ÉCLAT
MATHEMATICS JOURNAL

Lady Shri Ram College
For Women
Volume-VIII
2016-2017

ADA LOVELACE

This journal is dedicated to Augusta Ada King-Noel, Countess of Lovelace, one of the most prominent female mathematicians who quickly gained fame as the "Enchantress of numbers".

In addition to being the founder of scientific computing, Lovelace aimed at creating a mathematical model that she called a calculus of the nervous system. She also described an algorithm representing how the proposed mechanical general-purpose computer (the Analytical Engine) could be programmed to compute Bernoulli numbers and to solve complex problems. This is considered the first algorithm developed for implementation on a computer. In 1953, the Analytical Engine was recognized as an early model for a computer.

PREFACE

This edition of Éclat, 2016 - 2017 has been a remarkable journey taking us through twelve intense mathematics papers with incessant rounds of editing to bring out the final version. The journal, which was only an idea a few years back has become an integral part of the Mathematics department of the college. We hope that the readers will gain a deeper knowledge of mathematics. It has been a great challenge as well as a pleasure to put together research and insights into black and white.

This edition of clat contains four categories of mathematics namely, History of mathematics, Rigour in mathematics, Inter disciplinary aspects of mathematics and Extension of course content. We have tried our best to make the journal inclusive so that readers from all fields can find a suitable content of their interest. This journal along with increasing your mathematical understanding will also stimulate your thought process to come out with new ideas.

The publication of this journal would have been impossible without the enthusiasm of the department of Mathematics of our college. We sincerely thank the faculty of the Department of Mathematics, Lady Shri Ram College for Women, for guiding and supporting us throughout the year. We are open to any suggestions, corrections and submission from our readers.

We dedicate this volume of the journal to Augusta Ada King-Noel, Countess of Lovelace, one of the most prominent female mathematicians.

The Editorial Team

Anika Jain, B. Sc. (Hons.) Mathematics, 3rd Year
Esha Saha, B. Sc. (Hons.) Mathematics, 3rd Year
Aakshi Malasi, B. Sc. (Hons.) Mathematics, 2nd Year
Nikita Sobti, 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.

Crop Circles

Bhavna Phogaat and Mansi Yadav

Abstract

This paper revisits the controversial case of crop circles. A crop circle is a pattern created by flattening of a crop. Several theorists and researchers have attributed different theories for the effect. This paper highlights some of the theories and also paves way towards the geometrical reasoning of crop circles.

1   Introduction

Crop circles are patterns that appear in fields. The pattern is created when certain areas of the crops are tamped down, but others are left intact. The edge is so clean that it looks like it was created with a machine. Even though the stalks are bent, they are not damaged. Usually, the crop continues to grow as normal.

Sometimes, the patterns are simple circles: the most basic and common being the simple circle. They may also come in sets of two (doublets), three (triplets) or four (quadruplets). Circles also may be enclosed in a thin outer ring. In other instances, they are elaborate designs consisting of several interconnecting, geometric shapes. The stalks inside a crop circle are typically bent into what is known as a swirl pattern, and the circles may spin clockwise or counterclockwise. In patterns with several circles, one circle may spin clockwise and another counterclockwise. Even a single circle may contain two "layers" of stalks, each spinning in a different direction.

2   Theories on Crop Circles

The first crop circle was described in the 1678 news pamphlet as "Strange News Out of Hartfordshire", now more commonly known as "The Mowing Devil". It told the story of a farmer who made a deal with the devil to mow his field. While the woodcut illustration appears to show a demonic creature cutting the field in a circular pattern, the story indicates that the whole field was mysteriously cut and not just in a small section as is in the case of modern crop circles.

People who attempt to study these circles have coined a name for themselves: Cereologists. The word comes from the name of the Roman goddess of vegetation, Ceres. There are several theories held by cereologists who think crop circles are the result of some not well understood physical phenomena.

- **UFOs**

One theory is that crop circles are created by UFOs. This is because of how the first crop circle phenomenon was reported during modern times. In 1972, two witnesses, Arthur
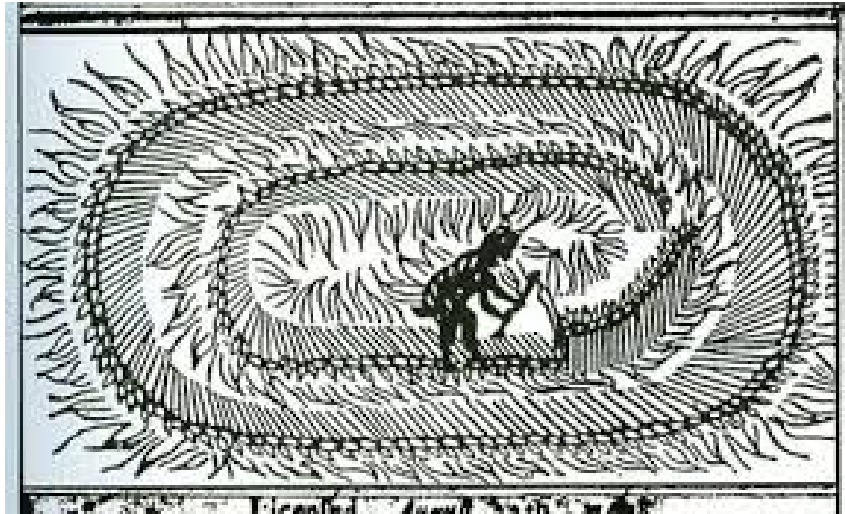
Figure 1: The Mowing Devil

Shuttlewood and Bryce Bond, sat on the slope of Star Hill near Warminster, England, hoping to catch a glimpse of the strange unidentified flying craft that had made that part of England a UFO Mecca for almost a decade. But what they witnessed on that moonlit night was something even more extraordinary:

"Suddenly, I heard a noise. It seemed as if something pushed down the wheat. That night the air was completely still. I looked around. The moon had just appeared, shining brightly. In front of my eyes I could see a great imprint taking shape. The wheat was forced down in a clockwise direction."

Some of the early, simple crop circles certainly do suggest fields that might have been flattened by the weight of a grounded flying saucer. However as the circles became more complex in shape, proponents of the UFO theory have had to modify their ideas suggesting that the marks left were due to the strange effect of the craft's drive force on the plants. Some even argued that the shapes are messages purposefully left by the alien spaceship's crew.

- **Earth Energy**

In the early 1990s, American biophysicist Dr. William Levengood discovered that crops in circles were damaged much in the same way as plants heated in a microwave oven. He proposed the idea that the crops were being rapidly heated from the inside by some kind of microwave energy. It was believed by some researchers that the earth created its own energy, which forms the circles. One possible form of earth energy is electromagnetic radiation. Other researchers say that the energy comes from under the ground or from the soil. Either the energy is natural, such as a fungus that attacks the crops and causes their stems to bend over, or it is a by-product of something man-made, such as bombs that exploded during World War II.

- **Wind**

One scientific theory says that crop circles are the result of an unusual weather effect. Dr. Terence Meaden of the Tornado and Storm Research Organization (TORRO) in Wiltshire, England, says that small currents of swirling winds called vortices (similar to "dust devils") create crop circles which are charged with energy. He called this the Plasma Vortex Theory which can be defined as a spinning mass of air which has accumulated a significant fraction of electrically charged matter. The spinning columns force a burst of air down to the ground, which flattens the crops. Vortices are common in hilly areas such as parts of southern England. When dust particles get caught up in the spinning, charged air, the effect is similar to that of a ball lightning, but larger and longer lasting, which may explain the UFO-like glowing lights many witnesses have seen near crop circles.

- **Man made**

In 1991, two hoaxers, Bower and Chorley, approached a British newspaper acknowledging their actions and announced that they had made hundreds of crop circles since 1978. They proceeded to demonstrate to the world how they had done it. However after several hours, the result was an embarrassing mess. A second attempt proved just as pathetic. They had also claimed never to have been active in the Avebury area, but this had in fact been the prime circle-making area since 1988. When asked questions like - Who made those in other parts of the country? How come they had never been caught? They had no answer. How had they created the anomalies, the mathematical precision? The more the questions were asked, the more they backed down their claims. Robert Van der Broeke soon came out as a self-acclaimed paranormal medium. But he too was soon accused of fraud.
Colin Andrews, cereologist and author of the book, *Circular Evidence*, admits that about 80 percent of crop circles are probably man-made, but says that the other 20 percent are probably the work of some "higher force." An interesting fact was that the phenomenon showed intelligent reaction to the thoughts and theories of the researchers. When they said that formations were caused by freak winds on sides of hills, the crop circles moved to flat, open areas. When men armed with planks of wood were suggested as the culprits, circles developed in oil seed rape (canola), one of the most brittle plants. If balloons were to blame, they appeared under high voltage wires.

3 Mathematics of Crop Circles

In 1980 when the phenomenon caught everyone's interest, suddenly there were two circles in a field, lying symmetrically side-by-side. In 1981 signs of intelligence were manifested when a 52-foot circle was flanked by two smaller circles exactly half its size, aligned perfectly north-south. In 1983 the first quintuplet design appeared - a large central circle surrounded by four smaller circles within the standing crop, precisely aligned to the four cardinal points. Four of the circles were rotated clockwise, the fifth counter-clockwise.
By 1988, hundreds of crop circles had been documented and researched. And every year they had grown exponentially, always developing and splitting in structure as if suggesting some sort of language, more complex as the years progressed and at the pace in which the research teams picked up on the subtle clues left on the fields. Circles developed a simple ring, then double rings. Then Celtic crosses, where the four 'satellite' circles were connected by a thin band, too narrow even for a small child to thread without disturbing the crop. In 1988, during a BBC interview inside a new formation, the cricket-like warble

that had been heard several times throughout the years was captured on tape. Eventually analysed by NASA's Jet Propulsion Laboratory, it was measured as 100 bpm, at a frequency of 5.2kHz, and mechanical in nature. A bird or insect was also out of the question.

In one bizarre case at Corhampton, three single circles were placed mathematically correct within an equilateral area, making it the first crop formation to exhibit a musical diatonic ratio.

Furthermore, its physical features were extraordinary. Here was a series of circles that had been previously flattened but now the stems were lifting back to the light of the sun in a selective manner, in three separate groups. In the first group, plants were bending on the node nearest to the ground. In the second group plants were lifting up on the node half-way up the stem, and in the third group plants were bending at the node nearest the head. From the air, they had grown into a pattern consisting of seven concentric rings and forty-eight spokes.

The most striking feature of the circles is the frequency with which they occur. In 1990, over 700 crop circles appeared in Britain. Soon after 1990, the circles became more elaborate. More complex crop patterns, called pictograms, emerged. Crops could be made to look like just about anything – smiling faces, flowers or even words. Some of the more sophisticated patterns are based on mathematical equations.

Astronomer and former Boston University professor Gerald S. Hawkins studied several crop circles and found that the positions of the circles, triangles and other shapes were placed based on specific numerical relationships. In one crop circle that had an outer and inner circle, the area of the outer circle was exactly four times that of the inner circle. The specific placement of the shapes indicates that, whoever the circle makers were, they had an intricate knowledge of Euclidean geometry (the geometry of a flat surface introduced by the mathematician Euclid of Alexandria).

4 Dr. Gerald Hawkins' Theorems

Dr. Gerald Hawkins used the principles of Euclidean geometry to prove the four theorems derived from the relationships among the areas depicted in the Crop Circle patterns. The Five Circles Theorems are only the geometrical portion of the surprise that many of the numbers regularly turning up in his analyses were diatonic ratios. Diatonics are mathematical equivalents of musical notes as played, for example, on the white keys of a piano. Each of these theorems requires diatonic ratios, which are based on the ratios of areas and diameters between the circles as given in Figure 2.

- **Theorem 1**

The ratio of the diameter of the triangles circumscribed circle to the diameter of the circles at each corner is 4 : 3.

- **Theorem 2**

For an equilateral triangle, the ratio of the areas of the circumscribed and inscribed circles is 4 : 1. The area of the ring between the circles (the annulus) is 3 times the area of the inscribed circle.

- **Theorem 3**

For a square, the ratio of the areas of the circumscribed and the inscribed circles is $2 : 1$. If a second square is inscribed within the inscribed circle of the first, and so on to the mth square, then the ratio of the areas of the original circumscribed circle and the innermost circle is $2^m : 1$.

- **Theorem 4**

For a regular hexagon, the ratio of the areas of its circumscribed and inscribed circle is $4 : 3$.

- **Theorem 5**

Theorem 5 is a general theorem from which theorem I-IV can be derived and only the triangle, square and hexagon will give a diatonic ratio from the circumscribed and inscribed circles.
In the figure below, the triangle changes shape as the circles expands and contracts to touch the sides. The diagram generates the four crop circle theorems.
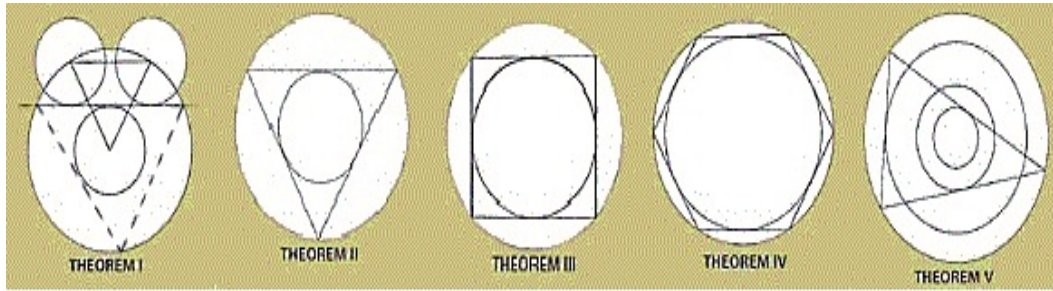

Figure 2: Dr. Gerald Hawkins' Theorems

5 Application of the Theorems

In figure 3, O is the center of circles 1, 2, 3 and equilateral triangle ADE. ABC is also equilateral with height AD. The moon has center D, radius AD. But OB is also a height of triangle ABC, therefore circle 3 with radius OB is the same size as the moon. Circle 2 is tangent to the moon on OD produced, and circle 1 is the exterior circle of the hexagon tangent to circle 2. This construction fits the crop formation to within the limits of measurement, and we can find the areas of the circles exactly. They give diatonic ratios. From 1 to 2 we get a ratio of $4 : 3$, and from 1 to 3 we get closely a ratio of $10 : 3$.
In Figure 4, the Crop Theorem II and then two applications of Theorem III make the area of the outer circle 16 times bigger than the area of the inner disc: $4 * 2 * 2 = 16$, which is musical note C in the 5th octave.
And in Figure 5, the pattern contains crop theorems, but it is embellished with 3 paws, 3 legs and 6 spokes. Because of the fitted square, the ring gives a ratio of $2 : 1$ by Theorem III. The equilateral triangle and central circle depicts Theorem II.
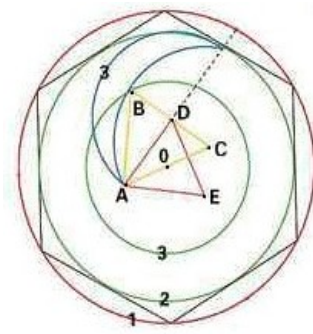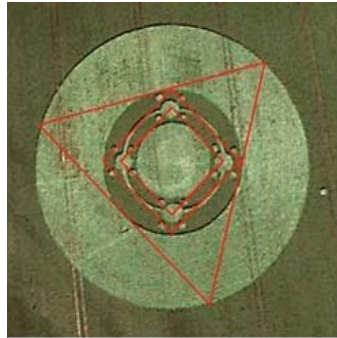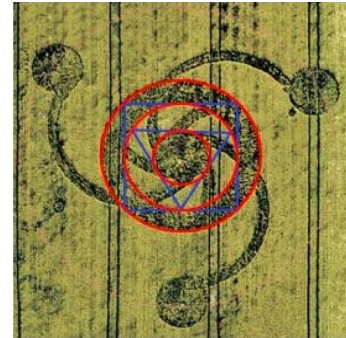
| Figure 3 | Figure 4 | Figure 5 |

6   Conclusion

Crop circles have been described as falling "within the range of the sort of things done in hoaxes". Although obscure, natural causes or alien origins of crop circles are suggested by fringe theorists, there is no scientific evidence for such explanations and human causes are considered consistent for all crop circles. Believers and naysayers each have their own theories, but the truth remains elusive.

