being the criminal's anchor point. Thus, the researchers used concepts of Bayesian analysis to tune the existing models to achieve the required aim.

3.2 Mike O' Leary's model using Bayesian analysis

Mike O' Leary, a professor at Towson University, United States, questioned the fact that the geographic targeting model produced only a score, when what is actually required is a probability. So, he developed a model of Geographic Profiling, making use of the laws of Bayesian probability. O' Leary's model on Geographic Profiling[5] allows us to find out the probability that a criminal is based at a certain anchor point given the crimes he or she has committed using Bayesian analysis.

3.2.1 Bayesian analysis of a single crime

Let $x \in R$ with $x = (x_1, x_2)$, be the crime location on a 2D spatial grid, $z \in R$ with $z = (z_1, z_2)$, be the anchor point of the criminal and let $\alpha \in R$ be the criminal's average distance where he or she had committed crimes or offences.
Suppose the criminal has committed only one crime at location x. Then using Bayes theorem (2), we can find out the probability for the anchor point z,

$$P(z, \alpha|x) = \frac{P(x|z, \alpha) \cdot P(z, \alpha)}{P(x)} \tag{5}$$

where,
1. $P(z, \alpha|x)$ is the posterior distribution, that is the probability of the location being the criminal's anchor point keeping in account the parameter α indicating average crime distance given the criminal has committed a crime at location x.
2. $P(x|z, \alpha)$ is the likelihood distribution, that is the probability of x being the crime location given z is the anchor point.
3. $P(x)$ is the marginal distribution that is the probability of x being the location where the certain crime has been committed.
4. $P(z, \alpha)$ is the prior distribution that is the probability that the criminal has an anchor point z and an average crime distance α before considering a series of crimes.
With the assumption that z is independent of the average crime distance α we get,

$$P(z, \alpha) = P(z) \cdot P(\alpha)$$

where,
1.P(z) is the prior probability function that gives us the distribution of anchor points before we incorporate a series of crimes.
2.$P(\alpha)$ is the probability function for providing us with the prior distribution of the criminal's average crime committed distance before we incorporate a series of crimes.
Hence, we get,

$$P(z, \alpha|x) \propto P(x|z, \alpha).P(z, \alpha)$$

for the probability distribution of the anchor point when a single crime had been given.

3.2.2 Bayesian analysis of crime series

This analysis involves estimating the probability for the anchor point 'z' given the crime series $x_1, x_2, x_3, \ldots x_n$ Bayes Theorem implies:

$$P\left(z, \alpha|x_1, x_2, x_3, \ldots, x_n\right) = \frac{P\left(x_1, x_2, x_3, \ldots x_n|z, \alpha\right).P(z, \alpha)}{P\left(x_1, x_2, x_3, \ldots, x_n\right)}$$

with the same definitions given in section 3.2.1
Assuming that the sites where the crimes or offences were committed are independent we can reduce

the formula to

$$P\left(x_1, x_2, x_3, \ldots, x_{\mathrm{a}}|z, \alpha\right) = P\left(x_1|z, \alpha\right).P\left(x_2|z, \alpha\right)\ldots P\left(x_{\mathrm{n}}|z, \alpha\right)$$

substituting we get,

$$P\left(z, \alpha|x_1, x_2, x_3, \ldots, x_n\right) \propto P\left(x_1|z, \alpha\right).P\left(x_2|z, \alpha\right)\ldots P\left(x_n|z, \alpha\right) \cdot P(z) \cdot P(\alpha)$$

Since we are only interested in z we take the conditional distribution and get,

$$P\left(z, \alpha|x_1, x_2, x_3, \ldots, x_n\right) \propto \int_0^\infty P\left(x_1|z, \alpha\right) \cdot P\left(x_2|z, \alpha\right)\ldots P\left(x_n|z, \alpha\right) \cdot P(z) \cdot P(\alpha)d\alpha$$

While developing this model two fundamental assumptions were made:
1. z is independent of α i.e. average distance the criminal travels is independent of the criminal's anchor point.
2. Criminal's choice of crime sites are pairwise independent.
Now, what remains is deciding a model for offender behaviour $P(x|z, \alpha)$

3.2.3 Simple model for offender behaviour

The following model [5] take an assumption that the criminals or the offenders have the same average distance of committed offences ' α ' known in advance.
If it is assumed that the criminal chooses a target location based on the Euclidean distance from the location of offence to the offender's anchor point, then with (normal) bi-variate distribution [6] we get,

$$P(x|z, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2}|x - z|^2\right)$$

since there are two random variables involved, what we get is a two dimensional normal distribution with the mean at the anchor point and standard deviation $\sigma = \sqrt{\frac{2}{\pi}}\alpha$.
Suppose, offender's distance decay effect also follows bivariate normal distribution with mean z and variance σ^2, so

$$D(x|z, \alpha) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|x - z|^2}{2\sigma^2}\right)$$

The offender's travelling distance is the is the distance between the anchor point and the crime location, $r = |x - z|$, then its density function is

$$f(r|\sigma) = 2\pi r D(x|z, \alpha) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

rather than using σ^2 as the parameter, we will use α. Now average distance is given by

$$Er = \int_0^\infty r f(r|\sigma)dr = \int_0^\infty \frac{r^2}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

Integrating by parts we get,

$$Er = \int_0^\infty \exp\left(-\tfrac{r^2}{2\sigma^2}\right) dr = \sqrt{\tfrac{\pi}{2}}\sigma$$
$$\alpha = \sqrt{\tfrac{\pi}{2}}\sigma$$
$$\sigma = \sqrt{\tfrac{2}{\pi}}\alpha$$

Therefore by replacing σ with α, we get

$$P(x|z,\alpha) = \frac{1}{4\alpha^2}\exp\left(-\frac{\pi}{4\alpha^2}|x-z|^2\right)$$

Assuming that all the offenders have the same average distance 'α' we obtain the posterior distribution as a product of normal distributions,

$$P\left(z|x_1,x_2,x_3,\ldots,x_n\right) = \left(\frac{1}{4\alpha^2}\right)^n \exp\left(-\frac{\pi}{4\alpha^2}\sum_{n=1}^{n}|x-z|^2\right)$$

3.2.4 Shortcomings of Mike O' Leary model

There was a slight problem with Mike O'Leary's model. The problem was that he assumed that the criminal has only one anchor point. But in reality this rarely happens to be the case. An anchor point can be a home, a workplace, a park or for that matter sometimes all of them. So what we obtain is a probability surface, but we consider only one place as our anchor point.

4 Comparative Study

1. While the Criminal Geographic Targeting Model assigns a hit score to the concerned area, the Mike O Leary's Model provides the probability of the area being the criminal's anchor point.

2. The Criminal Geographic Targeting Model uses Manhattan Metric System (3.1) to calculate distance. Mike O Leary's model uses the concept of Euclidean distance given by the formula:

$$d(p,q) = \sqrt{(p_1-q_1)^2 + \ldots\ldots + (p_n-q_n)^2}$$

where $p = (p_1, p_2, \ldots p_n)$ and $q = (q_1, q_2, \ldots q_n)$ represents two points in Euclidean $n-space$

3. While Mike O Leary's Model restricts itself by assuming that the criminal has only one anchor point, the Criminal Geographic Targeting Model takes into account the possibility of multiple anchor points.

5 Summary

The discussed models provide a general framework of how spatial features of crime scenes can be used to determine the probable area of either criminal's residence or operation area. It explained how the concept of Bayesian statistics helped in assigning a particular area a probability, of it being criminal's anchor point. Though Geographic Profiling techniques on its own cannot solve crimes, its effective implementation helps in figuring out the hunting area, thus narrowing down the suspects and facilitating police investigation.
To deal with the inadequacies of the Criminal Targeting Model given by Kim Rossmo and Mike O' Leary's Model, researchers have modified the existing models by using the concept of Dirichlet Process Mixture Model which assigns probability scores to multiple anchor points. Other developments in this field include efforts to incorporate psychological aspects and processes to integrate the

concept of mental map with the spatial details of crime sites. Mental Map refers to the internal representation and perception of the world developed by individuals. This would lead to an enhanced understanding of spatial patterns followed by serial criminals.

References

[1] Hobbs N. T., Hooten, M.B., *Bayesian Models: A Statistical Primer for Ecologists*, Princeton University Press, 2015

[2] Lawson A. B., *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, 3rd edition, Chapman and Hall/CRC, 2018

[3] Carlin B. P., Louis, T. K., *Bayesian Methods for Data Analysis* 3rd edition, Chapman and Hall/CRC, 2008

[4] Rossmo D. K., *Geographic Profiling*, CRC Press, 2000

[5] O'Leary M., *A New Mathematical Approach to Geographic Profiling*, December 14, 2009

[6] Morrow J., *Yet Another Mathematical Approach to Geographic Profiling*, February 22, 2010

[7] Morrow J., *Why Crime Doesn't Pay: Locating Criminals Through Geographic Profiling*, February 22, 2010

(Mentor) Dr. Sucheta Nayak, Assistant Professor, Department Of Mathematics, Lady Shri Ram College For Women, New Delhi
suchetanayakk@gmail.com

Osheen Khare, B.Sc(H) Mathematics, 2nd Semester, Lady Shri Ram College For Women, New Delhi
osheenkhare2001@gmail.com

Sumaiya Ahmed, B.Sc.(H) Mathematics, 2nd Semester, Lady Shri Ram College For Women, New Delhi
sumaiyaahmedlsr@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (34-43)

Brent-Dekker Method: An Attempt to Decode

Dr. M. Singh, J. Sharma and S. Agrawal

Abstract

This paper discusses solving a root-finding problem with special emphasis on one of the most applicable methods used in many computer algebra systems and programming languages: "The Brent-Dekker Method". Its merits, demerits and comparison with other methods have also been discussed. Program codes for the latter in wxMaxima 17.05.0 are also incorporated in the paper.

Keywords: Root-finding Method, Approximation, Iteration, Convergence, Error Tolerance, Interpolation

1 Introduction

In mathematics, finding a root of an equation is a fundamental problem. A *root* of an equation means, an x satisfying the equation $f(x) = 0$, where f is a real-valued function. Such an x is also called a *zero* of f. We all are familiar with finding zeros of polynomials of small degree by factorization. However, in real-life situations, more complex equations like:

$$f(x) = x + e^x = 0;\ f(x) = 2x^5 - 10x + 5 = 0,$$

often arise [4]. Hence, we need different specialized techniques to evaluate their roots with maximum accuracy. In this paper, we shall be explaining some of these simple methods[1] and assessing how they can also be combined to create hybrid root-finding methods, in particular the **Brent-Dekker Method**. The next section briefly summarizes the basic terminology and concepts which are further being used in the paper.

2 Prerequisites

- **Iterative Method.** It is a technique to solve mathematical problems by successively following some mathematical argument(s) or concept(s) to generate a sequence of approximate solutions to the problem, by using an initial guess(es). The n^{th} approximation of the solution is derived from the previous iteration(s). Finally, an approximate solution of the problem (i.e., solution with some error) is obtained after a finite number of iterations.

[1] Most methods are applicable when f is algebraic or transcendental (involving trigonometric, hyperbolic, transcendental functions or their inverses).

- **Measuring Convergence Speed.**

  (i) **Rate of Convergence (RoC).** Let $\langle r_n \rangle$ be a sequence that converges to a number r. If there exists a sequence $\langle \beta_n \rangle$ converging to zero and a positive constant λ independent of n such that

  $$|r_n - r| \leq \lambda |\beta_n| \tag{1}$$

  for sufficiently large values of n, then, we say $\langle r_n \rangle$ converges to r with rate of convergence $O(\beta_n)$. A sequence with RoC $O(\frac{1}{n^2})$ converges slower than a sequence having RoC $O(\frac{1}{2^n})$ because $\langle \frac{1}{n^2} \rangle$ converges to zero slower than as $\langle \frac{1}{2^n} \rangle$ converges to zero.

  (ii) **Order of Convergence (OoC).** Suppose $\langle r_n \rangle$ is a sequence that converges to r. Let $e_n = r_n - r$, if there exist positive constants λ and α such that

  $$\lim_{n \to \infty} \frac{|e_{n+1}|}{|e_n|^\alpha} = \lambda,$$

  then the OoC of the sequence $\langle r_n \rangle$ is α with asymptotic error constant λ.
  The order of convergence measures the effectiveness with which each iteration reduces the error in approximation.
  An iterative method is said to be of order α if OoC of its generated sequence is α. When $\alpha = 1$, $\alpha = 2$, the OoC of iterative method is called linear, quadratic respectively.
  Note 1. Non-integral values of α are also possible.

- **Intermediate Value Theorem (IVT).** *Let f be a continuous real-valued function defined over the closed interval $[p, q]$ and let d be any real number that lies between the values $f(p)$ and $f(q)$. Then, there exists a real number c with $p < c < q$ such that $f(c) = d$.*

- **Root-finding Techniques.** Basic techniques for root-finding are iterative methods which fall into two categories:

  (i) **Simple Enclosure Methods.** They involve finding an interval in which root is guaranteed to exist and systematically shrinking its size, using IVT in each iteration. Convergence to root is guaranteed, but such methods have a slow rate of convergence.

  (ii) **Fixed Point Iteration Schemes.** They exhibit rapid convergence when properly constructed, but require stronger conditions to guarantee convergence to root.

- **Interpolation.** Suppose, we have a set of points (x_i, f_i) for $i = 0, 1, ..., k$, where f_i denotes the value of some function (usually unknown) f at x_i. We are interested in approximating the value of f at some x, where $x_i < x < x_{i+1}$, for some i, $0 \leq i < k$. To do this, we try to find a simpler function h that in some sense approximates the value. Determining the function h by enforcing the conditions $f(x_i) = h(x_i)$ at each x_i is known as Interpolation.
  There are various forms of interpolation. We shall be using Lagrange Interpolation, which interpolates the given set of k points using a polynomial having degree at most k. The Lagrange form of interpolating polynomial is given by

  $$P_k(x) = \sum_{i=0}^{k} L_{k,i}(x) f_i \quad \text{where} \quad L_{k,j}(x) = \prod_{i=0, i \neq j}^{k} \frac{x - x_i}{x_j - x_i}.$$

In particular, the Lagrange form of quadratic interpolating polynomial is given by

$$P_2(x) = \frac{x-x_1}{x_0-x_1}\,\frac{x-x_2}{x_0-x_2}\,f_0 + \frac{x-x_0}{x_1-x_0}\,\frac{x-x_2}{x_1-x_2}\,f_1 + \frac{x-x_0}{x_2-x_0}\,\frac{x-x_1}{x_2-x_1}\,f_2. \tag{2}$$

- **Error Tolerance.** Suppose r_n is the approximation of a root r of $f(x) = 0$ after the n^{th} iteration of an iterative root-finding technique. We choose a real number $\varepsilon > 0$ and terminate the iterations as soon as $|r_n - r| < \varepsilon$. However, most of the times, root is unknown in the problem, hence, we check $|r_n - r_{n-1}| < \varepsilon$. Such an ε is called the Error Tolerance.

Note 2. Throughout the paper, we will assume that f is a continuous real-valued function defined on $[p, q]$ and $f(x) = 0$ has at least one real root. We shall solve this equation with a given error tolerance ε.

3 Root-finding Methods

In this section, we shall discuss bisection, secant and inverse quadratic interpolation methods to find a root of $f(x) = 0$ for a continuous function f defined on $[p, q]$.

3.1 Bisection Method ([1]).

It is a simple enclosure method. We first find p and q such that $f(p)$ and $f(q)$ have opposite signs. Since zero lies between those values, therefore, a zero of f is guaranteed to exist in the interval $[p, q]$ by IVT. Let $p_1 = p$, $q_1 = q$ for the first iteration. Then, we systematically shrink the size of the interval in each iteration by using IVT to determine which half of the interval contains a root. In the n^{th} iteration, the enclosing interval is $[p_n, q_n]$ and the approximation of a root after n^{th} iteration is $m_n = \frac{p_n+q_n}{2}$. m_n is used as one of the endpoints for the next enclosing interval. The iteration is terminated if $|m_n - m_{n-1}| < \varepsilon$, where ε is error tolerance, otherwise, IVT is used to determine which of the two sub-intervals $[p_n, m_n]$ or $[m_n, q_n]$ contains a root and hence the next enclosing interval $[p_{n+1}, q_{n+1}]$ is obtained for finding a better approximation of a root.

For a continuous function f on the interval $[p, q]$ with $f(p)f(q) < 0$, the bisection method generates a sequence of approximations $\langle m_n \rangle$ which converges to a root $m \in (p, q)$ with the property $|m_n - m| < \frac{q-p}{2^n}$. By (1), the inequality suggests that the sequence of approximations generated by the bisection method has RoC $O(\frac{1}{2^n})$. It can also be verified that the OoC of bisection method is 1. This method requires one new function evaluation per iteration. Though bisection method guarantees convergence to a root, however, it does not guarantee its uniqueness. Also, the requirement that an interval $[p, q]$ be found such that $f(p)f(q) < 0$ implies that this method cannot be used to locate the roots of even multiplicity[2].

[2] A root r of the equation $f(x) = 0$ is said to be a root of multiplicity t if f can be written in the form $f(x) = (x - r)^t g(x)$, where $\lim_{x \to r} g(x) \neq 0$. A root is called multiple if $t \geq 2$.

3.2 Secant Method ([1]).

It is a fixed point iteration scheme. Two initial approximations, say p and q are chosen. For the first iteration, denote $p_0 = p$ and $p_1 = q$. The x-intercept of the secant joining $(p_0, f(p_0))$ and $(p_1, f(p_1))$ gives the next approximation of a root, say p_2. Repeating this process, approximation at n^{th} iteration is given by

$$p_{n+1} = p_n - \frac{p_n - p_{n-1}}{f(p_n) - f(p_{n-1})}\ f(p_n). \tag{3}$$

Iterations are terminated when $|p_{n+1} - p_n| < \varepsilon$, where ε is the error tolerance. Until then, we continue to approximate a root by the x-intercept of the secant joining $(p_{n+1}, f(p_{n+1}))$ and $(p_n, f(p_n))$.

The OoC of secant method is approximately 1.618 for a simple root[3] and is approximately 1 for a multiple root. Secant method usually converges faster than most of the simple enclosure methods. Also, it does not require the knowledge of the derivative of the function whose zero is being approximated. Thus, it requires only one new function evaluation per iteration. However, secant method diverges if we get $f(p_n) = f(p_{n-1})$ in any of the iterations. Also, for a multiple root, the initial guesses of a root influence the performance of the secant method. Hence, it is important to use accurate initial guesses.

3.3 Inverse Quadratic Interpolation (IQI) ([2]).

Suppose we have three points, viz. x_{n-2}, x_{n-1} and x_n with their function values, $f(x_{n-2}) = y_{n-2}$, $f(x_{n-1}) = y_{n-1}$ and $f(x_n) = y_n$, for some function $y = f(x)$. We can interpolate them quadratically, resulting in a parabola. The intersection of this parabola with the x-axis would represent the new root estimate. However, when the parabola has complex roots, it would not intersect the x-axis. Hence, we employ inverse quadratic interpolation, i.e., we fit the points with a parabola in y instead of using a parabola in x, by assuming that f has an inverse quadratic function I_2, then $x_{n-2} = I_2(y_{n-2})$, $x_{n-1} = I_2(y_{n-1})$ and $x_n = I_2(y_n)$. This results in creating a sideways[4] parabola, which always intersects the x-axis.
Using Lagrange quadratic interpolation (2), we get the inverse quadratic interpolation of f as

$$I_2(y) = x_{n-2}\frac{y - y_{n-1}}{y_{n-2} - y_{n-1}}\ \frac{y - y_n}{y_{n-2} - y_n} + x_{n-1}\frac{y - y_{n-2}}{y_{n-1} - y_{n-2}}\ \frac{y - y_n}{y_{n-1} - y_n} + x_n\frac{y - y_{n-2}}{y_n - y_{n-2}}\ \frac{y - y_{n-1}}{y_n - y_{n-1}}.$$

Substituting $y = 0$ in the above equation, results in the following recursion formula

$$x_{n+1} = x_{n-2}\frac{y_{n-1}\ y_n}{(y_{n-2} - y_{n-1})(y_{n-2} - y_n)} + x_{n-1}\frac{y_{n-2}\ y_n}{(y_{n-1} - y_{n-2})(y_{n-1} - y_n)} + x_n\frac{y_{n-2}\ y_{n-1}}{(y_n - y_{n-2})(y_n - y_{n-1})}. \tag{4}$$

This method requires one new function evaluation per iteration. Its OoC is approximately 1.839. For well-behaved functions[5], this method saves about 0.5 function evaluations per zero on an average, as compared to secant method. However, if by any chance, two of the function values y_{n-2}, y_{n-1} or y_n coincide, the algorithm fails completely. Thus, inverse quadratic interpolation is rarely used as a stand-alone algorithm [10].

[3] A root is called simple if it is of multiplicity one.
[4] A horizontal parabola, i.e., a parabola that opens to the right or to the left.
[5] Functions which are Lipschitz continuous, hence do not exhibit abrupt variations [12].

4 The Brent-Dekker Method

The bisection method converges slowly but convergence is guaranteed. The secant method is faster, but convergence is not guaranteed as it may involve division by zero in a computation. In 1969 [5], Dekker gave the idea to use secant method and switch to bisection when needed. Below, we explain Dekker's method.

