ÉCLAT
MATHEMATICS JOURNAL

Lady Shri Ram College For Women
Volume X
2018-2019

HYPATIA

This journal is dedicated to Hypatia who was a Greek mathematician, astronomer, philosopher and the last great thinker of ancient Alexandria. The Alexandrian scholar authored numerous mathematical treatises, and has been credited with writing commentaries on Diophantuss Arithmetica, on Apolloniuss Conics and on Ptolemys astronomical work. Hypatia is an inspiration, not only as the first famous female mathematician but most importantly as a symbol of learning and science.

PREFACE

Derived from the French language, Éclat means brilliance. Since it's inception, this journal has aimed at providing a platform for undergraduate students to showcase their understanding of diverse mathematical concepts that interest them. As the journal completes a decade this year, it continues to be instrumental in providing an opportunity for students to make valuable contributions to academic inquiry.

The work contained in this journal is not original but contains the review research of students. Each of the nine papers included in this year's edition have been carefully written and compiled to stimulate the thought process of its readers. The four categories discussed in the journal - History of Mathematics, Rigour in Mathematics, Interdisciplinary Aspects of Mathematics and Extension of Course Content - give a comprehensive idea about the evolution of the subject over the years.

We would like to express our sincere thanks to the Faculty Advisors of the department, who have guided us at every step leading to the publication of this volume, and to all the authors who have contributed their articles for this volume. We hope that Éclat continues to motivate students to go beyond the prescribed limits of the text and to explore many more avenues in the field of mathematics. We are open to any suggestions, corrections and submissions from our readers.

This journal is dedicated to Hypatia who was a Greek mathematician, astronomer, philosopher and the last great thinker of ancient Alexandria.

The Editorial Team

Kushagri Tandon, B. Sc. (Hons.) Mathematics, 3rd Year
Namrata Lathi, B. Sc. (Hons.) Mathematics, 3rd Year
Khuisangmi Konghay, B. Sc. (Hons.) Mathematics, 2nd Year
Lipika Parekh, B. Sc. (Hons.) Mathematics, 2nd Year

CONTENTS

History of Mathematics

Mathematics is one of the oldest academic discipline involving stimulating and intriguing concepts. It is far beyond the ken of one individual and to make any contribution to the evolution of ideas, and understanding of the motivation behind the ideas is needed. The section covers the genesis of mathematical ideas, the stream of thought that created the problem and what led to its solution. The aim is to acquaint the readers with historically important mathematical vignettes and make them inured in some important ideas of Mathematics.

Euler's Identity

Nayana Nair

Abstract

Euler's identity is the paradigm of mathematical beauty. This equality connects five fundamental mathematical constants through three basic arithmetic operations in an elegant equation that displays the powerful connection between them. This paper discusses the significance of Euler's identity and the various approaches in explaining the same.

1 Leonard Euler

Euler's strokes of genius laid the groundwork for most of the mathematics we have today. He was born in Basel, Switzerland in 1707 and was tutored by Johann Bernoulli at the University of Basel, who instantly discovered his talent for mathematics. Euler spent most of his career in St. Petersburg and Berlin. His powers of memory and concentration were legendary. Despite his complete loss of vision in 1771, his research continued unabated and his talent to perform prodigious mental computations proved indispensable. Like Beethoven who wrote music he never heard, Euler created mathematics he never saw. Euler's incredibly voluminous and diverse work fills 74 massive volumes of the *Opera omnia*. Euler left hardly an area of mathematics untouched, putting his mark on such diverse fields as analysis, number theory, mechanics, hydrodynamics, cartography, topology, weights, measures and even the theory of music, to name a few.

2 Euler's Elegant Equation

$$e^{i\pi} = -1$$

This peculiar equation first appears in Euler's *Introductio*, published in Lausanne in 1748. A unique fusion of exponential and trigonometric, real and complex, this identity is considered as one of the most remarkable identities in all of mathematics and is filled with cosmic beauty. It was from this that Euler deduced strange consequences such as $i^i = \frac{1}{\sqrt{e^\pi}}$ about which the Harvard mathematician Benjamin Peirce is reported to have said, *Gentlemen, we have not the slightest idea of what this equation means, but we may be certain that it means something very important.* One of Euler's greatest achievements was introducing a connection between exponential functions that shoot off to infinity as x becomes large and trigonometric functions that oscillate between the values -1 and 1 with the help of complex numbers.

Figure 1: Leonard Euler

This equation has had a significant impact on the world of mathematics and other sciences, changing many different areas of work and study, and is considered as one of his greatest and most significant contributions.

2.1 Derivation

While Euler was solving the differential equation, $y^{(2)} + y = 0$, he found four solutions $y = \sin x$, $y = \cos x$, $y = e^{ix}$ and $y = e^{-ix}$. This strange mix of trigonometric and exponential equations showed that these trigonometric functions must be some combination of e and the imaginary number i. Many continuous, differentiable functions can be written as a series called a Taylor series. This series for the exponential function e^x is still valid if x is a complex number. For *sine*, *cosine* and e^x the series is as follows:

$$\begin{aligned}
\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \\
\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \\
e^x &= 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \frac{x^6}{6!} + \frac{x^7}{7!} + \cdots
\end{aligned}$$

Replacing x with ix in the expansion of e^x we get the following series:

$$\begin{aligned}
e^{ix} &= 1 + ix + \frac{(ix)^2}{2} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \frac{(ix)^6}{6!} + \frac{(ix)^7}{7!} + \cdots \\
&= 1 + ix - \frac{x^2}{2} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \cdots \\
&= \left(1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right) \\
&= \cos x + i \sin x
\end{aligned}$$

For $x = \pi$

$$e^{i\pi} = -1$$

Since $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$, we can conclude that the value of $\lim_{n\to\infty}(1 + \frac{ix}{n})^n$ tends closer and closer to -1. This is equivalent to n repeated multiplications in the complex plane with the final value coinciding with -1.

2.2 Theoretical Uses

Euler's identity has several applications within mathematics. It allows standard trigonometric functions to be expressed in terms of exponential functions as $\cos x = \frac{e^{ix}+e^{-ix}}{2}$ and $\sin x = \frac{e^{ix}-e^{-ix}}{2i}$ and can be used in verifying the addition and subtraction formulas for *sine* and *cosine*.

2.2.1 Logarithms of Negative Numbers

Euler's identity gave a greater understanding of complex numbers. The unusual idea of finding the logarithm of a negative number became a reality as from the identity it follows that $\ln(-1) = i\pi$. Any odd multiple of π will give an expression for $\ln(-1)$, i.e $\ln(-1) = i(2n-1)\pi$ where $n \in \mathbf{Z}$.

2.2.2 Complex Numbers Raised to Complex Numbers

The bizarre yet remarkable consequence of this identity is that it allows us to compute complex powers of complex numbers which turn out to be real numbers. For example substituting $x = \frac{\pi}{2}$ in the identity we get $e^{\frac{i\pi}{2}} = i$, now raising this to i we get

$$i^i = e^{\frac{-\pi}{2}} \approx 0.2078796$$

These repercussions of Euler's identity allows us to manipulate complex numbers and has profound implications in complex number theory.

3 Modern Applications

Euler's identity has extensive applications in electronics and engineering. It is implemented in linear, time invariant input-output boxes also known as LTI boxes which takes values of $x_n(t)$ and output the values of $y_n(t)$ through multiplication by some constant which could be a complex number. With the help of Euler's formula, scientists are able to understand planetary retrograde motion which is the apparent motion of a planet in a direction opposite to that of other bodies in a system with respect to a particular point, at complex values of time which earlier would have seemed illogical.

4 Conclusion

Euler's identity brought about a paradigm shift in the imaginary field and its relation with the real world. This identity has a profound impact on nearly every field of

theoretical mathematics as well as in real world problem solving. From formulating the basic laws of fluid mechanics to remarkable research in number theory and analytical mechanics, Euler's indispensable and influential contributions to various branches of mathematics make him a true mathematical pioneer.

References

[1] Dunham, W., *Euler: The Master of Us All*, The Mathematical Association of America (1999)

[2] Maor, E., *e: The Story of a Number*, The Princeton University Press (1994)

[3] Mallett, T., Christensen, S. and Hetrick, S.(2009), *Euler's Identity: Concerning the real, the unreal and the infinite*, Washington State University

[4] https://www.usna.edu/Users/math/meh/euler.html

NAYANA NAIR, B.Sc(H) MATHEMATICS, 2nd SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
nayana.nair391@gmail.com

Decoding The Enigma: Story Of Turing and his Machine

Tanvi Vohra

Abstract

This paper attempts to trace the journey of the great mathematician Alan Turing and the development of his machine and the effect on the Nazis and the Government caused by its invention. The paper further focuses on the working of the Bombe and explaining the logic behind it.

1 Early Life and History

Turing was born on June 23, 1912 at 2, Warrington Crescent, London W9. He was truely a gifted being who influenced the development of computer science and brought forward the concept of Algorithms and Computation with the help of his machine. Until his father's retirement, he and his elder brother John were fostered in various English homes - where originality and discovery were given least preference. Science was his extra-curricular passion. Turing then went on to study higher mathematics at King's College, Cambridge in 1931. Besides his interests and proficiency in Mathematics, Turing was also an olympic level runner.

Figure 1: Alan Turing

It is believed that the stimulus for effective communication and competition came only from a friend, Christopher Morcom, who gave Turing a vital period of intellectual companionship. He found himself powerfully attracted to Christopher and by 1933, his homosexuality became a definitive part of his identity. At a young age of 22, he was elected to a fellowship at King's in recognition of his research in probability theory and in the following year, Turing won a Smith's Prize for the same. He learnt from the 1935 lecture course of M.H.A Newman (the Cambridge topologist) that a further question posed by Hilbert still lay unsolved.

2 Second World War

In the year 1938-39, Turing continued to live on his King's college fellowship while he secretly worked part-time for the British cryptanalytic department. On September 3, when the British declared war, he took up full-time work at the wartime cryptanalytic headquarters, Bletchley Park. The work they performed here was limited to decoding the messages sent by the Germans through their machine "Enigma". Turing believed that in order to defeat a machine, human minds alone were not sufficient. One of these ideas took the form of a machine called the Bombe. "Hut 8", where he worked along with the other cryptanalists, became the most critical section of Bletchley Park.

3 The Inspiration

3.1 Hilbert's Decision Problem

The aim of all the mathematicians of that time was embedding a sequence of mathematical and logical operations into a mechanism so that it could think on its own. In 1928, David Hilbert, a German Mathematician, cut to the heart of the matter by bringing forward a challenge known as *Entscheidungsproblem* or the decision problem. Hilberts challenge was to find a general, methodical plan or procedure that takes an input in the form of a statement (or a string of symbols) in the given language. Further, to answer the decision problem, the output must be true whenever the statement is true, and false, whenever it is false. This method is known as the decision method. In 1936, Alan Turing and Alonzo Church independently showed that, in general, the decision problem has no resolution, proving that no consistent formal system of arithmetic has an effective decision method. This problem inspired Turing, which led him to invent computer programs. It was in the course of his work on the *Entscheidungsproblem* that he invented the universal Turing machine.

3.2 The Church-Turing Thesis

The Church - Turing thesis states that anything humanly computable is also computable by the universal Turing Machine. Church defined some functions as being the Lambda definable functions the values of which could be calculated by repeated substitution. In 1936, Turing showed that this thesis was equivalent to his own and showed that every Lambda definable function was determinable by the Turing Machine and vice versa. In a review of Turings

work, Church acknowledged the superiority of Turings formulation of the thesis over his own.

4 The Machine and its Elements

In the mid 90s, calculating machines were very popular, but there were very less or rather, no thinking machines. Turing attempted to build one such machine, which could think on its own and arrive at conclusions. To begin with, Turing imagined a machine as a human being inside a closed box, with a pencil and a stack of paper. Along with it, a book of instructions was also to be provided. The machine would be able to perform different set of functions only by changing the instruction set. Next, he went on to define algorithms in the form of symbols instead of mechanisms.
The elements of the machine were as follows:

1. A simple scanner head which could move up and down and read individual symbols each at a time. This is known as the 'read operation'. The scanner would also have an inbuilt printer in order to print symbols as instructed.

2. A set of predefined instructions or algorithms on the basis of which the machine would work. This is what Turing referred to as *"The Instruction Book"*.

3. A thin role of paper on which the symbols would be printed.

The most crucial decision was to create the instruction book. It was a simple book which contained different states (or pages). Each page or state had three columns:

1. The if (aware) column or the observation column

2. The Do or the operational column

3. The Next State or page

Column 1 contains a list of symbols. The awareness symbol matches with any one symbol listed here. Column 2 defines the instructions to be performed corresponding to each symbol in column 1. Column 3 lists the next page to jump to when the instructions specified to step 1 are performed.

Table 1: State (Page) of The Instruction Book

IF (Awareness)	**Do (instructions)**	**Next State (page)**
1	type "3", move right	State 3
2	move 1 right	State 2
3	move 3 left	State 28
.	.	.
.	.	.
.	.	.

5 The Process

The first part of the process is the observation part. The scanner would scan the entered problem and interpret the symbol. At each time, the scanner is pointing at only one symbol, named as the 'awareness symbol'. Once the machine is aware about the symbol, the next step is to locate this symbol in the first column of the state (or the page) to find a match. After the match is found, the operator now has to move to the second column, that is, the operational column. It should be noted that all the instructions in the second column would mainly involve 2 things:-

- Writing symbols
- Moving the scanner to a different position.

The machine now has to simply follow the instruction corresponding to the awareness symbol.
The last column, known as the Next State column, lists the next page to jump to. In this instruction book, each page is a new step.
Turing called these steps 'states' of the machine. Meanwhile, the scanner is still pointing on a symbol, which would now be the awareness. The machine is now on a new state and again, the awareness symbol is located in the first column of this state. Like before, the instruction corresponding to the awareness symbol is performed and the machine moves to the next state. This process continues state after state with the machine printing symbols or moving the position of the scanner as instructed by the operational column. The machine finally comes to a state of rest, when it arrives at a result and halts. This is when the final output is obtained.
To make it simpler, Turing thought that the writing down of the symbols as

well as the instruction book could all be written in one big piece of paper. He further arranged this paper in the form of a long array, leading to his famous simplification of a machine, which has 2 basic elements - a role of tape and a scanner printer head.

6 The Ending

Turing made a number of vital contributions to the war effort, and helped his old logic teacher Max Newman to develop the world's first programmable, electronic computer. After a year of Government Mandated hormonal therapy, on 8th June, 1954, the great mathematician was found dead on his bed. A half eaten, cyanide coated apple alongside his bed proved that this was a result of inhaling cyanide fumes. He was only 41 years old. In 2013, Queen Elizabeth II granted Turing a Posthumous Royal Pardon, honoring his unprecedented achievements. Historians estimated that breaking Enigma shortened the war by more that 2 years, saving around 14 million lives. It remained a government held secret for more than 50 years. Turing's work inspired generations of research into what scientists called 'Turing Machines'. Today, they are known as computers.

References

[1] Heaton, L., *A Brief History Of Mathematical Thought- Key concepts and where they come from*, Robinson, 2015.

[2] Hodges, A., *Alan Turing: The Enigma*, Princeton University Press, 5th December 2014.