References

[1] Journal of Scientific Exploration, Vol.28, No.1,pp.17-33,2014 0892-3310/14

[2] https://www.scribd.com/doc/96150452/Crop-Circle-Theorems

[3] http://www.wikipedia.org

[4] http://www.cropcirclesecrets.org/

[5] http://science.howstuffworks.com/science-vs-myth/unexplained-phenomena/question735.htm

[6] http://www.unmuseum.org/cropcir.htm

BHAVNA PHOGAAT, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
bhav3095@gmail.com

MANSI YADAV, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
mansiyadav199523@gmail.com

Breakthrough in Mathematics

Aakshi Malasi

Abstract

The history of Mathematics is nearly as old as humanity itself. Since antiquity this discipline has been fundamental to advances in engineering, astronomy and philosophy. This paper discusses important discoveries and works of eminent Mathematicians such as Euclid, Euler, Gödel, Fermat and Fourier that have been imperative in shaping the way Mathematics is perceived today.

1   Euclid's Elements

It is considered by many to be one of the most beautiful works in the history of Mathematics. This mathematical and geometric treatise, consisting of 13 books, is attributed to the ancient Greek mathematician Euclid which he developed around 300 BC. It has subsequently led to the logical development of geometry and other branches of Sciences. The books covers Euclidean geometry and the ancient Greek version of elementary number theory. The work also includes an algebraic system that has become known as geometric algebra, which is powerful enough to solve many algebraic problems, including the problem of finding the square root of a number. The 'Elements' have been studied for the last 24 centuries in many languages written originally in Greek and then translated to Arabic, Latin, and many modern languages. The 'Elements' started with 23 definitions, five postulates, and five "common notions," and systematically built the rest of plane and solid geometry upon this foundation. The five Euclid's postulates are:

1. It is possible to draw a straight line from any point to another point.

2. It is possible to produce a finite straight line continuously in a straight line.

3. It is possible to describe a circle with any centre and radius.

4. All right angles are equal to one another.

5. If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the straight lines (if extended indefinitely) meet on the side on which the angles which are less than two right angles lie.

The influence of Euclid's elements is such that it is still considered a masterpiece in the application of logic to Mathematics. Scientists Nicolaus Copernicus, Johannes Kepler, Galileo Galilei, and Sir Isaac Newton were all influenced by the Elements, and applied their knowledge of it to their work. Mathematicians and philosophers, such as Thomas Hobbes, Baruch Spinoza, Alfred North Whitehead, and Bertrand Russell, have attempted to create their own foundational "Elements" for their respective disciplines, by adopting the axiomatised deductive structures that Euclid's work introduced.

2 Euler's Identity

Given by the Swiss mathematician Leonhard Euler, it is one of the most remarkable discoveries in Mathematics. Euler was an 18th-century born mathematician who developed many concepts that are integral to modern mathematics. Having written 886 papers and many books, Euler is one of the most prolific mathematicians of all times. Our knowledge of trigonometric and exponential functions tells us that

$$e^{ix} = \cos x + i \sin x$$

is the equation that connects real and complex numbers. For $x = 1$ the equation reduces to $e^{i\pi} = -1$ rearranging the terms gives us $e^{i\pi} + 1 = 0$ which is Euler's identity.

This is an incredible link among the most important numbers in mathematics. Back in 1988, a Mathematical Intelligence poll voted Euler's identity as the most beautiful feat of all of mathematics. This result brings together the dream team of numbers, five of the most important mathematical constants. The number 0 was invented around the 7th century by the Indian mathematician Brahmagupta and is also the additive identity. The number 1 was the first number to be invented making it the father of mathematics. 1 has its own importance in Mathematics as the multiplicative identity. π which is the ratio of circumference of a circle to its diameter is often considered to be the face of geometry. The next number in Euler's identity is Euler's number itself, e. Deceiving in its simplicity Euler's number is the base of the natural algorithm and is approximately equal to 2.71828. It is the limit of $(1 + 1/n)^n$ as n approaches infinity, an expression that arises in the study of compound interest. Euler started to use the letter e in 1727 or 1728, in an unpublished paper on explosive force in cannons, and the first appearance of e in a publication was Euler's Mechanica(1736). The reason why Euler introduced this number was to describe the natural phenomenon of 100% growth. The number i, defined as the square root of -1 is the most fundamental of all the imaginary numbers. All around the world, Euler's identity is considered to be stunning in nature. The great physicist Richard Feynman called the identity "one of the most remarkable, almost astounding, formulas in all of mathematics" and mathematics professor Keith Devlin said like a painting that brings out the beauty of the human form that is far more than just skin deep, Euler's Equation reaches down into the very depths of existence."

3 Gödel's Incompleteness Theorem

It is next in the list of important mathematical breakthroughs. At the beginning of the 20th century, mathematician David Hilbert presented a list of 23 of the most important unsolved problems in mathematics. Second on his list was to prove that the axioms of arithmetic are consistent i.e. free from internal contradictions. It may appear obviously true, yet in a groundbreaking 1931 paper that unified logic, mathematics and philosophy, Gödel is widely believed to have proven Hilbert's problem in the negative. These theorems proven by Kurt Gödel in 1931, are important both in mathematical logic and in the philosophy of Mathematics. **Gödel's First Incompleteness Theorem** first appeared as " theorem VI" in Gödel's 1931 paper. The formal theorem which is written in highly technical language maybe paraphrased in English as

'Any effectivley generated theory capable of expressing elementary arithmetic cannot be both

Figure 1: Leonhard Euler

consistent and complete. In particular, for any consistent, effectively generated formal theory that proves certain basic arithmetic truths, there is an arithmetical statement that is true, but not provable in theory.'

Gödel's **second incompleteness theorem** can be paraphrased as
'For any formal effectively generated theory T including basic arithmetical truths, and also certain truths about formal provability, if T includes a statement of its own consistency, then T is inconsistent.'

An imperfect but useful analogy is found within the liar paradox. In this paradox we begin with a machine that can be fed any statement and outputs whether that statement is true or false with unfailing accuracy. Now consider inputting the statement "this statement is false". The machine could output neither "true" nor "false" without producing a contradiction - the equivalent of an undecidable proposition. Gödel's two incompleteness theorems are among the most important results in modern logic, and have deep implications on various issues. They concern the limits of provability in formal axiomatic theories.

These results have had a deep impact on the philosophy of mathematics and logic. There have been attempts to apply the results also in other areas of philosophy such as philosophy of mind.

4 Fast Fourier Transform(FFT)

The importance of the discovery of the Fast Fourier Transform (FFT) on the modern computing age descends from the purpose of the Discrete Fourier Transform (DFT). The development of fast algorithms for DFT can be traced to Gauss's unpublished work in 1805 when he needed it to interpolate the orbit of asteroids Pallas and Juno from sample observations. The DFT is a transform first introduced by Fourier in the early 19th century that has the ability to break down signals (sound waves or wireless signals) into their component frequencies. Once a signal is transformed into its frequencies, often it can be manipulated in a much easier fashion. For example, a sound decomposed into its frequencies can have its high-frequency noises (which should be unnoticeable) filtered out thereby decreasing the noise and size of the signal without harming the quality. This is just one of a large amount

of DFT applications which range from data and image compression (by being able to discard the least noticeable frequencies) to Magnetic Resonance Imaging and many fields in between. All this was good from a theoretical standpoint, but the DFT and its inverse suffered from requiring a largely impractical amount of time to compute. If it were not for the invention of a FFT by J.W. Tukey and John Cooley in the 1960s, the DFT might have remained a footnote in history. However, their algorithm drastically reduced the time needed to compute the DFT and led to the ubiquity of its application across engineering and mathematical fields. There are many different FFT algorithms involving a wide range of mathematics, from simple complex-number arithmetic to group theory and number theory. The Fast Fourier Transform computes the DFT of N points in the naive way, using the definition, takes $O(N2)$ arithmetical operations, while an FFT can compute the same DFT in only $O(N \log N)$ operations. The difference in speed can be enormous, especially for long data sets where N may be in the thousands or millions. In practice, the computation time can be reduced by several orders of magnitude in such cases, and the improvement is roughly proportional to $N \log N$. This huge improvement made the calculation of the DFT practical; FFTs are of great importance to a wide variety of applications, from digital signal processing and solving partial differential equations to algorithms for quick multiplication of large integers.

5 Conclusion

This list of important mathematical discoveries, although highly subjective, is universally acknowledged by mathematicians as the most remarkable innovations in the history of Mathematics.
From its roots in ancient Mesopotamia and Greece to the mathematical revolutions of the middle ages to the complexity and abstraction of the modern era, Mathematics has come a long way. It has evolved all the way from simple counting, measurement and calculation, through the application of abstraction, imagination and logic, to the broad, complex and often abstract discipline we know today.

References

[1] God Created the Integers: The Mathematical Breakthroughs that Changed History by Stephen Hawking

[2] https://mic.com/articles/29778/pi-day-5-greatest-mathematical-discoveries-in-history

[3] http://www.science4all.org/article/eulers-identity/

[4] http://www.wikipedia.org

[5] https://fas.org/sgp/eprint/discovery.pdf

[6] http://storyofmathematics.com/

AAKSHI MALASI, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
aakshimalasi01@gmail.com

Fractals and Menger Sponge

Neha Petwal

Abstract

Extending beyond the typical perception of mathematics as a body of perplexing formulas, fractal geometry amalgamates art with mathematics to demonstrate that equations are more than just a collection of numbers. What makes fractals even more interesting is that they are the best existing mathematical descriptions of many natural forms, such as coastlines, mountains or parts of living organisms. In this paper we will talk about how Fractals, based on pure mathematics, has now become widespread and multidisciplinary unravelling its unique worth repeatedly.

1 Introduction

A fractal is a mathematical set that exhibits a repeating pattern displayed at every scale. It is also known as expanding symmetry or evolving symmetry. Fractals include the idea of a detailed pattern that repeats itself.

Two of the most important properties of fractals are self-similarity and non-integer dimension which usually exceeds the topological dimension. Hence, fractals are informally considered to be infinitely complex as they appear similar in all levels of magnification. As mathematical equations, fractals are usually nowhere differentiable. An infinite fractal curve can be conceived as winding through space differently from an ordinary line, still being a 1-dimensional line yet having a fractal dimension indicating that it also resembles a surface. The non-integer dimension is quite difficult to explain. Many natural phenomena are better described using a dimension between two whole numbers. So while a straight line has a dimension of one, a fractal curve will have a dimension between one and two, depending on how much space it takes up as it twists and curves. The more the flat fractal fills a plane, the closer it approaches two dimensions. Likewise, a "hilly fractal scene" will reach a dimension somewhere between two and three. So a fractal landscape made up of a large hill covered with tiny mounds would be close to the second dimension, while a rough surface composed of many medium-sized hills would be close to the third dimension. Some of the most prominent examples of fractals are the Koch curve, the Sierpinski triangle and many more.

2 History and Origin

Inspite of its close connection with computer techniques, work on fractals had started long ago by the British cartographers, who encountered the problem in measuring the length of Britain coast. The coastline measured on a large scale map was approximately half the length of coastline measured on a detailed map. The closer they looked, the more detailed

and longer the coastline became. Little did they realise that they had discovered this wondrous concept of fractals.

Compared to the Euclidean geometry, which has a long history for more than 2000 years, Fractal geometry is very new. The mathematical roots of the idea of fractals have been traced throughout the years as a formal path of published works, starting in the 17th century with notions of recursion, then moving through increasingly rigorous mathematical treatment of the concept to the study of continuous but not differentiable functions in the 19th century by the seminal work of Bernard Bolzano, Bernhard Riemann, and Karl Weierstrass.

In the 1970s, the work of a leading researcher Gaston Maurice Julia on the iteration of polynomials was revived and popularised by the Polish-born Benoit Mandelbrot. Inspired by Julia's work, and with the aid of computer graphics, IBM employee Mandelbrot was able to show the first pictures of the most beautiful fractals known today.

The term 'fractal' was first coined by mathematician Benoit Mandelbrot in 1975. Mandelbrot based it on the Latin fractus meaning "broken" or "fractured", and used it to extend the concept of theoretical fractional dimensions to geometric patterns in nature. Benoit Mandelbrot's famous book, 'The Fractal Geometry of Nature', was published in 1982.

3   Menger Sponge

One of the most significant instance of a fractal which grabbed attention is a 'Menger Sponge'. First described by Karl Menger in 1926 it is all about surface appearances. It's a purely theoretical shape that has infinite surface area and no volume whatsoever. And because of that, it doesn't occupy three dimensions. It manages to exist in fractional dimensions.
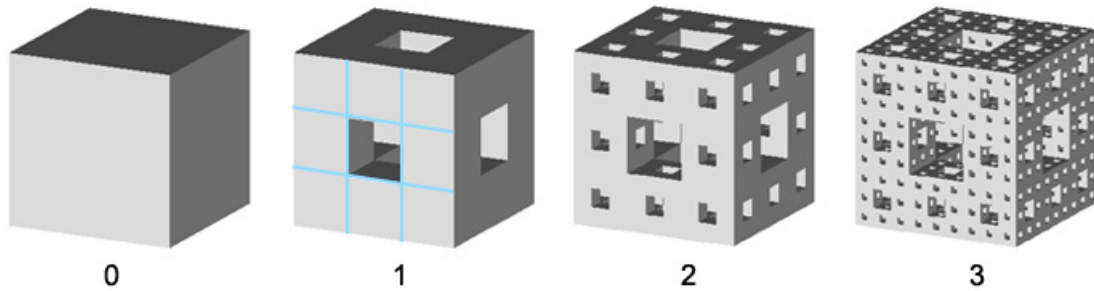

Figure 1: Menger Sponge

3.1   Construction

1. We begin with a cube (3×3).

2. Dividing each face of the cube into 9 squares, like a rubix cube it will subdivide the cube into 27 smaller cubes.

3. Now if we remove the smaller cubes in the middle of each face and also the cube in the center of the larger cube, leaving twenty smaller cubes will give us the level -1 of the menger sponge (resembling a void cube).

4. Repeating the above steps for each of the remaining smaller cubes will give us level -2 of the sponge and so on.

Now each of the blocks that make up the original hollow cube is also made up of a hollow cube, and the surface will begin to look pitted. Because the edges of each cube are left intact, the structure maintains its shape even as it gets more and more tiny hollows drilled into it. As the number of divisions reaches infinity, the whole thing becomes a kind of lattice with no volume inside, just surfaces of infinitely pitted and thinned walls.

4 Multidisciplinary Explorations of Fractals

Fractal geometry has permeated many areas of science such as astrophysics, biological sciences and has become one of the most important techniques in computer science.

1. Computer Graphic
   Many image compression schemes use fractal algorithms to compress computer graphics files to less than a quarter of their original size.

2. Astrophysics
   Astrophysicists believe that key to solve the mystery behind the existence of stars is the fractal nature of interstellar gas. Fractal distributions are hierarchical, like smoke trails or billowy clouds in the sky. Turbulence shapes both the clouds in the sky and the clouds in space, giving them an irregular but repetitive pattern that would be impossible to describe without the help of fractal geometry.

3. Biological Sciences
   Most natural objects, such as clouds and organic structures, resemble fractals. Self-similarity has been found also in DNA sequences. In the opinion of some biologists fractal properties of DNA can be used to resolve evolutionary relationships in animals. Scientists discovered that the basic architecture of a chromosome is tree-like; every chromosome consists of many 'mini-chromosomes', and therefore can be treated as fractal. In addition, fractals are used to predict or analyse various biological processes or phenomena such as the growth pattern of bacteria, the pattern of situations such as nerve dendrites, etc.

5 Conclusion

All in all, studying fractals is both a complicated yet an interesting branch of Mathematics. And yet despite all its intricacies, the various aspects of fractals never fail to astound us as one goes deeper into understanding them.

References

[1] https://www.fractal.org

[2] http://www.fractalfoundastion.org

[3] http://www.whatis.techtarger.com

[4] http://www.mathworld.wolfram.com

[5] https://en.wikipedia.org/wiki/fractal

NEHA PETWAL, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
nehapetwal1@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.

Brownian Motion and Applications in the Financial Market

Anika Jain and Eishita Yadav

Abstract

The mathematical theory of brownian motion has been applied in contexts ranging far beyond the movement of microscopic particles. The mathematical equations used to describe Brownian motion are the fundamental tools on which all the financial and derivate pricing models are based. The nature of derivate assets provide an interesting conduit for the analysis and application of Brownian motion and solving partial derivative equations, while maintaining real world applications. The purpose of the paper is to introduce the Brownian motion with its properties and the applications of the Black-Scholes option pricing model.

1 Brownian motion

1.1 A Brief History

Brownian motion is defined as the continuous and erratic random movement of microscopic particles in fluid as a result of continuous bombardment from molecules of the surrounding medium. It was first discovered in 1827 by the biologist Robert Brown while he was studying pollen particles floating in the microscope. He was however unable to determine the mechanisms that caused this motion. Later in 1905, Albert Einstein succeeded in stating the mathematical laws governing the movements of particles on the basis of the principles of kinetic molecular theory of heat. In 1923 Norbert Weiner, an American Mathematician established the mathematical existence of Brownian motion. He verified the existence of stochastic processes with the given properties.