4.1 Dekker's Method.

To find a root of an equation $f(x) = 0$ with a given error tolerance ε by Dekker's method, the following steps are to be executed:

- Find p and q, where q is a better guess of a root as compared to p, such that $f(p)f(q) < 0$ then IVT guarantees the existence of a root in the interval with endpoints p and q.
- For the first iteration, take $p_1 = p$, $q_1 = q$, $q_0 = p$. Then, proceed as follows:
  (i) If $f(q_1) = f(q_0)$, find the mid-point[6] of p_1 and q_1, i.e., $m = \frac{p_1+q_1}{2}$. Then, $q_2 = m$.
  (ii) If $f(q_1) \neq f(q_0)$, find the x-intercept of the secant joining the points $(q_0, f(q_0))$ and $(q_1, f(q_1))$, i.e., $s = q_1 - \frac{q_1 - q_0}{f(q_1) - f(q_0)} f(q_1)$. If s lies between q_1 and m, then $q_2 = s$, otherwise $q_2 = m$, where m is the mid-point of p_1 and q_1.
- If $f(p_1)f(q_2) < 0$, then $p_2 = p_1$, otherwise $p_2 = q_1$. Now, we have three points, namely, q_2: current guess of a root, q_1: previous guess of a root, and p_2: a point such that $f(p_2)f(q_2) < 0$, so that IVT guarantees the existence of a root in the interval with endpoints p_2 and q_2.
  Caution: Do not perceive that $p_2 < q_2$.
- If $|f(p_2)| < |f(q_2)|$, then, interchange the values of p_2 and q_2 to make q_2 a better guess of a root.
- Now, proceed in a similar manner for subsequent iterations. For the convenience of the reader, we are giving the n^{th} iteration. Suppose at the n^{th} iteration, we have three points, q_n: current guess of a root, q_{n-1}: previous guess of a root, and p_n: a point such that the interval with endpoints p_n and q_n contains a root.
  (i) If $f(q_n) = f(q_{n-1})$, then $q_{n+1} = m = \frac{p_n+q_n}{2}$, i.e., mid-point of p_n and q_n.
  (ii) If $f(q_n) \neq f(q_{n-1})$, find the x-intercept of the secant joining the points $(q_{n-1}, f(q_{n-1}))$ and $(q_n, f(q_n))$, i.e., $s = q_n - \frac{q_n - q_{n-1}}{f(q_n) - f(q_{n-1})} f(q_n)$. If s lies between q_n and m, then $q_{n+1} = s$, otherwise $q_{n+1} = m$.
- If $f(p_n)$ and $f(q_{n+1})$ have opposite signs, then $p_{n+1} = p_n$, otherwise $p_{n+1} = q_n$.
- If $|f(p_{n+1})| < |f(q_{n+1})|$, then the values of p_{n+1} and q_{n+1} are exchanged.
- We terminate the iterations when $f(q_{n+1}) = 0$ or $|q_{n+1} - q_n| < \varepsilon$.

Dekker's method, being a combination of the bisection and secant methods, is expected to have OoC between 1 and 1.618. Dekker's method is at least as good as bisection method. Also, it takes care of the drawbacks of secant method. For well-behaved functions, it exhibits fast convergence. However, sometimes convergence may be very slow and in rare cases may even be worse than linear.

[6] This is a case in general, but it will never be true for the first iteration in view of $f(p)f(q) < 0$.

4.2 Brent's Method.

In 1971, Brent further modified it to ensure better approximation of a root by combining IQI and Dekker's method, with a guarantee of convergence at least as fast as bisection method. It is named as Brent's Method. Since it was built heavily on Dekker's method, it is often called Brent-Dekker Method [2],[3].

To find a root of an equation $f(x) = 0$ with a given error tolerance ε by Brent's method, the following steps are to be executed:

- Similar to Dekker's method, each iteration of Brent's method involves three points: p_n, q_n and q_{n-1}. Here, if $|f(p_n)| < |f(q_n)|$, then, exchange the values of p_n and q_n so that q_n is a better guess than p_n. Now, after the exchange, if $q_{n-1} = q_n$, then, change the value of q_{n-1} by taking $q_{n-1} = p_n$.
- Now, let us see how q_{n+1} is computed in the n^{th} iteration of Brent's method.
  - (i) If $p_n = q_{n-1}$, then q_{n+1} is computed by secant method, i.e., by (3), for the points p_n and q_n.
  - (ii) If $p_n \neq q_{n-1}$, then q_{n+1} is computed by IQI, i.e., by (4), for the points p_n, q_n and q_{n-1}.
- However, to fix the final value of q_{n+1} obtained either from step (i) or (ii), further check the following inequalities (skip this step for the first iteration):
  - (i) If q_n is computed by bisection method, check

$$|q_{n+1} - q_n| < \left|\frac{q_n - q_{n-1}}{2}\right|,$$

    if this inequality is satisfied, then q_{n+1} remains unchanged, otherwise $q_{n+1} = \frac{p_n+q_n}{2}$.
  - (ii) If q_n is computed either by secant method or by IQI, check

$$|q_{n+1} - q_n| < \left|\frac{q_{n-1} - q_{n-2}}{2}\right|, \tag{5}$$

    if this inequality is satisfied, then q_{n+1} remains unchanged, otherwise $q_{n+1} = \frac{p_n+q_n}{2}$.
- Then p_{n+1} is computed in the same manner as in Dekker's method. We terminate the process when $f(q_{n+1}) = 0$ or $|q_{n+1} - q_n| < \varepsilon$.

Note 3: (5) can be verified only from second iteration onwards.

As Brent's method is a blend of bisection, secant and IQI methods, it is expected to have OoC between 1 and 1.839. The advantage of Brent's method is that it prevents divergence of the algorithm, and it usually converges faster than Dekker's method, especially for well-behaved functions.

5 Comparison Amongst the Aforementioned Methods

In this section, we give an example by which we attempt to compare all the above-discussed methods, namely, bisection, secant, IQI, Dekker's and Brent's method. To compare the methods, we approximate the zero of $f(x) = 2x^5 - 10x + 5$ with an error tolerance $\varepsilon = 0.00001$. From Figure 1, we can see that $f(x) = 0$ has three roots. However, in the following example, we shall be estimating the root 0.507 (as given approximately in the graph). Accordingly, we choose $p = 0$ and $q = 1$.

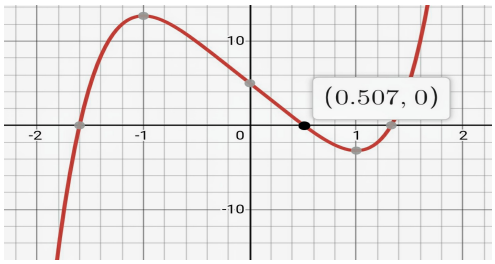


Figure 1: Graph of $f(x)$ plotted on Desmos Graphing Calculator.

Following are the program codes for the four prerequisite methods written by us in **wxMaxima 17.05.0**. For Brent's Method, the **Brentq** function in **python** is used [9].

5.1 Bisection Method.

```
bisec(a1,b1,n,e) := block([flag], flag:1, if (float(f(a1)*f(b1)>0))
then print("change values", return(""))
else for i:1 thru n do (if i>1 then m:p1, p1:(a1+b1)/2,
if f(a1)*f(p1)<0 then b1:p1 else a1:p1, print(float(p1)),
if (abs(p1-m)<e or f(p1)=0) then (flag:0, return(p1))),
if flag=1 then print(i, "increase n"))$
```

5.2 Secant Method.

```
secant(x1,x2,n,e) := block(flag:0, p[-1]:x1, p[0]:x2, for i:1 thru n do
(p[i]:p[i-1]-f(p[i-1])*(p[i-1]-p[i-2])/(f(p[i-1])-f(p[i-2])), print(float(p[i])),
if (abs(p[i]-p[i-1])<e or f(p[i])=0) then (flag:1, return(p[i]))),
if flag=0 then print("increase n"))$
```

5.3 Inverse Quadratic Interpolation.

```
‘sum = sum:0;
iqi3(x0,x1,x2) := block(sum:0, x[0]:x0, x[1]:x1, x[2]:x2,
L(i) := product(if k=i then 1 else (-f(x[k]))/(f(x[i])-f(x[k])), k, 0, 2),
for j:0 thru 2 do (sum:x[j]*L(j)+sum))$
iqi(x0,x1,x2,n,e):= block(flag:0, for i:1 thru n do (iqi3(x0,x1,x2),
s:float(sum), print(s), if (i>1 and min(abs(s-x0), abs(s-x1), abs(s-x2))<e)
then (flag:1, return("")), x0:x1, x1:x2, x2:s),
if (flag=0) then print("increase n"))$
```

5.4 Dekker's Method.

```
dekker(p,q,n,e) := block([flag,s], flag:1,
if float(f(p)*f(q))>0 then (print("change values"), return("")) else
(q[-1]:p, q[0]:q, p[0]:p,
for i:1 thru n do (if q[i-2]=q[i-1] then
q[i]:float((p[i-1]+q[i-1]))/2 else
(s:float(q[i-1]-f(q[i-1])*(q[i-1]-q[i-2])/(f(q[i-1])-f(q[i-2]))),
m:float((p[i-1]+q[i-1]))/2,
if ((q[i-1]<s and s<m) or (m<s and s<q[i-1])) then q[i]:s else q[i]:m),
if (f(q[i])*f(p[i-1])<0) then p[i]:p[i-1] else p[i]:q[i-1],
if abs(f(p[i]))<abs(f(q[i])) then (x:p[i], p[i]:q[i], q[i]:x),
print(i,q[i]), if(abs(q[i]-q[i-1])<e or f(q[i])=0) then
(flag:0, print("root",q[i]), return("")),
if(flag=1) then print("increase n")))$
```

5.5 Brent's Method.

```
!pip install pyroots
# define the function whose root you are searching
def f(x,a):
   print(x)
   return (2*x**5-10*x+5)+a
from pyroots import Brentq
brent = Brentq(epsilon=1e-5)
# solve the function in '[a, b]' while 'a' is equal to 0
result = brent(f, 0, 1, a=0)
print(result)
```

Table 1: Comparison of performance of five root-finding methods for $f(x)$ tabulated on Microsoft Excel

	BISECTION	SECANT	IQI	DEKKER	BRENT
2	0.500000000000	0.625000000000000	0.5758304195804	0.625000000000000	0.625000000000000
3	0.750000000000	0.420322671950184	0.4702674940839	0.420322671950184	0.456103835670199
4	0.625000000000	0.509816351625323	0.5068283422927	0.509816351625323	0.508653943248839
5	0.562500000000	0.506741582702468	0.5066784384865	0.506741582702468	0.506703699540827
6	0.531250000000	0.506678666036458	0.5066787213669	0.506678666036458	0.506678719744678
7	0.515625000000	0.506678721366829		0.506678721366829	
8	0.507812500000				
9	0.503906250000				
10	0.505859375000				
11	0.506835937500				
12	0.506347656250				
13	0.506591796875				
14	0.506713867188				
15	0.506652832031				
16	0.506683349609				
17	0.506668090820				
18	0.506675720215				

Table 1 shows the iteration-wise outputs generated by five root-finding methods using the above program codes. It is evident that Brent's method is at least as fast as other methods in approximating a root with the given error tolerance.

6 Conclusion

Brent's Method is widely used in computer algebra systems and programming languages to approximate a root of any function, for example, `fzero` function in MATLAB [7], SciPy optimize module in Python [8] and `uniroot` function in R [11], to name a few.

To get a faster algorithm, we have seen that Brent used an amalgamation of bisection, secant and IQI methods. During the process of research, we thought that the next idea could be the blend of Newton-Raphson method [2] with bisection, as Newton-Raphson method is faster than secant. Working in this direction, we found that some related literature is already available on this idea [6]. We are really enthusiastic about continuing the research further.

Acknowledgement

The authors thank Prof. R. P. Brent for his sincere efforts in helping them write this paper.

References

[1] Bradie B., *A Friendly Introduction To Numerical Analysis*, Pearson, 2007.

[2] Chapra S. C., *Applied Numerical Methods with MATLAB*, Third Edition, McGraw-Hill Education, 2012.

[3] Brent R. P., *An algorithm with guaranteed convergence for finding a zero of a function*, The Computer Journal, Volume 14, Issue 4, 1971, pp. 422–425.

[4] Chow T. Y., *What is a Closed-Form Number?*, American Mathematical Monthly 106, No. 5, 1999, pp. 440-448.

[5] Dekker T. J., *Finding a zero by means of successive linear interpolation*, Constructive aspects of the fundamental theorem of algebra, 1969, pp. 37-48.

[6] Kim J., Noh T., Oh W., Park S., Hahm N., *An improved hybrid algorithm to bisection method and Newton-Raphson method*, Applied Mathematical Sciences, Volume 11, No. 56, 2017, pp. 2789-2797.

[7] Moler C., Zeroin, Part 1: Dekker's Algorithm, *MathWorks*, October 12, 2015, https://blogs.mathworks.com/cleve/2015/10/12/zeroin-part-1-dekkers-algorithm.

[8] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.brentq.html.

[9] *GitHub*, https://github.com/pmav99/pyroots/blob/master/pyroots/brent.py.

[10] Mas M., Inverse quadratic interpolation in Julia, *GitHub*, Jun 27, 2016, https://mmas.github.io/inverse-quadratic-interpolation-julia.

[11] https://www.uni-muenster.de/IT.BennoSueselbeck/s-html/helpfiles/uniroot.html.

[12] https://users.wpi.edu/~walker/MA500/HANDOUTS/LipschitzContinuity.pdf.

(Mentor) Dr. Monika Singh, Assistant Professor, Department Of Mathematics, Lady Shri Ram College for Women, New Delhi
monikalsr05@gmail.com

Jaya Sharma, B.Sc.(H) Mathematics, 6th Semester, Lady Shri Ram College for Women, New Delhi
forjayasharma@gmail.com

Shagun Agrawal, B.Sc.(H) Mathematics, 6th Semester, Lady Shri Ram College for Women, New Delhi
shagunsarita@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (44-52)

Mathematics behind Blockchain

Mr. Y. Singh, P. Ganguly and Y. Jain

Abstract

This paper is a general exploration of blockchain, particularly in Bitcoin, and the mathematics behind its working and the challenges it has to overcome to evolve as the new high technology.

Keywords: Blockchain, Asymmetric Cryptography, Hash Functions, Digital Signatures, Bitcoin

1 Introduction

Blockchain is a peer-to-peer network technology, which works computationally using mathematical algorithms. It is a decentralised, distributed ledger that is used to maintain a continuously growing list of records, called blocks. It stores all the transactions in a secure, verifiable and a transparent way.

The features of this technology makes it distinctive from the conventional methods of maintaining a ledger which is usually under the governance of a central entity. Information added to a block is simultaneously accepted and viewed by all the users ensuring cross border transparency. This distributed approachability ensures trusted and secured transactions. The data stored in the blocks is also immutable. The time-stamping of data, decentralisation and cryptographic hashes make hacking impossible. Once information is secured in a block, it is highly immune to tampering. When data is meddled with, all the other nodes in the network get alarmed, which immediately prevents or blocks the hacker.

Blockchain also provides a clear show of all information entered, yet with the desired permissibility. The transactions always appear in a public key format which is traceable, yet the identity is not directly revealed during the presence or addition of a transaction. All these essential functions are possible due to the presence of mathematics. This paper focuses on mathematics behind the working of Bitcoin which is the most common example that is intrinsically linked to the blockchain technology.

2 Pre-requisite Knowledge

The following technical terms are necessary to understand the working of a blockchain:

- **Bitcoin**: Bitcoin is a crypto-currency developed by Satoshi Nakamoto [9] in the year 2008, which is based on the mechanism of blockchain. It was one of the first applications of blockchain.

- **Block**: The block is the main component in the blockchain to store data. It consists of the previous hash, timestamp, data, nonce and current hash.
- **Nonce**: Nonce is short for *number only used once*. The nonce is required to be attached to a hash to secure a block. It is generally a particular number of zeroes which are supposed to be added to the solved hash function. Solving the nonce requires a trial and error method.
- **Timestamp**: A timestamp records the exact time at which the transactions take place, and are recorded into the blocks.
- **Node**: A node can be defined as an individual player in a distributed system. Nodes form the infrastructure of a blockchain. They spread, store and preserve blockchain data.
- **Mining**: The process of mining is the main activity which uses cryptographic algorithms to construct blocks in a blockchain mechanism.
- **Miners**: Miners are people with high mathematical knowledge who help in the process of mining.
- **Encryption**: Conversion of a plain message into cipher text or a coded form.
- **Decryption**: Conversion of a cipher text to a readable format, that is, converting it into the original message.
- **Private Key**: It is a randomly generated number which is held privately by the users in a blockchain. No authorised access should be granted to the private key. It is used to decrypt messages.
- **Public Key**: It is freely available and published by the owner of the private key. Any message which has to be sent to the owner of the private key can be encrypted using the published public key.
- **Euler's Phi Function**: For positive integer n, let $Z_n = \{0, 1, ..., (n-1)\}$ be a subset of integers. The number of integers in the set whose *greatest common divisor* (gcd) with n is 1, called value of the Euler's phi function at n and is presented by $\phi(n)$. If n, m are positive integers such that $gcd(m, n) = 1$ then $\phi(mn) = \phi(m)\phi(n)$. Also, if p is prime then $\phi(p) = p-1$.
- **Group**: A group, symbolically presented as $(G, *)$, is a non- empty set G of elements together with a *binary operation*[1] '*' which satisfies three properties, namely Associative Property, Existence of Identity and Existence of Inverse [13].
- **Order of an element of a Group**: The order of an element g of a group $(G, *)$ is the smallest positive integer n such that $g * g * ..... * g$ (n times) = identity of the group. When binary operation is '+' and order of g is n, we write ng = identity of group $(G, *)$ [13].
- **Cyclic Group**: A group $(G, *)$ is said to be cyclic if there exists an element g in $(G, *)$ which can generate[2] all the elements of $(G, *)$ [13].

[1] Binary operation on a set '$\mathbf{A}$' is a function from $\mathbf{A} \times \mathbf{A}$ to $\mathbf{A}$.
[2] g is called the generator of G, and $G = \{ng \mid n \in \mathbf{Z}\}$.

- **Finite field**: A finite field is a finite set with two binary operations '+' & '×'. This algebraic structure satisfies certain properties and all calculations (broadly speaking) are supposed to be done in the field. In this work, we are dealing with finite field $(Z_p, \bigoplus_p, \bigotimes_p)$ where $Z_p = \{0, 1..., (p-1)\}$, p is prime, $\bigoplus_p$ is addition modulo p and $\bigotimes_p$ is multiplication modulo p [13].

3 Mathematics behind Blockchain

Cryptography, meaning hidden codes, is the science which uses mathematical algorithms to encrypt and decrypt messages for safe communication between individuals. The primary functions of cryptography are key exchange, non-repudiation, authentication, confidentiality and integrity of data. In the following paragraphs, we explain the concepts of public-key cryptography and the two most vital used concept of cryptography - hashing and digital signatures. We use the names Alice and Bob, the names generally used by mathematicians, to indicate two individuals sharing information over a network.

In cryptography, we start with the unencrypted data, known as plaintext. Plaintext is then encrypted into ciphertext, which in turn is decrypted into plaintext again. It can also be written as $C = E_k(P)$; $P = D_{k'}(C)$, where P: plaintext, C: ciphertext, E: encryption method, D: decryption method, and k and k' are encryption and decryption keys respectively.

3.1 Hash Function

A hash function converts a message of any arbitrary length to a fixed length. It formulates a fingerprint of some data. In case the data is changed or altered, the hash output immediately changes. The fingerprint is termed as *message digest*. Let h be the hash function of a message m of arbitrary length, then the corresponding message digest y is, $y = h(m)$. If the message is even slightly changed from m to m', then a new message digest is generated, $y\prime = h(m')$. Further, $y \neq y'$ can be checked to verify the alteration of the input. Thus, a keyed hash function also known as a *Message Authentication Code* (MAC) family, consists of four important components: (M, Y, K, H), where M is a set of possible messages, Y is a finite set of possible message digests, K is a finite set of possible keys and H is set of hash functions. For each $k \in K$, there is a hash function $h_k \in H$, where $h_k : M \to Y$.

The security of hash functions largely depends on the following three properties:

1. **The One Way Property/The Pre-image Resistant Property**

   Given a hash function $h : M \to Y$, and an element $y \in Y$, for a given possible message digest y, the problem of pre-image lies in computing an $m \in M$, such that $y = h(m)$. Finding the pre-image 'm' for a given 'y' is computationally infeasible. Such resistance of finding the pre-image is known as the *one way property*/the *pre-image resistant property* of hashing.

2. **Weak Collision Resistance/The Second Pre-image Resistant Property**

   Given a hash function $h : M \to Y$, and $m \in M$, for a given message m, the problem of second pre-image lies in computing a second message m', where $m \neq m'$, such that $h(m) = h(m')$. If this is valid, then a pair (m', h (m)) is generated. For security purposes this problem should be difficult to solve, otherwise the hashes of two messages will collide with each other. This situation is called the *weak collision resistance*/the *second pre-image resistant property* of hashing.

3. **The Property of Strong Collision Resistance**

   Given a hash function $h : M \to Y$, the problem of collision resistance lies in finding $m, m' \in M$, and $m \neq m'$, such that $y = h(m) = h(m')$. If this is valid, then two pairs are generated, (m, y) and (m', y). Again, such calculation is computationally infeasible and insecure. Therefore the prevention of such collision is known as the *property of strong collision resistance*.

   For example, in Bitcoin, **SHA 256**[3], of the SHA-2 (Secure Hash Algorithm 2) family, is used to secure the blocks. For example, the text 'I am writing a paper' is converted into the hash SHA 256 as: $8c4ed7e15026cb0f3620cb7e2c691e7ba8f312fb737895a8795c76508b72f2b8$, which is a 256-bit, 64 digit long hash value.

3.2 Digital Signatures

Digital signatures bind the data to its source. It assures the identity of the recipient and the sender. The sender of the message has sole ownership over his/her digital signature, and every user possesses a unique signature. The signature algorithm consists of two parts - a *signing algorithm* and its *verification mechanism*.