[3] www.turing.org.uk/publications

[4] http://turingtrust.co.uk

TANVI VOHRA, B.Sc.(H) MATHEMATICS, 2nd SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
vohratanvi5@gmail.com

Rigour in Mathematics

This section introduces advance Mathematics to the readers aiming at high standards of proofs. It stimulates interest and lays the foundation for further studies in different branches.

Represenation of Integers as Sums of Squares

Kushagri Tandon

Abstract

This paper explores a problem of existence of an integer $g(k)$, such that each positive integer can be written as the sum of no more than a fixed number $g(k)$ of kth powers.The resulting problem called "Waring's Problem", based on an assertion made by Edward Waring in 1770 is discussed here for a particular case of $k = 2$.

1 The Waring's Problem

In his book, *Meditationes Algebraicae (1770)*, **Edward Waring** stated that each positive integer is expressible as a sum of at most 9 cubes, also a sum of at most 19 fourth powers and so on. This assertion is that of representation of positive integers as sums of a fixed number s of non-negative kth powers.

It is plainly impossible to represent all integers if s is too small, for example if $s = 1$ and if $s < k$. The number of values of x_1 for which $x_1^k \leq n$, does not exceed $n^{1/k} + 1$; and so the number of set of values $x_1, x_2, \cdots, x_{k-1}$ for which $x_1^k + \cdots + x_{k-1}^k \leq n$ does not exceed $(n^{1/k} + 1)^{k-1} = n^{(k-1)/k} + O(n^{(k-2)/k})$. Hence most numbers are not representable by $k - 1$ or fewer kth powers. The question is that whether, for a given k, there is any fixed $s = s(k)$ such that $n = x_1^k + x_2^k + \cdots + x_s^k$ is soluble for every n. Hence, if all numbers are representable by s kth powers, there is a least value of s for which this is true. The least value of s is denoted by $g(k)$. In other words, for a given k, a number $g(k)$ is sought such that every $n > 0$ can be written as sum of no more than fixed number $g(k)$ of kth powers, where $g(k)$ depends only on k i.e. $n = a_1^k + a_2^k + \cdots + a_{g(k)}^k$ where the a_i are non negative integers, not necessarily distinct.

We shall show that $g(2) = 4$. It is known that $g(3) = 9$; every number is representable by 9 or fewer cubes, and every number except $23 = 2^3 + 2^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3$ and $239 = 4^3 + 4^3 + 3^3 + 3^3 + 3^3 + 3^3 + 1^3 + 1^3 + 1^3$ can be represented by 8 or fewer. The evidence indicates that only 15 other numbers, of which the largest is 454, require so many cubes as 8; and that 7 suffice from 455 onwards.

As far as $k \geq 6$ is concerned, it has been established that the formula

$$g(k) = [(3/2)^k] + 2^k - 2$$

where $[x]$ represents the greatest integer function i.e. $[x] = max\{z : z \in \mathbb{Z}, z \leq x\}$ holds except possibly for a finite number of values of k. There is evidence which suggests that this expression is correct for all k.

We define $G(k)$ as the least value of s for which it is true that all sufficiently large numbers i.e. all numbers with at most a finite number of exceptions, are representable by kth powers. Clearly $G(k) \leq g(k)$. Exact values of $G(k)$ are known when $k = 2$ and $k = 4$, namely, $G(2) = 4$ and $G(4) = 16$. Below, are given some known values and estimates for the first few $g(k)$ and $G(k)$:

$g(2) = 4$	$g(3) = 9$	$g(4) = 19$	$g(5) = 37$	$g(6) = 73$	$g(7) = 143$
$G(2) = 4$	$4 \leq G(3) \leq 7$	$G(4) = 16$	$6 \leq G(5) \leq 17$	$9 \leq G(6) \leq 24$	$8 \leq G(7) \leq 33$

2 Sums of Two Squares

In this section we state and prove some important results to find the necessary and sufficient conditions that a positive integer be representable as the sum of two squares. In doing so we show that certain numbers cannot be represented as sums of two squares and hence $g(2) \neq 2$.

Lemma 2.1. *If m and n are each the sum of two squares, then so is their product mn.*

Proof. If $m = a^2 + b^2$ and $n = c^2 + d^2$ for integers a, b, c, d, then

$$\begin{aligned} mn &= (a^2 + b^2)(c^2 + d^2) \\ &= a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 \\ &= a^2c^2 + b^2d^2 + 2abcd - 2abcd + a^2d^2 + b^2c^2 = (ac + bd)^2 + (ad - bc)^2 \end{aligned}$$

■

Lemma 2.2. *(Thue) Let p be a prime and let $\gcd(a, p) = 1$. Then the congruence*

$$ax \equiv y \pmod{p}$$

admits a solution x_0, y_0, where

$$0 < |x_0| < \sqrt{p} \text{ and } 0 < |y_0| < \sqrt{p}$$

Proof. Define $S = \{0, 1, 2, \cdots [\sqrt{p}]\} \times \{0, 1, 2, \cdots [\sqrt{p}]\}$. S has $([\sqrt{p}] + 1)^2$ ordered pairs. Now $\sqrt{p} < [\sqrt{p}] + 1 \implies p < ([\sqrt{p}] + 1)^2$.

So, S has more than p elements. As (x, y) varies over S, there are $([\sqrt{p}] + 1)^2 > p$ expressions of the type $ax - y$.

Since, $ax - y$ is an integer, and every integer is congruent modulo p to exacty one of $\{0, 1, \cdots, p-1\}$. So, the pigeonhole principle[1] guarantees that at least two members of S must be congruent modulo p. Take the pairs $(x_1, y_1) \neq (x_2, y_2)$ such that $ax_1 - y_1 \equiv ax_2 - y_2 \pmod{p}$ i.e. $a(x_1 - x_2) \equiv (y_1 - y_2) \pmod{p}$.

Setting $x_0 = x_1 - x_2$ and $y_0 = y_1 - y_2$, it follows that x_0 and y_0 provide a solution to the congruence $ax \equiv y \pmod{p}$. Then $(|x_0|, |y_0|) \in S$. We shall show that $|x_0| \neq 0$ and $|y_0| \neq 0$, so that $|x_0|, |y_0| \in \{1, 2, \cdots, [\sqrt{p}]\}$.

Suppose that $|x_0| = |x_1 - x_2| = 0$.

$$\therefore x_1 = x_2 \implies a(x_1 - x_2) = 0 \equiv y_1 - y_2 \pmod{p} \implies p|(y_1 - y_2)$$

Since $y_1, y_2 \in \{0, 1, \cdots, [\sqrt{p}]\}$, we know that $y_1 < p$, $y_2 < p$. Therefore, $y_1 - y_2 = 0$ i.e. $y_1 = y_2$. Hence, $(x_1, y_1) = (x_2, y_2)$, which is a contradiction. Therefore, $|x_0| \neq 0$. Similarly, suppose that $|y_0| = 0$.

$\therefore y_1 - y_2 = 0 \implies a(x_1 - x_2) \equiv 0 \pmod{p} \implies p|a(x_1 - x_2)$

$\implies p|(x_1 - x_2) (\because \gcd(a, p) = 1) \implies x_1 - x_2 \equiv 0 \pmod{p}$

Using the above logic we get $x_1 = x_2 \implies (x_1, y_1) = (x_2, y_2)$, which is a contradiction.

Hence, $0 < |x_0| < \sqrt{p}$ and $0 < |y_0| < \sqrt{p}$. ■

Lemma 2.3. *If p is an odd prime then $x^2 \equiv -1 \pmod{p}$ has a solution iff $p \equiv 1 \pmod{4}$*

Proof. If p is an odd prime then either $p \equiv 1 \pmod{4}$ or $p \equiv 3 \pmod{4}$.

Suppose $x^2 \equiv -1 \pmod{p}$ has a solution say, $x \equiv a \pmod{p}$ so that, $a^2 \equiv -1 \pmod{p}$. Because, $p \nmid a$, by Fermat's little theorem,

$$1 \equiv a^{p-1} \equiv (a^2)^{(p-1)/2} \equiv (-1)^{(p-1)/2} \pmod{p} \tag{1}$$

. If $p \equiv 3 \pmod{4}$, then $4|(p-3) \implies p = 4k + 3$, for some $k \in \mathbb{Z}$. So, $(p-1)/2 = 2k + 1$ is odd. Therefore, $(-1)^{(p-1)/2} = -1$. By (1), $1 \equiv -1 \pmod{p} \implies p|(1 - (-1))$ i.e. $p|2$ which is not possible, since p is an odd prime. Therefore, $p \equiv 1 \pmod{4}$.

On the other hand, suppose that $p \equiv 1 \pmod{4}$. We observe that,

$$\begin{gathered} p - 1 \equiv -1 \pmod{p} \\ p - 2 \equiv -2 \pmod{p} \\ \vdots \\ (p+1)/2 \equiv -(p-1)/2 \pmod{p} \end{gathered}$$

[1] If n objects are placed in m pigeonholes and if $n > m$, then some pigeonhole will contain at least two objects

Since, p is an odd prime, by Wilson's theorem $(p-1)! \equiv -1 \pmod{p}$. So,

$$\begin{aligned} -1 &\equiv 1 \cdot 2 \cdot 3 \cdots (p-1)/2 \cdot (p+1)/2 \cdots (p-1) \\ &\equiv 1 \cdot 2 \cdot 3 \cdots (p-1)/2 \cdot (-(p-1)/2) \cdots (-3) \cdot (-2) \cdot (-1) \qquad (2) \\ &\equiv (-1)^{(p-1)/2} \{((p-1)/2)!\}^2 \pmod{p} \end{aligned}$$

Now, $p \equiv 1 \pmod{4} \implies 4|(p-1) \implies p = 4k+1$, for some integer $k \implies (p-1)/2 = 2k$ which is even. So, by (2), $\{((p-1)/2)!\}^2 \equiv -1 \pmod{p} \implies x_0^2 \equiv -1 \pmod{p}$ where $x_0 = ((p-1)/2)!$. Hence, the congruence $x^2 \equiv -1 \pmod{p}$ has a solution $x \equiv ((p-1)/2)! \pmod{p}$ ■

Theorem 2.4. *An odd prime p is expressible as a sum of two squares iff $p \equiv 1 \pmod{4}$*

Proof. If p is an odd prime then either $p \equiv 1 \pmod{4}$ or $p \equiv 3 \pmod{4}$. Suppose that p can be written as sum of two squares. We claim that no prime p such that $p \equiv 3 \pmod{4}$ can be written as sum of two squares.

Suppose $p \equiv 3 \pmod{4}$. For any integer a, a is congruent modulo 4 to exactly one of $\{0, 1, 2, 3\}$. Consequently, $a^2 \equiv 0$ or $1 \pmod{4}$. So, for arbitrary integers a and b, $a^2 + b^2 \equiv 0, 1,$ or $2 \pmod{4}$ Since, $p \equiv 3 \pmod{4}$, the equation $p = a^2 + b^2$ is impossible. Therefore, $p \equiv 1 \pmod{4}$.

For the converse, assume that $p \equiv 1 \pmod{4}$. By Lemma 2.3, integer $a = ((p-1)/2)!$ is a solution of the quadratic congruence $x^2 \equiv -1 \pmod{p}$. Now $\gcd(a, p) = 1$, so the congruence $ax \equiv y \pmod{p}$ admits a solution x_0, y_0 for which the conclusion of Lemma 2.2 holds i.e. $0 < |x_0| < \sqrt{p}$ and $0 < |y_0| < \sqrt{p}$. As a result,

$$-x_0^2 \equiv a^2 x_0^2 \equiv (ax_0)^2 \equiv y_0^2 \pmod{p} \text{ i.e. } x_0^2 + y_0^2 \equiv 0 \pmod{p}$$

Thus,

$$x_0^2 + y_0^2 = kp \qquad (3)$$

for some integer $k \geq 1$. Since $0 < |x_0| < \sqrt{p}$ and $0 < |y_0| < \sqrt{p}$, we obtain $0 < x_0^2 + y_0^2 < 2p$, implication of which is that in equation (3), $k = 1$.

Consequently, $x_0^2 + y_0^2 = p$. ■

Remark 2.5. (Counting a^2 and $(-a)^2$ as the same) Any prime p of the form $4k+1$ can be represented uniquely (aside from the order of the summands) as a sum of two squares.

Theorem 2.6. *A number n is representable as the sum of two squares if and only if all prime factors of n of the form $4k+3$ have even exponents in the standard form of n. In other words, let the positive integer n be written as*

$n = N^2m$, where m is square-free. Then n can be represented as the sum of two squares if and only if m contains no prime factor of the form $4k + 3$. [2].

Proof. Suppose that m has no prime factor of the form $4k + 3$. If $m = 1$ then $n = N^2 + 0^2$. Hence, we are done. By Fundamental theorem of Arithmetic $m > 1$ can be written as $m = p_1p_2 \cdots p_r$ where the p_i's are distinct primes. Then each p_i is either equal to 2 or an odd prime. If p is an odd prime then either $p \equiv 1 \pmod 4$ or $p \equiv 3 \pmod 4$. Since, m has no prime factor of the form $4k + 3$, therefore, if p_i is an odd prime, then it is of the form $4k + 1$. By Theorem 2.4 and the fact that $2 = 1^2 + 1^2$, each p_i can be written as a sum of squares. By Theorem 2.1, m is a product of sums of two squares hence there exists integers x and y satisfying $m = x^2 + y^2$. We get

$$n = N^2m = N^2(x^2 + y^2) = (Nx)^2 + (Ny)^2$$

Thus, n is a sum of two squares.

On the other hand, assume that n can be represented as the sum of two squares say for some integers a and b, $n = a^2 + b^2 = N^2m$.

Let p be any odd prime divisor of m (we assume without loss of generality that $m > 1$). If $d = \gcd(a, b)$, then $a = rd$, $b = sd$, where $\gcd(r, s) = 1$.

So, $d^2(r^2 + s^2) = N^2m$. By hypothesis m is square free, hence $d^2|N^2$. But, then

$$r^2 + s^2 = \left(\frac{N^2}{d^2}\right) m = tp$$

for some integer t. Therefore, $r^2 + s^2 \equiv 0 \pmod p$. Now, $\gcd(r, s) = 1 \implies$ either $\gcd(r, p) = 1$ or $\gcd(s, p) = 1$. [3] Without loss of generality assume that $\gcd(r, p) = 1$. So, the congruence $rx \equiv 1 \pmod p$ has a unique solution modulo p. Let r' satisfy the congruence $rr' \equiv 1 \pmod p$. Multiplying the equation $r^2 + s^2 \equiv 0 \pmod p$ by $(r')^2$, we obtain $(r')^2r^2 + (r')^2s^2 \equiv 0 \pmod p$ i.e. $(rr')^2+(sr')^2 \equiv 1+(sr')^2 \equiv 0 \pmod p$ i.e. $(sr')^2 \equiv -1 \pmod p$. Thus, the congruence $x^2 \equiv -1 \pmod p$ has a solution. Thus, by Lemma 2.3 $p \equiv 1 \pmod 4$.

Since p was an arbitrary odd prime divisor of m and $p \equiv 1 \pmod 4$, there is no prime of the form $4k + 3$ that divides m. ■

Remark 2.7. We say that $n = x^2 + y^2$ is a *primitive* representation of n if $\gcd(x, y) = 1$, and otherwise an *imprimitive* representation.