1.2 Definition

A real valued stochastic process $\{B(t) : t \geq 0\}$ is called a (linear) Brownian motion with start in $x \in \mathbb{R}$ if the following holds:

- $B(0) = x$

- the process has independent increments, i.e. for all times $0 \leq t_1 \leq t_2 \leq .... \leq t_n$ the increments $B(t_n) - B(t_n - 1), B(t_n - 1) - B(t_n - 2), ..., B(t_2) - B(t_1)$ are independent random variables

- for all $t \geq 0$ and $h > 0$, the increments $B(t + h) - B(t)$ are normally distributed with expectation zero and variance h

- Almost surely, the function $t \to B(t)$ is continuous

15

1.3 Local path properties of Brownian motion

- **Hölder continuity** Brownian paths $t \to B_t$ are continuous $t \to B_t$ is Hölder continuous of order α for every $\alpha < \frac{1}{2}$. That is, $\forall\, T > 0$, $0 < \alpha < 1/2$, $\exists$ a constant C $= C(T, \alpha)$ s.t. $\forall\, t, s < T$

$$\mid B_t - B_s \mid \leq C \mid t - s \mid^\alpha$$

- **Differentiabilty** Brownian paths are nowhere Hölder continuous of order $\alpha > \frac{1}{2}$. In particular Brownian paths are nowhere differentiable.

  Consider a small increment $B(t + dt) - B(t)$. By the defination of brownian motion , this is normally distributed with mean 0 and variance dt.

  Thus,

$$E(\mid B(t + dt) - B(t) \mid^2) = dt$$

  i.e the typical size of an inrement $\mid B(t + dt) - B(t) \mid$ is about $\sqrt{dt}$. Now as $dt \to 0$, $\sqrt{dt} \to 0$, which is consistent with the continuity of paths. However, if we try to take the derivative:

$$\frac{dB_t}{dt} = \lim_{t \to 0} \frac{B(t + dt) - B(t)}{dt}$$

  then we see that when dt is small, the absolute value of the numerator looks like $\sqrt{dt}$ which is much larger than dt. Therefore the limit does not exist.

- **Local maxima and minima** For a continuous function f:$[0, \infty) \to \mathbb{R}$, a point t is a local (strict) maximum if $\exists\, \epsilon > 0$ s.t. $\forall\, s \in (t - \epsilon, t + \epsilon), f(s) \leq f(t)(f(s) < f(t))$. Then for almost all paths, the set of local maxima for the Brownian path B is countable and dense in $[0, \infty)$, each local maxima is strict, and there is no interval on which the path is monotone.

- **Points of increase and decrease** A point t is a point of increase if $\exists\, \epsilon > 0$, s.t. $\forall\, s \in (0, \epsilon), f(t - s) \leq f(t) \leq f(t + s)$. Then for almost all paths, the Brownian path B has no points of increase or decrease.

2 Geometric Brownian Motion

Consider a single stock, with price $S(t)$, which varies with time. The value of the option deriving from that stock should have a market value that is a function of S and t. Let this be $D(t) = V(t, S(t))$. In the world of finance, the most significant descriptor of the profitability of an asset is its rate of return. In order to describe the perturbations of the return on a share of stock, we will model it with the geometric brownian motion.

2.1 Definition

A stochastic process $S(t)$ is said to follow a Geometric Brownian Motion if it satisfies the following stochastic differential equation:

$$dS_t = uS_t dt + \sigma S_t dW_t$$

where W_t is a Weimer Process (Brownian motion) and σ, μ are constants. μ is called the percentage drift and σ is called the percentage volatility. So, consider a Brownian motion trajectory that satisfies this differential equation, the right hand side term $\mu S_t dt$ controls the trend of this trajectory and the term $\sigma S_t dW_t$ controls the random noise effect in the rajectory. Since its a differential equation, we want to find the solution. By technique of separation of variables, the equation becomes:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$$

Then take integration on both the sides

$$\int \frac{dS_t}{S_t} = \int (\mu dt + \sigma dW_t) dt$$

Since dS_t relates to the derivative of $\ln S_t$, so the next step involving the Ito calculus and arriving the following equations

$$\ln(\frac{dS_t}{S_t}) = (u - \frac{1}{2}\sigma^2)t + \sigma W_t$$

Taking the exponential on both sides and plugging the initial condition S_0, we obtain the solution. The analytical solution of this geometric brownian motion is given by:

$$S_t = S_0 \exp((\mu - \frac{\sigma^2}{2})t + \sigma W_t)$$

In general, the process above is of solving a stochastic differential equation, and in fact, geometric brownian motion is defined as a stochastic differential equation. Thus given the constant μ and σ, we are able to produce a Geometric Brownian Motion solution through out time interval. Note that given drift rate and volatility rate, we can represent GBM solution in the form

$$S(t) = S_0 \exp(X(t))$$

where

$$X(t) = (\mu - \frac{\sigma^2}{2})t + \sigma W_t$$

2.2 Markov Chain Properties of Geometric Brownian Motion

Now we show that Geometric Brownian Motion satisfies the Markov Chain property, given the $S(t) = S_0 \exp(X(t))$ definition above, we have

$$S(t+h) = S_0 \exp(X(t+h)) = S_0 \exp(X(t)) + X(t+h) - X(t)$$

$$= S_0 \exp(X(t)) \exp(X(t+h)) - X(t) = S_t \exp(X(t+h)) - X(t)$$

Thus the future states $S(t+h)$ depends only on the future increment of the Brownian Motion, namely $X(t+h)X(t)$, which is independent.

3 Black Scholes Model

The Black Scholes formula developed by Fischer Black and Myron Scholes in 1973 was revolutionary in its impact on the financial industry. Their seminal paper, "The Pricing of Options and Corporate Liabilities" was published in the Journal of Political Economy. Today, many of the techniques and pricing models used in finance are rooted in the ideas and methods presented by these two men. From the viewpoint of innovation theory, the Black-Scholes option pricing formula is of interest because it constitutes a paradigm change. Before the publication of the formula, finance was dominated by the paradigm of risk-based pricing models rooted in an equilibrium (neoclassic) thinking with the Capital Asset Pricing Model (CAPM) being the most prominent example.
The Black-Scholes option pricing formula marks an important step towards contemporary finance with its arbitrage-based pricing models and the related paradigm of arbitrage theory. The formula is a scientific as well as an economic innovation: Using the Black-Scholes option pricing formula it was possible for the first time to compare the price resulting from supply and demand with the analytical value of an option.

3.1 Brief History

Fisher Black started working to create a valuation model for stock warrants. This work involved calculating a derivative to measure how the discount rate of a warrant varies with time and stock price. The result of this calculation held a striking resemblance to a well-known heat transfer equation. Soon after this discovery, Myron Scholes joined Black and the result of their work is a startlingly accurate option pricing model. Black and Scholes can't take all credit for their work, in fact their model is actually an improved version of a previous model developed by A. James Boness in his Ph.D. dissertation at the University of Chicago. Black and Scholes' improvements on the Boness model come in the form of a proof that the risk-free interest rate is the correct discount factor, and with the absence of assumptions regarding investor's risk preferences.

3.2 Assumptions of the Black and Scholes Model

- **The stock pays no dividends during the option's life:** Most companies pay dividends to their share holders, so this might seem a serious limitation to the model considering the observation that higher dividend yields elicit lower call premiums. A common way of adjusting the model for this situation is to subtract the discounted value of a future dividend from the stock price.

- **European exercise terms are used:** European exercise terms dictate that the option can only be exercised on the expiration date (American exercise term allow the option to be exercised at any time during the life of the option). This limitation is not a major concern because very few calls are ever exercised before the last few days of their life because the remaining time value on the call option would be lost.

- **Markets are efficient:** This assumption suggests that people cannot consistently predict the direction of the market or an individual stock. The market operates continuously with share prices following a continuous Itô process which is a Markov

process ("one where the observation in time period t depends only on the preceding observation") in continuous time.

- **No commissions are charged:** Usually market participants do have to pay a commission to buy or sell options or pay some kind of small fee to floor traders. The fees that individual investor's pay is more substantial and can often distort the output of the model.

- **Interest rates remain constant and known:** This model uses the risk-free rate to represent this constant and known interest rate even though there is no such thing in reality.

- **Returns are lognormally distributed:** This assumption suggests, returns on the underlying stock are normally distributed, which is reasonable for most assets that offer options.

3.3 The Black Scholes Formula

The Black-Scholes option pricing formula is a pricing algorithm for European call options on stocks without dividends. The key idea behind the model is to hedge the option by buying and selling the underlying asset in just the right way and, as a consequence, to eliminate risk.

$$C = SN(d_1) - Ke^{-rt}N(d_2)$$

where, C =Theoretical call premium

S =Current Stock Price

t =Time until option expiration

K =Option Striking Price

r =Risk-free Interest Rate

N =Cumulative Standard Normal Distribution

s =Standard Deviation of Stock Returns

$$d_1 = \frac{ln(\frac{S}{K}) + (r + \frac{s^2}{2})t}{s\sqrt{t}}$$

$$d_2 = d_1 - s\sqrt{t}$$

The first part, $SN(d_1)$, derives the expected benefit from acquiring a stock outright. This is found by multiplying stock price S by the change in the call premium with respect to a change in the underlying stock price $N(d_1)$. The second part of the model, $Ke^{(-rt)}N(d_2)$, gives the present value of paying the exercise price on the expiration day. The fair market value of the call option is then calculated by taking the difference between these two parts.

3.4 Definitions

- Delta=$N(d_1)$

  Delta is a measure of the sensitivity the calculated option value has to small changes in the share price.

- Gamma=$\frac{d^2C}{dS^2} = \frac{e^{\frac{-d^2}{2}}}{S\sigma\sqrt{2\Pi T}}$ Gamma is a measure of the calculated delta's sensitivity to small changes in share price.

- Theta=$\frac{dC}{dt} = \frac{e^{\frac{S\sigma}{\frac{d^2}{2}}}}{2\sqrt{2\Pi T}} - \frac{rE}{e^{rT}}*N(d-\sigma\sqrt{T})$ Theta measures the calculated option value's sensitivity to small changes in time till maturity.

- Vega=$\frac{S\sqrt{T}}{\frac{e^{\frac{d^2}{2}}}{\sqrt{2\Pi}}}$ Vega measures the calculated option value's sensitivity to small changes in volatility.

- Rho=$\frac{TE}{e^{rt}}N(d-\sigma\sqrt{T})$

3.5   After the Black and Scholes Model

Since 1973, the original Black and Scholes Option Pricing Model has been the subject of much attention. Many financial scholars have expanded upon the original work. In 1973, Robert Merton relaxed the assumption of no dividends. In 1976, Jonathan Ingerson went one step further and relaxed the the assumption of no taxes or transaction costs. In 1976, Merton responded by removing the restriction of constant interest rates. The results of all of this attention, that originated in the autumn of 1969, are alarmingly accurate valuation models for stock options.

References

[1] http://www.frankfurt-school.de/clicnetclm/fileDownload.do?goid=000000531416AB4

[2] https://www.cs.princeton.edu/courses/archive/fall09/cos323/papers/blackscholes73.pdf

[3] http://bradley.bradley.edu/ arr/bsm/pg04.html

[4] Wikipedia, *Brownian Motion* , Wikipedia,2006.

[5] https://www.stat.berkeley.edu/ peres/bmbook.pdf

[6] M.Kozdron,*A random look at Brownian Motion*,Duke University,2002.

Further Reading

[1] http://www.math.uchicago.edu/ may/VIGRE/VIGRE2010/REUPapers/Turner.pdf

ANIKA JAIN, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
anikajain611@gmail.com

EISHITA YADAV, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
ey4172@gmail.com

Lévy Processes

Nikita Sobti

Abstract

This paper describes the importance of Lévy processes. It discusses the two main important building blocks of Lévy Processes: Poisson Processes and Brownian Motion and the application of Lévy Processes in real life.

1   Introduction

Going back to Merton's model and till the present date, various types of models that include stochastic processes that involve discrete movements called jumps have been studied in financial mathematics. In the last decade, researchers of major banks started to accept jump diffusions and Lévy processes as a valuable tool in their day-to-day modelling. The increasing interest to jump models in finance is mainly due to the following reasons:

1. In a model with continuous paths like a diffusion model, the price process behaves locally like a Brownian motion and the probability that the stock moves by a large amount over a short period of time is very small, unless one fixes an unrealistically high value of volatility. Therefore, in such models the prices of short term Out of The Money (OTM) options should be much lower than what one observes in real markets. On the other hand, if stock prices are allowed to jump, even when the time to maturity is very short, there is a non-negligible probability that after a sudden change in the stock price the option will move in the money.

2. From the point of view of hedging, continuous models of stock price behaviour generally lead to a complete market or to a market which can be made complete by adding one or two additional instruments, like in stochastic volatility models. Since in such a market every terminal payoff can be exactly replicated, options are redundant assets, and the very existence of traded options becomes a puzzle. The mystery is easily solved by allowing for discontinuities: in real markets, due to the presence of jumps in the prices, perfect hedging is impossible and options enable the market participants to hedge risks that cannot be hedged using the underlying only.

3. From a risk management perspective, jumps allow to quantify and take into account the risk of strong stock price movements over short time intervals, which appears non-existent in the diffusion framework.

4. The strongest argument for using discontinuous models is simply the presence of jumps in observed prices. Figure 1 depicts the evolution of the DM/USD exchange rate over a two-week period in 1992, and one can see at least three points where the rate moved by over 100 bp within a 5-minute period. Price moves like these ones clearly cannot

be accounted for in a diffusion model with continuous paths, but they must be dealt with if the market risk is to be measured and managed correctly.
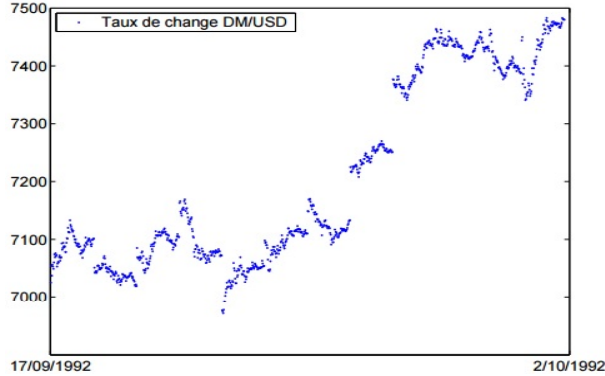

Figure 1: Jumps in the trajectory of DM/USD exchange rate, sampled at 5-minute intervals.

The main empirical motivation of using Lévy processes in finance comes from fitting asset return distributions. Consider the daily (either continuous or simple) returns of S&P 500 index (SPX) from Jan 2, 1980 to Dec 31, 2005. We plot the histogram of normalised (mean zero and variance one) daily simple returns in Figure 1, along with the standard normal density function. The max and min (which all occurred in 1987) of the normalised daily returns are about 7.9967 and -21.1550. Note that for a standard normal random variable $Z, P(Z < -21.550) \approx 1.4 * 10^{-107}$ as a comparison the whole universe is believed to have existed for 15 billion years or $5 * 10^{17}$ seconds.

Clearly the histogram of SPX displays a high peak and two asymmetric heavy tails. This is not only true for SPX but also for almost all financial asset prices, e.g. world wide stock indices, individual stocks, foreign exchange rates, interest rates. In fact it is so evident that a name *"leptokurtic distribution"* is given, which means the kurtosis of the distribution is large. More precisely, the kurtosis and skewness are defined as

$$K = E\left(\frac{(X-\mu)^4}{\sigma^4}\right), S = E\left(\frac{(X-\mu)^3}{\sigma^3}\right)$$

For the standard normal density $K = 3$, and if $K > 3$ then the distribution is called *leptokurtic* and the distribution will have a higher peak and two heavier tails than those of the normal distribution. For the SPX data, the sample kurtosis is about 42.23. The skewness is about -1.73; the negative skewness indicates that the left tail is heavier than the right tail. The classical geometric Brownian motion model, which models the stock price as $S(t) = S(0)e^{\mu t + \sigma W_t}$ with W_t being the standard Brownian motion, is inconsistent with this feature, because in this model the return, $\ln\left(\frac{S(t)}{S(0)}\right)$, has a normal distribution. Lévy processes, among other processes, have been proposed to incorporate the leptokurtic feature.
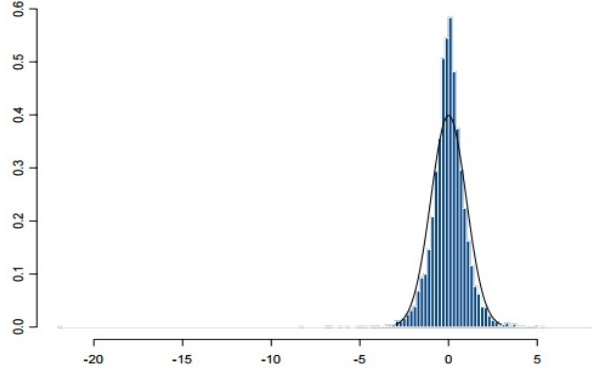
22

Figure 2: Comparison of the histogram of the normalized daily returns of S&P 500 index (from Jan 2, 1980 to Dec 31, 2005) and the density of N(0,1). The feature of a high peak and two heavy tails (i.e. the leptokurtic feature) is quite evident.