If Alice wants to send a message m to Bob, she will sign it with a signing algorithm S_k (say), where k is the secret private key shared between both of them. The output $S_k(m)$ can be verified with a public verification algorithm V_k (say). A signature pair (m, S), where m is the input message and S the signature, the verification algorithm answers in 'True' or 'False', after checking the validity of S with respect to m. Digital signatures can be summarised as containing the five important parts (M, D, K, S, V), where M is a finite set of messages, D is a finite set of signatures, K is a finite set of keys. S and V are sets of signing algorithms and verification algorithms respectively. For every $k \in K$, there exists a signing algorithm $S_k \in S$, and a verification algorithm $V_k \in V$, where $S_k : M \to D$ and $V_k : M \times D \to \{True, False\}$, where

$$V_k(m, d) = \begin{cases} True & d = S_k(m) \\ False & d \neq S_k(m) \end{cases}$$

for all $m \in M$ and $d \in D$. The pair (m, d) is called the *signature message*.

For a message m, only Alice (sender) has the access to produce a signature pair (m, S), where S is Alice's signature for which $V_k(m, S) = True$. If a third party computes the same pair (m, S) where S was not signed originally by Alice, then such a situation is called *forgery of the signature*.

[3]SHA-256 is a cryptographic hash function which converts a message of any bit size to a 256-bit long hash code.

3.3 Types of Cryptography

There are two types of cryptography - symmetric and asymmetric cryptography. Since blockchain is based on the idea of asymmetric cryptography, a prime focus is given to various types of asymmetric cryptography in this section.

Asymmetric cryptography or public-key cryptography was developed by Whitfield Diffie, Martin Hellman and Ralph Merkle [12]. A public-key cryptosystem is an advancement from the symmetric one-key system. It is based on the idea of two keys to make communication safer and more secure. In an asymmetric system, the receiver of the message owns two keys, a public key: k_{public} which is shared with everyone and a private key: $k_{private}$ which is kept by himself/herself. Any message which is to be sent to an individual, is first encrypted using their public key k_{public} and then decrypted by the receiver using his own private key $k_{private}$. An important thing to note in this system is the use of a **one-way encryption function**; that is, once the message gets encrypted, it is highly infeasible for anyone else to decrypt it, unless the private key of the receiver is being used.

Furthermore, there are majorly three types of public-key crypto algorithms which are used in blockchain, but the discussion is restricted to only two major cryptosystems, namely, RSA and elliptical curves.

3.3.1 RSA System/the Integer Factorisation Method:

The RSA system short for Rivest – Shamir – Adleman [2], was introduced in the year 1977. It is also known as the *integer factorisation method* since it is based on the idea that two large prime numbers can be multiplied easily, but factoring the product is computationally difficult. Under the RSA system, the steps to derive keys are as following:

Step 1: For generating a public key, two very large prime numbers a and b (say) are taken, and their product $c = ab$ is computed. Then, the Euler's phi-function is calculated as $\phi(c) = (a–1)(b–1)$.

A random number (say) 'e' is then selected, where $e \in \{1, 2, \ldots., \phi(c-1)\}$, such that gcd $(e, \phi(c)) = 1$. The public key generated is the pair (e, c).

Step 2: The private key d is derived from the public key, and the equation of deriving 'd' is

$$d\ e\ \equiv\ 1\ mod\ \phi(c). \tag{1}$$

Suppose Alice wants to send message m to Bob. To do this, she will use the public key of Bob (say) $(e_{\mathrm{B}}, c_{\mathrm{B}})$. The encrypted message will be r, obtained by the following congruence equation

$$m^{e_{\mathrm{B}}} \equiv r \quad mod\ c_{\mathrm{B}}.$$

Bob will decrypt the message by using his own private key (say) d_{B} with the congruence equation

$$r^{d_{\mathrm{B}}} \equiv m \quad mod\ c_{\mathrm{B}}.$$

One may wonder, how does $r^{d_B}\ mod\ c_{\mathrm{B}}$ give the original message m? The answer is, Euler's Theorem and equation (1) are used in the above calculations, which makes it possible to obtain m.

Now, the next task is to make the message authentic. That will be done by Alice's digital signature (say) S. For that, Alice will use her own private key (say) d_{A}. To compute S, following congruence equation is used $S \equiv m^{d_{\mathrm{A}}} \ mod\ c_{\mathrm{A}}$. Bob will verify Alice's signature by using Alice's public key $(e_{\mathrm{A}}, c_{\mathrm{A}})$, that is, by $m \equiv S^{e_{\mathrm{A}}} \ mod\ c_{\mathrm{A}}$.

3.3.2 Elliptical Curve Cryptography

The concept of elliptical curves has been added to the public-key cryptosystem family much later than the RSA. Blockchain uses the elliptical curve digital signature algorithm which is why a brief explanation of elliptical curves and its working over finite fields is given below, but for deeper knowledge one can look into [4], [11].

Elliptical Curves over Real Numbers

A non-singular[4] elliptical curve is defined to be a set of points $(x, y) \in \mathbf{R} \times \mathbf{R}$, where $\mathbf{R}$ is a set of real numbers which satisfies the equation:

$$y^2 = x^3 + ax + b \tag{2}$$

along with a special point $\mathbf{O}$ called the *point of infinity*[5]. The constants in equation (2), that is 'a' and 'b', should satisfy the equation $4a^3 + 27b^2 \neq 0$, where $a, b \in \mathbf{R}$

Elliptical Curve over Finite Field ($\mathbf{Z}_p$)

Elliptical curves in $\mathbf{R}$ can often cause an error due to the presence of irrational numbers, and storing it efficiently in the memory of a system becomes a problem. Therefore the concept of finite field is used to construct elliptical curves. Elliptical curve over finite field $\mathbf{Z}_p$ is defined to be a set of points $E = \{(x, y)|(x, y) \in \mathbf{Z_p} \times \mathbf{Z}_p\} \cup \{\mathbf{O}\}$ lying on the curve satisfy the equation

$$y^2 \equiv x^3 + ax + b \ (mod\ p)$$

along with a special point $\mathbf{O}$, where p is prime, $p \neq 2, 3$, a and b are constants in $\mathbf{Z}_p$. The curve is assumed to be non-singular, that is $4a^3 + 27b^2 \not\equiv 0 \ mod\ p$. There are two important properties that elliptical curves posses, namely, *point addition* and *point doubling*. The formulas are discussed as follows:

Given two points on the curve, (say) $A(x_1, y_1)$ and $B(x_2, y_2)$, addition of the points A and B, when $A \neq B$, gives a third point C(x_3, y_3), given by

$$x_3 \equiv t^2 - x_1 - x_2 \ (mod\ p)$$

$$y_3 \equiv t(x_1 - x_3) - y_1 (mod\ p)$$

where t is the slope of the line joining points A and B. When A = B, the sum is called *point doubling*, in this case C(x_3, y_3), is given by

$$x_3 \equiv ((3x_1 + a)/2y_1)^2 - 2x_1 (mod\ p)$$

$$y_3 \equiv ((3x_1 + a)/2y_1)(x_1 - x_3) - y_1 (mod\ p).$$

[4] A non-singular curve E is the curve that has no cusps or self-intersections.

[5] Special point $\mathbf{O}$ has been added because it will play the role of identity of the group as E forms a group w.r.t some operation.

There are certain parameters which are important for key generation for the signature and verification algorithm. These are as follows:

1. Bounds for the cardinality of elliptic curve E over $\mathbf{Z}_p$ are given by Hasse's theorem. According to this theorem $p+1-2\sqrt{p} \leq C(E) \leq p+1+2\sqrt{p}$, where $C(E)$ denotes cardinality of E over $\mathbf{Z}_p$. **Hasse theorem** gives the upper and lower bounds of the cardinality of the curve, whereas for counting the exact number of points in an elliptical curve over a finite field, **Schoof's theorem** is used. One may refer to [3] for a detailed study.

2. Elliptic curve E over $\mathbf{Z}_p$ is always either a cyclic group with identity $\mathbf{O}$ or the product of two cyclic groups with respect to point addition [14].

Elliptic Curve Digital Signature Algorithm (ECDSA)

Let $\mathbf{O} \neq g \in E$ and order of g is n, that is if $g = (x_0, y_0)$ then $g + g + g$ .....+ $g = \mathbf{O}$.
For key generation using elliptical curves, the following steps are implemented:

Step 1: A random integer 'd' is selected from the set {1,2,....,(n-1)}, it is the desired private key.

Step 2: Then $Q = dg$ is computed.

Q is the desired public key. The computation of an ECDSA uses the above parameters along with a hash function h to encrypt message.

The following steps will be followed to derive a signature pair for message m:

Step 1: Let a random integer, (say) i be selected such that $1 \leq i \leq n$–1 and $gcd(i, n) = 1$.

Step 2: Let the product $ig = (x, y)$.

Step 3: Let $r \equiv x \ mod \ n$, if r turns out be 0, go back to Step 1 and choose other i.

Step 4: Compute $i^{-1} \ mod \ n$.

Step 5: Then compute $S \equiv i^{-1}(h(m) + dr) \ mod \ n$, if $S = 0$, go back to Step 1.

Step 6: The desired signature of message m is the pair (r, S).

Now the signature can be verified by the steps given below:

Step 1: First, it needs to be verified that r and S lie in the set {1,2,....,(n-1)}.

Step 2: $w \equiv S^{-1} \ mod \ n$ is computed.

Step 3: $u_1 \equiv h(m)w \ mod \ n$ and $u_2 \equiv rw \ mod \ n$ are computed.

Step 4: Compute $T = u_1 g + u_2 Q$, where $g \in E$ and Q is public key, (say) $T = (x_1, y_1)$.

Step 5: If $T = \mathbf{O}$ then the signature gets rejected, else calculate $v \equiv x_1 \ mod \ n$, then the signature is valid if and only if $v = r$.

4 Conclusion

Bitcoin uses the Elliptic Curve Digital Signature Algorithm (ECDSA) to sign transactions. The process of hashing and verification of signatures to provide security is performed by the miners. Once data is transmitted by a person, the miner nodes in the network start the process of creating blocks which require solving hash functions to encrypt a block. Along with the hash code of previous block, a time stamp, a guessed nonce attached in front of a hash, and with the imputed information a block is created. Proof of work and authenticity is ensured by the miners by solving such signature verification algorithms.

Blockchain provides a secure way to transfer digital assets without an intermediary. It can process title deeds, smart contracts and facilitate transactions. It can also solve many problems discovered in early attempts at online voting. In the healthcare sector, it plays an important role of decentralising research data and reducing administrative overheads. Blockchain can also merge all the sources of personal identification into one open source, and secure record. This would reduce theft and loss of documentation while protecting the individuals right to privacy. However, it is extremely essential for the network to be motivated to work under ethical standards. Once, and if and only if, these standards are adhered to, Blockchain technology could reach its potential to improve businesses, democratising the global economy, and helping support more open and fair societies. Thus, it is clear that the world is ready for blockchain but. But the question is, are we?

References

[1] Menezes A.J., Van Oorschot P. C. and Vanstone S. A., *Handbook of Applied Cryptography.*, CRC Press, Boca Raton, Florida, USA, 1997.

[2] Rivest R.L, Shamir A., and Adleman L., *A method for obtaining digital signatures and public-key cryptosystems.*, Communications of the ACM, 21(2):120–126, February 1978.

[3] Schoof R., *Elliptical curves over finite fields and the Computation of square roots mod P, Mathematics of Computation*, American Mathematical Society, 1985.

[4] Silver J.H, *The Arithmetic of Elliptical Curves,*, The university of Michigan, CRC-Press, 1995.

[5] Pelzl J. and Paar C., *Understanding Cryptography: A Textbook for Students and Practitioners*, Springer Science and Business Media, 2009.

[6] Koblitz N., *Introduction to Elliptical curves and Modular forms*, University of Michigan, Springer-Verlag, 1984.

[7] Hankerson D., Vanstone S. and Menezes A.J., *Guide to elliptic curve cryptography*, Springer Science and Media, 2006.

[8] Blake I., Seroussi G. and Smart N. P., *Elliptical curves in cryptography*, Cambridge University press, 1999.

[9] Nakamoto S., *Bitcoin:A Peer-to-Peer Electronic Cash System*, 2008.

[10] Kikwai B.K., *Elliptic Curve Digital Signatures and Their Application in the Bitcoin Cryptocurrency Transactions*, International Journal of Scientific and Research Publications, Volume 7, Issue 11.

[11] Koblitz N., *Elliptic curve crypto systems*, American Mathematical Society, 1987.

[12] Diffie W., Hellman M. E., *New directions in Cryptography*, IEEE Transactions on Information Theory, 1976.

[13] Gallian.A.J., *Contemporary Abstract Algebra*, University of Minnesota, Houghton Miffin Company, 1998.

[14] Silverman.H.J., *An Introduction to the Theory of Elliptic Curves*, Brown University, NTRU Cryptosystems, Inc., 2006

(Mentor) Mr. Yograj Singh, Assistant Professor, Department of Mathematics, Lady Shri Ram College for Women, New Delhi
yograjchauhan26@gmail.com

Priyana Ganguly, B.Sc.(H) Mathematics, 2nd Semester, Lady Shri Ram College for Women, New Delhi
priyanaganguly@gmail.com

Yashita Jain, B.Sc.(H) Mathematics, 2nd Semester, Lady Shri Ram College for Women, New Delhi
yashitajain2155@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (53-60)

A Mathematical Approach to Global Positioning System

Dr. S. Nayak and M. Chitranshi

Abstract

Global Positioning System (GPS) is a satellite-based radio navigation system formed by 24 satellites orbiting the earth and their corresponding receivers on the earth. This utility detects and transmits time and location based information through a GPS receiver, located anywhere on or near the earth's surface. This paper discusses the mathematics involved in the tracking of a GPS signal.

Keywords: GPS, Pseudo-ranges, Gaussian Elimination

1 Introduction

The Global Positioning System (GPS) is a functional Global Navigation Satellite System. A receiver's location, movement, speed etc. can be retrieved by the signals transmitted by satellites orbiting our earth. At present, 24 such satellites are orbiting around the earth. They are spaced 19,300 kilometres above the earth's surface. There are six orbits at 60° around the earth and 55° at the Equator [5]. Satellites are evenly spread out so that four satellites are accessible via direct line of sight from anywhere on the globe. GPS is used in various realms of our day to day life [1]. For example, navigation, land surveying, cartography, estimating travel and - more recently - in exercise science time etc. GPS is also useful in the scientific study of earthquakes and synchronization of telecommunications networks. It provides a precise time reference [2].

The position of a GPS receiver is calculated by measuring the distance between itself and three or more GPS satellites. Each satellite is equipped with an atomic clock. When first switched on, GPS devices undergo an initialization period, during which they receive signals from the satellites and synchronize the GPS clock with the satellite's atomic clock. GPS devices constantly receive and analyze electromagnetic signals from the satellites, calculating precise distance (range) to each satellite being tracked. GPS devices use trilateration [4], a mathematical technique to determine user position, speed and elevation. Data from a single satellite provides a general location of the point within a circular range. The possibilities are narrowed down to one specific point with the introduction of three satellites. The more number of satellites used for observation, the more accurate is the position obtained [3].

This paper comprises 4 sections. In Section 2, we discuss the mathematical formulation of determining positions. In Section 3, we discuss the computational method used to determine positions and provide its MATLAB code. Lastly in Section 4, we summarise our findings.

2 Determining Positions

Classically, positions are determined by the time taken by a signal transmitted from the satellite to reach a receiver. Owing to the satellite and receiver clock offsets and a number of factors such as reflection of waves in the Troposphere and Stratosphere, there are accuracy errors in the time measured by this procedure and the distances thus obtained are referred to as *pseudo-ranges* [2].

Applying the concept of trilateration, if the distance from a point on the earth (a receiver) to three satellites are known along with the satellite locations, then the location of the point can be obtained. Data from one satellite gives a circular range of probable locations. Two satellites give out two possibilities, that is, the two intersection points of the virtual circle around them. A third satellite provides the exact location which is the overlapping area of three circles as shown in Figure 1. To further enhance the accuracy, more satellites may be taken into consideration.

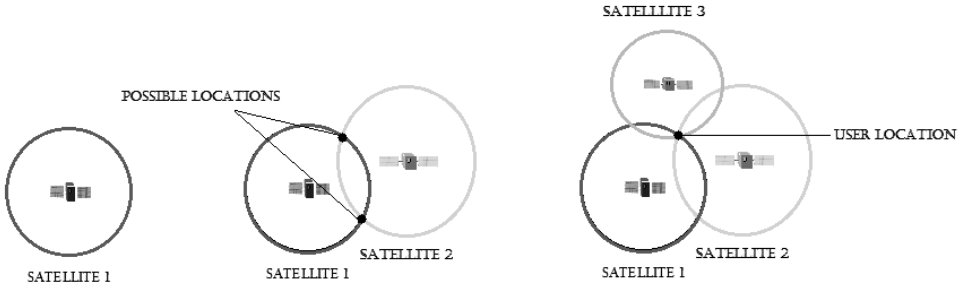


Figure 1: Trialteration

The basic measurement made by a GPS receiver is the time required for a signal to navigate from a GPS satellite to the receiver. As the signal travels at the speed of light c, this time interval can be converted to distance simply by multiplying it by c [5].
To start our analysis, we assume that the clock in the receiver is synchronised with the clock in the satellite. This assumption is however erratic [7]. When a GPS receiver is switched on, its clock will in general be mis-synchronised by an indefinite amount with respect to the satellite clocks. One of the prime reasons for this mis-synchronisation is that the clocks in the satellite are usually Cesium or Rubidium atomic clocks[1] whereas the ones in the receiver (watches, mobile phones) are crystal watches.

A timing error as small as that of a millisecond can result in an error in position of about 300 kilometres, which is clearly an intolerable amount.
If the error in noting time (or time difference) can be determined, then the pseudo-range can be corrected and the actual position of the receiver determined.

Let the error in time be dT

$$\Delta = cdT$$

$R_i (i = 1, 2, 3, 4)$ are the pseudo-ranges. The actual range r_i is given by the equation $r_i = R_i - \Delta;\ i = 1, 2, 3, 4$.

[1] Atomic clocks are clock devices regulated by vibrations of an atom or molecule. They are used for their high accuracy and precision in various fields, especially in satellite navigation

Let's say we have two transmitters. Imagine a sphere around each transmitter whose radius is the distance of the transmitter (satellite) to the receiver. There are two possible locations at the two points of intersection of these circles. Now consider three transmitters. We repeat the same process of drawing spheres around them. We can narrow down our possibilities to one location now, thus getting a precise location. The more satellites we use, the more accurate position we get. Theoretically, at least three satellites are required to determine a position. In practical view, we take a fourth satellite to account for the receiver clock error.

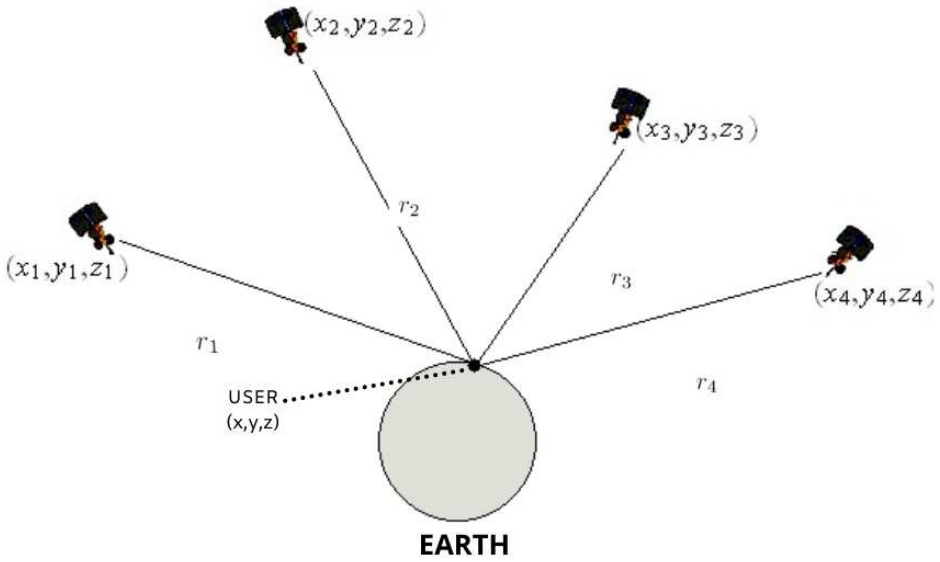


Figure 2: Observer and satellite positions

Suppose the point whose location is to determined has the coordinates (x, y, z) as shown in Figure 2. The intersection of the aforementioned imaginary spheres is the exact location of the point (x, y, z) on the earth. The set of basic equations for computing user position, using pseudo-ranges is given by:

$$\begin{aligned}
(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 &= r_1^2 \\
(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 &= r_2^2 \\
(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2 &= r_3^2 \\
(x - x_4)^2 + (y - y_4)^2 + (z - z_4)^2 &= r_4^2
\end{aligned}$$

Where x_i, y_i, z_i are the known coordinates of the satellites in the space and x, y, z are the unknown

coordinate of the observer on the earth. We now subtract the first of these equations from each of the other three equations [8]. This eliminates quadratic terms in x, y, z and leads to the following set of non-homogeneous linear equation in x, y, z :

$$\begin{array}{rcrcrcl} 2(x_2-x_1)\,x & + & 2(y_2-y_1)\,y & + & 2(z_2-z_1)\,z & = & r_1^2-x_1^2-r_2^2+x_2^2 \\ 2(x_3-x_1)\,x & + & 2(y_3-y_1)\,y & + & 2(z_3-z_1)\,z & = & r_1^2-x_1^2-r_3^2+x_3^2 \\ 2(x_4-x_1)\,x & + & 2(y_4-y_1)\,y & + & 2(z_4-z_1)\,z & = & r_1^2-x_1^2-r_4^2+x_4^2 \end{array} \tag{1}$$

This system can be solved by the method of Gaussian elimination thus giving us the coordinates (x, y, z).