[2] An integer is said to be square-free if it is not divisible by the square of any integer greater than 1. Hence, an integer $n > 1$ is square-free if and only if n can be factored into a product of distinct primes

[3] if $\gcd(r, p) \neq 1$ and $\gcd(s, p) \neq 1$, then $p|r$ and $p|s$. So $\gcd(r, s) \geq p$, which is a contradiction since, $\gcd(r, s) = 1$

Note. **We can conclude from this section that not all positive integers are representable as sum of two squares. Therefore,** $g(2) \neq 2$.

3 Sums of Three Squares

In this section we show that certain numbers cannot be represented as sums of three squares and hence conclude that $g(2) \neq 3$.

Consider the equation $N = x^2 + y^2 + z^2$. Then each of x, y and z has two possibilities for parity (either even or odd), which yield eight possible sets of values for N. Thus we consider the sum modulo 8.

Remark 3.1. For any integer n if n is even, then $n^2 \equiv 0$ or $4 \pmod 8$. We now consider the congruences modulo 8 when n is odd.

Lemma 3.2. *If n is odd, then $n^2 \equiv 1 \pmod 8$*

Proof. n is odd implies that $n+1$ and $n-1$ are even.

Consider the equation $n^2 - 1 = 4\left(\frac{n+1}{2}\right)\left(\frac{n-1}{2}\right)$. Now, $\left(\frac{n+1}{2}\right)$ and $\left(\frac{n-1}{2}\right)$ are consecutive integers. So either $\left(\frac{n+1}{2}\right)$ or $\left(\frac{n-1}{2}\right)$ is even. Therefore, $n^2 \equiv 1 \pmod 8$. ■

Remark 3.3. The possible congruences for a sum of three squares is given by,

$$x^2 + y^2 + z^2 \equiv \begin{cases} 0 \text{ or } 4 & \text{if } x, y, z \text{ are even} \\ 1 \text{ or } 5 & \text{if } x, y \text{ are even} \\ 2 \text{ or } 6 & \text{if } x \text{ is even} \\ 3 & \text{if } x, y, z \text{ are odd} \end{cases} \pmod 8$$

Since 7 is not a residue modulo 8, no number of the form $8k + 7$ can be represented as a sum of three squares.

Remark 3.3 is sufficient to prove that $g(2) \neq 3$. But, to show which numbers are representable we give stronger results.

Theorem 3.4. *No positive integer of the form $N = 4^n(8m + 7)$ can be represented as sum of three squares.*

Proof. It is clear from the Remark 3.3 that the equation $a^2 + b^2 + c^2 = 8m + 7$ is impossible. Now, let us suppose that $4^n(8m + 7)$, where $n \geq 1$, can be written as

$$4^n(8m + 7) = a^2 + b^2 + c^2$$

Then each of the integers a, b, c must be even (by using Remark 3.3 since, N is even and $4|N$). Putting $a = 2a_1, b = 2b_1, c = 2c_1$, we get

$$4^n(8m+7) = (2a_1)^2+(2b_1)^2+(2c_1)^2 = 4a_1^2+4b_1^2+4c_1^2 \implies 4^{n-1}(8m+7) = a_1^2+b_1^2+c_1^2$$

If $n - 1 \geq 1$, the arguement may be repeated until $8m + 7$ is eventually represented as sum of three squared integers, which contradicts the result of Remark 3.3. ■

Note. The following theorem states that the condition of Theorem 3.4 is also sufficient in order that a positive integer be realizable as the sum of three squares, however we shall not give the proof since it is beyond the scope for inclusion here.

Theorem 3.5. *A positive integer N is a sum of three squares if and only if N is not of the form $4^n(8m + 7)$.*

4 Sums of Four Squares

We show in this section that $g(2) = 4$. The first explicit reference to the fact that every positive integer can be written as the sum of four squares, counting 0^2, was made by **Bachet (in 1621)** and he checked this conjecture for all integers up to 325. Fifteen years later, **Fermat** claimed that he had a proof however, he gave no details. **Euler** discovered the fundamental identity that allows one to express the product of two sums of four squares as such a sum, and the crucial result that the congruence $x^2 + y^2 + z^2 \equiv 0 \pmod{p}$ is solvable for any prime p. A complete proof of the four-square conjecture was published by **Lagrange** in 1772, who acknowledged his indebtedness to the ideas of Euler. The next year, Euler offered a much simpler demonstration, which is essentially the version presented here.

Lemma 4.1. ***(Euler)*** *If the integers m and n are each the sum of four squares, then mn is likewise so representable.*

Proof. If $m = a_1^2 + a_2^2 + a_3^2 + a_4^2$ and $n = b_1^2 + b_2^2 + b_3^2 + b_4^2$ for integers a_i, b_i, then

$$\begin{aligned} mn &= (a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) \\ &= (a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4)^2 + (a_1b_2 - a_2b_1 + a_3b_4 - a_4b_3)^2 \\ &+ (a_1b_3 - a_2b_4 - a_3b_1 + a_4b_2)^2 + (a_1b_4 + a_2b_3 - a_3b_2 - a_4b_1)^2 \end{aligned}$$

■

Lemma 4.2. *If p is an odd prime, then the congruence $x^2+y^2+1 \equiv 0 \pmod{p}$ has a solution x_0, y_0 where $0 \leq x_0 \leq (p-1)/2$ and $0 \leq y_0 \leq (p-1)/2$.*

Proof. Consider the sets,

$$S_1 = \left\{1+0^2, 1+1^2, 1+2^2, \cdots, 1+\left(\frac{p-1}{2}\right)\right\}, S_2 = \left\{-0^2, -1^2, -2^2, \cdots, -\left(\frac{p-1}{2}\right)\right\}$$

Let if possible, $1+x_1^2 \equiv 1+x_2^2 \pmod{p}$ for some $x_1, x_2 \in \left\{0, 1, 2, \cdots, \left(\frac{p-1}{2}\right)\right\}$ such that $x_1 \neq x_2$ then either $x_1 \equiv x_2 \pmod{p}$ or $x_1 \equiv -x_2 \pmod{p}$. Now, the latter consequence is not possible, since $0 < x_1 + x_2 < p$ (unless $x_1 = x_2 = 0$). Thus, $x_1 \equiv x_2 \pmod{p} \implies x_1 = x_2$ (since $x_1, x_2 \in \left\{0, 1, 2, \cdots, \left(\frac{p-1}{2}\right)\right\}$), which is a contradiction since $x_1 \neq x_2$.

Thus, no two elements of the set S_1 are congruent modulo p. By similar reasoning no two elements of S_2 are congruent modulo p. So, together S_1 and S_2 contain $2\left[1+\frac{1}{2}(p-1)\right] = p+1$ integers. By pigeonhole principle, some integer in S_1 must be congruent modulo p to some integer in S_2 i.e. there exist x_0, y_0 such that $1 + x_0^2 \equiv -y_0^2 \pmod{p}$ where $0 \leq x_0 \leq (p-1)/2$ and $0 \leq y_0 \leq (p-1)/2$. ■

Corollary 4.3. *Given, an odd prime p, there exists an integer $k < p$ such that kp is the sum of four squares.*

Proof. According to the Lemma 4.2, there exist integers x_0, y_0 satisfying $0 \leq x_0 < \frac{p}{2}$ and $0 \leq y_0 < \frac{p}{2}$ such that

$x_0^2+y_0^2+1 \equiv 0 \pmod{p} \implies p|(x^2+y^2+1) \implies x_0^2+y_0^2+1^2+0^2 = kp$ for some integer k

Now, $x_0 < \frac{p}{2}$, $y_0 < \frac{p}{2} \implies kp = x_0^2 + y_0^2 + 1 < \frac{p^2}{4} + \frac{p^2}{4} + 1 < \frac{p^2}{2} + 1 < p^2 \implies k < p$, as asserted. ■

Theorem 4.4. *Any prime p can be written as the sum of four squares.*

Proof. The result clearly holds for $p = 2$, since $2 = 1^2 + 1^2 + 0^2 + 0^2$. Now, we shall show the result for odd primes.

From Corollary 4.3, given the odd prime p, there exists an integer $k < p$ such that kp is the sum of four squares.

Let k be the smallest positive integer such that kp is the sum of four squares; say

$$kp = x^2 + y^2 + z^2 + w^2 \tag{4}$$

We shall show that $k = 1$. We start by showing that k is an odd integer. To the contrary assume that k is even. Then x, y, z, w are all even; or all odd; or two are even and two are odd.

In any of the above cases, we may rearrange the variables so that

$$x \equiv y \pmod 2 \text{ and } z \equiv w \pmod 2$$

Thus, $2|(x-y), 2|(z-w) \implies \frac{1}{2}(x-y), \frac{1}{2}(z-w) \in \mathbb{Z}$. Also, $\frac{1}{2}(x-y) + y, \frac{1}{2}(z-w) + w \in \mathbb{Z}$ Thus,

$$\frac{1}{2}(x-y), \frac{1}{2}(x+y), \frac{1}{2}(z-w), \frac{1}{2}(z+w) \in \mathbb{Z}.$$

Now, $\left(\frac{x-y}{2}\right)^2 + \left(\frac{x+y}{2}\right)^2 = \frac{x^2+y^2}{2}$ and $\left(\frac{z-w}{2}\right)^2 + \left(\frac{z+w}{2}\right)^2 = \frac{z^2+w^2}{2}$. Hence,

$$\frac{1}{2}(kp) = \frac{1}{2}(x^2+y^2+z^2+w^2) = \left(\frac{x-y}{2}\right)^2 + \left(\frac{x+y}{2}\right)^2 + \left(\frac{z-w}{2}\right)^2 + \left(\frac{z+w}{2}\right)^2$$

is a representation of $(k/2)p$ as a sum of four squares, which is a contradiction to the fact that k was the least integer such that kp is a sum of four squares. Hence, our assumption was false. Therefore, k is odd. Now, to show $k = 1$.

Assume to the contrary that $k \neq 1$. Now, each of x, y, z, w are congruent modulo k to one of $\{0, 1, 2, \cdots, k-1\}$. Since, k is an odd integer hence is at least 3, so

$$\begin{gathered} 0 \equiv 0 \pmod k \\ 1 \equiv 1 \pmod k \\ \vdots \\ (k-1)/2 \equiv (k-1)/2 \pmod k \\ (k+1)/2 \equiv -(k-1)/2 \pmod k \\ \vdots \\ k-2 \equiv -2 \pmod k \\ k-1 \equiv -1 \pmod k \end{gathered}$$

Hence, it is possible to choose integers a, b, c, d such that

$$a \equiv x \pmod k \quad b \equiv y \pmod k \quad c \equiv z \pmod k \quad d \equiv w \pmod k \tag{5}$$

and

$$|a| < \frac{k}{2} \quad |b| < \frac{k}{2} \quad |c| < \frac{k}{2} \quad |d| < \frac{k}{2} \tag{6}$$

Thus, by using equation (4),(5)

$$a^2+b^2+c^2+d^2 \equiv x^2+y^2+z^2+w^2 \equiv 0 \pmod{k} \implies a^2+b^2+c^2+d^2 = nk \quad (7)$$

for some nonnegative integer n. Now, using equation (6),

$$0 \leq nk = a^2+b^2+c^2+d^2 < 4\left(\frac{k}{2}\right)^2 = k^2 \quad (8)$$

If $n = 0$, then equation (7) implies that $a^2 = b^2 = c^2 = d^2 = 0 \implies a = b = c = d = 0$ and, in consequence, equation (5) implies that

$$k|x, k|y, k|z, k|w \implies x = kx', y = ky', z = kz', w = kw'$$

for some integers x', y', z', w'.

Hence by (4), $kp = (kx')^2 + (ky')^2 + (kz')^2 + (kw')^2 = k^2((x')^2 + (y')^2 + (z')^2 + (w')^2) \implies k^2|kp \implies k|p$ which is not possible since $k < p$ and by assumption, $k \neq 1$. Thus, $n \neq 0$.

Now, from equation (8), the relation $nk < k^2$ allows us to conclude that $n < k$. Thus, $0 < n < k$. Combining the equations (4), (7) and using lemma 4.1, we get

$$\begin{aligned} k^2np &= (kp)(kn) = (x^2+y^2+z^2+w^2)(a^2+b^2+c^2+d^2) \\ &= r^2+s^2+t^2+u^2 \end{aligned}$$

where

$$\begin{aligned} r &= xa+yb+zc+wd \\ s &= xb-ya+zd-wc \\ t &= xc-yd-za+wb \\ u &= xd+yc-zd-wd \end{aligned}$$

Now, by equations (5), (7)

$$r = xa+yb+zc+wd \equiv a^2+b^2+c^2+d^2 \equiv 0 \pmod{k}$$

Similarly, $s \equiv t \equiv u \equiv 0 \pmod{k}$. Thus $k|r$, $k|s$, $k|t$, $k|u$.

So, $np = \left(\frac{r}{k}\right)^2 + \left(\frac{s}{k}\right)^2 + \left(\frac{t}{k}\right)^2 + \left(\frac{u}{k}\right)^2$, where $r/k, s/k, t/k, u/k$ are all integers.

Because, $0 < n < k$ we therefore arrive at a contradiction to the choice of k as the least integer for which kp is a sum of four squares.

So, our assumption that $k \neq 1$ was false. Hence, $k = 1$ which implies that by equation (4) $p = x^2+y^2+z^2+w^2$ i.e. p can be written as sum of four squares. ■

Theorem 4.5. *(**Lagrange**) Any positive integer n can be written as the sum of four squares, some of which may be zero.*

Proof. Clearly, the integer 1 is expressible as $1 = 1^2 + 0^2 + 0^2 + 0^2$, a sum of four squares. Asuume that $n > 1$ and let $n = p_1 p_2 \cdots p_r$ be the prime factorization of n into (not necessarily distinct) primes. By Theorem 4.4 each p_i is realizable as a sum of four sqaures, and the Euler's identity (Lemma 4.1) permits us to express the product of any two primes as a sum of four squares. This, by induction, extends to any finite number of prime factors, so applying the identity $r - 1$ times, we obtain the desired representation of n as sum of four squares. ■

Note. We conclude in this section that every positive integer can be represented as a sum of four squares. Since, $g(2) \neq 2$ and $g(2) \neq 3$ therefore, $g(2) = 4$.

5 Related Problems and Recent Developments

Lagrange's theorem motivated the more general problem of representing each positive integer as a four-variable expressions of the form $ax^2 + by^2 + cz^2 + dw^2$ where a, b, c, d are given positive integers. In **1916**, Indian mathematician **Srinivasa Ramanujan** presented 53 such "universal quadratics", four of which had been previously known. For instance, the expression $x^2 + 2y^2 + 3z^2 + 8w^2$ yields all positive integers.

In **2005**, **Manjul Bhargava** proved that there are only 204 of the desired quadratics. Finally in completion to the question, **Bhargava and Jonathan Hanke** found a particular set of 29 positive integers that will serve as a check for any quadratic expression. If the quadratic expression can represent each of those 29 integers, it can represent all positive integers.

Another problem that has attracted considerable attention is whether an nth power can be written as a sum of n nth powers, with $n > 3$. Progress was first made in 1911 with discovery of the smallest solution in fourth powers, $353^4 = 30^4 + 120^4 + 272^4 + 315^4$. In fifth powers, the smallest solution is $72^5 = 19^5 + 43^5 + 46^5 + 47^5 + 67^5$.