2 Definition of Lévy Processes

Stochastic processes are collections of random variables X_t, $t \geq 0$ (meaning $t \in [0,\infty)$ as opposed to $n \geq 0$ by which means $n \in N = \{0,1,2,\ldots\}$). For us, all X_t , $t \geq 0$, take values in a common state space, which we will choose specifically as $R(or [0,\infty)$ or R^d for some $d \geq 2$). We can think of X_t as the position of a particle at time t, changing as t varies. It is natural to suppose that the particle moves continuously in the sense that $t \longmapsto X_t$ is continuous, or that it has jumps for some $t \geq 0$:

$$\triangle X_t = X_{t+} - X_{t-} = \lim_{\varepsilon \to 0} X_{t+\varepsilon} - \lim_{\varepsilon \to 0} X_{t-\varepsilon}$$

We will usually suppose that these limits exist for all $t \geq 0$ and that in fact $X_{t+} = X_t$,i.e. that $t \to X_t$ is *right-continuous with left limits* X_{t-} for all $t \geq 0$ almost surely. The path $t \to X_t$ can then be viewed as a *random right-continuous function*.

A real-valued (or R_d -valued) stochastic process $X = (X_t)_{t \geq 0}$ is called a *Lévy process* if

1. the random variables $X_{t0}, X_{t1} - X_{t0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent for all $n \geq 1$ and $0 \geq t_0 < t_1 < \ldots < t_n$ (independent increments),

2. $X_{t+s} - X_t$ has the same distribution as X_s for all s, $t \geq 0$ (stationary increments),

3. the paths $t \to X_t$ are right-continuous with left limits

3 Poisson Processes

One of the important building blocks of Lévy processes are poisson processes. An $N(\subset R)$-valued stochastic process $X = (X_t)_{t \geq 0}$ is called a *Poisson process* with rate $\lambda \in (0,\infty)$ if X satisfies (1)-(3) and

4. $P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k \geq 0, t \geq 0$ *(Poisson distribution)*

In layman terms, if the waiting times between buses at a bus stop are exponentially distributed, the total number of buses arrived up to time t is a Poisson process. The trajectories of a Poisson process are piece-wise constant (right-continuous with left limits or RCLL), with

jumps of size 1 only. The jumps occur at times T_i and the intervals between jumps (the waiting times) are exponentially distributed. However for financial applications, it is of little
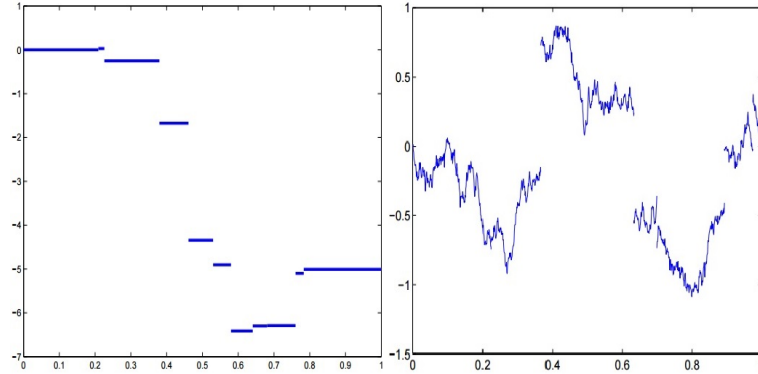

Figure 3: Left: sample path of a compound Poisson process with Gaussian distribution of jump sizes. Right: sample path of a jump-diffusion process (Brownian motion + compound Poisson).

interest to have a process with a single possible jump size. Thus we introduce the compound Poisson process:

The compound Poisson Process It is a generalisation where the waiting times between jumps are exponential but the jump sizes can have an arbitrary distribution. More precisely, let N be a Poisson process with parameter λ and $\{Y_i\}_{\{i \geq 1\}}$ be a sequence of independent random variables with law f. The process

$$X_t = \sum_{i=1}^{N_t} Y_i$$

is called *compound Poisson process*. Its trajectories are RCLL and piece-wise constant but the jump sizes are now random with law f. (Fig.3)
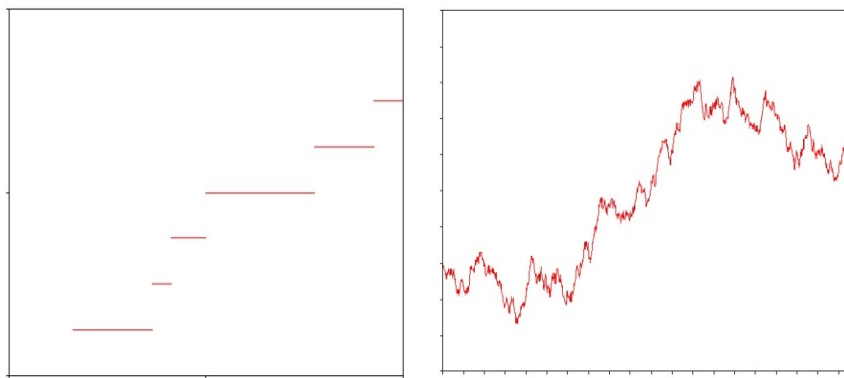

Figure 4: Poisson Process and Brownian Motion.

4 Applications of Lévy Processes

Population models

Branching processes are generalisations of birth and-death processes where each individual in a population dies after an exponentially distributed lifetime, but gives birth not to single children,but to twins, triplets, quadruplet etc. To simplify, it is assumed that children are only born at the end of a lifetime. The numbers of children are independent and identically distributed according to an offspring distribution q on $\{0, 2, 3, \ldots\}$. The population size process can jump downwards by 1 or upwards by an integer. It is not a Lévy process but is closely related to Lévy processes and can be studied with similar methods. There are also analogues of processes in $[0, \infty)$, so-called continuous-state branching processes that are useful large-population approximations.

Insurance ruins

A compound Poisson process with positive jump sizes can be interpreted as a claim process recording the total claim amount incurred before time t. If there is linear premium income at rate $r > 0$, then also the gain process is a Lévy process.

References

[1] http://www.columbia.edu/ sk75/levy.pdf

[2] http://www.proba.jussieu.fr/pageperso/tankov/wccfpaper.pdf

[3] http://www.stats.ox.ac.uk/ winkel/ms3b10.pdf

NIKITA SOBTI, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
nikitasobti@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.

Cost Time Trade-off in Three Axial Sums' Transportation Problem

Lakshmisree Bandopadhyaya
Typed by Shubhra Aggarwal

Abstract

The bi-objective Cost-time Trade-off Three Axial Sums' Transportation Problem is shown to be equivalent to a single-objective standard Three Axial Sums' problem, which can be solved easily by the existing efficient methods. The equivalence is established for some specially defined solutions termed as Lexicographic optimal solutions with minimum pipe-line.

1 Introduction

In 1961, Charnes and Cooper [3] discussed an approach to the solutions of managerial level problems involving multiple conflicting objectives (or goals). In 1962, Ignizio [8] was the first person to study the application of goal-programming to an engineering design problem. While working on the Saturn/Apollo antenna designing program (the U.S. moon-landing mission), which had to satisfy a number of conflicting objectives, he extended the original goal programming concept of Charnes and Cooper to a non-linear model. Later, Ignizio [9, 10] developed further extensions of the goal programming approach. In 1965, Y.Iziri [11] proposed the inclusion of the concept of "preemptive priorities". He suggested that a priority be given to each objective or a set of commensurable objectives in the problem. In practice, this concept is achieved through finding the lexicographic minimum of an ordered vector. The Three Axial Sums' Problem was first defined by E.D. Schell [13] in 1955. An efficient solution method was suggested by A. Corban [4] in 1964. The problem of minimizing the duration of transportation has been studied by many authors like Hammer [6, 7], Garfinkel and Rao [5], Szwarc [16], Bhatia, Swaroop and Puri [1, 2], Sharma and Swaroop [15], Seshan and Tikekar [14] and Prakash [12]. Some of these authors have also tried to unify the two problems by giving high and low priorities to the two objectives. This is what has come to be known as the cost-time trade-off problem. In the present paper cost-time trade-off has been studied in a bi-objective Three-Axial Sums' Transportation problem. It is established that such Trade-off Three-index Problems are equivalent to a single objective standard Three-Axial Sums' Problem. The reduction to a single objective form has been achieved by assigning suitable weights to the objectives. The weights have been clearly defined to give high and low priorities to cost and time respectively. The equivalence is established for some specially defined solutions termed as lexicographic optimal solutions with minimum pipeline.

27

2   Theoretical development

A cost-time trade-off Three Axial Sums' problem is :

$$P(1): \begin{cases} Minimize\ C = \sum_I \sum_J \sum_K c_{ijk}x_{ijk} \\ \quad T = max\ t_{ijk} : x_{ijk} > 0 \\ subject\ to \sum_J \sum_K x_{ijk} = a_i, i \in I \\ \quad \sum_I \sum_K x_{ijk} = b_j, j \in J \\ \quad \sum_I \sum_J x_{ijk} = c_k, k \in K \\ \quad x_{ijk} \geq 0, \end{cases}$$

where *I, J, K* are respectively the index-sets for the warehouses, markets and the modes of transport.

a_i is the availability at the ith warehouse,

b_j is the demand at the jth market,

c_k is the capacity of the kth mode of transport,

t_{ijk} is the total time taken by the kth mode in transporting from i^{th} warehouse to j^{th} market.

c_{ijk} is the per unit cost of transporting from i^{th} warehouse to j^{th} market by the kth mode.

x_{ijk} is the total amount transported from i^{th} warehouse to j^{th} market by the kth mode. The method of solving P(1) is to optimise the cost minimisation transportation problem in P(1) by any of the methods suggested by Schell, [13], or Corban [4]. Then locating all the alternate optimal solutions corresponding to the cost, choose the one with the minimum time. The corresponding solution that is obtained is the optimal solution is the optimal solution of P(1) with cost as the first priority and the time as second priority.

The triplet (X, C, T) will be used to denote the optimal solution $X = \{x_{ijk}\}$ of P(1) with cost as first priority and time as second, where C is the optimal cost and T is the corresponding minimum time.

It will be shown that the problem P(1) with cost as first priority and time as second can be reduced to the following single objective Three-Axial Sums' problem.

$$P(2): \begin{cases} Minimize\ Z = M_0 \sum_I \sum_J \sum_K c_{ijk}x_{ijk} + \sum_{s=1}^q (M_s \sum_{L_s} x_{ijk}) \\ subject\ to \sum_J \sum_K x_{ijk} = a_i, \\ \quad \sum_I \sum_K x_{ijk} = b_j, \\ \quad \sum_I \sum_J x_{ijk} = c_k, \\ \quad x_{ijk} \geq 0, \end{cases}$$

where $L_1, L_2, \ldots, L_q$ and $M_0, M_1, \ldots, M_q$ are defined below: Let

$$\max_{I,J,K} t_{ijk} = \beta_1, \qquad \min_{I,J,K} t_{ijk} = \beta_q$$

The set $\{t_{ijk} : i \in I, j \in J, k \in K\}$ is sub-divided into subsets $L_1, L_2, \ldots, L_q$ satisfying the following two conditions:

(i) L_u contains the t_{ijk}'s having the same numerical value, $u = 1, 2, \ldots, q$;

(ii) $\beta_1 > \beta_2 > \ldots > \beta_q$.

$M_0, M_1, \ldots, M_q$ are   the   priority   factors   associated   with $\sum_I \sum_J \sum_K c_{ijk}x_{ijk}$, $\sum_{L_1} x_{ijk}, \sum_{L_2} x_{ijk}, \ldots, \sum_{L_q} x_{ijk}$ respectively, $\sum_{L_s} x_{ijk}$ denoting the sum of all those x_{ijk}'s

with corresponding $t_{ijk} \in L_s$. $M_0, M_1, \ldots, M_q$ are all positive and they are so defined, that the sign of $\sum_{k=0}^{q} \alpha_k M_k$ is the same as the non-zero α_k, with smallest subscript k in it, irrespective of the other α_k's.

Definition: Lexicographically Optimal Solution with Minimum Pipeline

An optimal solution (X^*, C^*, T^*) of the cost-time Trade-off Three Axial Sums' Transportation problem $P(1)$, will be called lexicographically optimal with minimum pipeline, if for any solution (X, C^*, T^*), either

$$\sum_{L_s} x^*_{ijk} < \sum_{L_s} x_{ijk},$$

where $T^* \in L_s$, $X^* = \{x^*_{ijk}\}$, $X = \{x_{ijk}\}$, or if

$$\sum_{L_s} x^*_{ijk} = \sum_{L_s} x_{ijk}$$

and T_1^* is the second largest time for both the solution X and X^*, so that $T_1^* \in L_{s+1}$, then

$$\sum_{L_{s+1}} x^*_{ijk} < \sum_{L_{s+1}} x_{ijk}.$$

The equivalence of the problem $P(1)$ with cost as first priority and the time as second with the problem $P(2)$ is established, through the following theorem and its converse.

THEOREM. *If $(X^* = \{x^*_{ijk}\}_{I \times J \times K}, C^* < T^*)$ is a lexicographically optimal solution of $P(1)$ with minimum pipeline, then it is also optimal for $P(2)$.*

PROOF. Obviously X^* is feasible for $P(2)$, as the feasibilty conditions of $P(1)$ and $P(2)$ are same.

Now we show X^* is optimal for $P(2)$.

Let $X = \{x_{ijk}\}_{I \times J \times K}$ be any feasible solution of $P(2)$ with corresponding cost as C and time T.

Now $C \geq C^*$, as X is also feasible for $P(1)$.

CASE(I) $C > C^*$.

That is,

$$\sum_I \sum_J \sum_K c_{ijk} x_{ijk} > \sum_I \sum_J \sum_K c_{ijk} x^*_{ijk}$$

which implies

$$\sum_I \sum_J \sum_K c_{ijk}(x_{ijk} - x^*_{ijk}) > 0. \tag{1}$$

Now

$$
\begin{aligned}
Z - Z^* &= M_0 \sum_I \sum_J \sum_K c_{ijk}(x_{ijk} - x^*_{ijk}) + \sum_{u=1}^{q} M_u \sum_{L_u} (x_{ijk} - x^*_{ijk}) \\
&= \sum_{u=0}^{q} M_u \alpha_u.
\end{aligned}
$$

Since $\alpha_0 > 0$ (using (1)), $\quad\quad Z - Z^* > 0$ (by definition of $M_0, M_1 \ldots, M_q$)

$\Rightarrow Z > Z^*$

$\Rightarrow X^*$ has the minimum objective function value compared to any feasible solution of $P(2)$.

$\Rightarrow X^*$ is optimal for $P(2)$.

CASE(II) $C = C^*$

This shows that X^*, X are two alternate optimal solutions of $P(1)$ with respect to cost minimisation. The optimality of (X^*, C^*, T^*) suggests that T^* is the minimum of all the times corresponding to the alternate optimal solutions with respect to cost C^*.

Thus $T^* \leq T$. In the case $T^* < T, T^* \in L_s$ and $T \in L_p$ then $s > p$ and

$$L_1 = L_2 = \ldots = L_{s-1} = \phi \text{ for } X^* = \{x^*_{ijk}\},$$
$$L_1 = L_2 = \ldots = L_{p-1} = \phi \text{ for } X = \{x_{ijk}\}$$

Now $Z - Z^* = \sum_{k=1}^{q} M_k \sum_{L_k} (x_{ijk} - x^*_{ijk})$ (as $C = C^*$) has the sign of $\sum_{L_p} (x_{ijk} - x^*_{ijk}) = \sum_{L_p} x_{ijk}$.

Hence

$$Z - Z^* > 0$$
$$\Rightarrow Z > Z^*$$
$$\Rightarrow X^* \text{ is optimal for } P(2).$$

Again in the case $T^* = T$, we have $x = \{x_{ijk}\}$ as a feasible solution of $P(1)$ with the same cost C^* and same time T^* as the optimal solution X^*.

The definition of lexicographic optimality suggests that either (i) or (ii) holds.