3 Numerical method

Solving linear equations is one of the most important applications of linear algebra. The concept finds its applications in various fields, viz., mathematics, statistics, computer science, physics. A number of methods have been devised for the same. Gaussian elimination is one of them. One application is the GPS technology. The system of equations (1) obtained in Section 2 can be solved by this technique. It is vital to note here that this method fails if the square matrix is singular [2].

Let the linear system be denoted by $\mathbf{Ax} = \mathbf{b}$ and there be m equations consisting of n unknowns. The algorithm followed in Gauss elimination includes finding the rank of matrix and calculating the inverse using what is called the augmented matrix, denoted as:

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

Elementary row operations are performed to bring the augmented matrix to an upper triangular form, denoted as:

$$\left[\begin{array}{cccc|c} c_{11} & c_{12} & \cdots & c_{1n} & d_1 \\ 0 & c_{22} & \cdots & c_{2n} & d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{mn} & d_m \end{array}\right] \tag{3}$$

To understand the implementation of this technique in GPS technology better, let us consider an example for our system of equation (1) obtained in Section 2. Suppose that the four satellites

[2] Singular matrix is one which is not invertible, that is, its determinant is zero

under consideration yield the following coordinates (in thousand metres):

$$\begin{array}{ll} x_1 = 25,851 & x_3 = 3,641 \\ y_1 = -12,144 & y_3 = 94,785 \\ z_1 = 2,719 & z_3 = 9,332 \\ x_2 = 1,245 & x_4 = 35,632 \\ y_2 = -1,369 & y_4 = 7,741 \\ z_2 = 2,867 & z_4 = 8,234 \end{array}$$

Let the clock error dt be 0.001 millisecond.

$$\Delta(\text{in metres}) = cdT = 299,792,458 * 0.001/1000 = 2,99.7924$$

Suppose that, in addition, the pseudo-ranges registered by the GPS devices (in thousand metres) are given by:

$$R_1 = 23,785$$

$$R_2 = 22,254$$

$$R_3 = 21,364$$

$$R_4 = 25,747$$

Using the equations $r_i = R_i - \Delta$ $(i = 1, 2, 3, 4)$ given in Section 2 we get the actual distances as:

$$\begin{aligned} r_1 &= R_1 - \Delta = 23,485.2076 \\ r_2 &= R_2 - \Delta = 21,954.2076 \\ r_3 &= R_3 - \Delta = 21,064.2076 \\ r_4 &= R_4 - \Delta = 25,447.2076 \end{aligned}$$

Susbstituting these values in equation (1), we get:

$$\begin{array}{rcrcrcr} -49212x & + & 21550y & + & 296z & = & -5971567431.3288 \\ -44420x & + & 213858y & + & 13226z & = & -547163185.8008 \\ 19562x & + & 39770y & + & 11030z & = & 505288560.3776 \end{array} \tag{4}$$

The system thus obtained can be rewritten as:

$$\left[\begin{array}{ccc|c} -49212 & 21550 & 296 & -5971567431.3288 \\ -44420 & 213858 & 13226 & -547163185.8008 \\ 19562 & 39770 & 11030 & 505288560.3776 \end{array}\right] \tag{5}$$

Applying row operations, we get:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0.02317501252 & 12115.9978816 \\ 0 & 1 & 0.0666584091 & -41.9463377536 \\ 0 & 0 & 7925.64547449 & 269943615.67 \end{array}\right] \tag{6}$$

Thus the values of x, y and z are 11326.6682687, 2312.29921047 and 34059.5118137 respectively. Hence $(11326.6682687, 2312.29921047, 34059.5118137)$ are the coordinates of the observer with respect to a fixed known origin on the earth.

3.1 MATLAB Code

We provide the MATLAB code [6] of the method of Gaussian Elimination:

```
   fileID = fopen(filename.txt','r');
formatSpec = '%f';
P = fscanf(fileID,formatSpec);
n = P(1);
a = zeros(n,n);b = zeros(n,1);
x = zeros(n,1);
col = 1;row = 1;i = 2;
U=zeros(n); L=zeros(n);
while row<=n
   while col<=n
     a(row,col) = P(i);
     col = col+1; i = i+1;
   end
   b(row,1) = P(i);
   col = 1;row = row + 1;i = i+1;
end
A = [a b];
[m,n]=size(A);
for j=1:m-1
   for z=2:m
    if A(j,j)==0
     t=A(j,:);A(j,:)=A(z,:);
      A(z,:)=t;
    end
   end
    for i=j+1:m
     A(i,:)=A(i,:)-A(j,:)*(A(i,j)/A(j,j));
   end
   end
   x=zeros(1,m);
   for s=m:-1:1
     c=0;
     for k=2:m
      c=c+A(s,k)*x(k);
```

```
  end
  x(s)=(A(s,n)-c)/A(s,s);
end
fclose(fileOP);
```

4 Summary

The mathematics concerning satellite navigation has posed many mathematical challenges such as weighing the options of trilateration and triangulation in pin pointing user location, computing and approximating clock error. These concerns have been overcome by some of the best mathematicians of all time. In this paper, we have discussed formulating navigating equations using trilateration and pseudo-ranges and computing them using Gauss elimination. Various methodologies in addition to Gauss elimination have been devised to formulate and compute navigation equations for a GPS. Gauss-Jordan elimination method is an advancement in Gauss elimination. The least square (LS) standard approach to approximate solutions of overdetermined systems is another method to solve the navigation equations wherein the unknown coefficients are found based on minimizing the sum of squares of the errors made in the results of every single equation. This method will also serve to find out positions of high velocity objects. As the velocity of a moving body increases, the accuracy of the position decreases. Thus, the least square approach will be able to calculate position even in movement by high velocity bodies like airplanes or satellites.

GPS is now used in a variety of commercial and research applications such as environmental exposure, farming, ecology and driving assessment. Advancements in technology have led to development of portable GPS units which possess adequate memory to store positional data over time, thus offering to obtain location information at low costs.

The technology, however, still possess certain limitations. They fail to note position indoors, in concrete building or under heavy tree canopy, particularly in urbanised areas. In addition to this, the GPS signal is influenced by a number of other factors such as atmospheric conditions and local obstructions which can generate error in calculating position.

References

[1] Brinker R. C., Elfick M. H., Fryer J. G. and Wolf P. R., *Elementary surveying*, 8th edition, 1994.

[2] Langley R.B., *The Mathematics of GPS*, GPS World, Volume 2, No. 7, 1991, pp. 45–50.

[3] Agrawal D. P., Bajaj R. and Ranaweera S. L., *GPS: location-tracking technology*, Computer, 35(4), 2002, pp. 92-94.

[4] Kalman D., *An underdetermined linear system for gps*, The College Mathematics Journal 33, No. 5, 2002, pp. 384-390.

[5] Thompson R. B., *Global Positioning System: The Mathematics of GPS receivers*, Mathematics Magazine, 71(4), 1998, pp. 260-269.

[6] Register A. H., *A guide to MATLAB object-oriented programming*, CRC Press, 2007.

[7] Maddison R. and Mhurchu C.N., *Global positioning system: a new opportunity in physical activity measurement*, International Journal of Behavioral Nutrition and Physical Activity 6.1, 6(73), 2009.

[8] Mama M., *Mathematical Modelling of The Global Positioning System Tracking Signals*, School of Engineering, Blekinge Institute of Technology, 2008.

(Mentor) Dr. Sucheta Nayak, Assistant Professor, Department Of Mathematics, Lady Shri Ram College for Women, New Delhi
suchetanayak@lsr.edu.in

Mihika Chitranshi, B.Sc. (H) Mathematics, 4th Semester, Lady Shri Ram College for Women, New Delhi
mihichi1699@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (61-69)

Ellsberg Paradox and Utility Theory

Mr. Kuldeep, S. Arora and T. Borah

Abstract

The paper aims to discuss the Ellsberg Paradox in detail where people's choices violate the postulates of Subjective Expected Utility (SEU). In essence, it explores the mathematics involved in decision making and the probability one holds while expecting a particular outcome. Further elaboration on The Prisoner's Dilemma, an application of Game Theory, aims to decipher the hypothesis of risk based outcome without prior knowledge. Various comprehensive strategies used by players validate Ambiguity Aversion and corroborate risk analysis, thereby relating the philosophy of decision making with probability, and economics of *risk vs ambiguity.*

Keywords: Ellsberg Paradox, The Prisoner's Dilemma, Subjective Expected Utility, Ambiguity Aversion, Game Theory, Degrees of Belief

1 Introduction

Decision-making under certainty follows that the agent/person is aware of the outcomes of his/her decisions whereas decisions made without certainty imply that the agent lacks conviction about the outcomes. Decisions without certainty are subdivided into decisions under uncertainty, in which the agent is unaware even of the outcome probabilities and decisions taken under risk, wherein the agent is not assertive of the outcomes, but can predict the outcome probabilities associated with each decision. Experimental evidence between two players involved in a game is furthermore explored by analysing their strategies in assessing the opponent's capabilities. Their rationale in decision making and Keynes' *"Degrees of Belief"* [5] linked to various aspects of philosophical logic would allow us to make inferences based on relative and strategic ambiguity in correlation to the existence of Ellsberg Paradox [1]. The gist of this argument lies on the direct applications of dilemmas, risk and constructive outcomes in reality.

The Ellsberg Paradox

Frank Knight, a distinguished economist, established the difference between risk and uncertainty with a thought experiment involving two people attempting to draw a red or a black ball from an urn containing both red and black balls: One person knows that there are red and black balls, but is unaware of the numbers of each; while the other knows that there are three red balls and one black ball. The latter faces a decision under quantifiable risk, while the former faces a decision under unquantifiable uncertainty, often referred to as *ambiguity*. The empirical demonstration of the Ambiguity Aversion [2] effect, based on Knight's experiment mentioned above, is called the Ellsberg Paradox. As his primary examples, Ellsberg gave two thought experiment decision problems.

We shall consider the **Two-Colour/Two Urn Ellsberg Paradox:** Two urns are filled with identical red and black balls. Urn A contains 50 balls each of red and black colour, and Urn B contains an unknown ratio of 100 red and black balls. A person has to choose a colour (red or black) and an urn (A or B) and then draw a ball randomly from the chosen urn. If the ball turns out to be same as the chosen colour, the person wins. A majority of decision makers prefer Urn A to the ambiguous Urn B, the reason being, risk involved in drawing a ball from Urn A is known.

Consider a case where the decision maker chooses the colour red and Urn A. Since probability of drawing a red ball from Urn A is $\frac{1}{2}$, the decision maker must, according to the assumptions of SEU[1][3] theory, have assigned less than $\frac{1}{2}$ probability to drawing a red ball from Urn B, otherwise Urn A should not be preferred for drawing a red ball. Since the ball to be drawn must be either red or black, and both their probabilities must sum to unity, it is implied that the subjective probability of decision-maker to draw a black ball from Urn B must be greater than $\frac{1}{2}$. The SEU theory states that the decision maker should decline the prospect of drawing a red ball from Urn A and instead prefer that of drawing a black ball from Urn B. Under such circumstances, it is deemed that the decision-maker fails to maximise SEU and in turn violates its postulates. Thus, Ambiguity Aversion violates certain assumptions of the SEU theory.

2 The Experiment

The fundamental rationality assumption of Game Theory[2] [4] indicates that players always favour strategies that maximize their individual SEU, relative to their knowledge at that time. According to the SEU theory, people have logical preferences and probability judgements, and after evaluating

[1]Subjective Expected Utility is the subjective opinion of the decision-maker in making the most out of what is available of an opportunity in times of risk.

[2]Game Theory is the study of games which consists of strategic interactions among rational decision-makers.

all the information available they make decisions that they deem best. Interpretation of SEU can also be done in terms of Revealed Preference Theory[3], which was pioneered by renowned economist, Paul Samuelson. It states that a player who chooses option O_A and rejects option O_B reveals a preference of O_A over O_B and a higher utility for O_A than O_B. However, it must also be understood that choices are not the only indicators of preferences - other factors also influence choices.

Expected utility can be calculated as the sum of product of probabilities of events and their respective utility amounts.

$$E[u(x)] = \sum p_i.u(x_i)$$

Let X be a finite payoff space, The set δX of outcomes on X is the set of all functions $p : X \longrightarrow [0, 1]$ such that

$$\sum_{x \in X} p(x) = 1$$

According to the Revealed Preference theory a preference $p \succeq q$ can be made provided the payoff from the former is more than the latter i.e.

$$\sum_x p_x.u(x) \geq \sum_x q_x.u(x)$$

Common knowledge is a condition usually required in Game Theory. It demands that all the players know the the rules and strategy sets as well as payoff functions related to a game along with specific details. The common knowledge assumption asserts that all players must be aware of the information the other players have regarding rules and their utility functions. This assumption was introduced by David Kellog Lewis, an American philosopher, in 1969 and was first formulated mathematically by Robert Aumann, a well-known mathematician, in 1976. Models of risky, strictly determined and ambiguous strategic games were evaluated in terms of players' knowledge, so as to explore the various outcomes of Ambiguity Aversion experimentally. In few cases, common knowledge assumptions of Game Theory allow a player to predict other player's choice with certainty and rationality. This happens because the latter has a uniquely rational strategy according to SEU theory. In such circumstances, the player knows what the other will choose and therefore faces a decision under certainty, and the game is strictly determined. An example of such a game is depicted in Figure 1.

Let us consider two players, I and II. One of the rows is chosen by Player I, and one of the columns is chosen by Player II. Possible outcomes are in the form of four cells, with each cell indicating pairs of numbers (in units of utility) the payoffs to Players I and II, respectively, in that particular outcome.

[3]Revealed Preference Theory is a method to analyze choices and decisions made by individuals. This methodology is mainly employed to evaluate the influence of policies on consumer behaviour.

2.1 Case 1

Here if Player I chooses Strategy C and if Player II chooses Strategy D, then the former wins nothing. However, Player I wins 2 units if Player II chooses Strategy C. Player II has a dominant Strategy C over Strategy D i.e., Strategy C is more rewarding due to higher payoff. In more elaborate words, Player II wins 1 unit by choosing Strategy C but receives a zero payoff on choosing Strategy D, regardless of what Player I's choice.

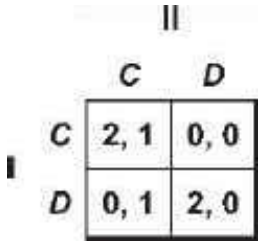


Figure 1: An unambiguous game where Player I predicts with certainty, the strategy employed by Player II. Each cell contains the payoffs to Player I (choosing between the rows) and to Player II (choosing between the columns) in that order; (source: Pulford B. D., Colman A. M., *Ambiguous games: Evidence for strategic ambiguity aversion*, The Quarterly Journal of Experimental Psychology, pp 1083-1100, 2007.)

Thus, common knowledge assumptions show that Player I knows what Player II will choose and can henceforth realise the outcome and its corresponding payoffs from choosing row 1 or 2. Because Player I seeks to maximise SEU, the outcome is therefore strictly determined, and the payoffs are predictable with certainty-subject only to the standard rationality and common knowledge assumptions. The game in Figure 1 is clearly unambiguous and hence is a model of a strictly determined game.

2.2 Case 2

In case of certain other games, Player II's strategy might remain unknown to Player I, but he/she can still try to predict with some confidence, by assigning subjective probabilities on how Player II will act. However, in such circumstances, Player I faces a risky situation and has to make an important decision. Figure 2 depicts the simplest game of this type. Here, there exists a payoff function that assigns Player II with a similar expected payoff in every single outcome of the game. This henceforth indicates Player II's indifference among the four possible outcomes. The given

assumptions do not specifically require Player II to choose between the strategies using randomisation, that is, by tossing a coin, and neither does it force Player I to believe that Player II might adopt any strategy considering equal probabilities (C or D).

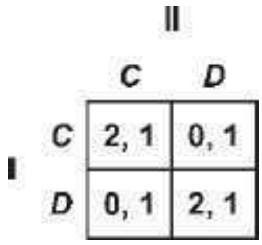


Figure 2: A risky game for Player I; (source: Pulford B. D., Colman A. M., *Ambiguous games: Evidence for strategic ambiguity aversion*, The Quarterly Journal of Experimental Psychology, pp 1083-1100, 2007.)

Game Theory suggests that a Player needs to be aware of his/her beliefs to make any strategical moves to proceed further in the game. It also requires for him to be consciously aware of his co-player's beliefs. Pertaining to the current circumstances, it is natural for Player I to assume that Player II has adopted equal probability for both strategies. It seems the most apt and likely behaviour put forward by Player II, in above mentioned conditions. Under this natural assumption, Player I faces a risky decision, that amounts to an expected payoff from Strategy C, choice of

$$(\frac{1}{2})(2) + (\frac{1}{2})(0) = 1$$

Similarly from Strategy D, choice of

$$(\frac{1}{2})(0) + (\frac{1}{2})(2) = 1$$

Such a game is considerably riskier in lieu of it being ambiguous.

2.3 Case 3

Now we shall consider the final possible case where the players are ignorant of their co-players preferences. These games are modelled owing to incomplete information. Modelling games of incomplete information was developed by Harsanyi (1967–1968) who introduced the theory of Bayesian games in which incomplete information is transformed into complete information by introducing a fictitious player representing *chance*. In a two player game in which Player I is ignorant of Player

II's preferences, the Harsanyi transformation defines multiple Player II types, each with a different payoff function representing the preferences that Player II might have. Player I's ignorance of Player II's preferences is modelled by specifying all the Player II preference patterns that Player I considers possible, with a probability assigned to each according to how likely Player I considers it to be. *Chance* makes the first move in a Bayesian game by selecting one of the Player II types, each type having a predetermined probability of being selected, according to the subjective probability that Player I assigns to it, and then Players I and II choose their strategies independently in the usual way. Player I is ignorant of which payoff matrix has been selected by *chance*, but the probabilities associated with the Player II types are assumed to be common knowledge in the game. This transformation thus enables one to apply standard analytic techniques used in games of complete information, eliminating the chances of a game of incomplete information. This notion is generalised for both players, provided they have incomplete information and is also applicable to the multiplayer games. The probabilities associated with Player II types are endogenous variables in Bayesian game theory, they are not inherent in the specification of a game but arise from a player's subjective response to it.

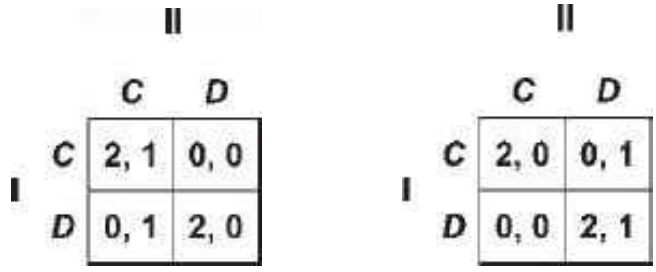


Figure 3: An ambiguous game where Player I can decide between the two payoff matrices which govern Player II's motives; (source: Pulford B. D., Colman A. M., *Ambiguous games: Evidence for strategic ambiguity aversion*, The Quarterly Journal of Experimental Psychology, pp 1083-1100, 2007.)

Figure 3 depicts a simple ambiguous game in which, Player I chooses under uncertainty. This models a game in which Player I counts on the possibility that Player II will be one of these two types but, does not know for sure their share of respective probabilities. Player I's payoffs are identical in both matrices, as in a conventional Bayesian game. The Player II type on the left has a strictly dominant C strategy that results in payoff of one, irrespective of Player I's choice. The Player II type on the right is certain to receive a payoff of one by choosing strategy D and payoff of zero on choosing C, irrespective of Player I's choice.The Player II type on the right shows a completely opposite scenario of that on left. Player I chooses a strategy without knowing Player

II's type, and therefore which payoff matrix applies, and in such a scenario Player I may be assumed to face an ambiguous choice with unknown probabilities.

3 The Prisoner's Dilemma

One of the most popular applications of Modern Game Theory is The Prisoner's Dilemma. It is stated as follows:-
Two members of a criminal gang are arrested and imprisoned. Each convict is placed in solitary confinement and has no means of communication with the other. The prosecutors cannot convict the pair on the principal charge, but they have enough evidence to convict them on a less prominent charge. Simultaneously, the prosecutors offer each prisoner a bargain. The prisoners are given the opportunity to either betray the other by testifying that the other committed the crime, or to cooperate with the other by remaining silent. The possible outcomes are:

- If both prisoners betray each other, each of them serves two years in prison;
- If one betrays the other but the other remains silent, then former will be set free and latter will serve three years in prison (and vice versa);
- If both prisoners remain silent, each of them will serve only one year in prison (on less prominent charge).

A \ B	B stays silent	B betrays
A stays silent	-1 / -1	0 / -3
A betrays	-3 / 0	-2 / -2

Figure 4: The payoff matrix for Prisoner's Dilemma; (source: http://law.stackexchange.com)

It is implied that the prisoners will have no obligation to defend or betray their partner and neither will they have an opportunity to interact with each other in the future. Their previous association would not hinder in the formation of their future reputations. The Prisoner's Dilemma offers the following probability of outcome. Since the reward of betrayal is much more superior than the idea of cooperation, each prisoner chooses his self-interest over mutual cooperation. The interesting and seemingly ironic part of this result indicates that individual reward satisfaction logically leads both prisoners in betrayal, whereas theoretically they would get better individual

reward if both of them silent. This again brings us to the key elements of our previous experiment, that an individual prefers risk of the known over the risk of the unknown. The possibility of an assumption that goes against humanity's deep rooted self-interest is out of question. It is presumed that each of the convicts will betray the other owing to rewarding self-interest, despite knowing the effort is in vain.
It is henceforth, predicted by the above lineage of experiments that the assumptions of game theory satisfy Ellsberg Paradox. A prisoner accepts the known probability of betrayal and rejects the lesser appealing (although ironically more rewarding) probability of cooperation. Despite the innumerable risks in making a significant decision, Ellsberg Paradox follows Ambiguity Aversion. It deviates from the ideal of SEU theory, and further helps us prove the winner of the debate titled "ambiguity v/s risk".