A related question is the *Euler's sum of powers conjecture.* **Leonhard Euler** conjectured that at least n nth powers are required to sum to an nth power, $n > 2$. A counterexample to this conjecture was provided by **L.J. Lander** and **T.R. Parkin** , $27^5 + 84^5 + 110^5 + 133^5 = 144^5$ as the smallest instance in which four fifth powers sum to a fifth power.

References

[1] Burton, David M., *Elementary Number Theory*, McGraw Hill Education (India) Private Limited, 2012.

[2] Hardy, G.H.; Wright, E.M., *An Introduction to the Theory of Numbers*. 4th ed. London: Oxford University Press, 1960. 297-316.

[3] Nathanson, Melvyn B., *Elementary Methods in Number Theory*, Graduate Text in Mathematics, Springer-Verlag New York, 2000.

[4] Bhaskar, J.(2008), *Sum of Two Squares*, University of Chicago (REU 2008).

[5] Wong, M.(2009), *Representing Integers as Sums of Squares*, University of Chicago (REU 2009).

[6] Lander, L.J.; Parkin, T.R. (1966), *Counterexample to Euler's Conjecture on Sums of Like Powers*, Bull. Amer. Math. Soc., Volume 72, Number 6 (1966), 1079.

KUSHAGRI TANDON, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
kushagritandon28@gmail.com

Counting Homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$

Khuisangmi Konghay

Abstract

This paper proves the results concerning the number of group homomorphisms and ring homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$. An example is given to illustrate how the formula proven in the above result is an excellent tool for finding the number of ring homomorphisms without having to go through intricate calculations.

1 Introduction

We begin by stating the following definitions.

Definition 1.1: Ring

A ring R is a set with two binary operations, addition (denoted by $a + b$) and multiplication (denoted by ab) satisfying the following properties:

1. $a + b = b + a$ for all $a, b \in R$. [Additive Commutativity]
2. $(a + b) + c = a + (b + c)$ for all $a, b, c \in R$. [Additive Associativity]
3. There is an element 0 in R such that $a + 0 = a$ for all $a \in R$. [Existence of Additive Identity]
4. For each $a \in R$, there exists an element $(-a) \in R$ such that $a+(-a) = 0$. [Existence of Additive Inverse]
5. $a(bc) = (ab)c$ for all $a, b, c \in R$ for all $a, b, c \in r$. [Multiplicative Associativity]
6. $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ for all $a, b, c \in r$. [Left and Right Distributivity]

Definition 1.2: Ring Homomorphism

A ring homomorphism ϕ from a ring R to a ring S is a mapping from R to S that preserves the two ring operations; that is, for all $a, b \in R$,

$$\phi(a + b) = \phi(a) + \phi(b) \text{ and } \phi(ab) = \phi(a)\phi(b).$$

2 Number of group homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$

Theorem: *The number of group homomorphisms from* $\mathbf{Z}_m$ *into* $\mathbf{Z}_n$ *is* $gcd(m, n)$.

Proof: The order of the image of a group homomorphism from $\mathbf{Z}_m$ into $\mathbf{Z}_n$ must divide both m and n.
Thus, it is a divisor of $gcd(m, n)$.
If k is a common divisor of m and n, then $\mathbf{Z}_n$ has a unique subgroup of order k.
Moreover, this subgroup has $\phi(k)$ generators.
Now, in order to specify a group homomorphism from $\mathbf{Z}_m$ onto a subgroup H of $\mathbf{Z}_n$, it is necessary and sufficient to map the integer 1[1] to a generator of H.
Thus, we conclude that the number of group homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$ is simply $\sum_{k|gcd(m,n)} \phi_k$, where ϕ_k is the number of generators of *the*[2] subgroup H of $\mathbf{Z}_n$ having order k.

By a result from number theory, $\sum_{k|gcd(m,n)} \phi_k = gcd(m, n)$.
Thus, we have proved the above theorem.

3 Number of ring homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$

Next, we consider the corresponding problem for rings. Thus, we proceed to prove the following theorem that gives number of ring homomorphisms from the ring $\mathbf{Z}_m$ into the ring $\mathbf{Z}_n$.

Theorem: *The number of ring homomorphisms from* $\mathbf{Z}_m$ *into* $\mathbf{Z}_n$ *is* $2^{\omega(n)-\omega\left(\frac{n}{gcd(m,n)}\right)}$, *where* $\omega(n)$ *denotes the number of distinct prime divisors of* n.

Proof: As in the case of a group, a ring homomorphism is completely determined by its action on 1.

[1] Our choice of the integer 1 is based on the fact that it is a generator of the set $\mathbf{Z}_m$, since $gcd(1, m) = 1$.

[2] by Fundamental Theorem of Cyclic Groups

Now, 1 is an idempotent element in $\mathbf{Z}_m$, and so is $\phi(1)$ in $\mathbf{Z}_n$.

Let $n = q_1^{t_1} q_2^{t_2} \cdots q_s^{t_s}$ be the prime decomposition of n.
Then, by the Chinese Remainder Theorem, $\mathbf{Z}_n$ is naturally ring homomorphic to the direct sum: $Z_{{q_1}^{t_1}} \oplus Z_{{q_2}^{t_2}} \oplus \cdots \oplus Z_{{q_s}^{t_s}}$.
Also, we see that any ring homomorphism from $\mathbf{Z}_m$ into $\mathbf{Z}_n$ induces a ring homomorphism from $\mathbf{Z}_m$ into $\mathbf{Z}_{{q_i}^{t_i}}$ for $i = 1, 2, \ldots s$.

So, let $\phi(1) = a \in \mathbf{Z}_n$.
Then in the direct sum $a = (a_1, a_2, \cdots, a_s)$ where $a_i \in \mathbf{Z}_{{q_i}^{t_i}}$.
Then, each a_i is also an idempotent element in $\mathbf{Z}_{{q_i}^{t_i}}$ and hence $a_i = 0$ or $a_i = 1$.
Thus, there are **at most** 2^s ring homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$.
Also, note that a ring homomorphism is also a group homomorphism. Thus, the additive order of a_i divides m.

The converse of this is also true.
That is, suppose $(a_1, a_2, \cdots, a_s)$ is any member of the direct sum with $a_i = 0$ or $a_i = 1$ such that additive order of a_i divides m.
Then there is a ring homomorphism from $\mathbf{Z}_m$ into $Z_{{q_1}^{t_1}} \oplus Z_{{q_2}^{t_2}} \oplus \cdots \oplus Z_{{q_s}^{t_s}}$ which carries 1 to $a = (a_1, a_2, \cdots, a_s)$.

Thus, the number of ring homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$ is simply the *number of s-tuples* which meet the two conditions that $a_i = 0$ or $a_i = 1$ and additive order of $a_i \mid m$.

Now, the additive order of 0 is 1 and the additive order of 1 in $\mathbf{Z}_{{q_i}^{t_i}}$ is ${q_i}^{t_i}$.
Therefore, we may take $\mathbf{a_i = 0}$ **or** $\mathbf{a_i = 1}$ **when** $\mathbf{{q_i}^{t_i} \mid m}$ and $\mathbf{a_i = 0}$ **when** $\mathbf{{q_i}^{t_i} \nmid m}$.
We state, without proving, that

$${q_i}^{t_i} \nmid m \text{ if and only if } q_1 \mid \tfrac{n}{gcd(m,n)}.$$

Therefore, denoting the number of distinct prime divisors of the integer k by $\omega(k)$, we conclude that the number of ring homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$ is $2^{\omega(n) - \omega\left(\frac{n}{gcd(m,n)}\right)}$.
Thus, we have proved the above theorem.

We now demonstrate the application of this formula with an example.

Example: *Compute the number of ring homomorphisms*

$$\phi : \mathbf{Z}_{20} \to \mathbf{Z}_{30}$$

We will first solve the problem by finding all possible (distinct) homomorphisms.

Let $\phi : \mathbf{Z}_{20} \to \mathbf{Z}_{30}$ be a ring homomorphism.
Let $\phi(1) = a$, then for $x \in \mathbf{Z}_{30}$

$$\phi(x) = \phi(1.x) = \phi(1).\phi(x) = a.x$$

where $a \in \mathbf{Z}_{30}$.
We know that a ring homomorphism must be a group homomorphism as well.
Therefore, the order of a must divide both 30 and 20.
That is, $|a| \mid 30$ and $|a| \mid 20$.
Thus, $|a|$ is a common divisor of 20 and 30, so $|a| \in \{1, 2, 5, 10\}$.
Since we considered $\phi(1) = a$, therefore, we can also say that the image of 1 has order $1, 2, 5$ or 10.

Order (n)	Elements having order n
1	0
2	15
5	6, 12, 18, 24
10	3, 9, 21, 27

Moreover, since 1 is an idempotent, so is $\phi(1) = a$. That is, $a^2 = a$.
Computing the squares of these elements $(mod 30)$, we find that
$\mathbf{0^2 = 0}$;
$\mathbf{15^2 = 15}$;
$\mathbf{6^2 = 6}$, $12^2 = 24$, $18^2 = 24$, $24^2 = 6$;
$3^2 = 9$, $9^2 = 21$, $\mathbf{21^2 = 21}$, $27^2 = 9$;

Therefore, we have **four idempotent elements** and hence, the following four ring homomorphisms:
$\phi(x) = 0$, $\phi(x) = 6x$, $\phi(x) = 15x$, $\phi(x) = 21x$;

We now consider the above formula.
Here, $m = 20$ and $n = 30$.

Now, $\frac{n}{gcd(m,n)} = \frac{30}{gcd(20,30)} = \frac{30}{10} = 3$
That is, $\omega\left(\frac{n}{gcd(m,n)}\right) = \omega(3) = 1$. Also, $\omega(n) = \omega(30) = 3$.
Hence, $\mathcal{N}(\phi : \mathbf{Z}_m \to \mathbf{Z}_n) = 2^{\omega(n)-\omega\left(\frac{n}{gcd(m,n)}\right)} = 2^{\omega(30)-\omega(3)} = 2^{3-1} = 2^2 = 4$.

Thus, using this formula, it is verified that four homomorphisms exist from $\mathbf{Z}_{20}$ into $\mathbf{Z}_{30}$.
Observation: *The number of group homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$ is the same as the number of group homomorphisms from $\mathbf{Z}_n$ into $\mathbf{Z}_m$. However, the corresponding statement for rings is not true.*

4 Conclusion

This paper provides formulas to calculate the the number of group homomorphisms and ring homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$, which proves to be an excellent tool without having to go through intricate calculations.

References

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

[2] Gallian, J. A. and Buskirk, J.V., *The Number of Homomorphisms from $\mathbf{Z}_m$ Into $\mathbf{Z}_n$*, The American Mathematical Monthly, Vol. 91, No. 3 (Mar., 1984), pp. 196-197.

[3] Kowkas, Sultan A. I., *On The Number of Ring Homomorphisms Over Certain Rings*, Birzeit University, 16th December 2014.

KHUISANGMI KONGHAY, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
konghaykhuimi@gmail.com

Extension of Course Contents

A great deal of learning happens beyond the formal coursework. This section hence aims to provide a creative, fertile setting for productive research that goes beyond the confines of classroom, and precincts of syllabi. It strengthens and expands the existing knowledge and adds interests to the course and provides an experience of trans-formative learning.

Symmetry Relations Between Crystals

Richa Sharma

Abstract

This paper discusses one of the central concepts in crystallography, namely symmetry. The arrangement of atoms inside a crystal has translational symmetry, which leads us to a type of group where the elements are linear transformations and the operation is composition. These define space groups which, in combination with lattice types, describe all of the possible crystal types.

1 Introduction

In the early days of crystal structure determinations, it became clear that the laws governing the packing of atoms and molecules be understood, that crystal structures have to be classified and ordered, and that relations between them must be recognized. Symmetry is indispensable for the determination and description of a specific crystal structure. Related crystal structures often have different space groups, and the relations between them result from group-subgroup relations between their space groups. Hence, in this article, the emphasis will be on the notion that group theory - in the crystallographic context - is merely a mathematical formalism that describes symmetry.

Principles of symmetric behaviour in Crystal Chemistry :

- In the crystal structure, the arrangement of atoms reveals a pronounced tendency towards the highest possible symmetry.
- Counteracting properties due to special properties of the atoms may prevent the attainment of the highest possible symmetry. However, in most cases the deviations from the ideal symmetry are small.
- During phase transitions and solid state reactions which results in products of lower symmetry, the higher symmetry of the starting material is indirectly preserved by the formation of oriented domains.
- Atoms of the same kind tend to occupy equivalent positions.

Depending on chemical composition, the kind of chemical bonding, electronic configuration, relative size of the atoms, pressure, temperature etc., there exists one energetically most favourable surrounding for atoms of a given kind, which all of these atoms strive to attain. According to quantum theory, atoms of the same kind are indistinguishable, but in a crystal this is only ensured if they are symmetry-equivalent.

2 Subgroups of space groups

A space group is the symmetry group of a configuration in space, usually in three dimensions. In crystallography, space groups are also called the *crystallographic* or Fedorov groups, and represent a description of the symmetry of the crystal. The symmetry operations are the group elements that make up the space group. Removal of some of the symmetry operations results in one of the subgroups. If there exists no intermediate subgroup between the space group and one of its subgroups, then this subgroup is a *maximal subgroup*, where the index is the factor by which the number of symmetry operations are reduced (this factor is always a prime number or some power of a prime number).

3 Atomic and wyckoff positions

It is by no means an easy task to recognize if two differently documented structures are alike or not. For all subgroups, except $I\,m\,3\,\bar{m}$[1] and $I\,a\,3\,\bar{d}$, there always exist several different equivalent sets of atomic coordinates for the exactly same crystal structure with an unchanged setting of the space group. For the group G, the number of equivalent coordinate sets is e; e is the index of G in its *Euclidean normalizer* $N_E(G)$. The infinitely large set of symmetry-equivalent points in a space group is called an **orbit**. If the corresponding coordinates are completely fixed by symmetry, the orbit is identical with the **Wyckoff position**. If however, one or more coordinates are variable, the Wyckoff position comprises infinitely many orbits; they differ in the variable coordinate. A Wyckoff position is designated by the Wyckoff label, for e.g., 4*c*. Here, 4 is the multiplicity; it shows how many points belonging to an orbit of the Wyckoff position in question are contained in one unit cell and c is an alphabetical label according to the listing of the *International Tables*. Many space groups have several equivalent Wyckoff positions that commonly make up a **Wyckoff set**. Between the points of an orbit and the corresponding points of a subgroup there exists a one-to-one relation. Both orbits have the same magnitude. Upon symmetry reduction, a Wyckoff position will either split into several symmetry-independent positions, or its site symmetry is reduced, or both happen. The relations of the Wyckoff set can be obtained with the computer program WYCKSPLIT, which is accessible via the inter-

[1]This is a Hermann−Mauguin notation. In geometry, Hermann−Mauguin notation is used to represent the symmetry elements in point groups, plane groups and space groups.

net at the Bilbao Crystallographic Server. It requires the input of the space group, subgroup, basis transformation and origin shift; it does not work for non-conventional settings unless transformation matrices are given to convert standard settings.