(i) $\sum_{L_s} x^*_{ijk} < \sum_{L_s} x_{ijk}$, in which case, $\sum_{L_s} (x_{ijk} - x^*_{ijk}) > 0$.

Now since $L_1 = L_2 = \ldots = L_{s-1} = \phi$ for X^* and $L_1 = L_2 = \ldots = L_{s-1} = \phi$ for X,

$$
\begin{aligned}
Z - Z^* &= \sum_{u=1}^{q} M_k \sum_{L_s} (x_{ijk} - x^*_{ijk}) \\
&= M_s \sum_{L_s} (x_{ijk} - x^*_{ijk}) + \ldots + M_q \sum_{L_q} (x_{ijk} - x^*_{ijk}).
\end{aligned}
$$

Hence $Z - Z^*$ has the sign of $\sum_{L_s} (x_{ijk} - x^*_{ijk}) > 0$, which implies $Z > Z^*$, and hence X^* is optimal for $P(2)$.

(ii) $\sum_{L_s} x^*_{ijk} = \sum_{L_s} x_{ijk}$. In this case, if the second largest time for X^* and X are same, say T_1, then $T_1 \in L_{s+1}$ and $\sum_{L_{s+1}} x^*_{ijk} < \sum_{L_{s+1}} x_{ijk}$ and

$$
\begin{aligned}
Z - Z^* &= \sum_{k=1}^{q} M_k \sum_{L_k} (x_{ijk} - x^*_{ijk}) \\
&= M_{s+1} \sum_{L_s+1} (x_{ijk} - x^*_{ijk}) + \ldots + M_q \sum_{L_q} (x_{ijk} - x^*_{ijk})
\end{aligned}
$$

has the sign of $\sum_{L_s+1} (x_{ijk} - x^*_{ijk}) > 0$. Therefore $Z - Z^* > 0$. Hence X^* is optimal for $P(2)$. Again if the second largest times of X^* and X are different, say T_1 and T_2 respectively, with $T_1 < T_2$, and if $T_1 \in L_m$ and $T_2 \in L_n$ then $m > n$,

$$
\begin{aligned}
Z - Z^* &= \sum_{u=1}^{q} M_u \sum_{L_u} (x_{ijk} - x^*_{ijk}) \\
&= M_n \sum_{L_n} (x_{ijk} - x^*_{ijk}) + \ldots M_q \sum_{L_q} (x_{ijk} - x^*_{ijk}).
\end{aligned}
$$

Since $L_1 = L_2 \ldots = L_{m-1} = \phi$ for X^*, $L_1 = L_2 \ldots = L_{n-1} = \phi$ for X, $Z - Z^*$ has the sign of $\sum_{L_n}(x_{ijk} - x^*_{ijk}) = \sum_{L_n}(x_{ijk})$, that is $Z - Z^* > 0$ and so X^* is optimal for $P(2)$.
So in all the cases with various sub cases it has been proved that (X^*, C^*, T^*) is optimal for $P(2)$.

THEOREM.(Converse): If $X^* = \{x^*_{ijk}\}$ is an optimal solution of $P(2)$ with $\sum_I \sum_J \sum_K c_{ijk} x^*_{ijk} = C^*$ and T^* is the corresponding time, then (X^*, C^*, T^*) a lexicographically optimal solution of $P(1)$ with minimum pipeline.

PROOF. Obviously (X^*, C^*, T^*) is feasible for $P(1)$. Suppose $X \neq X^*$ is an optimal solution for $P(1)$ with costs C and time T. Then $C \leq C^*$ implies

$$\sum_I \sum_J \sum_K c_{ijk} x_{ijk} \leq \sum_I \sum_J \sum_K c_{ijk} x^*_{ijk}.$$

It is claimed that $C = C^*$; for if $C < C^*$, then

$$Z - Z^* = M_0 \sum_I \sum_J \sum_K c_{ijk}(x_{ijk} - x^*_{ijk}) + \sum_{u=1}^{q} M_u \sum_{L_u}(x_{ijk} - x^*_{ijk})$$

(by the definition of priority factors $M_0, M_1 \ldots, M_q$) has the sign of $\sum_I \sum_J \sum_K c_{ijk}(x_{ijk} - x^*_{ijk})$ which is less than zero (as $C < C^*$). But $Z - Z^* < 0$ implies $Z < Z^*$ which contradicts the optimality of X^* for $P(2)$. Thus any optimal solution of $P(1)$ must have its optimal cost as C^*.
Now, if (X^0, C^*, T^0) is any alternate optimal solution of $P(1)$, then $T^0 \leq T^*$. In the case $T^0 < T^*$ and $T^0 \in L_p$ then $s > p$ and $Z^0 - Z^* = \sum_{u=1}^{q} M_u \sum_{L_u}(x^0_{ijk} - x^*_{ijk})$ has the sign of

$$\sum_{L_p}(x^0_{ijk} - x^*_{ijk}) = -\sum_{L_p}(x^*_{ijk})$$

[as $L_1 = L_2 = \ldots = L_{p-1} = \phi$ for X^* and $L_1 = L_2 = \ldots = L_{p-1} = \ldots =$
$L_{s-1} = \phi$ for X]. Hence $Z^0 - Z^* < 0$, which contradicts the optimality of X^*.
Therefore $T^0 = T^*$. Thus any optimal solution must have optimal cost C^* and time T^*. This implies X^* is optimal for $P(1)$. Again X^* is lexicographically optimal with minimum pipeline, as any solution (X, C^*, T^*) of $P(1)$ must satisfy $\sum_{L_s} x^*_{ijk} \leq \sum_{L_s} x_{ijk}$, for otherwise $\sum_{L_s} x^*_{ijk} > \sum_{L_s} x_{ijk}$ will contradict the optimality of X^*.
Such a work finds its application in all problems of routing finished goods from factories to retailers using different modes of transport. A factory owner is generally satisfied when his goods are delivered at a minimum cost. The minimisation of time of delivery is his second concern. The present model will be most useful in all such cases.

References

[1] H.L.Bhatia, K.Swarup and M.C.Puri, "Time-cost trade-off in a transportation problem", *Opsearch* **13** (1976) $129 - 142$.

[2] H.L.Bhatia, K.Swarup and M.C.Puri, "A procedure for the time minimizing transportation problem", *Indian J. Pure Appl. Math.* **8** (1977) $920 - 929$.

[3] A.Charnes and W.W.Cooper, *Management models and industrial applications of linear programming, Vols 1 and 2* (Wiley, New York, 1961).

[4] Adrian Corban, "A multi-dimensional transportation problem", *Rev.Roum.Math.Pure and Appl.***9** (8)(1964) $721 - 735$.

[5] R.S.Garfinkel and M.R.Rao, "The bottleneck transportation problem", *Nav.Res.Log.Quart.* **18** (1971) $465 - 472$.

[6] P.L.Hammer, "Time minimizing transportation problems", *Nav.Res.Log.Quart.* **16** (1969) $345 - 357$.

[7] P.L.Hammer, "Time minimizing transportation problems", *Nav.Res.Log.Quart.* **18** (1971) $487 - 490$.

[8] J.P.Ignizio, "S-II Trajectory Study and Optimum Antenna Placement", North American Aviation Report S1D-63, Downey, California, 1963.

[9] J.P.Ignizio, *Goal programming and extensions* (Health(Lexington Books), Lexington, Mass, 1976).

[10] J.P.Ignizio, "Goal programming: A tool for multi-objective analysis", *J.Oper.Res.Soc.* **29** (II) (1978) $1109 - 1119$.

[11] Y.Iziri, *Management goals and accounting for control* (Rand Mcnally, Chicago, 1965).

[12] S.Prakash, "On minimizing the duration of transportation", *Proc.Ind.Acad.Sci.(Math.Sci)* **91** (1982) $53 - 57$.

[13] E.D.Schell, "Distribution of a product by several properties", in *Proceedings of the 2nd Symposium in Linear programming, DCS/Comptroller H.Q.U.S.A.F., Washington D.C.* (1955).

[14] C.R. Seshan and V.G. Tikekar, "On Sharma-Swarup algorithm for time minimizing transportation problems", *Proc.Ind.Acad.Sci.(Math.Sci)* **89** (1980) $101 - 102$.

[15] J.K. Sharma and K.Swarup, "Time minimizing transportation problem", *Proc.Ind.Acad.Sci.(Math.Sci)* **86** (1977) $513 - 518$.

[16] W.Szwarc, "Some remarks on the time minimizing transportation problem", *Nav.Res.Log.Quart.* **18** (1971) $473 - 485$.

LAKSHMISREE BANDOPADHYAYA, RETIRED ASSOCIATE PROFESSOR OF MATHS FROM DESHBANDHU COLLEGE, PRESENTLY GUEST LECTURE AT LADY SHRI RAM COLLEGE FOR WOMEN
lakshmisree1950@gmail.com

Typed by SHUBHRA AGGARWAL, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
aggarwal.shubhra@yahoo.co.in

Understanding Higher Dimensions
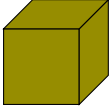
Devika Sharma

Abstract

This paper aims at understanding the popular mathematical concept of dimension of an object. It explores and builds on the intuitive idea of 'dimension'. We use this to improve our visualization of the $4D$ cube that featured in the recent sci-fi titled Interstellar.

1   Introduction

Often, in our casual conversations, we use the words $2D$ or $3D$ when describing an object. But what do we understand from this reference? In fact, now we even talk of objects that live in dimensions higher than 3! This has been particularly popularized by the use of the term 'spacetime' by physicists when talking about gravity, black holes or the big bang. Have you ever asked yourself or explained to someone what makes a $2D$ film different from a $3D$ one? What changes in your vision and perception when you put on the special glasses to watch a $3D$ film? Even though the screen that projects the film continues to be the same.

It would surprise you to observe that when we are shown an object and asked about its dimension, we instantly blurt out the correct answer but stumble when someone asks us to logically justify it. Let us begin by answering a few simple warm up questions to measure our intuitive understanding of the concept - dimension of an object. We will then use math as a language to extract the precise definition of our intuitive understanding.

What do you think is the dimension of our world? The first answer, without much conscious thought is 3. And it is, in fact, the correct answer. What do you think is the dimension of a line, a square, or a cube? And the answer is 1, 2 and 3, respectively. Why? How are we deciding this? If you think for a moment, you would agree that we are roughly defining the dimension (of the object) as the degree of freedom of movement. If this is the explanation we abide by, what does it tell us about the dimension of the dot? This information has been put together in the table below. The first column states the object of discussion, the second column gives its dimension and the third column gives a corresponding visual.

Objects	Dimension	
Our world	3	
A dot	0	
A line	1	
A square	2	
A cube	3	

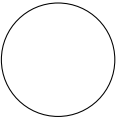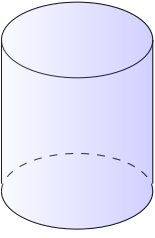| Objects	Dimension	Degrees of freedom in the ambient space
Our world	3	left $\longleftrightarrow$ right (length) up $\longleftrightarrow$ down (breadth) backward $\longleftrightarrow$ forward (width/depth)
A dot	0	None
A line	1	left $\longleftrightarrow$ right (length)
A square	2	left $\longleftrightarrow$ right (length) up $\longleftrightarrow$ down (breadth)
A cube	3	left $\longleftrightarrow$ right (length) up $\longleftrightarrow$ down (breadth) backward $\longleftrightarrow$ forward (width/depth)

What do we mean by the degree of freedom of movement? When we say that our world is 3-dimensional, we understand the movements as shifting left to right, moving forward and backward and jumping up and down. This gives us 3 independent degrees of movement. This is because our world sits in $\mathbb{R}^3$. A square lives in $\mathbb{R}^2$ and therefore one can either move left to right or jump up and down. In a line that lives in $\mathbb{R}$ you can only move left to right. But you can do none of this in a dot. You're stuck to just one point! This is why the dot is 0-dimensional. The left to right movement is called the *length*, the jumping up and down is called the *breadth* or the *height*, while the forward-backward movement is called the *width* or the *depth*. This information is summarised in the table above. This brings out the mathematical definition of the *dimension of an object*.
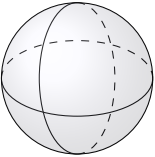
2 Definition

Definition 1. *An object X has dimension $n \geq 0$ if n is the smallest integer such that X*

*can be embedded in $\mathbb{R}^n$.*

If we agree that the above is a precise definition of our intuitive understanding of the dimension, then how do we explain the following examples? A circle that lives in the two dimensional space $\mathbb{R}^2$ but looks like a line looped back to itself. Thus, even though $\mathbb{R}^2$ has two independent degrees of freedom, we are only allowed to move along the clockwise or the anti-clockwise direction. From our first definition, it should be a 2-dimensional object while it is, in fact, just a 1 dimensional object. See the other two 'misleading' examples of the cylinder and the sphere in the table below. These exercises make us realize that counting the degrees of freedom of movement *in the object* gives us a more accurate answer to the dimension of an object. It roughly means that you look around to see how your neighbourhood looks!

Object	Visual	Ambient space	DoF	Dimension
Circle		$\mathbb{R}^2$	clockwise $\longleftrightarrow$ anti-clockwise	1
Cylinder		$\mathbb{R}^3$	clockwise $\longleftrightarrow$ anti-clockwise up $\longleftrightarrow$ down	2

| Object	Visual	Ambient space	DoF	Dimension
Sphere		$\mathbb{R}^3$	Latitudinal Longitudinal	2

Definition 2. *We say that an object (manifold) X is an n-dimensional (topological) space if every point $x \in X$ has a neighbourhood (open set) U that looks like the Euclidean space $\mathbb{R}^n$, for some integer $n \geq 0$.*

Let's check the precision of this definition with a few more examples. What is the dimension of a torus? A torus is the shape of the doughnut and can be constructed from a rectangular strip by patching it up in a certain way, as shown in the figure below.
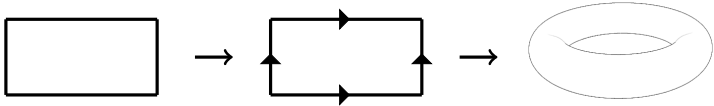

A priori, it sits in $\mathbb{R}^3$ but every point has a neighbourhood that looks like an open set on the rectangle. Hence, the torus is 2-dimensional. How about the the Möbius strip, that is

constructed as follows?
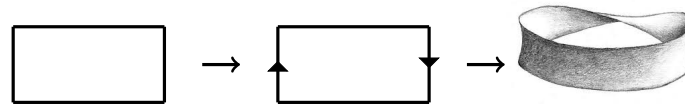
Even though, this looks like a slightly complex object, the fact that it is a 2-dimensional object follows from the same logic applied to the case of the cylinder or the torus above. An interesting difference between a cylinder and a Möbius strip is the way the latter connects the front and the back of the rectangular strip it is made from. Try drawing a central line on the Möbius strip and run along till it meets the starting point to complete a circle. Deconstruct the strip to its original rectangular form. You will see a line drawn in the middle of the rectangular strip on both sides. This is a property of the Möbius strip that makes it a preferable shape over the cylinder for practical purposes. It gives us double the surface area in the same amount of material used. This was the shape of the tape running through the tape recorder or VCR cassettes.

Another interesting object is the Klein's bottle. It's not possible to embed the Klein's bottle in the three dimensional world. The $3D$ depiction of the Klein's bottle makes it look like a self intersecting object, while it is not. See the picture of the dissected Klein's bottle below. What is the dimension of the Klein's bottle?
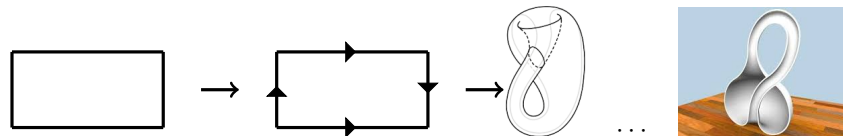
Thus the definition of the dimension of an object we arrived at is precise. This gives a convincing reason to understand the importance of math in our lives now. Math is a language used to precisely and concisely write down what we mean. It is a language that is free of bias. Solutions come from fearless fantasies. A guess and use of a mesh of logic and pre-existing results to back up your belief is all that is required. The world then scrutinises your argument before accepting it as the universal truth.

How do you think our world would look if we were to add another dimension to it? Physicists have already done it and they call it 'spacetime'. But how do we wrap our head around their claim that our world lives in a 4-dimensional world? Since we can't imagine the fourth dimension right away, we shrink our world to a 2-dimensional world and understand our perception of 3-dimensional beings.

3   Flatland - A Romance of Many Dimension

A cue comes from a book written by an English school master Edwin Abbott in 1884 titled 'Flatland - A romance of many dimension'. It is amusing that this was a satirical novel written to comment on the hierarchy of the Victorian culture and the author did not intend it to be an examination of the mathematical concept of dimension.