4 Conclusion

A "logical relation" between one proposition and another essentially means to have a logical conclusion presented from certain set of propositions. In situations where the first proposition neither logically infers nor logically excludes the second is defined as *Keynes' probability*. Therefore, the probability of a proposition is defined for a given set of premises, provided it has a *rational belief* attached to it. The notion of *degree of belief*[4] is not personal or subjective, any more than its counter extremities of logical necessity or logical impossibility.
Interactive behaviour of the players, their social and conscious probabilities of winning is ultimately also a valid conclusive argument in Logic. Ambiguous games have behavioural and psychological characteristics that distinguish them from the known-risk games on which the entire edifice of orthodox game theory is based. In this experiment, strategy choices differed significantly between the known-risk and ambiguous versions of the game. Each premise of strategy involved brims a series of moves in the player's mind which also semantically follows an unpredictable risk. This risk is bound to cause ambiguity in decision-making, thereby validating the existence of Ellsberg Paradox, Game Theory and Concept Of Utility.

References

[1] Ellsberg D., *Risk, Ambiguity and Decision*, Ph.D. thesis, Harvard Univesity, 1962.

[4] *degrees of belief* are basically evidential probabilities, defined in terms of dispositions to gamble at certain odds

[2] Ellsberg D., *Risk, Ambiguity, and the Savage Axioms*, Quarterly Journal of Economics, Volume 75, pp 643-669, 1961.

[3] Allias M., Hagen O., *Expected Utility Hypotheses and the Allais Paradox*, Springer, First Ed., 1979.

[4] Neumann J. V.. Morgenstern O., *Theory of Games and Economic Behavior*, Princeton University Press, Third Ed., 1953.

[5] Keynes J. M., *A Treatise on Probability*, Macmillan, London, 1921.

[6] Knight F., *Risk, Uncertainty, and Profit*, Houghton Mifflin Co., Boston, 1921.

[7] Pulford B. D., Colman A. M., *Ambiguous games: Evidence for strategic ambiguity aversion*, The Quarterly Journal of Experimental Psychology, pp 1083-1100, 2007.

[8] http://faculty.wcas.northwestern.edu/ msi661/Ambiguity-06-29-2013.pdf

[9] http://personal.lse.ac.uk/bradleyr/pdf/Ellsbergs

[10] http://www.closemountain.com/papers/ellsberg.pdf

(Mentor) Mr. Kuldeep, Assistant Professor, Department Of Mathematics, Lady Shri Ram College For Women, New Delhi
kuldeepkr09@gmail.com

Shubhi Arora, B.Sc.(H), Mathematics, 2nd Semester, Lady Shri Ram College For Women, New Delhi
shubhiarora787@gmail.com

Tanya Borah, B.Sc.(H) Mathematics, 2nd Semester, Lady Shri Ram College For Women, New Delhi
borah.tanyaa@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (70-78)

Information Security Using Abstract Algebra

Dr. S. K. Yadav, A. Mittal and R. Chandel

Abstract

Information security is of utmost importance in a digitally interconnected world and demands constant evolution to be at par with security threats. This paper talks about an important domain of Information Security: Error Detection in Man-Machine interactions, by constructing error-detecting codes using concepts of abstract algebra. The fundamental concepts used in the paper are: ring of integers modulo n and finite fields. The methods explored here are exemplary of the application of abstract algebra in real life.

Keywords: Rings, Fields, Codes, Codewords, Errors, Error Detecting Codes, Optimal Error Detecting Codes

1 INTRODUCTION

"The computer lets you make more mistakes, faster than any invention in human history"

- Mitch Ratcliffe

Data and information (see [3], [4], [5]) are the engines of the 21st century, and computers are the medium of accessing and manipulating information. However, as is the case of all valuable commodities, with the abundance of information comes the issue of information security. In today's world, techniques of cryptography and cipher are widely used for the purpose of security, wherein a commonly understood message is labeled or represented as another symbol based upon some undisclosed rule. In technical terms this is called a **code** (see [4]).
A traditional example of such labeling is the branding of cattle whereas technologically advanced applications are: bar codes on goods, codes on credit cards, passports etc. It becomes imperative that these methods of encryption are foolproof. Irrespective of the precision and accuracy of modern-day computers, man-machine communication is affected by errors caused by human operators. Often, these errors occur while entering characters using a keyboard. In this paper, we will discuss the following type of errors involving man-machine communication.
Consider the word "form" that has to be typed by the user. The user can make various typing errors (see [4]). Here, we look at following three:

- The character 'o' is mistyped as 'a' to produce "farm". This is called a **substitution error.**
- The character 'm' is omitted or an extra character 'u' is inserted between 'r' and 'm' to produce "for" or "forum" respectively. This is called a **deletion/insertion error**.
- The adjacent characters 'r' and 'o' are interchanged to produce "from". This is called a **transposition error**.

None of the above errors that occur to the word "form" may be detected since they lead to legal words. In order to detect these errors, the words of the language should be encoded and we need codes/encryption

techniques for this purpose. Such codes are called **Error Detecting Codes**. In this paper, we will consider codes of fixed length. Insertion/deletion errors cause a change in the length of the code and hence can easily be detected. We will talk about single substitution (S-error) and single transposition (T-error) errors. The codes which can detect both these types of errors are called **ST- Error Detecting Codes** (see [4]).
The arrangement of this article is as follows: In Section 2, we record some preliminaries and notations used in the article. In Section 3, we define and construct Optimal ST-error detecting codes as well as look at the condition under which such a code may be generated. We also define the check character in this section. In Section 4, we describe various cases and methods for constructing optimal ST-error detecting codes using rings, fields and other concepts from abstract algebra.

2 SOME PRELIMINARIES AND NOTATIONS

In this section, we record some preliminaries and notations used in the article. Let $\mathbb{Z}_n$ be the ring of integers modulo n, that is, $\mathbb{Z}_n = \{0,1,2,\ldots,n-1\}$, where the number of elements in $\mathbb{Z}_n$ is n (see [1], [2]). In our article, elements of $\mathbb{Z}_n$ will be referred to as **characters**.
We denote a set of m-tuples of elements from $\mathbb{Z}_n$ by C, and we say that C is a collection of codes of length 'm' over $\mathbb{Z}_n$, that is, $C \subseteq \mathbb{Z}_n^m$. The elements of the set C are called **codewords**. The cardinality of the set C is denoted by M. Since $\mathbb{Z}_n$ has n characters, we call C an ***n*-ary code**.
Under above assumptions and notations, the criteria for ST-error detection is as follows:
(i) The collection of codes C is said to detect single substitution error if and only if replacing a single character in a codeword of C produces a tuple not belonging to C. That is, $\tau_0\tau_1\ldots\tau_{i-1}\tau_i\tau_{i+1}\ldots\tau_{m-1} \in C$ implies that $\tau_0\tau_1\ldots\tau_{i-1}\tau_i'\tau_{i+1}\ldots\tau_{m-1} \notin C$ for $0 \leq i \leq m-1$ and $\tau_i' \neq \tau_i$.
For example: let $n = 3$ and $m = 4$. Consider codeword $0121 \in C$ implies that the tuple formed by substituting '0' at the second place is $0021 \notin C$.
(ii) The collection of code C can detect a single transposition error if and only if interchanging two adjacent distinct characters of a codewords in C produces a tuple not belonging to C. That is, $\tau_0\tau_1\ldots\tau_{i-1}\tau_i\tau_{i+1}\ldots\tau_{m-1} \in C$ implies that $\tau_0\tau_1\ldots\tau_{i-1}\tau_{i+1}\tau_i\ldots\tau_{m-1} \notin C$ for $0 \leq i \leq m-2$ and $\tau_i \neq \tau_{i+1}$.
For example: let $n = 3$ and $m = 4$. Consider codeword $0121 \in C$ implies that the tuple formed by transposing the characters in second and third positions is $0211 \notin C$.
Such errors lead to tuples not belonging to C and are consequently detected (see [3], [4]).

Example 2.1. Let $n = 4$, then we have $\mathbb{Z}_4 = \{0,1,2,3\}$. Consider code C with codewords of length $m = 3$ over $\mathbb{Z}_4$ that has 16 codewords. That is,

$$C = \{010, 112, 120, 131, 321, 230, 002, 101, 332, 211, 222, 313, 033, 300, 203, 021\}.$$

In this code, if we choose any codeword and introduce a single substitution or transposition error in it, the resulting tuple will not belong to C. For instance, consider the codeword 131 in C. Suppose a substitution error changes 3 to 2. The resulting tuple 121 does not belong to C. Similarly, consider the last codeword 021. Suppose a transposition error interchanges 2 and 1. The resulting tuple 012 does not belong to C. □

The reader may verify that this holds true for any single substitution or transpositon error in any codeword of C. Hence code C is an ST-error detecting code. Practically, it is neither feasible nor required to verify this for each case. Instead, the error-detecting capability of the code is intrinsic to the method of construction of the code, which we will explore later.

3 Optimal ST-Error Detecting Code

In this section, we define and construct Optimal ST-error detecting codes as well as look at the condition under which such a code may be generated. If each codeword in code C can be used to uniquely represent one message, then a code with M codewords can encode M messages. We are interested in maximizing the number of messages that can be encoded using an n-ary code.
This number M depends on the length m of the tuple, as well as cardinality n of $\mathbb{Z}_n$. Hence, M can be denoted as $M(n,m)$. The code with the maximum possible value of $M(n,m)$ is said to be an **Optimal ST-Error Detecting Code**. We will now see what is the maximum limit on $M(n,m)$.

Theorem 3.1 (see [4])**.** $M(n,m) \leq n^{m-1}$.

Proof. We will prove this theorem by contradiction. Let C be an ST-error detecting code defined over $\mathbb{Z}_n$ with tuples of length m. We define a set T of all possible tuples of length m over $\mathbb{Z}_n$. The cardinality of T is n^m. Out of these tuples, M codewords belong to the set C. Suppose C has $M > n^{m-1}$ codewords.
We now partition T into classes based on the following rule: Each tuple in a class agrees on the first $m-1$ characters. This will result in n^{m-1} disjoint classes since there exist n^{m-1} permutations for the first $m-1$ characters. Given that the number of codewords $M > n^{m-1}$, we can conclude that at least one class contains two codewords. However, since they belong to the same class, these two codewords would agree upon the first $m-1$ characters and differ only in the last character. This implies that a single substitution error would change one codeword to the other. This contradicts the fact that C is an ST-error detecting code. Hence, our assumption stands incorrect and $M \leq n^{m-1}$. □

To understand this proof better let us consider an example.

Example 3.2. Let $n = 3$, then we have $\mathbb{Z}_3 = \{0,1,2\}$. Consider code C with codewords of length $m = 3$ and cardinality M. Let T be the set of all possible tuples of length 3 over $\mathbb{Z}_3$, then cardinality of the set T is $3^3 = 27$. Therefore,

$$\begin{aligned} T \quad = \quad & \{000,001,002,010,011,012,100,101,102,110,111,112,020,021,022,200,201,202,120,121, \\ & 122,210,211,212,220,221,222\}. \end{aligned}$$

We partition this set based on the rule: *Each tuple in the class agrees on the first* 2 *characters.* Here, we denote the class by the notation $[-]$. Therefore, we get the following $3^2 = 9$ equivalence classes:

$$[00] = \{000,001,002\}, [01] = \{010,011,012\}, [10] = \{100,101,102\},$$
$$[11] = \{110,111,112\}, [02] = \{020,021,022\}, [20] = \{200,201,202\},$$
$$[12] = \{120,121,122\}, [21] = \{210,211,212\}, [22] = \{220,221,222\}.$$

Notice that tuples in the same class differ only in the last character, that is, differ by a single substitution error. Hence, at most one tuple from each class can belong to the code C. Therefore, the maximum number of codewords in C is equal to the total number of classes. Alternately, $M \leq 3^2$. Hence, we have verified the theorem. □

Now we define check character (see [4]) which we will use frequently in the next discussion. The purpose of this character is to determine the codewords which will belong to the set C as well as to check if any given tuple can be a codeword. Subsequently, we will look at an example of a check character.

Definition 3.3. A **check character** is a character in the codeword whose value depends on the values of other characters.

Example 3.4. Consider the ST-error detecting code C with $m = 3$ and $M = 9$ codewords. This code can encode 9 messages as follows: Let k_0k_1 be a message, where $k_0, k_1 \in \mathbb{Z}_2$. Then this message is encoded into $\tau_0\tau_1\tau_2$, where $\tau_0 = k_0$, $\tau_1 = k_1$ and

$$\tau_2 = -(k_0 + 2k_1) \bmod 3.$$

For example, if the message is 21, then $\tau_0 = 2$, $\tau_1 = 1$ and $\tau_2 = -(2 + 2 \times 1) \bmod 3$, that is, $\tau_2 = 1$. This message is encoded into the codeword $\tau_0\tau_1\tau_2 = 211$. In this example, τ_2 is a check character. □

In the Example 3.4, we have placed the check character at the end of the codeword. In general, we can place the check character anywhere in the codeword. However, in this article, we will be using constructions similar to the one described in Example 3.4.
Now, let us explore methods of constructing optimal ST-error detecting codes.

4 Construction of Optimal ST-Error Detecting Codes

In this section, we describe various conditions on n and methods for constructing optimal ST-error detecting codes using algebraic structures such as: rings, fields etc.
Consider an optimal n-ary ST-error detecting code. The technique of construction of this code depends on the number of characters n in $\mathbb{Z}_n$. Consequently, the following cases arise (see [4]):

Case 1: n is odd. Identify the set $\mathbb{Z}_n$ of n characters: $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$. Choose an integer $t \in \mathbb{Z}_n$ and keep it fixed. The code set C is defined in the following manner:

$$C = \{\tau_0\tau_1 \ldots \tau_{m-1} : \sum_{i=0}^{m-1} a_i\tau_i = t \bmod n\},$$

where

$$a_i = \begin{cases} 1, & \text{if } m-i \text{ is odd,} \\ 2, & \text{if } m-i \text{ is even.} \end{cases}$$

Here, the operation modulo n is mentioned explicitly in the definition. However, it is understood without explicitly mentioning, since the ring operations of $\mathbb{Z}_n$ are defined as modulo n.
Further, we define the check character in this construction. Consider the message: $\tau_0\tau_1 \ldots \tau_{m-2}$. Then the check character τ_{m-1} is given by

$$\tau_{m-1} = t - \sum_{i=0}^{m-2} a_i\tau_i \bmod n.$$

Let us now see whether C is an ST-error detecting code. Since a_i is either 1 or 2, therefore a_i is relatively prime to the odd number n. This implies that a_i^{-1} exists in $\mathbb{Z}_n$. Hence, the sum $\sum_{i=1}^{m-2} a_i\tau_i$ uniquely identifies the check character τ_{m-1} and any single substitution error can be detected.
Similarly, since $a_i - a_{i+1} = \pm 1$ is also relatively prime to odd number n, any single transpostion error can also be detected. Clearly, C is an n-ary code with the codewords of length m. Indeed, C agrees with Theorem 3.1 and is capable of containing a maximum of n^{m-1} codewords. This can be verified since there exist that many permutations for the first $m-1$ characters $\tau_0\tau_1 \ldots \tau_{m-2}$, and m^{th} check character τ_{m-1} is identified uniquely by given rule. Hence, C is an optimal n-ary ST-error detecting code for the case when n is odd.

Example 4.1. Consider $n = 3$, than we have $\mathbb{Z}_3 = \{0, 1, 2\}$. Let us construct a code with 3-tuples over $\mathbb{Z}_3$, that is, for $m = 3$. Let $t = 0$, then the above construction gives the following code :

$$C = \{\tau_0\tau_1\tau_2 : \sum_{i=0}^{2} a_i\tau_i = 0 \bmod 3\},$$

where τ_2 is the check character, and

$$a_i = \begin{cases} 1, & \text{if } m-i \text{ is odd,} \\ 2, & \text{if } m-i \text{ is even.} \end{cases}$$

Let $\tau_0\tau_1 = 12$, then we have

$$\begin{aligned} \sum_{i=0}^{2} a_i\tau_i &= 0 \bmod 3 \\ a_0\tau_0 + a_1\tau_1 + a_2\tau_2 &= 0 \bmod 3 \\ 1(1) + 2(2) + 1(\tau_2) &= 0 \bmod 3 \\ \tau_2 &= -2 \bmod 3 \\ \tau_2 &= 1 \bmod 3. \end{aligned}$$

Therefore, $\tau_2 = 1$ and the required codeword is $\tau_0\tau_1\tau_2 = 121$.
Using above procedure and the method of constructing optimal code described in Example 3.2, we get the following optimal code:

$$C = \{000, 011, 102, 022, 201, 110, 121, 212, 220\}.$$

Clearly, C is an optimal ST- error detecting code where $|C| = 3^{3-1} = 3^2 = 9$. □

Case 2: $n = 2^r$, where $r \geq 2$. First, we will see that the construction in Case 1 does not work if we construct the code C over $\mathbb{Z}_n$, that is, we cannot construct ST-error detecting code over $\mathbb{Z}_n$. But if we use the construction similar to Case 1, we can construct a single substitution error detecting code. We will try to understand by the following construction:

Construction 1: Identify the set $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ of n characters. Choose an element $t \in \mathbb{Z}_n$ and keep it fixed. The code set C is defined in the same manner as Case 1:

$$C = \{\tau_0\tau_1 \ldots \tau_{m-1} : \sum_{i=0}^{m-1} a_i\tau_i = t \bmod n\}.$$

Now, as reasoned in Case 1, this construction will be able to detect single substitution errors if and only if a_i is relatively prime to n. Since n is even, a_i must be odd to be relatively prime to n. Therefore, we can consider the following definition of a_i:

$$a_i = \begin{cases} 1, & \text{if } m-i \text{ is odd,} \\ 3, & \text{if } m-i \text{ is even.} \end{cases}$$

However, as reasoned in Case 1, in order to detect single transposition errors it is required that $a_i - a_{i+1}$ is relatively prime to n. But we have chosen a_i to be odd, and difference of any two odd numbers is always even. As a result, $a_i - a_{i+1}$ cannot be relatively prime to n, and this code cannot detect single transposition errors. This construction differs from Case 1 in the definition of a_i and is capable of detecting single substitution errors. Hence, this is an S-Error Detecting Code over $\mathbb{Z}_n$, where $n = 2^r$ and $r \geq 2$.

Example 4.2. Consider $n = 2^2 = 4$, then we have $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Let us construct codewords of length 3 over $\mathbb{Z}_4$, that is, for $m = 3$. Let $t = 1$, then the above construction gives the following code :

$$C = \{\tau_0\tau_1\tau_2 : \sum_{i=0}^{2} a_i\tau_i = 1 \bmod 4\},$$

where τ_2 is the check character, and

$$a_i = \begin{cases} 1, & \text{if } m-i \text{ is odd,} \\ 3, & \text{if } m-i \text{ is even.} \end{cases}$$

Let $\tau_0\tau_1 = 00$, then we have

$$\begin{aligned} \sum_{i=0}^{2} a_i\tau_i &= 1 \bmod 4 \\ a_0\tau_0 + a_1\tau_1 + a_2\tau_2 &= 1 \bmod 4 \\ 1(0) + 3(0) + 1(\tau_2) &= 1 \bmod 4 \\ \tau_2 &= 1 \bmod 4. \end{aligned}$$

Therefore, the required codeword is $\tau_0\tau_1\tau_2 = 001$.
Using above procedure, we get the following optimal code:

$$C = \{001, 012, 023, 030, 100, 203, 302, 111, 122, 133, 210, 313, 221, 232, 320, 331\}.$$

This code detects single substitution errors by definition. However, notice that certain pairs of codewords: 012 and 210, 023 and 203, 302 and 320; differ by a single transposition error. Hence, this code cannot detect single transposition errors. Infact, this construction is incapable of detecting all such errors where 0 is transposed with $n/2 = 2$. □

We have seen that code construction in Case 1 fails for $\mathbb{Z}_n$, when $n = 2^r$ and $r \geq 2$, that is, we can not find any ST-error detecting code over $\mathbb{Z}_n$, when $n = 2^r$ and $r \geq 2$. To resolve this drawback, we will now look at another construction.

Construction 2: In this construction, we will make use of finite fields for constructing an optimal ST-error detecting code. We know that finite fields exist if and only if their order is a prime or a power of a prime. For example, there exist finite fields of order 2, 3, 4, 5, 7, 8, 9, 11 and so on.
In this case, we consider field F of order $n = 2^r$, where $r \geq 2$. Let e be the multiplicative identity of F and a be a non-zero element of F different from e. Select any $t \in F$. Then the code given by

$$C = \{\tau_0\tau_1 \ldots \tau_{n-1} : \sum_{i=0}^{n-1} a_i\tau_i = t\},$$

where

$$a_i = \begin{cases} e, & \text{if } m-i \text{ is odd,} \\ a, & \text{if } m-i \text{ is even.} \end{cases}$$

This code detects ST-errors since all non-zero elements of a field are invertible, therefore e, a, and $e-a$ are invertible, that is, a_i and $a_i - a_{i+1}$ are invertible. Hence, the sum $\sum_{i=1}^{m-2} a_i\tau_i$ uniquely identifies the check character τ_{m-1}, and any single substitution or transposition error can be detected. We will see how this construction works with the help of an example.