4 Symmetry relations between crystal structures

Group-subgroup relations can be depicted by graphs in which the symbol for every group is connected with the symbols of its maximal subgroups. Two graphs are sufficient to present all kinds of group-subgroup relations between point groups. In the case of space groups, three kinds of maximal subgroups are distinguished: *translationengleiche* subgroups which have kept all translations but belong to a lower crystal class; *klassengleiche* subgroups having fewer translations but the same crystal class; *isomorphic* subgroups which belong to the same or the enantiomorphic space group type and have fewer translations, they are a special kind of klassengleiche subgroups. The kinds of translationengleiche subgroups can be depicted in thirty-seven graphs, and those of klassengleiche subgroups can be depicted in twenty-nine graphs. The number of isomorphic subgroups is always infinite. Minimal supergroups of space groups are more manifold than maximal subgroups. The symmetry group of an object in three-dimensional space is a *layer group* if it has translational symmetry only in two dimensions, and a *rod group* if it has translational symmetry only in one dimension. They are designated by modified Hermann-Mauguin symbols. This section highlights the different types of group-subgroup relations with the aid of simple examples.

4.1 Translationengleiche maximal subgroups

The space group $P\ b\ c\ a$ of PdS_2 is a translationengleiche maximal subgroup of $P\,a\,\bar{3}$, the space group of pyrite. The 3-fold axes of the cubic space group are lost, whereas the 2-fold screw axes parallel to the edges of the cube and the glide planes are retained, but they no longer are equivalent in the orthorhombic subgroup. As shown in Fig.1, the atomic coordinates have not changed much but the two structures differ (the c axis of PdS_2 being strongly stretched). Upon transition from $P\,a\,\bar{3}$ to $P\ b\ c\ a$ none of the occupied Wyckoff positions split, but their site symmetries are reduced. The relations between FeS_2,

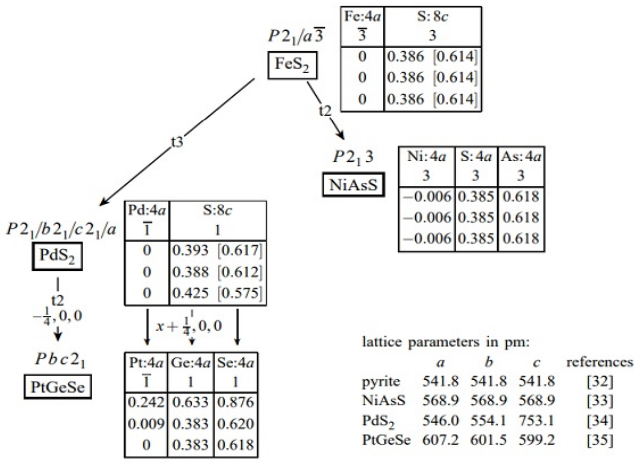


Figure 1: Bärnighausen tree for the structural family of pyrite. Coordinates in brackets refer to symmetry equivalent positions

PdS_2, $NiAsS$ and $PtGeS$ are summarized in Fig.1 using the Bärnighausen[2] tree method.

4.2 Klassengleiche maximal subgroups

Let us consider two variants of AlB_2 type as an example of Klassengleiche subgroups. AlB_2 has a simple hexagonal structure in the space group $P6/m\,m\,m$. The $ZrBeSi$ type has a similar structure, but the sheets consist of Be and Si atoms. As a consequence, the inversion centres in the middle of the six-membered rings cannot be retained, whereas those in the Al positions are retained in the Zr positions. This enforces a symmetry reduction to the Klassengleiche subgroup $P6_3/m\,m\,c$ with doubled c vector. $P6/m\,m\,m$ has two different Klassengleiche subgroups of the same type $P6_3/mmc$ with doubled basis vector c. The second one corresponds to $CaIn_2$. Here, the alternating shift of the atoms no longer permits the existence of mirror planes in the layers; however, neighbouring layers are mutually mirror-symmetrical. The calcium atoms are on the mirror planes, but no longer on inversion centres. The difference between the two subgroups $P6_3/m\,m\,c$ consists in the selection of the symmetry operations that are lost with the doubling of c. Among Klassengleiche subgroups

[2]A tree of group-subgroup relations between the involved space groups, now called a Bärnighausen tree. It's main concept is to start from a simple, highly symmetrical crystal structure and to derive more and more complicated structures by distortions and (or) partial substitutions of atoms.

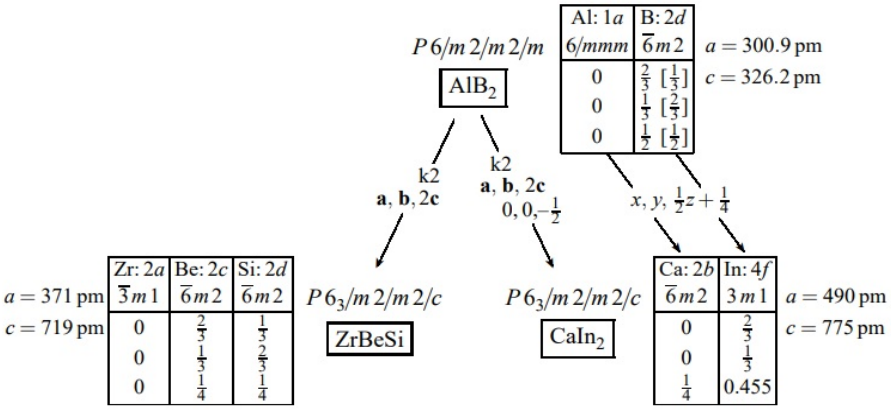


Figure 2: Two hettotypes of the AlB_2 type having the same space-group type and a doubled c axis, but different origin positions. Due to the doubling of c the z coordinates are halved. The origin shift of $\left(0, 0, -\frac{1}{2}\right)$ in the right branch refers to the lattice of the aristotype; as a consequence, $\frac{1}{4}$ has to be added to the z coordinates of the hettotype.

of index 2, there often exist two or four different subgroups of the same space-group type which differ in their origin positions. It is important to choose the correct one of them, with the correct origin shift. Note that an **aristotype** is a high-symmetry crystallographic structure type that can be viewed as an idealized version of a lower-symmetry structure, the lower-symmetry structure being the **hettotype**.

4.3 Isomorphic maximal subgroups

Isomorphic subgroups are a special kind of klassengleiche subgroups. The main particularity is that each space group has infinitely many isomorphic subgroups whose index agrees with the factor by which the unit cell has been enlarged. Consider the trirutile structure as an example. The space group of rutile, $P4_2/m\,n\,m$ has an isomorphic subgroup of index 3, but none of index 2. By triplication of c, it becomes possible to substitute the titanium atom positions of rutile by two different kinds of atoms in a ratio of 1:2, as in $ZnSb_2O_6$. Note that rutile and trirutile have different space groups of the same space-group type. A space group includes a **specific translational lattice** and is used to **designate the symmetry** of a given crystal structure. The space group type, however, is independent of the lattice metrics.

5 The space groups of two structures have a common supergroup

It is not necessary that two crystal structures be related by having a direct group-subgroup relation, they can be intimately related even when there exists a commom supergroup. The above mentioned structures of *NiAsS* and *PtGeSe* offer an example. In that case, the pyrite type corresponds to the common supergroup. Even if there is no known representative, it can be useful to look for a common supergroup. Using the modular way to put together symmetry relations set forth in the scheme discussed in the preceding sections, large family trees can be constructed. Headed by an aristotype, they show structural relationships among many different crystal structures.

6 Conclusion

In addition to showing relations between known structures types, one can also find subgroups of an aristotype for which no structures are known. Hence, for any space group appearing in the Bärnighausen tree, one can calculate how many different structure types are possible for a given chemical composition.

References

[1] Müller, U., *Symmetry Relationships Between Crystal Structures: Applications of Crystallographic Group Theory in Crystal Chemistry*, Oxford University Press, 2013.

[2] Powell, R.C., *Symmetry, Group Theory, and the Physical Properties of Crystals*, Springer Publication, 2010.

[3] Nazarov, M. (April, 2012), *Representation Theory of the Symmetric Group*, The University of Arizona, 1-33.

[4] https://en.wikipedia.org/wiki/Crystallographic point group

RICHA SHARMA, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
richa131999@gmail.com

Numerical Integration

Manishika Negi

Abstract

This paper discusses three techniques, namely Trapezoidal rule, Midpoint approximation and Simpson's rule for approximating the value of those definite integral functions whose antiderivative cannot be found. The Midpoint rule uses rectangles to approximate area under the curve, Trapezoidal rule uses straight lines while Simpson's rule uses parabolas.

1 Introduction

Numerical integration is the process of computing the value of a definite integral, $\int_a^b f(x)dx$, when the values of the integrand function, $y = f(x)$ are given at some tabular points. The most straightforward numerical integration technique uses the **Newton-Cotes formulae** (also called quadrature formulae), which approximate a function tabulated at a sequence of regularly spaced intervals by various degree polynomials. If the endpoints are tabulated, then the 2- and 3-point formulae are called the Trapezoidal rule and Simpson's rule, respectively. The term "numerical integration" first appeared in 1915 in the publication *A Course in Interpolation and Numeric Integration for the Mathematical Laboratory* by David Gibb.

2 Midpoint Approximation

This rule uses the midpoint m_k of each of the intervals i.e. the base of each rectangle, as the point at which to evaluate the function for the Riemann sum.

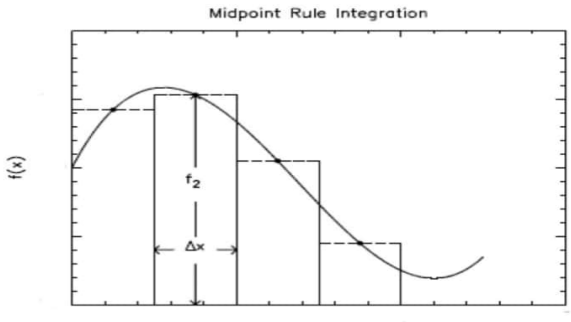
2.1 Derivation of formula

Consider using the Riemann sum to estimate a definite integral of a continuous function f as follows:

$$\int_a^b f(x)dx = \lim_{max\ h \to 0} \sum_{k=1}^{n} f(x_k)h_k$$

where h_k is the width of the kth subinterval of a partition $a = x_0 < x_1 < x_2 <$

$\ldots < x_n = b$ of $[a, b]$ into n subintervals and x_k denotes an arbitary point in the kth subinterval.



If we take all subintervals of the same width, so that $h = (b - a)/n$, then as n increases the Riemann sum will eventually be a good approximation of the definite integral. The midpoint approximation is defined as:

$$\int_a^b f(x)dx \approx h[y_{m_1} + y_{m_2} + \ldots + y_{m_n}]$$

where $y_{m_1}, y_{m_2}, \ldots, y_{m_n}$ are the values of f at the midpoints $m_1, m_2 \ldots m_n$ of the subintervals.

2.2 Midpoint Error Bounds

If f'' is continuous on $[a, b]$ and if K_2 is the maximum value of $|f''|$ on $[a, b]$ then.

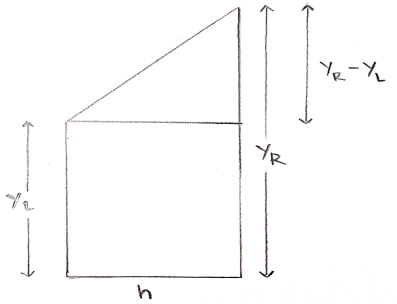
$$|E_M| \;=\; \left| \int_a^b f(x)dx - M_n \right| \;\leq\; \frac{(b-a)^3 K_2}{24n^2} \qquad (1)$$

3 Trapezoidal Rule

The general idea is to use trapezoids instead of rectangles to approximate the area under the graph of a function.

3.1 Derivation of formula

In the interval $[a, b]$, we subdivide it into n subintervals of equal width $h = (b-a)/n$. This gives rise to the partition $a = x_0 \leq x_1 \leq x_2 \leq \ldots \leq x_n = b$, where for each j, $x_j = a + jh$, $0 \leq j \leq n$. Moreover, we let $y_j = f(x_j)$; $0 \leq j \leq n$. That is, the vertical edges go from the x-axis to the graph of f.

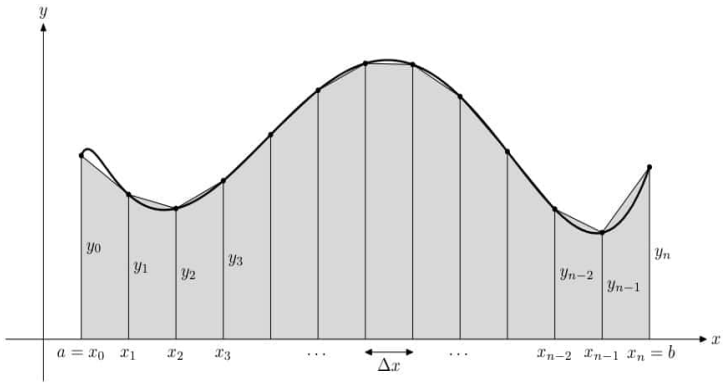


With reference to the sketch above,

$$\begin{aligned} \textit{Area of a trapezoid} &= \textit{Area of rectangle} + \textit{Area of triangle} \\ &= hy_L + \frac{h}{2}(y_R - y_L) \\ &= h(\frac{y_L + y_R}{2}). \end{aligned}$$

Now for n subintervals the trapezoidal approximation $\int_a^b f(x)dx$ is given by

$$\begin{aligned} T_n &= \frac{h}{2}(y_0 + y_1) + \frac{h}{2}(y_1 + y_2) + \ldots + \frac{h}{2}(y_{n-1} + y_n) \\ &= \frac{h}{2}(y_0 + 2y_1 + 2y_2 + \ldots + 2y_{n-1} + y_n) \\ &= \frac{h}{2}(y_0 + y_n + 2\sum_{j=1}^{n-1} y_j) \end{aligned}$$



Hence the trapezoidal rule is defined as :

$$\int_a^b f(x)\,dx \approx T_n = \frac{h}{2}(y_0 + y_n + 2\sum_{j=1}^{n-1} y_j)$$

3.2 Trapezoidal Error Bounds

If f'' is continuous on [a,b] and if K_2 is the maximum value of $|f''|$ on $[a, b]$, then

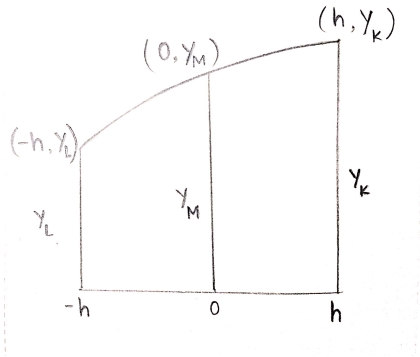
$$|E_T| \;=\; \left|\int_a^b f(x)dx - T_n\right| \;\leq\; \frac{(b-a)^3 K_2}{12n^2} \tag{2}$$

4 Simpson's Rule

Simpson's method replaces the slanted-line tops with parabolas. Though two points determine the equation of a line, three are required for a parabola. We now develop a formula for the area of a parabolic-top area-element, and the sum of such areas is to become the Simpson approximation.