Suppose we were flat landers. By this I mean, imagine we lived in a flat two dimensional board, and I was a square, as shown in the picture below. My movements and vision were both restricted to the flat plane I lived in. Other beings in my world were squares, circles,

triangles, lines and dots. I could tell each of them apart by moving around their contour and analyzing their bumps and bents. In the language of mathematics, I was familiar with objects who were zero, one or two dimensional.

One day I had a strange visitor. He came with a characteristic that I hadn't been exposed to in my life so far in the flatland. When I first saw him he was a circle. When I blinked my eye (!), he was a smaller circle and then a dot. From the dot he became a circle and grew to a certain size before shrinking back to a dot. See the figure below. He was constantly changing.
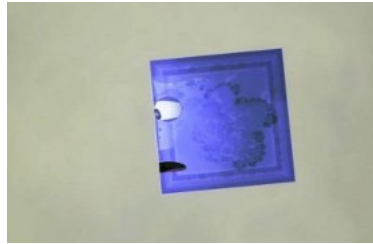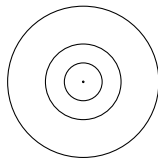
I asked circles of various radii in the flatland if the visitor was known to them. I asked the dot if he knew the visitor as a relative of his. But no one knew him. So I asked the visitor to reveal himself. The visitor said he was a sphere from a 3-dimensional world and was bouncing through our flatland! It is interesting to imagine what would the square see if objects like a disc, a prism, a vertical triangle or a Möbius band were to pass through his world. Continuing my conversation with the visitor, I asked him to describe his world in a language that I would understand. The visitor said if I was capable of looking up from my world I would see all of him at once. He said his three dimensional world was made up of an infinite stack of flatlands. At each point, in the upward and the downward direction from my flatland, if I was to attach a copy of the flatland, I'd be able to construct his world. This is how we imagine the fourth dimension. If we were to stick to Einstein's description of time as the fourth dimension, then the new world is nothing but our three dimensional world at each point in time.
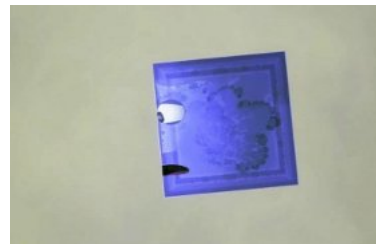
4   Conclusion

Why do we need to think of the fourth or higher dimensions? It appears in Einstein's theory of general relativity where he asserts that it is more natural to think of our world sitting in a 4-dimensional space called 'spacetime' if we want to understand a phenomena like gravity. The existence of higher dimensions must also be presumed in order to argue about these concepts rigorously. I conclude with the last interesting example of the *Tesseract*, the 4-dimensional cube.

Just as a cube is to a square, the tesseract is to the cube. The hyper surface of the tesseract is made of cubes, much similar to how each face of a cube is a square. For a visual experience, watch the following scene of the tesseract from the recent Hollywood film - Interstellar on the link:

https://www.youtube.com/watch?v=iJio07EtKYc

DEVIKA SHARMA, POST DOCTORAL FELLOW, WEIZMANN INSTITUTE OF SCIENCE, ISRAREL
devikasharma64@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.

Singular Value Decomposition in Image Processing

Esha Saha

Abstract

A matrix A can be decomposed in several ways such as decomposing it into lower and upper triangular matrices, popularly known as LU Decomposition. Singular Value Decomposition or SVD is one such decomposition of a real or complex matrix that has a wide application in data analysis. With the increase in demand for space for storing data, it is important to know how to reduce the size of the data without copromising with its quality significantly. SVD gives a simple yet effective method to do this which will be explored in the paper.

1 Singular Value Decomposition

The matrix multiplication of $Ax = y$ is simply rotation and stretching of the vector x to get vector y, the rotation and dilation matrix being given by A_r (A_r being unitary i.e. $A_r^{-1} = A^T$) and A_s respectively where, $A_r = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and $A_s = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix}$. A Singular Value Decomposition or SVD is the factorization of a real or complex matrix. To get SVD of a matrix A consider orthogonal unit vectors $v_1, v_2, ..., v_n$. Suppose matrix multiplication of $v_i s, i = 1, 2, ..., n$ with a matrix A yields $\sigma_1 u_1, \sigma_2 u_2, ..., \sigma_n u_n$ where $u_i s, i = 1, 2, ..., n$ are orthogonal unit vectors and $\sigma_i s, i = 1, 2, ..., n$ are the stretching parameters. Then the above procedure can be expressed as,

$$Av_j = \sigma_j u_j, \ 1 \le j \le n$$

These σ_is are known as the singular values and they are ordered i.e. $\sigma_1 \ge \sigma_2 \ge .... \ge \sigma_n > 0$. Then let,

$$V = \begin{bmatrix} v_1 & v_2 & ... & v_d \end{bmatrix}; V \text{ is } d \times d \text{ unitary matrix}$$

$$U = \begin{bmatrix} u_1 & u_2 & ... & u_n \end{bmatrix}; U \text{ is } n \times n \text{ unitary matrix}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & ... & 0 \\ 0 & \sigma_2 & ... & 0 \\ \vdots & & & \\ 0 & 0 & ... & \sigma_n \end{bmatrix}; \Sigma \text{ is } n \times d \text{ diagonal matrix}$$
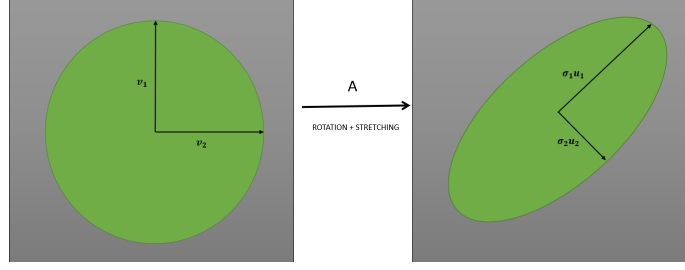
Figure 1: Geometrically representing SVD

$$\Rightarrow AV = U\Sigma$$
$$\Rightarrow AVV^* = U\Sigma V^* where V^* = V^{-1}$$
$$\Rightarrow A = U\Sigma V^*$$

Then we get the SVD of matrix A where A is a $n \times d$ matrix and V^* is the conjugate transpose of V. Since this paper aims to see the application of SVD in image processing, what SVD does here is finds orthogonal basis elements that minimizes the variance. So in a plane, it is a line fit and a line parallel to that(see Figure 1). What makes SVD extremely popular is that no matter how complex a matrix is, its SVD will exist.

2 Rank Reduction and Image Processing

Now that the decomposition of A is known, it is possible to reduce the rank of the matrix easily. The steps to reduce the rank of $A = U\Sigma V^*$ is given below:

Step 1: Keep only the first k rows of V^*.

Step 2: Take only the top k singular values. Note that since the singular values are ordered, all the significant ones will come at the top and hence taken into consideration.

Step 3: Keeping the first k columns of U compute $A_k = U_k\Sigma_k V_k^*$, thus getting a low rank approximation for A.

Coming to the application in image processing, the procedure to reduce the size of an image will be explored. A $m \times n$ pixels image can be represented as a $m \times n$ matrix. Each pixel has some level of black and white color given by some integer between 0 (black) and 255 (white). Each of these integers requires approximately 1 byte to store. Since a colour image has three components: red, green and blue; it requires more space to store. That is where SVD steps in. While SVD refers to the operation defined on matrices, Principal Component Analysis or PCA refers to the data analysis technique that is used. As a result of PCA, a reduced matrix can be obtained whose elements store the corresponding pixels of the compressed image. Thus the memory requires in storing is reduced without significantly compromising with the quality. Since images lead to large matrices, softwares such as MATLAB, Python, etc are used to perform PCA. One such demonstration has been done in following section.

3 Programming to Compress Image

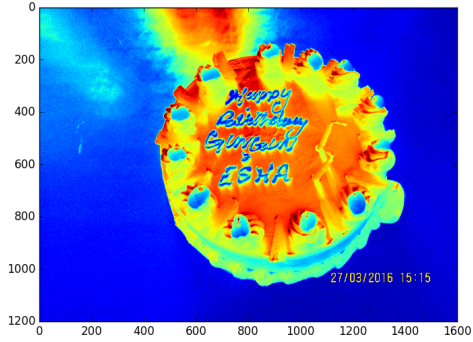Consider the image of the author's birthday cake (1200× 1600 pixels) given below in figure 2.
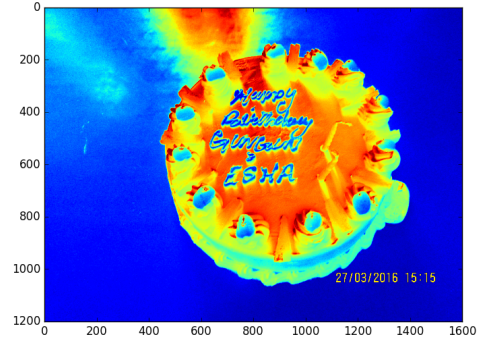

Figure 2: Original image (426 kb)

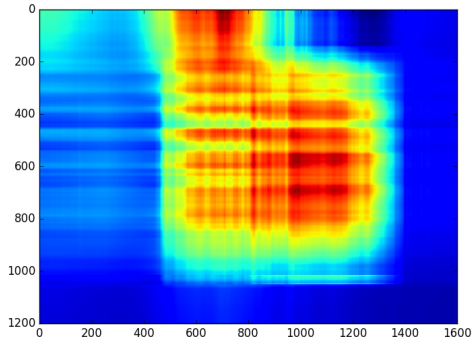Figure 3: SVD without rank reduction

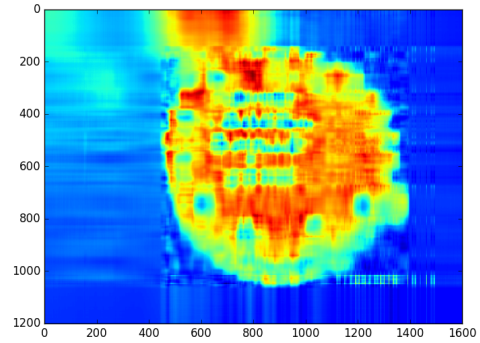Figure 4: Using 2 modes (222kb)

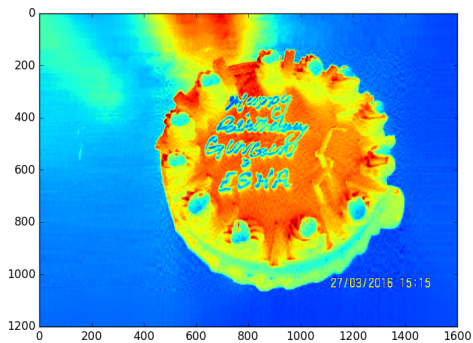Figure 5: Using 10 modes (356kb)
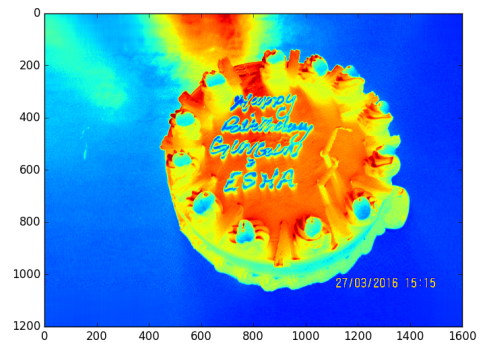

Figure 6: Using 70 modes (366kb)

Figure 7: Using 150 modes (389kb)

4 Interpretation

From the original figure, a SVD reconstruction without rank reduction gives the obvious result that is depicted in Figure 3 i.e. no reduction in file size or any change in the picture

quality. Considering only the first two singular values reduces both the quality and size to great extent (Figure 4). In figure 5 and 6, note an increase in both the size as well as the picture quality as modes increase to ten and seventy respectively. Using one - fifty modes, a reduced file size with an image quality almost as good as the original is achieved (Figure 7). Thus, SVD has successfully compressed the image.

4.1 Why only the first few modes are significant?

The plot in Figure 8 shows the extent to which each mode/ singular value is affecting the image quality. It clearly indicates that only the first 100-200 modes are affecting the image. That is why when only two or ten modes are taken, the image quality is extremely poor since many significant modes are getting neglected. However, with one - fifty modes the insignificant modes get neglected and hence do not affect the quality of the image.
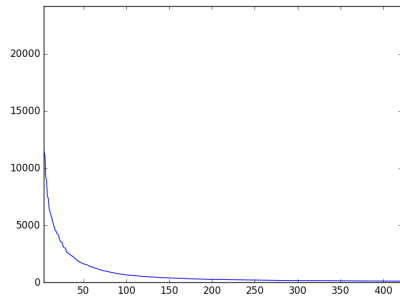
Figure 8
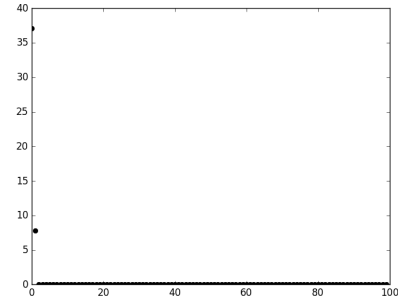
Figure 9

A similar example shows the reconstruction of a function using only two modes. Figure 9 indicates that only the first two modes are significant, thus the ones after that can be neglected. Figure 10 and 11 is the actual functional plot and reconstruction respectively.

Figure 10

Figure 11

5   Conclusion

Hence SVD is an extremely powerful tool that has wide applications. Apart from image procesing, it is also used in signal processing, solving system of linear equations, etc.

Python Code

SVD Reconstruction of image

```
1  import numpy as np
2  import matplotlib
3  from matplotlib import pyplot as py
4  from PIL import Image
5  im = Image.open("pr1.jpg")
6  im = im.convert("L")
7  mi=np.asarray(Image.open('pr1.jpg').convert('L'))
8  print(mi)
9  print(mi.shape)
10 u,s,v=np.linalg.svd(mi)
11 u1=np.array(u)
12 print(u1.shape)
13 i,j=np.indices(mi.shape)
14 s1=np.zeros([1200,1600])
15 s1[i==j]=s
16 print(s1.shape)
17 v1=np.array(v)
18 print(v1.shape)
19 A1=np.matmul(s1,v1)
20 print(A1.shape)
21 A2=np.matmul(u1,A1)
22 print(A2)
23 py.imshow(mi)
24 py.show()
25 py.plot(s)
26 py.show()
27 ##reducing the rank
28 u2=np.array(u1[:,:])
29 s2=np.array(s1[:,0:150])
30 v2=np.array(v1[0:150,:])
31 R1=np.matmul(s2,v2)
32 R2=np.matmul(u2,R1)
33 print(R2)
34 print(R2.shape)
35 py.imshow(R2)
36 py.show()
```
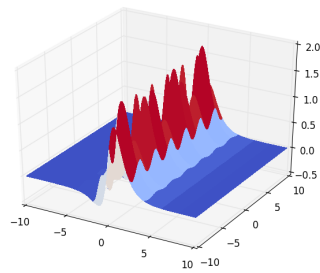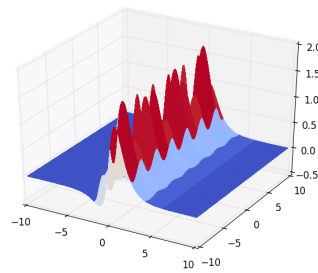
SVD Reconstruction of function

```
1  import numpy as np
2  from mpl_toolkits.mplot3d import Axes3D
3  import matplotlib.pyplot as plt
4  from matplotlib import cm
5  fig = plt.figure()
6  ax = fig.gca(projection='3d')
7  x= np.linspace( 10 ,10,100)
8  t=np.linspace( 10 ,10,100)
9  [T,X]=np.meshgrid(t,x)
10 Z=1/np.cosh(X)
11 P=Z*(1   0 .5*np.cos(2*T))+(Z*np.tanh(X))*(1   0 .5*np.sin(2*T));
12 # Plot the surface
13 ax.plot_surface(X,T,P, cmap=cm.coolwarm, linewidth=0, antialiased=False)
14 plt.show()
15 U, s, V = np.linalg.svd(P, full_matrices=False)
16 U1=np.array(U[:,0:2])
17 i,j=np.indices(P.shape)
```

```
18  s1=np.zeros([100,100])
19  s1[i==j]=s
20  s2=np.array(s1[0:2,0:2])
21  V1=np.array(V[0:2,:])
22  A1=np.matmul(s2,V1)
23  A2=np.matmul(U1,A1)
24  ax.plot_surface(X,T,A2, cmap=cm.coolwarm,
25  linewidth=0, antialiased=False)
26  plt.show()
27  plt.plot(s,'ko')
28  plt.show()
29  plt.plot(s[0:10],'ko')
30  plt.show()
```

References

[1] CS168: The Modern Algorithmic Toolbox Lecture 9: The Singular Value Decomposition (SVD) and Low-Rank Matrix Approximations, http://theory.stanford.edu/ tim/s15/l/l9.pdf

[2] Image Compression and Linear Algebra by Sunny Verma, Jakkam Phanindra Krishna, http://www.cmi.ac.in/ ksutar/NLA2013/imagecompression.pdf

[3] Using the Singular Value Decomposition for Image Compression, http://www.math.ucsd.edu/ jlobue/102/svd.pdf

[4] An Application of Linear Algebra to Image Compression Paul Dostert, http://math.arizona.edu/ brio/VIGRE/ThursdayTalk.pdf

[5] https://en.wikipedia.org/wiki/Principal_component_analysis

[6] https://docs.scipy.org/doc/numpy-1.10.0/index.html

[7] https://github.com/Dirivian/Jupyter_notebooks

ESHA SAHA, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
esha2703@gmail.com

Ruin Theory

Snigdha Jain

Abstract

This article gives an overview of ruin theory. It briefly discusses the probability models for the surplus process of an insurance portfolio and the mathematical tools used for the investigation of these models.