Example 4.3. Consider $n = 2^2 = 4$. Let $\mathbb{Z}_2[x]$ be the collection of all polynomials over $\mathbb{Z}_2$ in variable x. Let $I = < x^2 + x + 1 >$ be the Ideal of $\mathbb{Z}_2[x]$. Then quotient ring $\frac{\mathbb{Z}_2[x]}{I} = \{I, 1+I, x+I, x+1+I\}$ is a field of order 4. To see the construction of a finite field of order 4, refer [1] or [2]. Now, our field is $F = \frac{\mathbb{Z}_2[x]}{I}$ with zero element I and identity element $1+I$. For more clarity, we have the following Cayley tables for F:

+	I	$1+I$	$x+I$	$x+1+I$
I	I	$1+I$	$x+I$	$x+1+I$
$1+I$	$1+I$	I	$x+1+I$	$x+I$
$x+I$	$x+I$	$x+1+I$	I	$1+I$
$x+1+I$	$x+1+I$	$x+I$	$1+I$	I

and

$\times$	I	$1+I$	$x+I$	$x+1+I$
I	I	I	I	I
$1+I$	I	$1+I$	$x+I$	$x+1+I$
$x+I$	I	$x+I$	$x+1+I$	$1+I$
$x+1+I$	I	$x+1+I$	$1+I$	$x+I$

For simplicity, we denote the elements $I, 1+I, x+I, x+1+I$ of the field F by $0,1,2,3$ respectively. Note that we are not considering any map between the sets $\{I, 1+I, x+I, x+1+I\}$ and $\{0,1,2,3\}$. We are using $\{0,1,2,3\}$ only for the purpose of notation. Therefore, corresponding to the new notations we have the following tables.

+	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

and

$\times$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

.

Now, we have $n = 4$, $Q = \{0,1,2,3\}$. Let $m = 3$ and $t = 0$. Then,

$$C = \{\tau_0\tau_1\tau_2 : \sum_{i=0}^{2} a_i\tau_i = 0\},$$

where

$$a_i = \begin{cases} 1, & \text{if } 2-i \text{ is odd,} \\ 2, & \text{if } 2-i \text{ is even.} \end{cases}$$

Let $\tau_0\tau_1 = 23$, then by using above table we get,

$$\begin{aligned} \sum_{i=0}^{2} a_i\tau_i &= 0 \\ a_0\tau_0 + a_1\tau_1 + a_2\tau_2 &= 0 \\ 1(2) + 2(3) + 1(\tau_2) &= 0 \\ \tau_2 &= 3. \end{aligned}$$

Therefore, the required codeword is $\tau_0\tau_1\tau_2 = 233$. Using this procedure, we get the following optimal code:

$$C = \{000, 012, 101, 023, 202, 113, 122, 210, 221, 233, 320, 031, 303, 130, 311, 332\}.$$

Clearly, C is an optimal ST- Error Detecting Code where $|C| = 4^{3-1} = 4^2 = 16$. □

Remark 4.4. We know that if F is finite field, then order of F is a prime or a power of a prime. Therefore, above method can be extended for all $n = p^r$, where p is a prime number and $r \geq 1$ except when $n = 2$. In this article, we will not discuss code construction for the case when $n = 2$. (For the code construction of case $n = 2$, see [4].)

Case 3: $n = 2^s q'$, where $s \geq 2$ and $q' > 1$ is odd. Consider optimal ST-error detecting codes C' and C'' with tuples of same length m. Suppose C' and C'' are defined over sets Q' and Q'' having q' and q'' elements respectively. Define a set $Q = Q' \times Q''$ such that $|Q| = q = q'q''$, where each element is a 2-tuple: (q', q''), $q' \in Q'$, $q'' \in Q''$. On the set Q, we define q-ary code C as follows:

$$C = \{(\tau_0', \tau_0'')(\tau_1', \tau_1'') \ldots (\tau_{m-1}', \tau_{m-1}'') : \tau_0'\tau_1' \ldots \tau_{m-1}' \in C' \text{ and } \tau_0''\tau_1'' \ldots \tau_{m-1}'' \in C''\}.$$

Since $|C'| = M'$ and $|C''| = M''$, then $|C| = M'M''$. Since C' and C'' are optimal ST-error detecting codes, C is also an ST-error detecting code. For this case, assume q' is odd and $q'' = 2^s$, $s \geq 2$.
Let C' be a code with m-tuples over $\mathbb{Z}_{q'}$, as constructed in Case 1.

$$C' = \{\tau'_0\tau'_1\ldots\tau'_{m-1} : \sum_{i=0}^{m-1} a'_i\tau'_i = t' \bmod q'\},$$

where

$$a_i = \begin{cases} 1, & \text{if } m-i \text{ is odd,} \\ 2, & \text{if } m-i \text{ is even.} \end{cases}$$

Similarly, let C'' be a code with m-tuples as constructed in Case 2, Construction 2 for finite field of order q''.

$$C'' = \{\tau''_0\tau''_1\ldots\tau'_{m-1} : \sum_{i=0}^{m-1} a''_i\tau''_i = t''\},$$

where

$$a_i = \begin{cases} e, & \text{if } m-i \text{ is odd,} \\ a, & \text{if } m-i \text{ is even.} \end{cases}$$

Then the resulting code C is an optimal q-ary ST-error detecting code. Since C' and C'' are optimal codes of $(q')^{m-1}$ and $(q'')^{m-1}$ codewords, C is also an optimal ST-error detecting code with $(q'q'')^{m-1}$ codewords.

Example 4.5. Consider the codes constructed in Example 4.1 and 4.3 respectively. We have $m = 3$, $q' = 3$ and $q'' = 2^2 = 4$, this implies $n = q'q'' = 12$. We get :

$$C' = \{000, 011, 102, 022, 201, 110, 121, 212, 220\}$$

and

$$C'' = \{000, 012, 101, 023, 202, 113, 122, 210, 221, 233, 320, 031, 303, 130, 311, 332\}.$$

Then using construction in Case 3, we get the following code:

$$C = \{(\tau'_0, \tau''_0)(\tau'_1, \tau''_1)(\tau'_2, \tau''_2) : \tau'_0\tau'_1\tau'_2 \in C' \text{ and } \tau''_0\tau''_1\tau''_2 \in C''\},$$

that is,

$$C = \{000000, 001112, 110021, 002223, 220012, 111103, 112212, \ldots\},$$

which is an optimal ST-error detecting code and cardinality of C is: $12^{3-1} = 12^2 = 144$. □

5 CONCLUSION

In this article, we have discussed how to find an optimal n-ary ST-error detecting code of length m by using some concepts of abstract algebra. Mainly, we discussed three cases when (i) n is odd, (ii) $n = 2^r$, where $r \geq 2$, and (iii) $n = 2^s q'$, where $s \geq 2$ and $q' > 1$ is odd. In all three cases number of optimal n-ary ST-error detecting code is n^{m-1}. One can find optimal n-ary ST-error detecting code of length m for the cases when (iv) $n = 2$ and (v) $n = 2p$, where $p > 1$ and p is odd. For more detail one can see the paper written by A. G. Khaled in 1998 (see [4]).

References

[1] Gallian, J. A., *Contemporary Abstract Algebra*, Fourth Edition, Narosa Publishing House, 1999.

[2] Dummit, D. S. and Foote, R. M., *Abstract Algebra*, Englewood Cliffs, N. J., Prentice Hall, 1991.

[3] Verhoeff, J., *Error Detecting Decimal Codes*, Mathematisch Centrum Amsterdam, 1969.

[4] Khaled, A. G., *Detecting Substitutions and Transpositions of Characters*, The Computer Journal, Volume 41 No. 4, 1998.

[5] Pfleeger, C. P. and Pfleeger, S. L., *Security in computing*, Third Edition, Prentice Hall PTR, 2002.

[6] *Stack Exchange*, https://mathoverflow.net/questions.

(MENTOR) DR. SUNIL KUMAR YADAV, ASSISTANT PROFESSOR, DEPARTMENT OF MATHEMATICS, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
skymath.bhu@gmail.com

APURVAA MITTAL, B.SC.(H) MATHEMATICS, 4TH SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
gaurimittal.apurvaa@gmail.com

RAJSHREE CHANDEL, B.SC.(H) MATHEMATICS, 4TH SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
rajshreechandel@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (79-86)

Let's Play a Game

Dr. J. Darbari and S. Arya

Abstract

This paper focuses on combinatorial game theory, highlighting impartial games, taking the well-known game of Nim as a base example using Bouton's Theorem. It further delves into the Sprague-Grundy Theorem, which is applied to variants of the game of Nim. Using the Sprague-Grundy Theorem, optimal strategies that can be adopted by a player can be calculated and consequently the game can be solved.

Keywords: Game of Nim, Combinatorial Games, Bouton Theorem, Sprague Grundy Theorem

1 Motivation

Games have been around for centuries. It is human nature to indulge in competition even for the purposes of recreation. Some games are purely based on luck. A player's win is determined by the roll of a dice or flip of a coin. One can, at most, determine the probability of one's win in these games. But there are games where a player's win can be determined even before a single move is made. This paper explores games that can be solved using mathematics. Let us play a similar game, the rules of which are as follows:

- Two players take turns removing as many fruits as they want from any of the piles they choose
- The player must remove fruits from only one pile at a time and they must remove at least one piece of fruit
- The player to remove the last piece of fruit wins

Suppose there are three piles of fruits, containing 3 apples (named p), 4 mangoes (named q) and 5 bananas (named r) respectively. There are two players, A and B. Player A starts the game. The game unfolds as follows:

Player A takes 2 apples from p. Player B takes 3 bananas from r. Player A takes 1 mango from q. Player B takes 1 mango from q. Player A takes all remaining apples. Player B takes 1 mango from q. Player A takes 1 banana from r. Player B takes all remaining mangoes. Player A takes the last banana from r and wins the game. The same is represented in Figure 1.

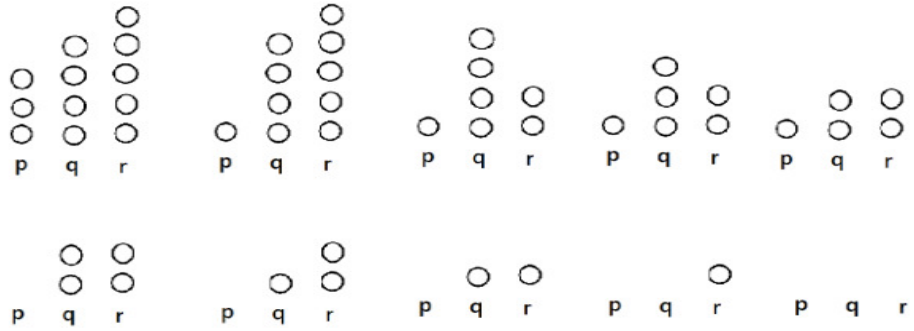


Figure 1: Step by step visualisation

This makes one wonder if Player A had an advantage because he/she started first? Will the first player win no matter what the initial configuration of the heaps are? Is there a way to figure out the outcome before the game is even played? Indeed. The game of Nim and any other combinatorial game can easily be solved using the Sprague Grundy Theorem. But first, what exactly is the game of Nim?

2 Introduction

The game of Nim [2] is the simplest impartial combinatorial game. The rules of the game are easy. Two players take turns removing (or 'nimming') as many stones as they want from any of the piles they choose. The player must remove stones from only one pile at a time and they must remove at least one stone. The player to remove the last stone wins. There can be n number of piles and each pile can have the same or different number of stones.

2.1 What are Combinatorial Games?

Combinatorial games are those which are played by two players, each taking turns alternatively and both players have perfect knowledge of the game. Thus, the players cannot move in a random fashion and chance does not act as a factor. These games must always end that is there shouldn't be a situation where the game is stuck in an endless loop. Depending on the rules of the game, there must be a win, a loss or a tie. It is widely accepted that in a normal combinatorial game, if a player cannot move, he/she loses, making the player who played last, the winner. Whereas in a misère combinatorial game, the player unable to make a move, wins. These few characteristics make combinatorial games different from those in classical game theory.

Combinatorial games are of two types - Impartial and Partisan [5]. An impartial combinatorial game is one in which either of the players can move any of the pieces on the board. This also implies that at any point in the game, both the players have all possible moves available from any position. On the other hand, partisan games are those in which each player has his/her set of pieces he/she is allowed to move. To illustrate, consider a classic game of chess, wherein one player only moves black pieces and the other moves only the white ones. Similarly, in a game of tic-tac-toe, one player

can either use crosses or knots. Hence, in partisan games there is a restriction on the movement of pieces in the game.

3 Theoretical Background

3.1 *P* and *N* Positions

P and N positions describe the game state at any given point of time during the play [3]. P positions are classified as ones where the previous player can win by playing optimally. It means that P positions are good to move to. Whereas, N positions are ones where the next player can win by playing optimally. N positions are bad to move to. Every position is either a P position or an N position. There is always a move to move from an N position to a P position whereas, it is not possible to move from a P position to another P position. Thus, from a P position it is only possible to move to an N position. To label all positions in a game, a few simple steps need to be followed:

- Every terminal position (the position where no more moves can be made) is labelled as a P position, as clearly the previous player wins
- All positions immediately leading to the terminal position must be labelled as N positions
- Find positions that lead to only N positions, label them as P positions
- Keep repeating the process till all positions are labelled

And this is the logic behind winning a combinatorial game. Your strategy must be to move to a P position. This leaves your opponent no choice but to move to an N position. You can now move to at least one P position. Since the terminal position is a P position, you eventually win.

Consider Figure 2 as a hypothetical game. We now try to label each node as a P position or an N position.

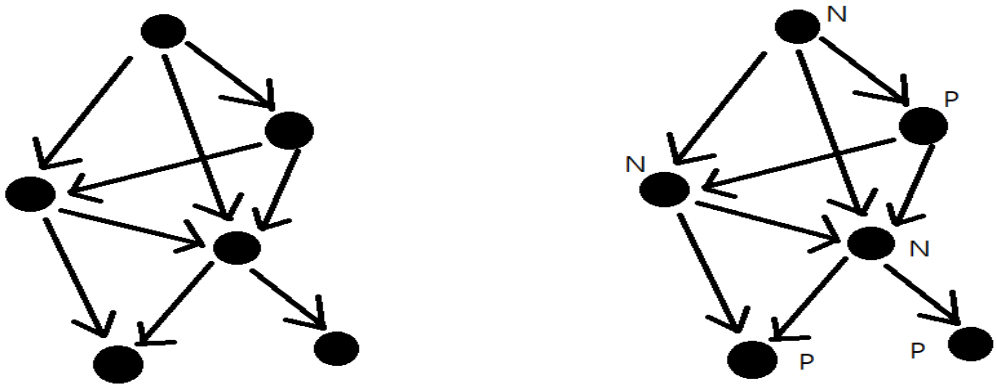


Figure 2: Labelling P and N positions

Coming back to the game of Nim, expressing the game state as an n-tuple. Consider a one pile version of the game. The only P position in this case is the terminal position. All others are N

positions. This is a trivial case. In the case of two piles, the P position is $(1, 1)$ implying that both piles have just one stone left. The next player can only remove one of the piles, leaving the other for you to remove, making you the winner. With the same logic, any position where both the piles have same number of stones (a, a) is a P position. Positions where one pile has one stone and the other has a greater number, $(1, a)$ will all be N positions as the next player has the opportunity to move the game state to $(1, 1)$ and then winning. Combining these cases, P positions are () and (a, a) while N positions are (a) and (a, b) where $a \neq b$. Moving onto games with the smallest pile of size t, we obtain P positions as shown in Table 1, adapted from [3]:

t	P positions
1	$(1,2,3),(1,4,5),(1,6,7)\ldots$
2	$(2,4,6),(2,5,7),(2,8,10)\ldots$
3	$(3,4,7),(3,5,6),(3,8,11)\ldots$
4	$(4,8,12),(4,9,13),(4,10,14)\ldots$
5	$(5,8,13),(5,9,12),(5,10,15)\ldots$

Table 1: P positions with smallest pile of size t

3.2 Nim-Sum

XOR also known as Exclusive OR or Exclusive Disjunction is a bit-wise operator, which means it operates on binary numbers (numbers whose base is 2). The XOR operator results in a 0, if even number of 1's are present in the input, and results in 1, if there are odd number of 1's in the input. It is denoted by $\oplus$. For example: Let us take two binary numbers, 00100100 and 00100001 and we get $00100100 \oplus 00100001 = 00000101$. The cumulative XOR value of the number of stones in each heap at any point in the game, is called the nim-sum at that point in the game [1].

Theorem 1 (Bouton)**.** *The theorem states that in a game of Nim of k piles namely,* $n_1, n_2, ..., n_k$ *is a* P *position if and only if* $n_1 \oplus n_2 \oplus ..... \oplus n_k = 0$ *[1].*

In simpler words, in a normal game of Nim, the first player has a winning strategy if the initial nim-sum of the sizes of the heaps is non-zero. If the nim-sum is zero, the second player has a winning strategy.

What if we make a change in the classic game of Nim? Instead of taking any number of coins, we impose a restriction that a player can take either 1, 2 or 3 coins only at each turn. To decode the winner, we use the Sprague-Grundy Theorem.

Theorem 2 (Sprague Grundy)**.** *The theorem states that if there is a composite game (more than one sub-game) made up of N sub-games and two players, A and B and if both A and B play optimally (i.e., they don't make any mistakes), then the player starting first is guaranteed to win if the XOR of the Grundy numbers of position in each sub-game at the beginning of the game is non-zero. Otherwise, if the XOR evaluates to zero, then player A will lose definitely, no matter what [2].*

3.3 Grundy Numbers

Grundy Numbers, commonly known as 'Nimbers' of a position, is a recursively defined function giving the minimum excludant value (MEX) of all possible game states after execution of that position. Minimum Excludant (MEX) is the smallest non-negative number not present in a set. Table 2, which has been adapted from [4], shows the Minimum Excludant (MEX) of some examples:

MEX($\emptyset$)	0
MEX($\{1, 2, 3\}$)	0
MEX($\{0, 2, 4, 6, \dots\}$)	1
MEX($\{0, 1, 4, 7, 12\}$)	2

Table 2: Minimum Excludant Values of some examples

Taking the case of our hypothetical game, we now determine the Grundy numbers for each node as shown in Figure 3.

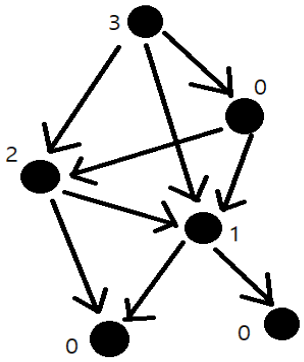


Figure 3: Labelling Grundy numbers

Now, to find the grundy numbers of the modified Nim game,
Grundy $(0) = 0$, as if there are no coins left to pick, the player loses.
Grundy (1) = MEX $(\{0\}) = 1$
Grundy (2) = MEX $(\{0, 1\}) = 2$
Grundy (3) = MEX $(\{0, 1, 2\}) = 3$
Grundy (4) = MEX $(\{1, 2, 3\}) = 0$
We see that if there are 4 coins left for the player, he/she is bound to lose. Also, Grundy of any number greater than or equal to 4 can be calculated as:
Grundy (n) = MEX $(\{$Grundy $(n-1)$, Grundy $(n-2)$, Grundy $(n-3)\})$.

This information has been summarised in the table below.

n	0	1	2	3	4	5	6	7	8	9	10
Grundy (n)	0	1	2	3	0	1	2	3	0	1	2

Table 3: Grundy numbers for modified game of Nim

Notice that all positions with Grundy = 0 are P positions and all positions with Grundy $\neq 0$ are N positions. Now with all our components ready let's apply the Sprague-Grundy theorem to a modified version of Nim, to predict the winner. Suppose there are three heaps with 3, 4 and 5 coins respectively. Player A and Player B take turns alternatively. On one's turn, a player can either remove 1, 2 or 3 coins from any one heap. The player to pick up the last coin wins. Firstly, we need to divide our game into sub-games, namely the three heaps of coins make 3 sub-games. Next, we calculate Grundy Numbers of each sub-game, using the table.
Grundy (3) = MEX ({Grundy (0), Grundy (1), Grundy (2)}) = MEX ($\{0, 1, 2\}$) = 3
Grundy (4) = MEX ({Grundy (1), Grundy (2), Grundy (3)}) = MEX ($\{1, 2, 3\}$) = 0
Grundy (5) = MEX ({Grundy (4), Grundy (3), Grundy (2)}) = MEX ($\{0, 3, 2\}$) = 1
Converting the Grundy Numbers to binary we evaluate $01 \oplus 00 \oplus 11 = (10)_2 = (2)_{10}$ which is non-zero. Hence, the player who starts the game will definitely win, given that both players play optimally. This implies that the game of Nim is solved, or in other words, the winner can be determined before the first move, provided that the players play optimally.

4 Model

The same can be implemented by a code in Java. This code checks all possible combinations of two heaps wherein each heap can have at most 10 coins. It calculates the number of times the first player, say A wins and also the number of times the second player, namely B wins. Moreover, it prints all the initial configurations of the two piles, as an n-tuple, in which Player A is the winner in a normal game of Nim, wherein the players are free to remove any number of coins and the player to remove the last coin is the winner. We also assume that both players play optimally, without making mistakes.