4.1 Derivation of formula

We consider a parabola $y = Ax^2 + Bx + C$ with its axis parallel to the y-axis and passing through three equally spaced points $(-h, y_L)$, $(0, y_M)$, and (h, y_R).



Substituting the three points into the equation of parabola gives the following three equations in the three unknowns A, B and C.

$$\begin{aligned} y_L &= Ah^2 - Bh + C \\ y_M &= C \\ y_R &= Ah^2 + Bh + C \end{aligned}$$

Solving these three equations (by adding the first to the last, and then by subtracting the last from the first), yields:

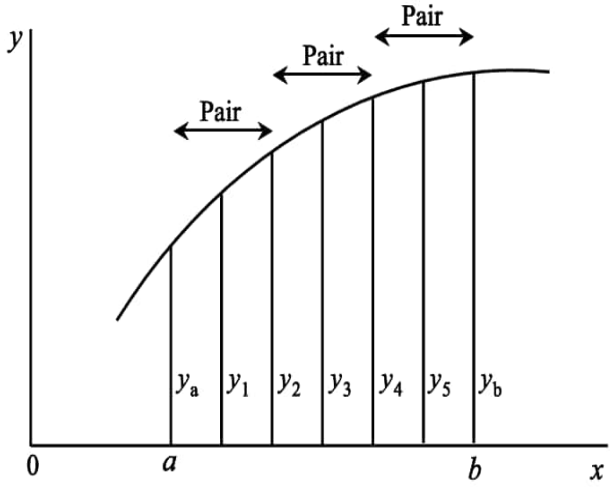
$$\begin{aligned} A &= \frac{y_L + y_R - 2y_M}{2h^2} \\ B &= \frac{y_R - y_L}{2h} \\ C &= y_M \end{aligned}$$

We now compute the area under the parabola $y = Ax^2 + Bx + C$ and above the interval [-h, h] for the values of A , B and C that we just found.

$$\begin{aligned} \int_{-h}^{h} (Ax^2 + Bx + C)dx &= (A\frac{x^3}{3} + B\frac{x^2}{2} + Cx)|_{-h}^{h} \\ &= \frac{1}{3}\left(2Ah^3 + 2Ch\right) \\ &= h(\frac{2}{3}Ah^2 + 2C) \end{aligned}$$

$$
\begin{aligned}
&= h(\frac{1}{3}(y_L + y_R - 2y_M) + 2y_M) \\
&= \frac{h}{3}(y_L + y_R - 2y_M + 6y_M) \\
&= \frac{h}{3}(y_L + y_R + 4y_M)
\end{aligned}
$$

Now, let n be an even positive integer, and suppose we divide an interval $[a, b]$ into n equal parts, each of length $h = (b - a)/n$. And suppose f is a function defined on $[a, b]$. As before, we label the resulting partition $a = x_0 \leq x_1 \leq x_2 \leq \ldots \leq x_n = b$, where for each j, $x_j = a + jh,\ 0 \leq j \leq n$. And again, we let $y_j = f(x_j),\ 0 \leq j \leq n$.



Next, start at the left endpoint a of the interval and erect a parabolic-top area-element on the first two subintervals. The base of this area-element goes from x_0 to x_2, and we use as vertical sides the lines that intersect the graph at (x_0, y_0) on the left and (x_2, y_2) on the right. The point (x_1, y_1) on the graph of f at the midpoint of the interval gives the third point we need to determine the parabola that forms the top of the area-element. From the formula we developed above, let
S_n = sum of the areas of $n/2$ parabolic-top area-elements.
Now,

$$
S_n = \frac{h}{3}\left[(y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + \cdots + (y_{n-2} + 4y_{n-1} + y_n)\right]
$$

$$= \frac{h}{3}\left[y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n\right]$$

Hence the Simpson's rule is defined as:

$$\int_a^b f(x)\,dx \approx S_n = \frac{h}{3}\left[(y_0 + y_n) + 4\sum y_{odd} + 2\sum y_{even}\right]$$

4.2 Simpson Error Bound

If $f^{(4)}$ is continuous on $[a, b]$ and if K_4 is the maximum value of $\left|f^{(4)}(x)\right|$ on $[a, b]$, then

$$|E_S| = |\int_a^b f(x)dx - S_n| \;\leq\; \frac{(b-a)^5 K_4}{180n^4} \tag{3}$$

If $f(x)$ is a polynomial of degree 3 or less, then we have $f^{(4)}(x) = 0$ for all x, so $K_4 = 0$ and hence $|E_S| = 0$. Thus, **Simpson's rule gives exact results for polynomials of degree** 3 **or less**.

5 A Comparision Of The All Methods

Of the three methods mentioned above, Simpson's rule generally produces more accurate results than the midpoint or trapezoidal approximation for the same amount of work. Let us express (1), (2) and (3) in terms of the subinterval width $h = \frac{b-a}{n}$
We get,

$$|E_M| \leq \frac{K_2}{24}(b-a)h^2$$

$$|E_T| \leq \frac{K_2}{12}(b-a)h^2$$

$$|E_S| \leq \frac{K_4}{180}(b-a)h^4$$

For Simpson's rule, the upper bound on the absolute error is proportional to h^4, whereas the upper bound on the absolute error for the midpoint and

trapezoidal approximations are proportional to h^2. Thus, reducing the interval width by a factor of 10, for example, reduces the error bound by a factor of $(10)^2 = 100$ for the midpoint and trapezoidal approximations but reduces the error bound by a factor of $(10)^4 = 10,000$ for Simpson's rule.
This suggests that, **as n increases, the accuracy of Simpson's rule improves much more rapidly than that of the other approximations**.

6 Conclusion

This article discusses the three basic numerical techniques that helps us to approximate the value of the definite integral which cannot be solved analitically i.e by using mathematical approach.
The Simpon's rule is the most efficient technique, primarily due to the use of parabolas to approximate each part of the curve. Also, since it uses quadratic polynomials, Simpson's rule actually gives exact results when approximating integrals of polynomials up to cubic degree. While the midpoint and trapezoidal approximations give exact results for polynomials of degree 1 or less.
Numerical methods can handle any complicated physical geometries which are often impossible to solve analytically.

References

[1] Anton, H., Bivens, I. and Davis, S., *Calculus 10th edition*, John Wiley and Sons (Asia), Singapore, 2013.

[2] https://nptel.ac.in/courses/122104018/node121.html

[3] https://https://math.dartmouth.edu/ m3cod/klbookLectures/406unit/trap.pdf

MANISHIKA NEGI, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
manishikanegi@hotmail.com

Non-Orientable Surfaces

Rajlaxmi Adwant
Simran
Lipika Parekh

Abstract

This paper discusses non-orientable surfaces, with special emphasis on the Möbius strip. It strives to explain the complexities associated with non-orientability and thus ventures into Topology.

1 Introduction

A non-orientable surface is any surface that contains a Möbius band, or, strictly speaking, a subset that is homeomorphic to the Möbius band. On a non-orientable surface, there is no way to consistently define the notions of 'right' and 'left' and anything which slides around a non-orientable surface will come back to its starting point as a mirror image. Non-orientable surfaces form two classes, those based on the real projective plane, which have odd Euler characteristic, and those based on the Klein bottle, which have even Euler characteristic. All surfaces with an odd Euler characteristic are non-orientable. All non-orientable surfaces can be obtained by starting with the projective plane or the Klein bottle and adding handles to them, where the number of handles are known as *the genus*. A non-orientable surface can not be embedded (i.e. mapped one-to-one) in three-space but can be immersed (mapped locally but not globally) there.

2 Manifold

A manifold is a topological space that locally resembles Euclidean space near each point. More precisely, each point of an n-dimensional manifold has a neighbourhood that is homeomorphic to the Euclidean space of dimension n. In this more precise terminology, a manifold is referred to as an n-manifold. A manifold is a shape. When one thinks of shapes, one normally thinks about circles or squares or triangles, but the definition of a shape needs to be more broad. After all, almost anything that can be imagined has a shape, even though that shape may not have a specific name.Shapes may also have different dimensions.For example, a square has two dimensions, height and width, and a cube has three dimensions, height, width, and depth, but clearly both are shapes. Thus a one-dimensional manifold (or one-manifold) is just a one-dimensional shape or a curve. A two-dimensional manifold (or two-manifold) is just a two dimensional shape or a surface.

3 Orientable Surfaces

Orientability is a property of surfaces in Euclidean space that measures whether it is possible to make a consistent choice of surface normal vector at every point. A choice of surface normal allows one to use the right-hand rule to define a "clockwise" direction of loops in the surface, as needed by Stokes' theorem for instance. More generally, orientability of an abstract surface, or manifold, measures whether one can consistently choose a "clockwise" orientation for all loops in the manifold. Equivalently, a surface is orientable if a two-dimensional figure such as a circle in the space cannot be moved (continuously) around the space and back to where it started so that it looks like its own mirror image. To understand what it means to be orientable, consider the sphere. Clearly, this surface has no mirror-reversing effect because nobody on earth has ever returned mirror-reversed after a long journey. (A mirror-reversed person would seem to have changed handedness and would see all of our writing backwards). So presumably a sphere is orientable. The sphere also has some nice properties.A person can walk in any direction on the sphere and he or she will end up where at the starting point.This also means that one cannot fall off of the sphere. When a surface or any manifold has these properties, the manifold is said to have no boundary.But, is not the sphere itself a boundary for the space it encloses?Remember that while talking about two-manifolds only the surface itself is being considered.The space that a sphere encloses may not even exist. For example, a two-dimensional being living in the surface of a sphere has no conception of the three-space that might exist outside his or her world.The entire world simply looks like one infinite plane. Therefore, a two-dimensional being living in the surface of the sphere would truly believe that his or her world had no boundary, even though it has a finite area.

4 Möbius Strip

The Möbius strip (also sometimes called a Möbius band) has a boundary just like the cylinder, but it has only one boundary, not two.Upon tracing along the edge of the Möbius strip, one will cover the entire boundary, eventually returning to the starting point. One way to represent the Möbius strip as a subset of$\mathbb{R}^3$is by using the parametrization:

$$\begin{aligned} x(r,\alpha) &= \cos(\alpha)\cdot\left(1+\frac{r}{2}\cos\frac{\alpha}{2}\right) \\ y(r,\alpha) &= \sin(\alpha)\cdot\left(1+\frac{r}{2}\cos\frac{\alpha}{2}\right) \\ z(r,\alpha) &= \frac{r}{2}\sin\frac{\alpha}{2} \end{aligned}$$

where $0 \leq \alpha < 2\pi$ and $1 \leq r \leq 1$.

This creates a Möbius strip of width 1 whose center circle has radius 1, lies in thexyplane and is centered at$(0,0,0)$.The parameterαruns around the strip while r moves from one edge to the other. Any surface (or any manifold at all) which has a Möbius band in its structure is non-orientable.

5 Construction of Non Orientable Surfaces

There is a general method for constructing non-orientable surfaces which proceeds as follows. Choose three homogeneous polynomials of positive even degree and consider the map:

$$\mathbf{f} = \left(f_1\left(x,y,z\right), f_2\left(x,y,x\right), f_3\left(x,y,x\right)\right) : \mathbb{R}^3 \to \mathbb{R}^3$$

Then restricting x, y and z to the surface of a sphere by writing

$$\begin{aligned} x &= \cos\theta \sin\phi \\ y &= \sin\theta \sin\phi \\ z &= \cos\phi \end{aligned}$$

A way to understand the construction of non orientable surfaces like the Möbius strip and the Real Projective Plane is elucidated through the following image.

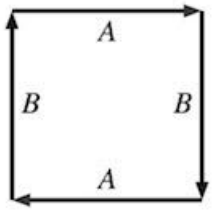


Figure 1

To make a Möbius band, twist the square so that the sides labeled A can be glued together with the arrows aligning. Then, to make the Möbius band into a projective plane, the same things must be done with the sides labeled B.

6 Other Non Orientable Surfaces

Some notable non-orientable surfaces are depicted below:

1. A Klein Bottle
2. Boy's Surface
3. Real Projective Plane

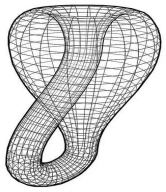


Figure 2

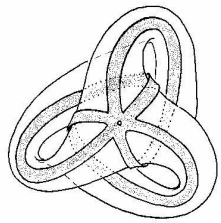


Figure 3

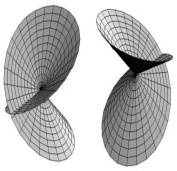


Figure 4

7 More about the Möbius Strip

Möbius (A.F. Möbius, 1790-1868) was a German mathematician and astronomer active in the 19^{th} century. He discovered it, so the Möbius strip is named after him. He was also an astronomer so he had an interest in the structure of space, and it is assumed that he devised the strip while exploring various notions about the boundaries of the universe. Stepping back for a moment into an earlier age, it was thought that the world was flat. The fact that the world is round was demonstrated by Columbus, who discovered the new world in 1492. Afterwards, Magellan made the first circumnavigation of the world, thus proving the fact. This also had a significant impact on geometry. Up until that time geometry was essentially Euclidean, and it was thought that all problems might be soluble with this well-integrated model. There was an approximate contemporary of Möbius known as *Lobachevsky* (1792-1856). He was a Russian mathematician who while pursuing research into parallel lines, validated a formulation of non-Euclidean geometry and presented his results. Non-Euclidean geometry is born from a rejection of the axiom of parallel lines. In

this geometry, the axiom of parallel lines is replaced by the notions that for a given straight line on a plane there are at least two straight lines no different from the given line that pass through a point which is not on the given line, the sum of the internal angles of a triangle is at least the sum of two right angles, and furthermore, straight lines are nitely closed, and divergences from Euclidean geometry can also be seen in the ordering of points on a straight line, etc. For example, consider a triangle NAB formed by the North Pole and two points on the equator. In this case the sum of the internal angles is larger than the sum of two right angles. The sum of the internal angles of a triangle drawn on a notepad is the sum of two right angles, but triangles on the Earths surface do not accord with Euclidean geometry. There is a relationship in the sense that non-Euclidean geometry is locally Euclidean. When Columbus and Magellan proved that the Earth is round, the mistake regarding the edge of the world was resolved. The ancient mystery regarding the nature of the boundary of the universe on the other hand remains. Space is vast, and being unable to reach the boundary, humans cannot resolve this problem absolutely.

References

[1] Nishiyama, Y., *Playing With Möbius Strip*, International Journal of Pure and Applied Mathematics, Volume 78, January 2012

[2] http://www.math.brown.edu/ banchoff/Yale/project15/math.htm

[3] http://www.daviddarling.info/encyclopedia/N/non-orientable_surface.html

[4] https://blogs.scientificamerican.com/roots-of-unity/a-few-of-my-favorite-spaces-the-real-projective-plane/

[5] http://webmath2.unito.it/paginepersonali/sergio.console/CurveSuperfici/AG11.pdf

[6] http://mathworld.wolfram.com/NonorientableSurface.html

[7] https://en.m.wikipedia.org/wiki/Orientability

RAJLAXMI ADWANT, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
rajlaxmiadwant@gmail.com

SIMRAN, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
simranbhola1999@gmail.com

LIPIKA PAREKH, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
plipika98@gmail.com

Interdisciplinary Aspects of Mathematics

Mathematics is just not a classroom discipline but a tool for organizing and understanding various concepts and applications. This section covers topics that delve into other disciplines, integrating the mode of thinking and knowledge of the respective discipline with Mathematics. The section hence highlights the cosmic scope of Mathematics, leveraging its amalgamation with other disciplines.