1   Introduction

Ruin theory uses mathematical models to describe an insurer's vulnerability to insolvency or ruin. In such models key quantities of interest are the probability of ruin, distribution of surplus immediately prior to ruin and deficit at time of ruin. A high probability of ruin indicates instability: measures such as reinsurance or raising some premiums should be considered, or the insurer should attract extra working capital. The probability of ruin enables one to compare portfolios with each other, but we cannot attach any absolute meaning to the probability of ruin, as it doesn't actually represent the probability that the insurer will go bankrupt in the near future. The ruin probability only accounts for the insurance risk, not the managerial blunders that might occur.

2   Surplus Process

Suppose at time 0 the insurer has an amount of money set aside for his portfolio. This amount of money is called the initial surplus (denoted by $U \geq 0$) which is needed because the future premium income on its own may not be sufficient to cover future claims (ignoring the expenses). The insurer's surplus at any future time $t(> 0)$ denoted by $U(t)$ is a random variable since its value depends on the claims experience up to time t.

$$U(t) = U + ct - S(t); S(t) : Expense \tag{1}$$

From Figure 1 we can see that claims occur at time T_1, T_2, T_3, T_4 and T_5 and at these times the surplus immediately falls by the amount of claim. Between claims the surplus increases at constant rate "c" per unit time. Insurer's surplus falls below 0 as a result of the claim at time T_3, thus showing that the insurer has run out of money and it is said that RUIN has occurred. The insurer will want to keep the probability of ruin as small as possible, or at least below a predetermined bound.

To be more precise, the following two probabilities are defined as:

$$\psi(U) = P[U(\tau) < 0, \text{for some } t, \, 0 < t < \infty], \tag{2}$$

$$\psi(U, t) = P[U(\tau) < 0, \text{for some } \tau, \, 0 < \tau], \tag{3}$$
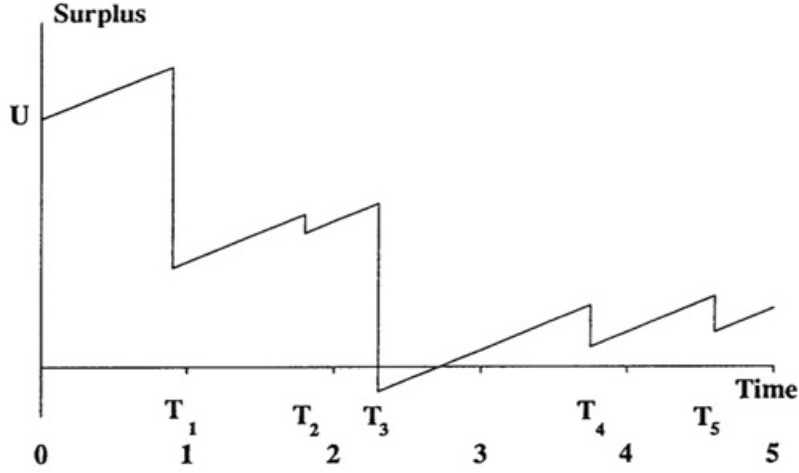
Figure 1

$\psi(U)$ is the probability of ultimate ruin whereas $\psi(U,t)$ is the probability of ruin within time t. Some obvious relationships are:

- The larger the initial surplus the lesser the probability that the ruin will occur may it be a finite or an infinite period of time.

- For the given initial surplus U, the longer the time period considered to check the ruin the higher will be the probability of ruin.

- The probability of the ultimate ruin can be approximated by the probability of ruin within finite time t provided t is sufficiently large.

3 Probability Of Ruin in Discrete Time

It is also possible to determine probability of ruin at discrete interval of time h in which following two discrete time probabilities of ruin are defined:

$$\psi_h(U) = P[U(t) < 0, \text{for some } t,\, t = h, 2h, 3h, ..] \qquad (4)$$

$$\psi_h(U,t) = P[U(\tau) < 0, \text{for some } \tau = h, 2h, ...t - h, t] \qquad (5)$$

Listed below are relationships between different discrete time probabilities of ruin for $0 \leq U_1 \leq U_2$ and for $0 \leq t_1 \leq t_2 \leq \infty$

- $\psi_h(U_2, t) \leq \psi_h(U_1, t)$

- $\psi_h(U_2) \leq \psi_h(U_1)$

- $\psi_h(U, t_1) \leq \psi_h(U, t_2) \leq \psi_h(U)$

46

4   Compound Poisson Process

Claim No. process: $N(t)_{t \geq 0}$

Claim amounts: $X_{i,\ 0 < i < \infty}$

Claim No. process is assumed to be a poisson process and aggregate claims are assumed to be compound poisson process $S(t)_{t \geq 0}$ The following three assumptions are made:

- The random variable $X_{i,0 < i < \infty}$ are independent and identically distributed.

- The random variable $X_{i,0 < i < \infty}$ are independent of (N(t)) for all t.

- The stochastic process $N(t)_{t \geq 0}$ is a poisson process whose parameter is denoted by λ.

For a compound Poisson process $S(t)$, the mean, the variance and the Moment Generating Function (MGF) of the process are given by:

$$E[S(t)] = \lambda t E(X) \tag{6}$$

$$Var[S(t)] = \lambda t E(X^2) \tag{7}$$

$$M_{S(t)}(r) = e^{\lambda t [M_X(r) - 1]} \tag{8}$$

5   Premium Security Loadings

The security loading is the percentage by which the rate of premium income exceeds the rate of claims outgo. So for the poisson process it is,

$$C = (1 + \theta)E(S) = (1 + \theta)\lambda m_1 \tag{9}$$

The insurer will need to adopt a positive security loading when pricing policies in order to cover expenses, profit, contingency margins and so on.

5.1   Technicality

A technical result is used concerning $M_x(r)$ (MGF of the individual claim amount distribution). It will be assumed through that there is some number γ $(0 < \gamma)$ such that $M_x(r)$ is finite for all $r < 0$ and:

$$lim_{r \to \infty} M_x(r) = \infty \tag{10}$$

6   Lundberg's Inequality

Lundberg's inequality states that:

$$\psi(U) \leq e^{-RU} \tag{11}$$

Where U is the insurer's initial surplus and $\psi(U)$ is the probability of ultimate ruin. R is a parameter associated with a surplus process known as the adjustment coefficient. Its value of R depends upon the distribution of aggregate claims and on the rate of premium income. Figure 2 shows the graph of both e^{-RU} and $\psi(U)$ against U when claim amounts are exponentially distributed with mean 1 and when the premium loading factor is 10%. It can be seen that, for large values of $U, \psi(U)$ is very close to the upper bound, so that $\psi(U) \approx e^{-RU}$.
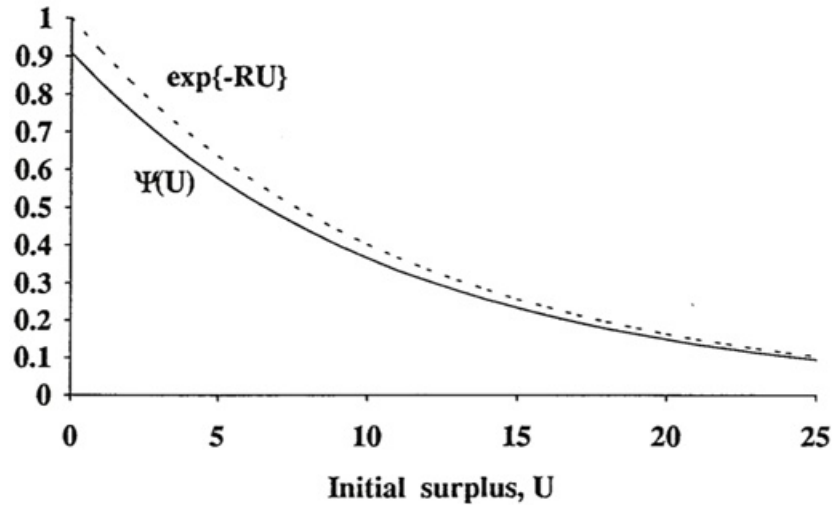
Figure 2

R can be interpreted as measuring risk. The larger the value of R, the smaller the upper bound for $\psi(U)$ will be. Hence, $\psi(U)$ would be expected to decrease as R increases. Basically, R is an inverse measure of risk. Larger the R smaller the ruin probability.
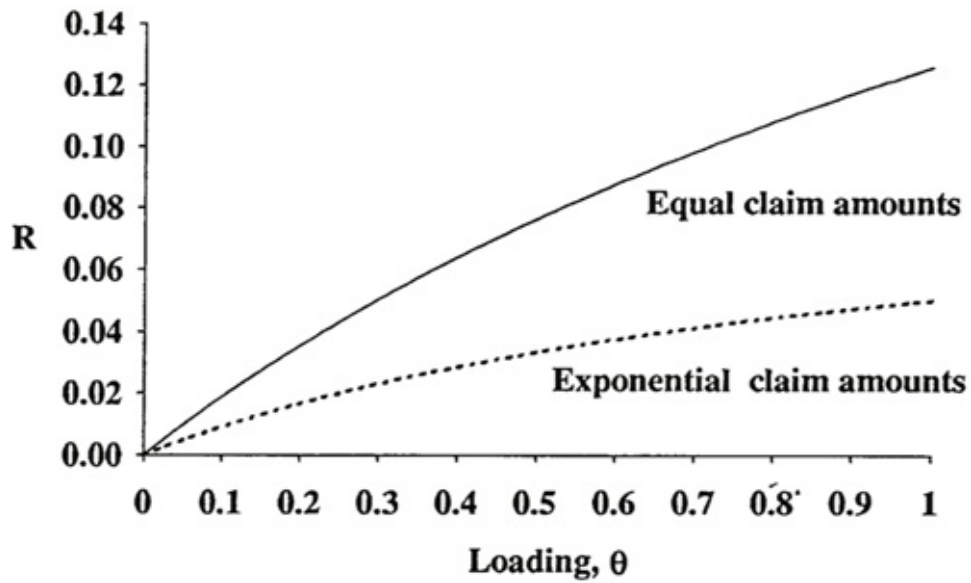
Figure 3

Figure 3 shows the graph of R as a function of loading factor θ when the claim amounts distribution is exponential with mean 10 and all claims are of amount 10. In both cases R is an increasing function of θ which justifies the relation $\psi(U) \approx e^{-RU}$.

48

7 Adjustment Coefficient - Compound Poison Process

The adjustment coefficient is a parameter associated with a surplus process which takes account of two of these factors:

- Aggregate claims

- Premium income

The adjustment coefficient is defined in terms of the Poisson parameter, the moment generating function of individual claim amounts and the premium income per unit time. The adjustment coefficient, denoted by R, is defined to be the unique positive root of:

$$\lambda M_X(r) - \lambda - cr = 0 \tag{12}$$

The value of the adjustment coefficient depends on the Poisson parameter, the individual claim amount distribution and the rate of premium income. However writing
$c = (1 + \theta)\lambda m_1$ gives,

$$M_X(r) = 1 + (1 + \theta)m_1 r \tag{13}$$

So, R is independent of the Poisson parameter and simply depends on the loading factor, θ, and the individual claim amount distribution. R does not depend on the Poisson parameter, the basic reason is that increasing the Poisson parameter speeds up the whole process, so that claims arise more quickly. This means that ruin, if it is going to happen, will happen sooner, rather than later. However it does not affect the probability that ruin does actually occur, when we are considering ruin at any time in the future.

8 Reinsurance And Ruin

One of the options open to an insurer who wishes to reduce the variability of aggregate claims from a risk is to effect reinsurance. A reduction in variability would be expected to increase an insurer's security, and hence reduce the probability of ruin. A reinsurance arrangement could be considered optimal if it minimises the probability of ruin. If a reinsurance arrangement can be found that maximises the value of the adjustment coefficient, the upper bound for the probability of ultimate ruin will be minimised. As the adjustment coefficient is a measure of risk, it seems a reasonable objective to maximise its value. Here, the effect on the adjustment coefficient of proportional and excess of loss reinsurance arrangements will be considered.

8.1 Maximising the adjustment coefficient under Proportional Reinsurance

Proportional Reinsurance means that the Primary insurer and Reinsurer share liabilities (i.e. sums insured) in a clearly defined proportion as described within the underlying treaty. Premiums and claims are also split up according to the respective share of the risk (i.e. proportionally). For example- If a reinsurer agrees to pay 40% of the claims, further if a claim of Rs. 1000 arises, 40% of Rs 1000, i.e., Rs. 400 will be paid by the reinsurer while

the rest, i.e., Rs. 600 will be paid by the insurer. The insurer's premium income per unit time, before payment of the reinsurance premium, is

$$(1+\theta)\lambda m_1. \tag{14}$$

The reinsurance premium is assumed as $(1+\xi)(1-\alpha)\lambda m_1$. The insurer's premium income, net of reinsurance, is $[(1+\theta)-(1+\xi)(1-\alpha)]\lambda m_1$. We assume $\xi \geq \theta$, if this were not true, it would be impossible for the insurer to pass the entire risk on to the reinsurer and to make a certain profit. For the insurer's premium income, net of reinsurance, to be positive:

$$(1+\theta) > (1+\xi)(1-\alpha) \tag{15}$$

$$\alpha > (\xi-\theta)/(1+\xi) \tag{16}$$

The insurer's net of reinsurance premium income per unit time must exceed the expected aggregate claims per unit time. Otherwise ultimate ruin is certain, i.e.,

$$[(1+\theta)-(1+\xi)(1-\alpha)]\lambda m_1 > \alpha\lambda m_1 \tag{17}$$

$$\alpha > (\xi-\theta)/\xi \tag{18}$$

8.2 Maximizing the adjustment coefficient under Excess of Loss Reinsurance

Excess of loss reinsurance is a type of reinsurance in which the reinsurer indemnifies the ceding company for losses that exceed a specified limit. Excess of loss reinsurance is a form of non-proportional reinsurance. For example- If the reinsurer sets the limit as Rs. 5000, then any claims arising with the value below Rs. 5000 will be fully paid by the insurer while as soon as the claim amount exceeds Rs. 5000, the remaining amount will be paid by the reinsurer irrespective of how large the claim amount might be. We consider the excess of loss reinsurance on the adjustment coefficient. We assume that:

- Insurer's premium income per unit time (before reinsurance) is $(1+\theta)\lambda m_1$

- Reinsurer's premium income per unit time is
  $(1+\xi)\lambda E(Z)(\xi \geq \theta) \text{ and } Z = max(0, X-M)$

- Insurer's premium income per unit time (net of reinsurance) is
  $C^* = (1+\theta)\lambda m_1 - (1+\xi)\lambda E(Z)$

References

[1] http://www.wikipedia.org

[2] http://www.insurance-institute.ru/content/20eb1c68.pdf

[3] http://web.sgh.waw.pl/ rlocho/Ruintheory.pdf

[4] http://www.mi.uni-koeln.de/ schmidli/vorl/Risk/vorl.pdf

[5] https://www.actuaries.org.uk

SNIGDHA JAIN, B.Sc.(H) MATHEMATICS, 6th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
snigdhajain285@gmail.com

Bayesian Game Theory

Akshita Bhat and Sanya Rastogi

Abstract

Game theory is the study of mathematical models of conflict and cooperation between intelligent rational decision-makers. Bayesian game is a game in which the players do not have complete information on the other players, they have beliefs with known probability distribution. In this paper we shall also highlight the concepts of signalling and Sheriff's dilemma.