4.1 Code in Java

```
import java.util.*;
import java.io.*;
class subdemo
{
public static void main(String args[])throws IOException
{
int heaps=2;
//creating arrays to store all the combinations
int A[]=new int[heaps];
String B[]=new String[heaps];
int C[]=new int[heaps];
//initialising counter variables
int nimsum=0;
int countA=0;
int countB=0;
//for loops to check for all possible combinations
for(int m=1; m <= 10; m++)
{
```

```
A[0]=m;
B[0]=Integer.toBinaryString(A[0]);
C[0]=Integer.parseInt(B[0]);
for(int n=m; n <= 10; n++)
{
A[1]=n;
//converting decimal numbers to their binary representation
B[1]=Integer.toBinaryString(A[1]);
C[1]=Integer.parseInt(B[1]);
for(int i=0; i<heaps; i++)
{
//calculating the XOR value
nimsum=nimsum^C[i];
}
if(nimsum==0)
{
countB++;
}
else
{
System.out.print("("+m+","+n+"), ");
countA++;
}
//resetting the counter
nimsum=0;
}
}
System.out.println();
//printing in the required format
System.out.println("Player A wins "+countA+" times.");
System.out.println("Player B wins "+countB+" times.");
}
}
```

4.2 Output

$(1,2),(1,3),(1,4),(1,5),(1,6),(1,7),(1,8),(1,9),(1,10),(2,3),(2,4),(2,5),(2,6),(2,7),(2,8),(2,9),$
$(2,10),(3,4),(3,5),(3,6),(3,7),(3,8),(3,9),(3,10),(4,5),(4,6),(4,7),(4,8),(4,9),(4,10),(5,6),$
$(5,7),(5,8),(5,9),(5,10),(6,7),(6,8),(6,9),(6,10),(7,8),(7,9),(7,10),(8,9),(8,10),(9,10).$
Player A wins 45 times.
Player B wins 10 times.

Similarly, after making simple modifications to the code a little we obtain results for a 3 pile game of Nim as Player A wins 212 times and Player B wins 8 times. Taking this further, in a 4 pile game, Player A wins 495 times and Player B wins 220 times. So indeed, every game does depend on the initial configuration of the number of heaps as well as the number of coins in each heap.

5 Conclusion

The practical strategy to win a game of Nim is to think in binary! Reduce the piles to two non-zero heaps with equal number of stones so that the nim-sum will be $x \oplus x = 0$. To make nim-sum zero, attempt to leave an even number of heaps with stones in powers of 2, starting with the largest power possible. Another way to describe the strategy is that we express all the pile heights in binary, and we want an even number of 1's in each binary place position. In simpler words, try to attain a balanced position, wherein there are even number of 1's. This will be a P position. Any move by your opponent will cause the game to go into an unbalanced state. There is always a move to move an unbalanced game to a balanced one.

References

[1] Siegel A.N., *Combinatorial Game Theory*, Graduate Studies in Mathematics, Volume 146, American Mathematical Society.

[2] Legner P.,*Combinatorial Game Theory: Analysis of the Game of Nim and some interesting Variants*, University of Cambridge, 2010.

[3] Chang A.,*Combinatorial Game Theory*, University of Chicago, 2011.

[4] https://www.geeksforgeeks.org/introduction-to-combinatorial-game-theory/

[5] https://www.cs.cmu.edu/afs/cs/academic/class/15859-f01/www/notes/comb.pdf

(Mentor) Dr. Jyoti Darbari, Assistant Professor, Department of Mathematics, Lady Shri Ram College for Women, New Delhi
jydbr@hotmail.com

Swasti Arya, B.Sc.(H) Mathematics, 4th Semester, Lady Shri Ram College for Women, New Delhi
swasarya@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (87-94)

Mathematical Explorations: Bridging the Gap Between School and College Mathematics

Dr. J. B. Ghosh

Abstract

This paper aims to bring out the vast difference in how mathematics has been understood by the students and teachers at various levels of education. It attempts to highlight the true sense or meaning of the subject and the ways in which students can be encouraged to explore the subject and to learn.

Keywords: Mathematical Explorations, Fractal Explorations, Sierpinski Triangle, Koch Snowflake, Nim, Tower of Hanoi, Spreadsheets, Mathematical Thinking

1 Introduction

Mathematics has for years been the common language for scientific expression. More so, it is universally recognized as the language for classification, representation and analysis in almost every field of human endeavor. It forms a significant part of a child's education at school and later may continue to be a part of her higher education, as there is hardly any field, which is not impacted by mathematics. Notwithstanding its importance in education, Mathematics continues to be the most feared subject and is usually perceived as being 'tough' or 'difficult'. 'Its not for everyone' or 'you either have it in you or you don't' are common responses which are often heard about mathematics. This perception about mathematics, at least at the school stage, is primarily because it is largely taught in an abstract manner without any connections to real life. Students (and even teachers) are often unaware of the beauty and relevance of the discipline in the myriad facets of life. The emphasis on applying procedures, developing manipulative skills to attain the one right answer and preparation for the board examinations leads students to develop a very blinkered approach to the subject. While some students are able to master the skills and sail through school mathematics, a large majority struggles and finally abandons the subject at the college level.

When students encounter mathematics in college, they often find themselves faced with a math quite different from the kind they encountered in school. Initially, in the first year of college, mathematics feels like a new discipline and this leads to considerable difficulties. As Keith Devlin, in his book on *Introduction to mathematical thinking*, points out that the transition from high school to college level mathematics is a difficult one. He attributes this to a change in emphasis.

> In high school, the focus is primarily on mastering procedures to solve various kinds of problems. This gives the process of learning very much a flavour of reading and absorbing the recipes of

a kind of a mathematical cookbook. At college, the focus is on learning to think in a different, specific way – to think like a mathematician [[1], pp.1].

While school mathematics focuses on applying standard methods and procedures, college mathematics requires students to develop the skills of solving new and perhaps non-standard problems. There is an emphasis on rigor and formal argumentation and proof begins to play a central role. Learning mathematics should be about developing mathematical thinking and the Position paper on teaching of mathematics of the National Curriculum Framework (National Council for Educational Research and Training [NCERT], 2006) articulates this very important aspect.

Clarity of thought and pursuing assumptions to logical conclusions is central to the mathematical enterprise. There are many ways of thinking, and the kind of thinking one learns in mathematics is the ability to handle abstractions [[5],pp.1].

Not only is there a shift in emphasis in mathematics learning as students move from school to college, there is also a shift in expectations from students pursuing mathematics in college. Keith Devlin aptly describes this gap between school and college mathematics by saying that success in school mathematics requires one to "think inside the box" whereas to succeed in college one needs to "think outside the box". How then should this gap be bridged? What can be done at the school stage so that this transition is easier for students?

The Position paper on teaching of mathematics makes a few useful recommendations, which can actually help bridge the gap between school and college mathematics. Firstly it states that mathematics teaching at all levels (of school) be made more student centered, so that students understand the basic structure of mathematics, learn to think mathematically and relate mathematics to life experiences. Secondly, it suggests a "shift from content to processes" and recommends that a pedagogy be evolved which emphasises the processes of learning mathematics such as visualization, estimation, approximation, use of heuristics, reasoning, proof and problem solving. Thirdly, it also recommends the use of technology in the form of computer software to enable students to visualize and explore mathematical concepts.

Carefully designed mathematical explorations can actually give students a glimpse of the kind of mathematical thinking that is required at the college level. In this article we will highlight the role of such explorations in engaging students in the processes of mathematical thinking. We shall illustrate two examples briefly, one in which secondary school students explored the topics of Fractals using a spreadsheet and another in which students analysed different games and in the process learnt new mathematical concepts.

2 Exploration of Fractals

The study of Fractals, though not included in school curricula, is a rich source of exploratory tasks. Fractal constructions can be used to introduce students to notions of iteration and recursion, which are important for mathematical thinking. The author conducted a series of fractal activities with students of grade 11 in which they explored various geometric fractals such as the Sierpinski triangle, Sierpinski Carpet, Koch Snowflake and Pythagorean tree and also created their own fractal

patterns. The activity series was primarily designed to introduce them to the topic of geometric sequences and to the concepts of self-similarity and fractal dimension.

1.jpg

Figure 1: The Sierspinski Triangle - stages 0, 1, 2 and 3; (source: https://en.wikipedia.org/wiki/Sierpiński_triangle)

The series of explorations began with the Sierpinski triangle construction, in which the midpoints of the sides of an equilateral triangle are joined, to obtain four smaller triangles after which the center triangle is removed. This resulting figure with three smaller equilateral triangles and a triangular 'hole' is referred to as stage 1. The recursive process of creating smaller equilateral triangles and removing the center triangle was repeated to obtain subsequent stages. Figure 1 shows stages 0, 1, 2 and 3 of the construction.

stage	no of shaded triangles	shaded area
0	1	1
1	3	0.75
2	9	0.5625
3	27	0.421875
4	81	0.31640625
5	243	0.237304688
6	729	0.177978516
7	2187	0.133483887
8	6561	0.100112915
9	19683	0.075084686
10	59049	0.056313515
11	177147	0.042235136
12	531441	0.031676352
13	1594323	0.023757264
14	4782969	0.017817948
15	14348907	0.013363461
16	43046721	0.010022596
17	129140163	0.007516947
18	387420489	0.00563771
19	1162261467	0.004228283
20	3486784401	0.003171212

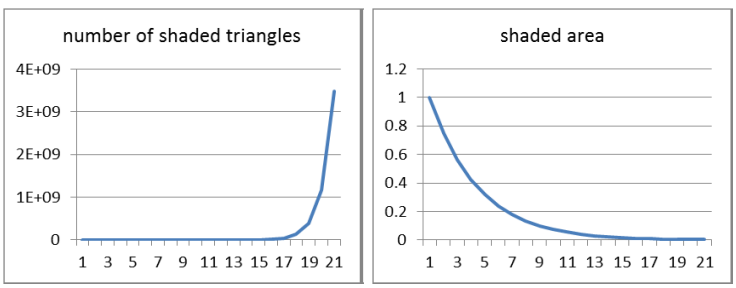
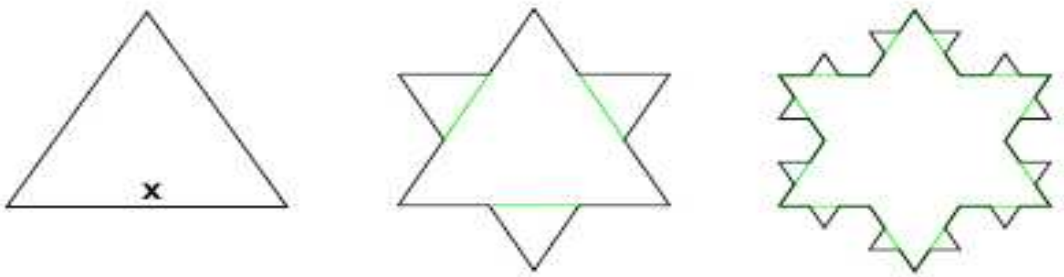


Figure 2: Numerical and graphical representations of the geometrical sequences arising from the Sierspinski triangle construction in a spreadsheet

Several tasks were assigned to students to steer their investigations. In the first task, they had

to count the number of shaded triangles at each stage and find a general formula for the nth stage. They were also required to compute the area of the shaded portions at each stage and generalise to the nth stage. Tabulating the results led to the sequences 1, 3, 3^2, 3^3,... and 1, $\frac{3}{4}$, $(\frac{3}{4})^2$, $(\frac{3}{4})^3$,... for the number of shaded triangles and shaded area respectively. Students also identified the explicit rules as 3^n and $(\frac{3}{4})^n$ for the generalised nth terms of these sequences. In the second task, students had to relate the formula of stage n with that of stage $n-1$ for both sequences. Writing the recursive relation $S_n = 3 \times S_{n-1}$ for the number of shaded triangles, took more scaffolding, but after that students came up with the recursive formula $A_n = \frac{3}{4} \times A_{n-1}$ for the area, more easily. The third task required students to describe what would happen as n, the number of stages, approached infinity. They conjectured that the number of shaded triangles would 'become very large'and some used the phrase 'will approach infinity'. For shaded area, many students said it would get 'lesser and lesser'. A numerical and graphical exploration on a spreadsheet revealed the 'big picture'of the fractal at higher stages. Students concluded that as the number of stages increases, the number of shaded triangles grows exponentially whereas the shaded area approaches 0 (See Figure 2).



3.png

Figure 3: Stages 0, 1 and 2 of the Koch Snowflake; (source: https://en.wikipedia.org/wiki/Koch_snowflake)

Another activity in the fractal series was the Koch snowflake exploration, which also turned out to be very rewarding. It enabled students to explore various patterns in the snowflake construction through pictorial, tabular and symbolic representations and also to make connections among them. Stage 0 of this fractal comprises an equilateral triangle. To obtain stage 1 each side of the equilateral triangle is trisected, the middle segment removed and smaller equilateral triangles of side one third of the original side length are raised on these. Thus, to obtain stage 1 four shorter line segments replace each side of the original equilateral triangle. This process was repeated on every line segment of stage 1 to obtain stage 2, and so on. Figure 3 shows stages 0, 1 and 2 of the Koch Snowflake.

Attributes such as number of line segments, perimeter and area (enclosed by the curve) were observed at each stage. Students came up with geometric sequences, recursive formula as well as explicit rules for each attribute. Computing the area enclosed by the snowflake at each stage was interesting as well as challenging. Since the snowflake grows at each successive stage, some additional area (comprising equilateral triangles) gets added to the previous stage. The additional area terms led to the geometric sequence $\frac{\sqrt{3}}{4} \times \frac{3}{4}$, $\frac{\sqrt{3}}{4} \times \frac{4}{3^3}$, $\frac{\sqrt{3}}{4} \times \frac{4^2}{3^5}$, and so on.

The total area (stage 0 onwards) was written out as $\frac{\sqrt{3}}{4} + \frac{\sqrt{3}}{4} \times \{\frac{1}{3} \times 1 + \frac{4}{3^2} + \frac{4^2}{3^4} + \cdots\}$ where the terms inside the bracket formed a geometric series with a multiplying factor $\frac{4}{9}$. Applying $S_\infty = \frac{a}{1-r}$

to these terms yields $S_\infty = \frac{1}{1-\frac{4}{9}} = 9/5$. Thus the area of the Koch snowflake, as n, the number of stages, approaches infinity, was obtained as $= \frac{\sqrt{3}}{4} + \frac{\sqrt{3}}{4} \times \frac{1}{3} \times \frac{9}{5} = \frac{\sqrt{3}}{4}(1 + \frac{3}{5}) = \frac{\sqrt{3}}{4} \times \frac{8}{5} = \frac{2\sqrt{3}}{5}$ which is approximately equal to 0.69. This was very fascinating to the students as they saw a real context for applying the formulae $S_\infty = \frac{a}{1-r}$ and were able to visualize a situation where an infinite process led to a steady value.

Students found the Koch snowflake exploration very rewarding as it enabled them to explore the topic of geometric sequences using pictorial, tabular, symbolic and graphical representations and also make connections among them. Ghosh (2016) presents a more detailed version of this activity (see [2]).

3 The Mathematics of Games

Games and puzzles provide ample opportunity for mathematical explorations. Here we shall briefly describe two games, which are popular among school students and lead to interesting investigations. The first one is a simple version of the game of NIM and the second one is the Tower of Hanoi (ToH) puzzle. These were explored by students of grade 9 and 11 respectively.

The NIM is a two–person game. Students were introduced to it as follows:
Assume that there are 21 stones on the table. Two players (player 1 and player 2) play the game by taking turns alternately. Player 1 starts the game. During a move a player can remove either one, two, or three stones. If a player is unable to make a move during his turn, he loses (this only happens when there are no stones left). Analyse the situation in which player 1 wins the game. This game is referred to as the (1,2,3)-NIM.

The grade 9 students played the game in pairs (using coins, tokens or simply by drawing in their notebooks). They observed that if there is one, two or three stones left in player 1's turn then he stands to win. However if there are four stones left then player 1 has no choice but to lose (for if he removes one, two or three stones, player 2 will remove the remaining stones and win). To scaffold their thinking further, students were asked to complete a table where they had to indicate 'W' for winning and 'L' for losing, against the number of stones left at the player 1's turn. Here n denotes the number of stones left at player 1's turn. Filling this table led to much excitement among students as they identified winning strategies for both players. The output is shown in figure 4. They realized that player 1 loses if n, the number of left over stones is a multiple of 4. However in all other cases player 1 stands to win. The L-W sequence also follows a pattern as the cycle LWWW repeats itself.

Students were then introduced to the concept of 'modulo n' and to the basics of modular arithmetic. Using this new notation they represented the winning strategies for the (1,2,3)-NIM as follows:

Player 1 wins if $n \equiv 1, 2, 3 \bmod(4)$
Player 2 wins if $n \equiv 0 \bmod(4)$

Following this students were asked to explore other versions of NIM such as (2,3)- NIM, (1,4)-NIM and (1,3,4)- NIM. In each case they came up with winning strategies and represented their findings using modular arithmetic. For example, (1,4)– NIM had the repeating cycle WLWWL of period 5. Also

(77).png

n	0	1	2	3	4	5	6	7	8	9
	L	W	W	W	L	W	W	W	L	W

n	10	11	12	13	14	15	16	17	18	19	20
	W	W	L	W	W	W	L	W	W	W	L

Figure 4: Winning positions for player 1 in the game of NIM

4.png 

Figure 5: A wooden model of the Tower of Hanoi puzzle; (source: https://sites.google.com/a/jodogyan.org/www/teachinglearningmaterials/teaching-learning-material-2)

Player 1 wins if $n \equiv 1, 3, 4 \bmod(5)$
Player 2 wins if $n \equiv 0, 2 \bmod(5)$

After exploring various versions of the game students came up with the conjecture that NIM games are periodic. This important conjecture leads to the theorem : For any $h_1, ..., h_n$ the games $(h_1, \ldots, h_n)$-NIM is periodic.

The Tower of Hanoi is an interesting puzzle, as it provides a wonderful context for introducing students to recursive and explicit thinking. In a study by the author, the puzzle was administered to 28 students of grade 11. To begin the activity, wooden models of the puzzle were distributed to the students in pairs. The model comprises three vertical pegs fixed on a wooden base and six wooden circular discs of reducing radii placed on one of the pegs (see Figure 5). The problem lies in shifting the tower of discs from one peg to another using the third peg as an intermediary, subject to two constraints - *only one disc can be moved at a time, and a bigger disc can never be placed on top of a smaller one while making the moves.* The shift has to be achieved using a minimum number of moves.

To explore the puzzle students were given the following tasks:

Task 1: Let n represent the number of discs and $T(n)$ the minimum number of moves required for shifting n discs. Find $T(n)$ for $n = 2$, 3 and 4. Observe the pattern as n varies from 2, 3, 4 etc.

Can you express $T(n)$ in terms of $T(n-1)$?

Task 2: Write an explicit rule for $T(n)$ in terms of n. (Hint: Find the nth term of the sequence of number of moves)

Task 1 was meant to lead students to discover the recursive rule for obtaining the minimum number of moves. Initially students explored the puzzle by trying different moves with the discs. When asked to estimate the minimum number of moves required to shift the 6 discs to another peg, the answers ranged from 30 to 74! The researcher then asked students to simplify the problem and attempt the moves with lesser numbers of discs. To start with, most pairs observed that $T(2) = 3$.

Further they labelled the pegs as X, Y and Z and worked with three discs manually. They observed that $3 + 1 + 3 = 7$ moves were required to shift three discs. For example, if the discs had to be moved from X to Y, the two smaller discs had to be moved from X to Z (to release the largest disc) in three moves, following which the largest disc could be shifted to Y in one move. Finally the two smaller discs (on Z) would require another three moves to be shifted back to Y. This observation led them to generalise the pattern to larger number of discs. For example, $T(4) = 7 + 1 + 7 = 2 \times 7 + 1 = 15$ and $T(5) = 15 + 1 + 15 = 2 \times 15 + 1 = 31$ and so on. The researcher helped them to express the recursive relation symbolically as $T(n) = 2 \times T(n–1) + 1$.

To obtain the explicit formula, students observed the sequence of minimum number of moves, that is, 1, 3, 7, 15, 31 and tried to find the nth term. The differences between successive terms - 2, 4, 8, 16... led them to conjecture that the nth term was related to the powers of 2. The observations $T(2) = 3 = 2^2–1$, $T(3) = 7 = 2^3–1$ helped them to arrive at the conjecture, $T(n) = 2^n–1$. To prove this conjecture mathematical induction was used. Students were familiar with the steps of induction, as the topic is a part of the grade 11 syllabus. The following steps were used to prove the explicit rule:
Step 1: $T(1) = 2^1–1 = 1$.
Step 2: Assuming $T(k) = 2^k–1$, for a natural number k.

To prove $T(k + 1) = 2^{k+1}–1$
$T(k+1) = 2 \times T(k) + 1 = 2 \times (2^k–1) + 1 = 2^{k+1} - 2 + 1 = 2^{k+1} - 1$. This proved to be very exciting for students as they experienced mathematical induction in a concrete way. Further, the puzzle, also led to interesting counting problems. For instance, students explored the number of arrangements of n discs on three pegs and concluded that the number of arrangements for n discs on three pegs is $3n$. The ToH puzzle in fact lends itself to investigations in multiple ways. Ghosh (2018) describes how explorations of the puzzle by first year college students led to Hanoi Graphs and fractal like structures within them (see [3]).

4 Conclusion

The explorations described in this article provided students ample opportunity to engage in the processes of mathematical thinking articulated in the Position Paper on Teaching of Mathematics. In the fractal explorations, they engaged in the process of visualizing patterns followed by generalization - recursive as well as explicit. In doing so, they developed an insight into fractal geometry and also

explored geometric sequences using multiple representations. In particular, pictorial representations helped to explore the iterative growth of the fractals and spreadsheets helped to visualize the process numerically and graphically. In fact, as suggested by Pea (1987), in these explorations, technology (in the form of spreadsheets) played the role of an 'amplifier' (see [4]). It quickly produced computations and graphs and hastened the process of observing the properties of the fractals at higher stages.