Online Ad Auctions

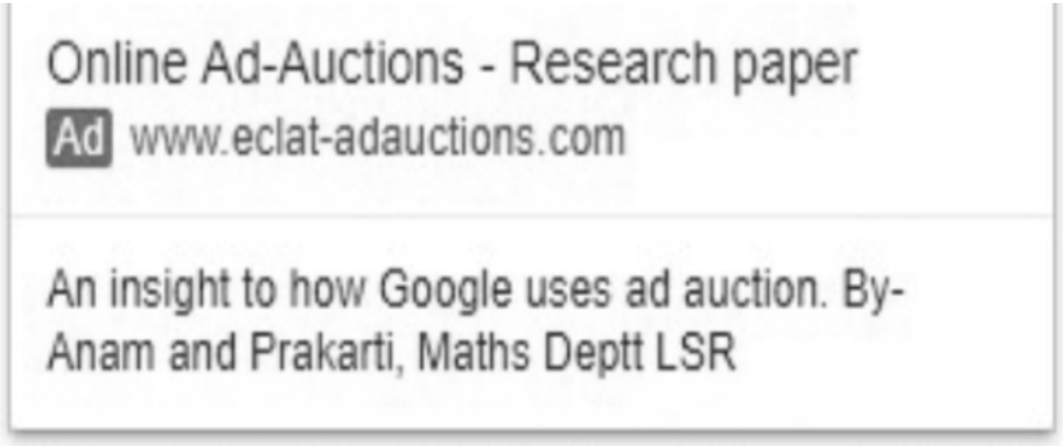
Anam Ali
Prakarti Walia

Abstract

This paper discusses how search engines use auctions to sell their ad spaces. Every time an ad qualifies to feature in a search, it goes through an ad auction. This paper gives a brief introduction and explains what an auction is, its basic rules, the concepts of game theory used in auctions and elementary price theory to analyze advertiser behaviour.

1 Introduction

Search engines use an auction system to sell their ad spaces. All search engines have a similar structure. They use an auction system to rank the ads that appear in the results page and to determine the cost for each ad click. The advertiser wants to show relevant ads so users click on them. Users want to see relevant ads, they don't want to be bothered with irrelevant material, and search engines want to make a good experience for both the user and the advertiser so they use their services in future.



2 How Ad Auction Works

Every time a query is made on a search engine, they run an ad auction and the auction is one for clicks so advertisers only have to pay when they receive

a click.
But in order to provide the best service, the ads that are shown to users should have high ad quality. Quality score has three components and the most significant one, by far, is the **click-through rate (CTR)**,
where,

$$CTR = \frac{no.\ of\ clicks\ ad\ received}{no.\ of\ times\ ad\ is\ shown}$$

The second largest component is the **relevance** - both that of a keyword to the ads as well as to the users' search query. Search engines determine relevance by analyzing the language in context of an ad or query and determining how well it relates to the keyword. They use relevance to make sure that only useful ads are displayed to users and prevents advertisers from paying their way on to a search that's unrelated to their product. The third component is the **landing page quality**. An ad is useful only if the landing page leads to help them find the information they are looking for. Ad rank for each ad is calculated by multiplying the quality score by the bid of the advertiser.

3 Auction Rules

The advertiser pays the search engine an amount determined by the bids of the other competing advertisers. The expected revenue received by the search engine is the price per click times the expected no. of clicks.
Let's start with some notations. Let there be $a = 1, \ldots, A$ index advertisers and $s = 1, \ldots, S$ index slots. Let (v_a, b_a, p_a) be the value, bid and price per click of the advertiser a for a particular keyword. Expected CTR of advertiser a in slot s (z_{as}) is the product of quality-effect (e_a) and the position-effect (x_s). So, $z_{as} = e_a x_s$

Second Price Auction
Rules of the Generalised Second Price auction used by major search engines are:

1. Each advertiser a chooses a bid b_a.
2. The advertisers are ordered by bid times predicted CTR $(b_a e_a)$.
3. The price that advertiser a pays for a click is the minimum required to retain its position.

4. If there are fewer bids than slots, the last bidder pays a reserve price r.

Consider a specific auction where every advertiser has a slot position. Assuming that there are no ties, the rules of auction implies $b_1e_1 > b_2e_2 > \ldots > b_me_m$, $m \leq the\, no.\, of\, possible\, slots$.

The price paid by the advertiser in slot s is the minimum necessary to retain its position so $p_se_s = b_{s+1}e_{s+1}$ which implies $p_s = \frac{b_{s+1}e_{s+1}}{e_s}$
For $s = 1$, $p_1 = \frac{b_2e_2}{e_1}$, in other words, price advertiser in slot 1 has to pay an amount equal to ad rank of advertiser below him divided by the quality score of advertiser in slot 1.
The price paid per click by the last advertiser is the reserve price if $m < S$ or determined by the bid of the first omitted advertiser $m = S$.

Example:
Assume there are 4 advertisers and 5 positions, the bids and ad quality is illustrated in the table given below:

Advertiser	Bid	Ad Quality	Ad Rank	Position	Price-per-click
a	4	1	4	4	min price
b	3	3	9	2	2.66
c	2	6	12	1	1.5
d	1	8	8	3	0.5

4 Nash Equilibrium

Auctions are studied through an application of Game Theory. *Nash Equilibrium* (NE) is a stable state where no player in the game can do any better by adopting another strategy. Make the following assumptions:

- no advertiser wants to exchange places with the other advertiser below or above him;
- all advertiser play their static best response.

Advertisers are interested in maximizing their surplus, which equals the value of clicks they receive minus the cost of those clicks. Let the ad quality

for all advertisers be the same, i.e.,

$$\mathbf{e_a} \equiv \mathbf{1}$$

In equilibrium, the advertiser in slot $s+1$ doesnt want to move up to slot s, so

$$(v_{s+1} - p_{s+1})x_{s+1} \geq (v_{s+1} - p_s)x_s$$

which implies,

$$p_s x_s \geq p_{s+1}x_{s+1} + v_{s+1}(x_s - x_{s+1}). \tag{1}$$

It is advertiser $(s+1)$'s value for those clicks that is relevant since that is the bid that the advertiser in slot s must beat. Denoting the reserve price paid by the last advertiser by p_m and solving inequality (1) repeatedly, we get,

$$\begin{aligned}
p_1 x_1 &\geq v_2(x_1 - x_2) + v_3(x_2 - x_3) + v_4(x_3 - x_4) + \cdots + p_m x_m \\
p_2 x_2 &\geq \quad + v_3(x_2 - x_3) + v_4(x_3 - x_4) + \cdots + p_m x_m \\
p_3 x_3 &\geq \quad\quad + v_4(x_3 - x_4) + \cdots + p_m x_m
\end{aligned}$$

The upper bound on total revenue:

$$\sum_s p_s x_s \geq v_2(x_1 - x_2) + 2v_3(x_2 - x_3) + \cdots + (m-1)p_m x_m.$$

The lower bound on total revenue:

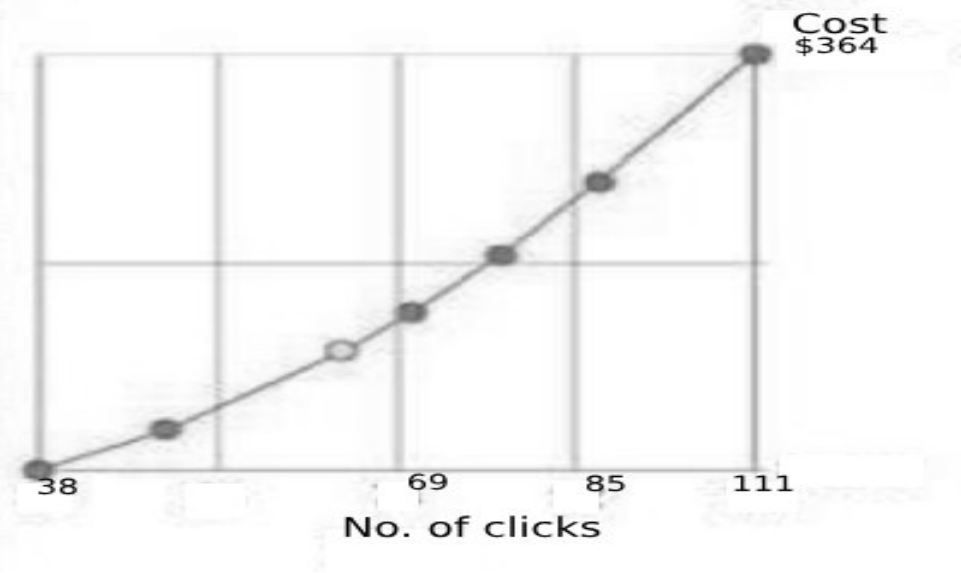
$$\sum_s p_s x_s \leq v_1(x_1 - x_2) + 2v_2(x_2 - x_3) + \cdots + (m-1)p_m x_m.$$

This calculation is used to determine a range of bids that satisfy the inequalities. The extreme points can be used to find the maximum and minimum equilibrium revenue preferred by advertisers and search engines respectively.

5 Elementary Price Theory

The theory of price uses the concept of supply and demand to determine the appropriate price point for a good or service. This theory posits that the point at which the benefit gained from those who demand the entity meets the seller's marginal costs, is the most optimal market price for that good.
In case of ad auctions, the advertisers need to find an **optimal behavior** in order to **maximize their surplus**. They look for the number of clicks that

would maximize their profit.
We know that the price-per-click (PPC) as well as the number of clicks increases on increasing the bid. So, the supply curve is upward sloping. The curve below shows how the cost changes with respect to the number of clicks:



The elasticity of the supply is dependent on the keywords. So, it is important that the advertisers bid optimally. To maximize their profit, the advertisers need to do marginal calculations. When the position of the advertisers changes, the PPC for all the clicks changes and not just the PPC for the extra clicks, hence, it is preferable to take the **marginal cost-per-click (MC)** which is the increase in total cost (c_a) divided by the increase in clicks (x_a). i.e.,

$$\mathbf{MC} = \frac{c_a - \hat{c}_a}{x_a - \hat{x}_a}$$

MC gives a much clearer representation of the cost of moving positions compared to CPC, which is an average measure.

The following table gives the MC for various advertisers:

Position	Bid	Clicks	Cost	CPC	Click Value	MC	Revenue	Profit
1	7.01	111	364	3.3	5	6.1	555	191
2	3.82	85	209	2.8	5	5.9	425	216
3	3.00	69	135	2.0	5	4.62	345	210
4	1.47	38	43.9	1.2	5	2.93	190	146

We see that the CPC in position 3 is \$2 whereas the MC is \$4.62 when it moves to position 2 as the change in the number of clicks is 16 $(85 - 69)$ and rise in cost is \$74 $(209 - 135)$. This MC can be used to find out the optimal bid.
The click value for all the advertisers is \$5. Thus, taking the MC into account, we see that the maximum profit is of \$216 in position 2 in comparison to a profit of \$191 in position 1. Hence, it is clearly **not optimal to be in first position.**

6 Conclusion

There is a growing literature concerning the design of online ad auctions. These auctions have an interesting, yet simple, theoretical structure as well as significant practical importance.

References

[1] Demange, G. and Gale, D., *The Strategy Structure of Two-sided Matching Markets*, Econometrica, July 1985, 53(4), pp. 873-888.

[2] Hansen, J. *The Economics of Search Engines - Search, Ad Auctions and Game Theory*, Copenhagen Business School, Applied Economics and Finance, Summer 2009.

[3] Edelman, B., Ostrovsky, M. and Schwartz, M. *Internet Advertising and the Generalized Second Price Auction*, American Economic Review, March 2007, 97(1), pp. 242-259.

[4] Varian, H. R., *Position Auctions*, International Journal of Industrial Organisation, University of California, December 2007, 25(7), pp. 1163-1178.

[5] *AdSense Support.*(2019).

ANAM ALI, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
aanamali18@gmail.com
PRAKARTI WALIA, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
prakartiwalia@gmail.com

Mathematics in Shoelaces

Anvita Jain
Shradha Rajpal

Abstract

Have you ever wondered if the tying of shoelaces involved mathematics? Have you ever heard of an n-lacing of a shoe? This paper will explore how combinatorics and elementary calculus can be used to answer these questions. It includes an introduction to a mathematical shoe, families of lacings and techniques of best lacing.

1 Introduction

Many are familiar with elementary calculus and combinatorics but only few have thought of its application in an activity as trivial as tying a shoe lace. How many of us have ever asked ourselves whether there are better ways to tie a shoelace than we commonly use? Have a look at one of your shoes. If the laces are tied, it must be either criss cross or straight lacing. There are many more ways of doing a task which are generally regarded as trivial daily routine habits. In fact, some unusual lacings are star lacing, serpent lacing, bowtie lacing, angel lacing etc.

2 Model of a Mathematical Shoe

A mathematical shoe consists of $2n$ eyelets, which are the points of intersection of two vertical lines and n equally spaced horizontal lines in the plane. In our model, we consider that the eyelets are perfectly aligned and the shoe exists only in one plane. The distance between the two vertical lines is fixed to be one and the distance between the two adjacent horizontal lines is referred to as h. The set of all eyelets contained in one of the vertical lines is known as a **column of eyelets** and the set of two eyelets contained in a horizontal line is known as a **row of eyelets**.

An **n-lacing** of our shoe is a closed path in the plane that consists of $2n$ line segments where the end points are $2n$ eyelets. We also assume that given an eyelet E, at least one of the two segments ending in it is not contained in the same column E. All the lacings satisfy the condition ensuring that every eyelet contributes towards pulling the two sides of the shoe together, so that

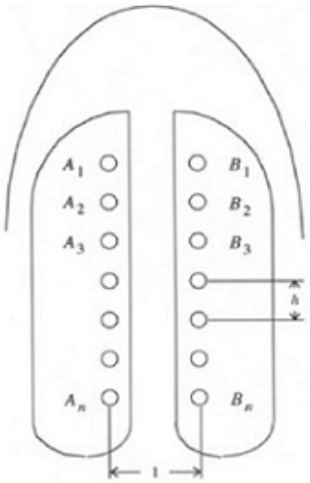


Figure 1: A Mathematical Shoe

there are no gaps between the lacings. If both the endpoints of a lace are contained in either of the columns then the segment is a *vertical* segment. A non vertical segment is known as a *diagonal* segment. A diagonal segment is *horizontal* if its endpoints have the same index. The *length* of a lacing is the sum of the lengths of its segments.

3 Types of Lacing

1. **Dense Lacing**: A lacing which does not contain any verticals is known as a dense lacing. In a dense lacing, the shoelace zizags back and forth between two columns of eyelets. For example: criss cross, star lacing.

2. **Straight Lacing**: A lacing which contains all the horizontals.

3. **Super Straight Lacing**: If all the non-horizontal segments are vertical in a straight lacing then it is said to be super straight lacing.

4. **One Column Lacings**: If two columns of eyelets are pulled together into one column, then it is referred to as one column lacing.