Devising winning strategies for the game of NIM accorded many opportunities for observing periodic patterns and generalizing them using modular arithmetic. The ToH puzzle engaged students in the process of estimating the minimum number of moves, abstracting a recursive pattern followed by an explicit rule and finally proving the results using mathematical induction. While attempting these explorations, students selected between representations and created new ones, simplified and generalised problems, made conjectures and created new questions for further exploration. More importantly the tasks within the explorations gave students access to higher level mathematical concepts and encouraged them to think 'out of the box'. The positive student feedback convinced the author about the importance of such explorations in developing the kind of mathematical thinking needed to prepare students for pursuing mathematics in college. Such explorations have the potential to familiarise students with the processes of mathematical thinking and introducing these in the school curriculumm may be one of the ways of bridging the chasm between school and college mathematics.

References

[1] Devlin K., *Introduction to mathematical thinking*, Published by Keith Devlin, 2012

[2] Ghosh J. B., *Algebraic Thinking through Koch Snowflake Constructions*, Mathematics Teacher, National Council of Teachers of Mathematics (NCTM), 2016, 109 (9): 693-99

[3] Ghosh J. B., *Tower of Hanoi: Exploring Multiple Representations*, Mathematics Teacher, National Council of Teachers of Mathematics (NCTM), 2018, 111 (6): 446-52

[4] Pea R. D., *Cognitive technologies in mathematics education*, A. H. Schoenfeld (Ed.), Cognitive science and mathematics education, Hilldale, NJ: Erlbaum, 1987, pp. 89–122

[5] Position paper of National Focus Group on Teaching of Mathematics, 2006, National Council of Educational Research and Training, http://www.ncert.nic.in/rightside/links/pdf/framework/nf2005.pdf

Dr. Jonaki B. Ghosh, Assistant Professor, Department of Elementary Education,
Lady Shri Ram College for Women, New Delhi
jonakibghosh@gmail.com

Éclat, A Mathematics Journal
Volume 11, 2019-20 (95-102)

Maryam Mizakhani: The Mathematical Genius

Dr. Anuradha, K. Konghay and L. Parekh

Abstract

This paper traces the life and contributions of Maryam Mirzakhani, the first Iranian mathematician and first female winner of the Fields Medal since its inception in 1936. We attempt to shed light on Maryam's ground-breaking research, while also highlighting glimpses of the modest soul behind all the brilliant mathematical artistry.

Keywords: Mizakhani, Fields Medal, Hyperbolic Geometry, Ergodic Theory

1 The Birth of a Mathematical Genius

Maryam Mirzakhani, the first and the only woman to receive the Fields Medal, was born on May 12, 1977 in Tehran. With a passion to read and write, she never thought she would become a mathematician. When Maryam finished her elementary school and joined the Tehran Farzanegan School, an all-girls school which was administered by Iran's National Organization for Development of Exceptional Talents, the Iran-Iraq war had just ended, and the country's education system was more stable after the Islamic Revolution in 1979 [ASV18]. She met her dear friend Roya Beheshti in middle school, who recalls Maryam in middle school to be well read with a passion for writing, who could get into a heated debate related to social and political issues and would not tolerate any kind of prejudice against women.

Figure 1.1 Maryam grew up in Tehran during the Iran-Iraq War; (source: https://www.quantamagazine.org/maryam-mirzakhani-is-first-woman-fields-medalist-20140812/)

Her interest in mathematics was first developed by her brother when he told her about Gauss's solution to the problem of adding numbers from 1 to 100. She was astonished by the solution [IntCM08]. Mirzakhani did poorly in mathematics in middle school for a couple of years. But later, with a combination of a motivating mathematics teacher and Maryam's enthusiasm to learn, mathematics became her strength. In Iran, there were separate schools for girls and boys, but this didn't mean lack of opportunities for girls. In fact, the principal of Farzanegan School, Ms. Haerizadeh, was a resolute woman who would go beyond boundaries to deliver opportunities to her students.

While still in high school, Maryam and Roya's passion for mathematics led them to participate in a summer workshop conducted by Sharif University which introduced students to college-level mathematics. This played an important role in Maryam's developing interest for mathematics. Roya and Maryam were the first women to represent Iran at **International Mathematical Olympiad** in 1994, Maryam won a gold medal in 1994 and 1995, and obtained a perfect score in 1995.

Figure 1.2 Maryam Mirzakhani with Roya Beheshti at International Mathematical Olympiad, 1994; (source: https://www.ams.org/journals/notices/201810/rnoti-p1221.pdf)

She attained her bachelor's degree in Mathematics from Sharif University, Iran. Despite the fact that in 1990, there were many regulations imposed related to gender segregation, the university provided a lot of opportunities like problem solving sessions and workshops to motivate each student. Also, her family, teachers and friends were very supportive and encouraged her to pursue her dreams. Before starting with her graduate work, she had published three papers [98, 96, 95] and co-authored a book [99] with her good friend Roya Baheshti. Mathematics had become an integral part of Maryam's life. Iranian President Hassan Rouhani believed that Maryam's accomplishment was a symbol of power of the ambitious Iranian women and young people on an international platform [Stan].

2 Pathway to Formal Research

2.1 Steely tenacity of an early researcher

Figure 2.3 (source: http://muslimmirror.com/eng/remembering-maryam-mirzakhani-1977-2017/)

Maryam built a strong background in combinatorics and algebra before joining Harvard University, Cambridge, USA for graduate work. There, she participated in informal seminars on geometry and dynamics organized by Professor Curtis McMullen, a 1998 Fields Medal awardee. Although she could not comprehend much of what was taught in these seminars, she was amazed how McMullen could capture the beauty of the subject and deliver to the students in an uncomplicated way. Maryam was an inquisitive student, she regularly prepared questions to ask McMullen, although there existed a language barrier. When Maryam gave her first talk on McShane's identity for the punctured tours in McMullen's seminar, she was nervous about delivering her lecture in English, but her zeal and excitement for the subject was evident in her performance. One of her classmates said, "Maryam's lecture style that day was unforgettable. She would race ahead of herself at times, unable to contain her excitement, and barrage of words was accompanied by large circular hand gestures." [NoAMS18]

Eventually her lectures were refined, and the content was powerful. Her in depth understanding of the subject led to a significant contribution in mathematics. She was focused and had the courage to question the unsolved. McMullen believed that she had fearless ambition [Stan].

2.2 Doctoral thesis in hyperbolic geometry

Under the guidance of McMullen, Maryam completed her doctoral thesis, titled **"Simple geodesics on hyperbolic surfaces and volume of the moduli space of curves"**. Her thesis deduced path breaking results in hyperbolic geometry. Hyperbolic geometry is a non-Euclidean geometry that satisfies all of Euclid's postulates except the parallel postulate (given any straight line and a point not on it, there exists exactly one straight line which passes through that point and never intersects the first line, no matter how far they are extended.). In fact, in hyperbolic geometry, for any straight line and a point not on it, there exist at least two straight lines that pass through that point and do not intersect the first line. The Hyperbolic Triangle, a triangle drawn on the surface of any plane of negative curvature, has its three internal angles that add up to less than π.

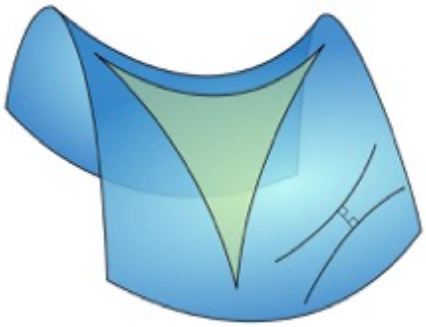


Figure 2.4 A hyperbolic triangle; (source: https://en.wikipedia.org/wiki/Hyperbolic_triangle)

Maryam gave the following formula to count the number of simple-closed[1] geodesics[2] of length $\leq L$ on a hyperbolic surface X of genus g, denoted $\sigma(X, L)$:

$$\sigma(\mathbf{X}, \mathbf{L}) = \mathbf{c}(\mathbf{X})\mathbf{L}^{6\text{g-}6} \text{ as } L \text{ tends to infinity}$$

$c(X)$ denotes the constant dependent on the hyperbolic structure of the surface. She concluded that the number of geodesics grow polynomially and depends on the surface. Through her work, she derived formulae to calculate the frequencies of different types of curves and constructed another proof to Witten's Conjecture.

Her thesis was eventually developed into three papers – a **proof of Witten's conjecture**, the **counting of simple closed geodesics on hyperbolic surfaces** [M-AM08] and the **volume of moduli space of curves** [MM07], which were published in three top journals of mathematics. Cumrun Vafa, Professor at Harvard University said, "Maryam beautifully exemplified that the pursuit of knowledge is a timeless, borderless, and yes, genderless adventure." [NoAMS18]

[1] A geodesic is said to be simple if it does not have any self-intersection points. A geodesic that starts and ends at the same point with the same direction is called a closed geodesic.

[2] A geodesic is a locally length-minimizing curve. It can be thought of as a generalization of the notion of a "straight line" from a plane to a surface, on which it represents in some sense, the shortest path between two points. A bug living in the surface and following a [geodesic] curve would perceive it to be straight.

3 From student to professor: Maryam's excellence continued

From 2004 to 2008, Maryam was a **Clay Mathematics Institute Research Fellow**[3] and an assistant professor at Princeton University, New York, USA. In 2008, she joined the Mathematics department of Standford, California, USA where she conferred the rank of Professor at an unusual young age of 31.

Figure 3.5 (source: https://www.claymath.org/library/annual_report/ar2008/08Interview. pdf)

She continued to pursue the study of geodesics in hyperbolic geometries. In a series of key papers, she developed theorems to characterize when such geodesics would be closed paths, like the route of an airplane circumnavigating Earth, or open, like the zigzag motion of light through a fiber optic cable. Such patterns deeply connected with the wind-tree model, billiards, the illumination problem, and diffusion in general.

She won the **AMS Blumenthal Award** in 2009 and the **Satter Prize** in 2013. Maryam was a pure mathematician who specialised in **Moduli Spaces, Teichmüller Theory, Hyperbolic Geometry, Ergodic Theory** and **Symplectic Geometry**. She was usually driven by in-depth understanding of the different complex structures based on abstract surfaces, rather than searching for it's applications.

4 Journey to the Fields Medal

One of Mirzakhani's most famous collaborative works was alongside Alex Eskin [EM] and, in part, Amir Mohammadi [EMM15]. The former collaboration was an effort to tackle one of the greatest open problems in their field.

[PH2017] The class of problems Mirzakhani was interested in dates back more than a century. In 1912, Austrian statistical physicist Paul Ehrenfest and his wife proposed the 'wind-tree' model as a way of understanding how impediments in a system affect diffusion (in this context, spreading out of particles due to their natural motion). They imagined a bounded forest that was empty except for regularly spaced trees - symbolized as rectangles forming a periodic pattern within a square lattice. If the wind entered the forest from a certain direction and scattered off various trees according to the law of reflection (incident angle equals reflected angle), then how quickly would nearby streams of air particles separate from each other and spread throughout the entire forest?

This problem is similar to a game of billiards wherein precise knowledge of initial conditions and configuration of the obstacles are key in tracking the behaviour of the billiard balls.

[3]The Clay Mathematics Institute (CMI), New Hampshire, USA is dedicated to increasing and disseminating mathematical knowledge

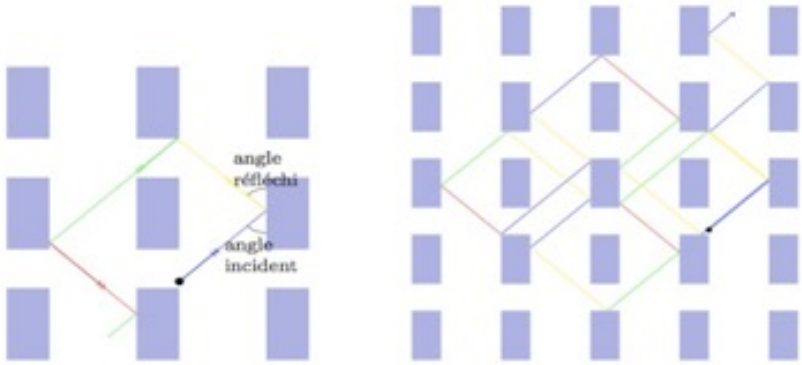


Figure 4.6 Billiard trajectories in periodic wind-tree model extracted from Vincent Delecroix's Ph.D. thesis; (source: https://tel.archives-ouvertes.fr/tel-00653165/document)

Another variation of this theme, called the **'illumination problem'**, was suggested by German-American mathematician Ernst Straus in the 1950s. Straus asked if a room with mirrored walls could be fully illuminated by a single point light source. The most recent development in this problem is contributed to Lelièvre, Monteil and Weiss (after previous breakthroughs in 1958 and 1995) [LMW16]. The results presented in their paper crucially rely on the breakthrough results of Eskin-Mirzakhani [EM] and Eskin-Mirzakhani-Mohammadi [EMM15]. [LMW16] proves that for any translation surface L, there are only finitely many points in L which are not illuminated by a given point in L. In terms of the original naively posed illumination problem, this means that if the room is polygonal with each angle a rational multiple of π, then no matter where the candle is placed, only finitely many locations in the room will not be illuminated [ASV18].

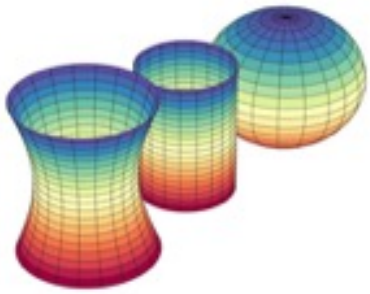


Figure 4.7 (L-R) Hyperboloid, Cylinder, Sphere; (source: https://en.wikipedia.org/wiki/Gaussian_curvature)

[PH17] Replace light rays with billiard balls or streams of air particles and it is easy to realize deep connections between these problems; each one involves geodesics (the shortest possible paths through certain spaces), the laws of reflection, and depends on the geometry and topology of the situation. Replace the flat room or surface with an even more complicated surface, such as a hyperboloid, and such problems become even more intriguing. We're commonly used to spheres (has positive curvature) and cylinders (has no curvature), but equally important are hyperboloids, which have negative curvature.

Maryam was not the kind to engage in a problem to extend another mathematician's work. Instead, she had a reputation for *engaging directly* with a hard scientific challenge and getting to the heart of the matter. She was able to perceive problems in their most general forms - and that was the genius of Maryam Mirzakhani. Her collaboration with Eskin [EM], which found applications to the billiard problem, the wind-tree puzzle, and the illumination problem, was dubbed the **'Magic Wand Theorem.'** Russian mathematician Anton Zorich described their work as "so beautiful and

powerful" that its applications would be "far beyond our current imagination." A lot of struggle went into proving these results. While progressing, seemingly near the end, they hit a roadblock. During this period, Maryam gave birth to her daughter Anahita. Even then, she continued working - and that was her *inner strength*.

She was invited to talk at the **International Congress of Mathematicians** in 2010, on the topic of "Topology and Dynamical Systems & ODE". She was also awarded the **2013 AMS Ruth Lyttle Satter Prize** in Mathematics and the **Simons Investigator Award 2013**. In 2014 she won a **Clay Research Award** in recognition of her 'many and significant contributions to geometry and ergodic theory, in particular to the proof of an *analogue of Ratner's theorem* on unipotent flows for moduli of flat surfaces'. Another major contribution of Mirzakhani is in connection with the *Thurston's earthquake flow*.

In August 2014, Maryam was awarded the **Fields Medal** (the highest honour in Mathematics) in Seoul at the International Congress of Mathematicians for **"her outstanding contributions to the dynamics and geometry of Riemann surfaces and their moduli spaces"**. She was the first Iranian and first female winner of the Medal since its inception in 1936. This was a breakthrough for the quest for equality in STEM [4] disciplines. In an interview published in Stanford News, Mirzakhani described her methods: "I don't have any particular recipe [for constructing novel proofs] ... It is like being lost in a jungle and trying to use all the knowledge that you can gather to come up with some new tricks, and with some luck you might find a way out." Thus, she remained as grounded and hardworking as ever.

5 The Legacy of Mirzakhani

As a student, researcher, professor and ultimately a pioneer in Mathematics, Maryam projected fierce intelligence and an even greater degree of humility. Despite her self-effacing self, Maryam's mathematics propelled her to become a role model worldwide - particularly for young girls and women in STEM fields. One may not have known Maryam personally, but it is easy to see from the accounts of her friends and colleagues that she was marked by a steely tenacity and fearless ambition. However, the towering graph of her career, with its breakthroughs and recognitions, was lined by the dark cloud of cancer; Maryam was diagnosed with breast cancer in 2013.

She always said, "beauty of mathematics only shows itself to the more patient followers". In 2016, Maryam Mirzakhani was made a member of the **National Academy of Sciences**, making her the *first Iranian woman* to be officially accepted as a member of the academy. In the same year, her cancer spread to her bones and liver. She fought the cancer with the same courage and spirit with which she attacked all the obstacles that her short life presented. After a year, Maryam lost the battle against cancer on 14th July 2017.

[4] Science, Technology, Engineering and Mathematics (**STEM**) is a term used to group together these academic disciplines

Figure 5.8 Maryam receives the Fields Medal from South Korean President Park Geun-hye; (source: https://www.dailymail.co.uk/news/article-4699868/Maryam-Mirzakhani-award-winning-mathematician-dies-40.html)

Maryam broke glass ceilings during her life, as a pioneering female mathematician, and even after her death - President Rouhani himself, broke a national taboo by publishing photos in which she appeared with her hair uncovered, a gesture that was widely noted in the press. As a result of advocacy carried out by the Women's Committee within the Iranian Mathematical Society, the International Council for Science agreed to declare Maryam Mirzakhani's birthday, **12 May**, as a day for **celebrating women in mathematics** in respect of her memory. This initiative was supported by Indian Women and Mathematics, among several organisations.

Mirzakhani had developed new geometric perspective to study moduli spaces which put earlier, seemingly unrelated, results into complete harmony. Her ground-breaking results also unleashed many new research directions for future generations of mathematicians to explore. Her work was highly theoretical in nature, but they were able to shed light upon problems in quantum field theory and had secondary applications to engineering and material science. Within mathematics, her works had implications on the study of prime numbers and cryptography.

The development of STEM field relies on extraordinary creativity, and limiting the creative potential to any subset - social, economic or gendered - would obstruct the development of these fields. Maryam was a brilliant mind whose legacy showed the fierce will of women on the path towards peaks of excellence. We remember Maryam's legacy as we celebrate her life as well as the contributions of all women in the ever-engaging beautiful field of Mathematics.

Figure 5.9 Maryam with her daughter, Anahita; (source: https://www.ams.org/journals/notices/201810/rnoti-p1221.pdf)

References

[NoAMS18] Maryam Mizakhani: 1977-2017, Notices of the AMS, November 2018, pp. 1221-1247.

[ASV18] Agarwal N., Shah R. and Venkataraman G., Maryam Mirzakhani, The Master Artist of Curved Surfaces, Resonance, Volume 23, Issue 3, March 2018, pp. 253-262.

[EM] Eskin A. and Mirzakhani M., Invariant and Stationary Measures for the SL(2, R) action on Moduli Space, arXiv 1302.3320, to appear in Publications Mathématiques de l'IHÉS.

[EMM15] Eskin A., Mirzakhani M. and Mohammadi A., Isolation, Equidistribution, and Orbit Closures for the SL(2, R) action on Moduli Space, Annals of Mathematics (2nd series), Volume 182, September 2015, pp. 673-721.

[M-AM08] Mirzakhani M., Growth of the number of simple closed geodesics on hyperbolic surfaces, Annals of Mathematics, 2008.

[M-IM07] Mirzakhani M., Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces, Inventionses Mathematicae, 2007.

[M-AMS07] Mirzakhani M., Weil-Petersson volumes and intersection theory on the moduli spaces of curves, Journal of the American Mathematical Society, 2007.

[Stan] https://news.stanford.edu/2017/07/15/maryam-mirzakhani-stanford-mathematician-and-fields-medal-winner-dies/

[PH17] https://www.forbes.com/sites/startswithabang/2017/08/01/maryam-mirzakhani-a- candle-illuminating-the-dark

[LMW16] Lelièvre S., Monteil T. and Weiss B., Everything is Illuminated, Geometry and Topology, Volume 20, Issue 3, 2016, pp. 1737–176.

[98] Mirzakhani M., A simple proof of a theorem of Schur, American Mathematical Monthly, Volume, Issue 3, 1998, pp. 260-262.

[96] Mirzakhani M., A Small Non-4-choosable Planar Graph, Bulletin of the Institute of Combinatorics and its Applications. Volume 17, 1996, pp. 15-18.

[95] Mirzakhani M., Decomposition of Complete Tripartite Graphs into 5-cycles, Combinatoncs Advances, Math. Appl., Kluwer Acad. Publication, Dordrecht, 1995, pp. 235-241.

[99] Elementary Number Theory, Challenging Problems (Farsi), Roya Beheshti, Maryam Mirzakhani, Fatemi Publishers, Tehran, Iran, 1999.

[IntCM08] https://www.claymath.org/library/annualreport/ar2008/08Interview.pdf

Dr. Anuradha, Associate Professor, Department of Statistics, Lady Shri Ram College for Women, New Delhi
sarkar.anuradha@gmail.com

Khuisangmi Konghay, B.Sc. (H) Mathematics, 6th Semester, Lady Shri Ram College For Women, New Delhi
konghaykhuimi@gmail.com

Lipika Parekh, B.Sc. (H) Mathematics, 6th Semester, Lady Shri Ram College For Women, New Delhi
plipika98@gmail.com

Contact Us:
Email Id: eclat.lsr@gmail.com