4 Properties of n-lacings:

1. For an n-lacing, say l, the following statements are equivalent: (1) The number of verticals in both the columns are equal; (2) The maximum

possible number of verticals in the lacing L are n, if n is even and it is $n-1$, if it is odd.
Proof: Let a denote the number of verticals in one column, say A; b denote the number of verticals in column B and d denote the diagonals. We consider that any eyelet can be the endpoint of at most one vertical and at least one diagonal. Therefore, column A consists of $2a$ eyelets in which exactly one diagonal ends and $n-2a$ diagonals in which only two diagonals end. Therefore, the number of diagonals are:

$$2a + 2(n-2a) = d = 2b + 2(n-2b)$$

Thus, $a = b$.

2. The existence of a superstraight lacing: If the value of n is even then a superstraight n-lacing exists.
Proof: Firstly to show that a superstraight lacing exists, we know that a serpent n lacings can be realized for any n which is even. Thus, a superstraight lacing exists for even values of n. A *superstraight n lacing* is a one which consists of $2n$ segments, n of which are horizontals and n are vertical. Therefore, by the above property, we conclude that n is even.

5 The Shortest Lacing

By the use of symmetries of the arrangement of eyelets, it is possible to find the shortest n-lacing. We claim that the bowtie lacing is the shortest way to tie a shoe-lace.

A bowtie lacing is made up by three different types of building blocks called an **end**, a **cross** and a **gap**. An **end** is defined as a horizontal segment that connects either the top pair or bottom pair of eyelets. For example, a crisscross n-lacing can be divided into two ends and $n-1$ crosses. The maximum number of gaps in an n lacing is $n/2$ if n is even, otherwise it is $(n-1)/2$. A *bowtie n-lacing* is an n-lacing that can be categorised into crosses, the maximum number of gaps and two ends. We can easily prove that if n is even, then there is exactly one bowtie n-lacing and there are exactly $(n+1)/2$ different bowtie n-lacings if n is odd. For example, the figure 6 below shows unique bowtie 4-lacing, whereas figures 7 and 8 represent two bowtie 5-lacings. Figure 8 which is a third bowtie 5-lacing ($3 = (5+1)/2$), is the horizontal mirror image of the bowtie 5-lacing.

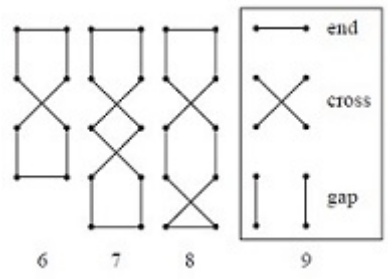


Figure 2: Types of Bowtie Lacings

6 The Strongest Lacing

When we try to pull on the ends of a shoelace, it acts like a pulley. In case of an ideal mathematical shoe, the tension along the shoe lace is a positive constant T. The component of this tension in the horizontal direction gives rise to a total tension T_h. This total tension T_h is sum of all horizontal components of T along the different segments of lacing. Then, **we claim that a strongest n-lacing is the one that maximises T_h.**

In case of a vertical segment, the horizontal component of T is zero and in case of a diagonal segment of length l, the horizontal component is T/l. Provided an n-lacing that contains verticals, it is always possible to find an n-lacing which is stronger by replacing two verticals contained in different columns by any suitable diagonals. This leads to the fact that any strongest n-lacing is dense. If this is the case and if $I_1, I_2, \ldots, I_{2n}$ represent the lengths of the different segments in a dense n-lacing, then what we have to maximise is the sum:

$$\sum_{1}^{2n} \frac{1}{l_i}$$

The solution of this problem consist of short segments as a small value of l_i gives rise to strongest lacing. The strongest n-lacing depends on the distance h between two adjacent rows of eyelets. For given value of $n > 2$ there exists a positive value h_n such that the strongest n-lacing are: (1) The crisscross n-lacing when $h < h_n$. (2) The crisscross n-lacing and the straight n-lacings when $h = h_n$. (3) The straight n-lacings when $h > h_n$.

The table below approximates the values of h_n for some values of n.

n	3	4	5	6	7	8	9	10
h_n	.9029	.7412	.6450	.3764	.3309	.4931	.4625	.4372

An interesting thing to note here is that for many real shoes having n pair of eyelets, the ratio of the distance between adjacent rows of eyelets to the distance between the columns is extremely close to h_n. This clearly implies that irrespective of our preference to lace straight or crisscross, the total horizontal tension maximises when we pull on the 2 ends of one of our shoelaces and thus **both crisscross and straight lacing are the strongest lacings**.

7 Conclusion

Our article is an attempt to make people think about the logic and the technique used in performing a trivial action such as tying a shoe lace. We could never have thought that our undone shoe laces can be so interesting and fun if we think about them. By boiling down a situation to its essentials-labelling, measuring and classifying, we set the environment for asking questions (like the strongest or shortest lace) whose answers surprise us.

References

[1] Polster, B., *The Shoelace Book: A Mathematical Guide to the Best (and Worst) Ways to Lace Your Shoes*, American Mathematical Society, 2006.

[2] NATURE, Volume 420, 5 December 2002

[3] https://www.ams.org/notices/200611/rev-adams.pdf

[4] https://plus.maths.org/content/shoelace-book

SHRADHA RAJPAL, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
shradharajpal99@gmail.com

ANVITA JAIN, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI.
jain.anvi90@gmail.com

Elliptic Curve Cryptography

Jaya Sharma

Abstract

This paper briefly discusses Elliptic Curve Crytography: an emerging form of public key cryptography, and the mathematics used in it.

1 What Is Public Key Cryptography?

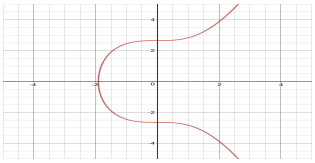
Public-key cryptography or asymmetric cryptography, is any cryptographic system that uses 2 keys:

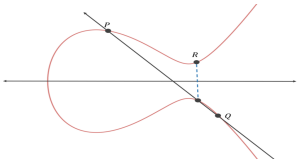
1. **Public Key**: A numerical value that any sender can use to encrypt data that is made available widely through a publicly accessible repository or directory.

2. **Private Key**: Whatever is encrypted with a public key may only be decrypted by its **corresponding private key** (known only to the receiver) and vice versa. It should be *computationally infeasible* to derive the private key corresponding to a given public key.

In the real world, the receiver of a message needs assurance that the message belongs to the sender and he should not be able to repudiate the origination of that message. This requirement is very crucial in business applications, since the likelihood of a dispute over exchanged data is very high. We use a **digital signature** for this purpose.
A *digital signature is a technique that binds a person/entity to the digital data.* This binding can be independently verified by receiver as well as any third party. Digital signature is a cryptographic value that is calculated from the data and a secret key known only by the signer.

One way to do Public Key Cryptography is using elliptic curves. This approach is called **Elliptic Curve Cryptography** (referred to as ECC henceforth).





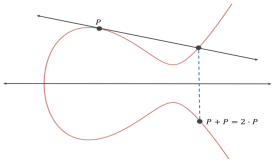


Figure 1: $y^2 = x^3 + 7$ Figure 2: $P + Q = R$ Figure 3: $P + P = 2P$

2 Elliptic Curves

Let $a, b \in \mathbb{R}$ be constants such that $4a^3 + 27b^2 \neq 0$. A ***non-singular*** elliptic curve consists of all the points that satisfy the equation:

$$y^2 = x^3 + ax + b$$

The condition $4a^3 + 27b^2 \neq 0$ is necessary and sufficient to ensure that the equation $y^2 = x^3 + ax + b$ has *three distinct roots*, which may be real or complex numbers. If $y^2 = x^3 + ax + b = 0$, then the corresponding elliptic curve is called a ***singular elliptic curve***[1]. All elliptic curves are symmetric about the x-axis. Figure 1 is the elliptic curve $y^2 = x^3 + 7$.

2.1 Point Addition

Elliptic Curves have an important property: **adding any 2 points on the curve yields a third point**. Suppose we want to add P and Q, as shown in Figure-2. We find a line going through the points and determine where it intersects the curve at a third point. Then we reflect that point about the x-axis to get the resultant point R.

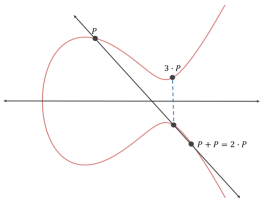


Figure 4: $P + 2P = 3P$

To do ECC, we specify the base point P on the curve and add that point to itself (taking special case of the tangent line passing through P). We get $2P$ (Figure-3), $3P$ (Figure-4) and **continue the process to get very large values.**

2.2 Speeding Up Point Addition

Let P be a point on the curve. Then, the following propoerty holds for point addition:

$$n.P + r.P = (n + r)P$$

where n and r are integers. Now if we want to compute x.P, where x is a random **256-bit** integer and can range from 0 to $1.1579209e + 77$, we can use this property to reduce the number of steps. First we compute the following finite sequence: $(2^0.P, 2^1.P, 2^2.P, \ldots, 2^{255}.P)$. It requires **255** point addition operations in total[1].
The binary expansion of x will *at most* contain 256 terms from the sequence (2^0 up to 2^{255}), which require 255 point operations to compute. Adding all these points to get x.P requires another 255 point additions. Hence, computing x.P will take maximum **255+255=510** point addition operations.

3 How Elliptic Curve Cryptography Is Done

3.1 The Public and Private Keys

Let X = x.P, where x is a random 256-bit integer, and P a random point on the curve. Now, if given X and P, can you find x?
It turns out that it is computationally infeasible. There is no known algorithm for determining x, which lies somewhere between 0 and $2^{256} - 1$. Hence it will take you about **2^{128} point operations** to compute x, no matter what your approach or starting point.
Even if you had the fastest supercomputer running since the beginning of the universe, you would still not have completed 2^{218} point addition operations (required on average to compute x)[2][2].Clearly, ***it is computationally infeasible to derive* x *if we know*** X. Hence we get our:

- **Private Key: x**
- **Public Key: X**

[1] as $2^n.P + 2^n.P = 2^{n+1}.P$, we get the next point by adding the current point to itself
[2] Even if you start in the middle

Now, x.P might result in a point whose x and y coordinates can't be stored in a 512-bit public key. So we define our elliptic curve over a finite field:

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

where p is a prime number.

3.2 Hash Functions

Hashing means *taking an input string of any length and giving out an output of a fixed length.* Consider SHA-256 (Secure Hashing Algorithm 256). Whatever the size of your input is, the output will always have a fixed **256-bits length**. So you can just remember the hash and keep track of any size input data.

3.2.1 Properties Of Cryptographic Hash Functions

1. **Deterministic:** This means that no matter how many times you parse a particular input through a hash function you will *always get the same result.*

2. **Quick Computation:** The hash function should be capable of returning the hash of an input quickly.

3. **Pre-Image Resistance:** The ideal cryptographic hash function has the property that it is **infeasible** to generate a message/data from its hash value *except by trying all possible messages* (using the **brute-force method**). Suppose you are dealing with a 128-bit hash. By using the brute-force method, on an average you will find the message after $2^{127} = 1.7 \times 10^{38}$ times. Hence it takes so long that it doesn't matter.

4. **The Avalanche Effect:** Even if you make a small change in your input, the changes that will be reflected in the hash will be *huge.*

5. **Collision Resistant:** Each input will have its own *unique hash.*

3.3 Proving we Know the Private Key x

We modify the property in section 2.2 to:

$$\text{hash}(m, r.P).n.P + r.P = (\text{hash}(m, r.P).n + r).P \quad (1)$$

We assign randomly chosen values to m and r and these are kept **private.**
Next we set $n.P$ =X.
n =x, since X=x.P. Let $r.P = R$ Hence (1) becomes:

$$\text{hash}(m, R).\text{X} + R = (\text{hash}(m, R).\text{x} + r).P \tag{2}$$

Let $s = (\text{hash}(m, R).\text{x} + r)$. So now we have:

$$\text{hash}(m, R).\text{X} + R = s.P \tag{3}$$

We claim:
We know x (private key) corresponding to X (the public key)=x.P *if and only if* we can provide working values of m, R and s that satisfy the equation (3).

For this claim to be true, two conditions need to be met:

1. **We know x ⇒ We can provide working values of m, R and s.**
   As m, r have been pre-assigned values, we can compute $R = r.P$ and s $= (\text{hash}(m, R).\text{x} + r)$ if we know x, and hence (3) is satisfied.

2. **We do NOT know x ⇒ We CAN NOT provide working values of m, R and s.**
   If we don't know x, we would have to solve the equation $\text{hash}(m, R).\text{X} + R = s.P$. Basically we have to find an input for a hash that has a specific value, which is **not possible** due to the **pre-image resistance** property of hash functions.

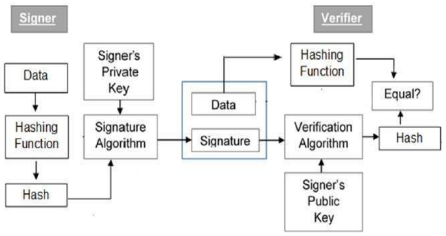
Therefore the only way to provide working values for m, R and s is by computing them using x. Hence we can prove that we know x by providing these correct values.

3.4 Does this give out any information about x?

m, R have nothing to do with x. Any third party could have the value of s, which could be any big integer. Hence, someone trying to get x from s would need to solve the equation $\text{x} = (s - r)/\text{hash}(m, R)$. But since they don't know r and can't compute it from $R = r.P$ (such a computation is infeasible as proven in section 3.1).
Hence **no potentially useful information** about x can be obtained from m, R and s.

3.5 Digital Signature

m, R and s can be used to prove that one x corresponding to X. So we let m be a message and R and s be **digital signatures for the message**. The verification process would only be successful if the specific message m plugged in (3) satisfies it. If a different value of m is plugged, then LHS of (3) would fail to be equal to RHS as s **was calculated using a different message**. Hence R and s form a digital signature for message m, and using them we prove that we know the private key corresponding to x.



3.6 An Illustration

ECC is used by the cryptocurrency Bitcoin. If you want to obtain a Bitcoin address, you generate a random 256-bit integer x, your private key. Then you compute the public key X = x.P using the parameters for the curve chosen. If you **hash your public key, you will obtain your address.**
When you want to send bitcoin from your address to another address, you create a transaction. You set m to the unsigned part of that transaction, and compute R and s from that m. Then you attach R and s to the transaction. After you broadcast your transaction, any node will be able to verify that m (the unsigned part of the transaction), R, and s satisfy hash(m,R).X + $R = s.P$. This, of course assumes that you also include X in your transaction, since your public key cannot be determined from your address.

4 Other Applications Of Elliptic Curves

Elliptic curves are also used in pseudo-random generators and other tasks. They are also used in several integer factorization algorithms that have applications in cryptography.

References

[1] Stinson, D. R., *Cryptography:Theory and Practice, Third Edition*, CRC, 2005.

[2] https://hackernoon.com/what-is-the-math-behind-elliptic-curve-cryptography-f61b25253da3

[3] https://www.tutorialspoint.com/cryptography/cryptography_digital_signatures.htm

[4] https://blockgeeks.com/guides/cryptographic-hash-functions

[5] https://en.wikipedia.org/wiki/Elliptic-curve_cryptography

[6] https://en.wikipedia.org/wiki/Cryptographic_hash_function

JAYA SHARMA, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN, NEW DELHI
forjayasharma@gmail.com