1 Introduction

A Bayesian game can be converted into a game of complete but imperfect information under the common prior assumption. In addition to the actual players in the game, there is a special player called Nature. Nature assigns a random variable to each player which could take values of types for each player and associating probabilities or a probability mass function with those types. Harsanyi's approach to modelling a Bayesian game in such a way allows games of incomplete information to become games of imperfect information. The type of a player determines that player's payoff function. The probability associated with a type is the probability that the player, for whom the type is specified, is that type. In a Bayesian game, the incompleteness of information means that at least one player is unsure of the type of another player.

Such games are called Bayesian because of the probabilistic analysis inherent in the game. Players have initial beliefs about the type of each player and can update their beliefs according to Bayes' rule as play takes place in the game, i.e. the belief a player holds about another player's type might change on the basis of the actions they have played. The lack of information held by players and modelling of beliefs mean that such games are also used to analyse imperfect information scenarios. In a Bayesian game (where players are modelled as risk-neutral), rational players are seeking to maximise their expected payoff, given their beliefs about the other players, in the general case, where players may be risk-averse or risk-loving, the assumption is that players are expected utility-maximising.

2 Bayesian Games and Bayesian Nash Equilibrium

2.1 Bayesian Games

In Bayesian games, we represent players' uncertainties about the game being played represented as a probability distribution over a set of possible games. We make two assumptions:

1. All possible games have the same number of agents and the same strategy space for each agent; they differ only in their payoffs.

2. The beliefs of the different agents are posteriors, obtained by conditioning a common prior on individual private signals.

2.2 Bayesian Nash Equilibrium

A Bayesian Nash equilibrium is defined as a strategy profile and beliefs specified for each player about the types of the other players that maximises the expected payoff for each player given their beliefs about the other players' types and given the strategies played by the other players.

3 Perfect Bayesian equilibrium

In game theory, a Perfect Bayesian Equilibrium (PBE) is an equilibrium concept relevant for dynamic games with incomplete information. A PBE is a refinement of both Bayesian Nash equilibrium (BNE) and sub-game perfect equilibrium (SPE). A PBE has two components - strategies and beliefs:

1. The **strategy** of a player in a given information-set determines how the player acts. The action may depend on the history, similar to a sequential game.

2. The **belief** of a player determines the node in that information-set the player believes that he is playing at. It may be a probability distribution over the nodes in the information-set.

The strategies and beliefs should satisfy the following conditions:

1. **Sequential rationality**: Each strategy should be optimal in expectation, given the beliefs.

2. **Consistency**: Each belief should be updated according to the strategies and Bayes' rule, in every path of positive probability (in paths of zero probability and off-the-equilibrium paths, the beliefs can be arbitrary).

Every PBE is both a SPE and a BNE, but the opposite is not necessarily true.

4 Sheriff's Dilemma

A sheriff faces an armed suspect. Both must simultaneously decide whether to shoot the other or not. The suspect can either be of criminal type or of civilian type. The sheriff has only one type. The suspect knows its own type and the Sheriff's type, but the Sheriff does not know the suspect's type. Thus, there is incomplete information, making it a Bayesian game. There is a probability p that the suspect is a criminal, and a probability 1-p that the suspect is a civilian; both players are aware of this probability.

The sheriff would rather defend himself and shoot if the suspect shoots, or not shoot if the suspect does not, even if the suspect is a criminal. The suspect would rather shoot if he is a criminal, even if the sheriff does not shoot, but would rather not shoot if he is a civilian, even if the sheriff shoots. Thus, the payoff matrix of this Normal-form game for both players depends on the type of the suspect. We assume that payoffs are given as in Figure 1.

Figure 1: Payoff Matrix of Civilian and Criminal

If both players are rational and both know that both players are rational and everything that is known by any player is known to be known by every player, play in the game will be as follows according to perfect Bayesian equilibrium:

When type is "civilian", the dominant strategy for the suspect is to not shoot, and when type is "criminal", the dominant strategy for the suspect is to shoot; we can thus remove the alternative strictly dominated strategy. Given this, if the sheriff shoots, he will have a payoff of 0 with probability p and a payoff of -1 with probability $1 - p$, i.e. an expected payoff of $p - 1$; if the sheriff does not shoot, he will have a payoff of -2 with probability p and a payoff of 0 with probability $1 - p$, i.e. an expected payoff of $-2p$. Thus, the Sheriff will always shoot if $p - 1 > -2p$, i.e. when $p > \frac{1}{3}$.

5 Application

5.1 Bayesian Games

A major application of Bayesian games is in auctions, which are historically and currently common method of allocating scarce goods across individuals with different valuations for these goods. This corresponds to a situation of incomplete information because the valuations of different potential buyers are unknown.

Different types of auctions and terminology:

1. First price sealed bid auctions: It is similar to English auctions, but in the form of a strategic form game; all players make a single simultaneous bid and the highest one obtains the object and pays its bid.

2. Second price sealed bid auctions: It is similar to first price auctions, except that the winner pays the second highest bid.

Model of auction: A valuation structure for the bidders (i.e., private values for the case of private-value auctions). A probability distribution over the valuations available to the bidders.

Let us focus on first and second price sealed bid auctions, where bids are submitted simultaneously. Each of these two auction formats defines a static game of incomplete

information (Bayesian game) among the bidders.

Bidders are risk neutral , they are interested in maximising their expected profits. This model defines a Bayesian game of incomplete information, where the types of the players (bidders) are their valuations, and a pure strategy for a bidder is a map We will characterise the symmetric equilibrium strategies in the first and second price auctions. Once we characterise these equilibria, then we can also investigate which auction format yields a higher expected revenue to the seller at the symmetric equilibrium.

5.2   Bayesian Nash Equilibrium

1. The first application of signalling games to economic problems was Michael Spence's Education game. Only workers with high ability are able to attain a specific level of education without it being more costly than their increase in wage. In other words, the benefits of education are only greater than the costs for workers with a high level of ability, so only workers with a high ability will get an education.

2. It can be used to solve problems of traffic control. Experimental game theory can employ human subjects to test out traffic behaviour when the human drivers are responsible for the travel route choices that are made.

3. Other studied areas include cases of resource allocation in wireless networks without any central control, packet routing in wireline networks, and the traffic flow of commands in computer circuits.

6   Conclusion

In our daily life, there exist settings where a player may not have complete information regarding nature or style about his adversary. Some of these settings are bidding in auctions, political campaign of a candidate and prices charged by a firm on a particular product where bayesian game theory and bayesian nash equilibrium can be applied.

References

[1] Games of incomplete information by Jonathan Levin

[1] https://www.ocw.mit.edu.com

[2] http://www.wikipedia.org

[3] http://www.sites.duke.edu

AKSHITA BHAT, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
akshitabhat1997@gmail.com

SANYA RASTOGI, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
sanya.rastogi2007@gmail.com

Game of Life

Simran Bhatia and Rashi

Abstract

We have probably heard of cellular automata. The most celebrated example among them is John Conway's "Game of Life". They are, in general, a sort of beads game played over an infinite lattice (grid). At each moment, a lattice point may be empty or occupied by a bead of one among a number of colours. At the next moment, the state of that lattice point is renewed according to some function of the state of the nearest neighbour points. This function specifies the rules of the game. Our problem is to find rules that will cause the beads to organise themselves into generalised "surfaces" in a continuous domain. We will try to study the behaviour of related patterns around us through its applications.

1 Introduction

The Game of Life is a cellular-automaton, zero player game, developed by John Conway in 1970. The game is played on an infinite grid of square cells, and its evolution is only determined by its initial state. The fact that the rules for the Game of Life are so short and clear makes it a great project for initiating novice programmers. Since Life was invented by John Conway in 1970, it has distracted many amateur and recreational mathematicians, leading to the discovery of many interesting patterns. For example, Paul Chapman showed that Life is Turing complete, and so it is in principle possible to translate any computer into a set of initial conditions. The rules of the game are simple, and describe the evolution of the grid:-

- **Birth**: a cell that is dead at time t will be alive at time $t + 1$, if exactly 3 of its eight neighbors were alive at time t.

- **Death**: a cell can die by:

  - Overcrowding: if a cell is alive at time $t + 1$ and 4 or more of its neighbours are also alive at time t, the cell will be dead at time $t + 1$.

  - Exposure: If a live cell at time t has only 1 live neighbour or no live neighbours, it will be dead at time $t + 1$.

- Survival: a cell survives from time t to time $t + 1$ if and only if 2 or 3 of its neighbours are alive at time t.

Instead of 8 neighbours the neighbourhood is now best described by a radius r, and a cell having $(2r + 1)^2 - 1$ neighbours. The rules can be arbitrarily complex, but for the start it is sensible to consider only such rules that can be described by two intervals. They are called "birth" and "death" intervals and are determined by two values each. These values can be given explicitly as the number of neighbours or by a filling, a real number between 0 and 1.

2   Model Description

There are two main ideas required for this generalisation: First, the infinite grid of cells is replaced with an effective grid that is obtained by averaging over a disk. Second, the transition functions are replaced by a series of differential equations derived from a smooth interpolation of the rules for the discrete Game of Life.

To explain how an effective grid works, first consider what would happen if we replaced the infinite discrete grid in the game of life with a time-dependent continuous real scalar field, $f : \mathbf{R^2} \times \mathbf{R^2} \to \mathbf{R}$ on the plane. Now here is the trick: instead of thinking of this is an infinite grid of infinitesimal cells, we give the cells a small but finite length. To do this, pick any small positive real number h which will act as the radius of a single cell (i.e. the Planck length for the simulation). Then we define the state of the *effective cell* at a point x as the average of the field over a radius h neighbourhood around x, (which we call $M(x,t)$ following the conventions in Rafler's paper):

$$M(x,t) = \frac{1}{\pi h^2} \int_{0 \leq |y| \leq h} f(x-y,t)\, dy$$

Now for each cell, we also want to compute the effective number of cells in its neighbourhood. To do this, we use the same averaging process, only over a larger annulus centred at x. By analogy to the rules in the original Game of Life, a reasonable value for the outer radius is about $3h$. Again, following Rafler's conventions we call this quantity $N(x,t)$:

$$M(x,t) = \frac{1}{8\pi h^2} \int_{0 \leq |y| \leq 3h} f(x-y,t)\, dy$$

3   Computer Implementation

As simple as the theoretical model is, it is not immediately obvious, how to implement it on a computer, as a computer cannot handle infinitesimal values, continuous domains, etc. But it can handle real numbers in the form of floating point math, and as it turns out, this is sufficient. We also can model the continuous domain by a square grid, the ideal data structure for computation. So we will be able to implement our function $f(x,t)$ as a float array.

And that the next state of f is computed via the update rule:

$$f(x,t+1) = S\left(N(x,t), M(x,t)\right)$$

Where:-

$$\sigma(x,a,\alpha) = \frac{1}{1 + e^{\left(\frac{-4}{\alpha}\right)(x-a)}}$$

$$\sigma_n(n,a,b) = \sigma(n,a,\alpha_n)(1 - \sigma(n,b,\alpha_n))$$

$$\sigma_m(m,a,b) = \sigma(m,0.5,\alpha_m)(1 - \sigma(m,b,\alpha_m))$$

$$S(n,m) = \sigma_n(n, \sigma_m(m,b_1,d_1), \sigma_m(m,b_2,d_2))$$

And we have 6 (or maybe 7, depending on how you count) parameters that determine the behaviour of the automata:

- $[b_1, b_2]$: The fraction of living neighbours required for a cell to stay alive (typically $\left[\frac{2}{8}, \frac{3}{8}\right]$).

- $[d_1, d_2]$: The fraction of living neighbours required for a cell to be born (typically $\frac{3}{8}$).

- σ_m : The transition smoothness from live to dead (arbitrary, but Rafler uses $\sigma_m \approx$ 0.148).

- σ_n : Transition smoothness from interval boundary (again, arbitrary but usually about $\sigma_n \approx 0.028$).

- h : The size of the effective neighbourhood (this is a simulation dependent scale parameter, and should not effect the asymptotic behaviour of the system).

4 Generalisation to 3D

In a 3-dimensional Life, each cell has 26 neighbours as opposed to 2 and 8 for 1-dimensional and 2-dimensional Life respectively. Possible variations of $3D$ Life abound, but most yield patterns that either expand too quickly or shrink too rapidly. Loosely speaking, one may define a valid $3D$ Life as one that loosely speaking satisfies the following two properties:

1. Supports a glider gun.

2. Exhibits bounded growth

First it is useful to introduce some terminologies. Let w and x be the minimum and maximum number of adjacent cells that need to be alive to sustain a cell that is currently alive in the next stage, and y and z be the minimum and maximum number of adjacent cells that need to be alive to make a cell that is currently dead alive in the next stage respectively. Then, given any dimension k, we can represent the rules governing any variation of Life succinctly as $wxyz$. For instance, Conway's Game of Life is governed by the rule 2333.

4.1 Rule 2333

This is the Game of Life, which follows two familiar and simple rules for deciding which cells will live and die from generation to generation.
If a cell is dead at time t, it comes alive at time $t+1$ if and only if the number of living cells in the squares immediately surrounding the cell is greater than the lowest fertility threshold (fl) and less than the highest fertility cutoff (fh). If a cell is alive at time t, it dies at time $t+1$ if and only if the number of living cells in the squares immediately surrounding the cell is greater than the highest environmental support (eh, i.e., "overcrowding") or less than the lowest environmental support (el, i.e., "undernourishment").
Surprisingly complicated patterns of behaviour by these cells can emerge based on just these two simple rules. In addition, by varying the four variables: el, eh, fl, fh, we can obtain different versions of Life, each with its unique behaviour and properties. We characterise the particular version by using four octal numbers for the four variables and stringing them together. The particular version of Life popularised by John Horton Conway is also the default one shown here: *Rule* 2333. This means that a cell will die if it is surrounded by less than 2 or more than 3 living cells, and a cell will only be born if its space is surrounded by exactly 3 living cells. 2333 Life has received much attention from mathematicians and computer scientists. Many interesting patterns, such as the Glider, the Glider Gun, the Spaceship, have been discovered.

5 Applications

Cellular automata along with Game of Life yield a surprisingly diverse range of practical applications, three of which are mentioned below:-

1. Music
   Runxt Life is a $3.99 generative music application created for the iOS platform, based on the cellular automaton theory "Conway's Game of Life". To use the sequencer, you draw single cells or a predefined pattern stamp on a two dimensional grid. This pattern will evolve according to the rules of Conway's Game of Life after you press play or step forward. Five different colours each represent a different sound or channel. Runxt Life uses the output of the grid, affected by a few parameters, to either generate sounds or send MIDI notes to your computer with the help of the Runxt Life OSC to MIDI application. In this application you can configure which colour is sent to which MIDI channel.
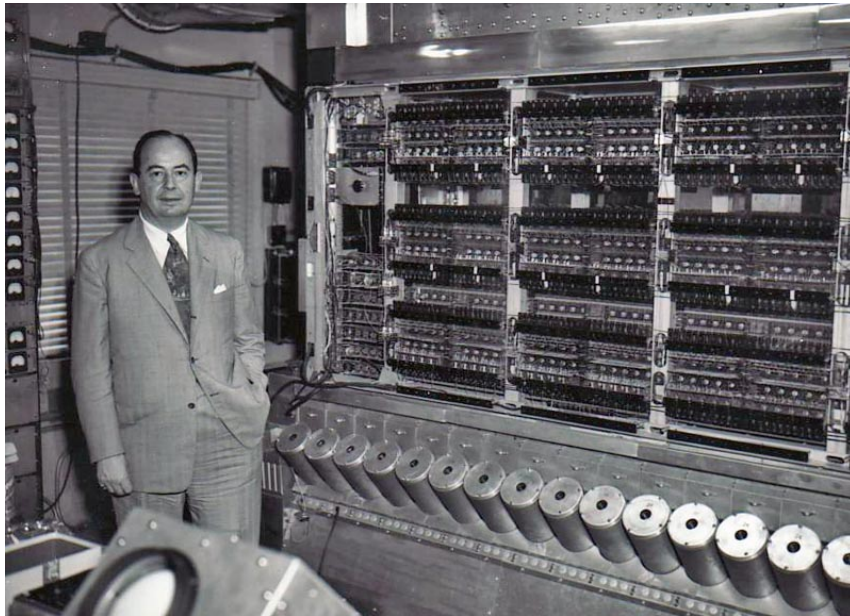

Figure 1: Turing Machine

2. Turing Machine
   A Turing Machine as in Figure 1 is a mathematical device invented by Alan Turing which has a potentially infinite tape to hold the input data and to store the results. It was determined that cellular automata, as they appear in the Game of Life, have the same computational capacity as Turing machines. The Church-Turing thesis that states:
   "No method of computation carried out by a mechanical process can be more powerful than a Turing machine."
   Therefore, as the Game of Life is Turing complete, it is one of the most powerful models of computation. In other words, no mechanical form of computation can solve

a problem that a Turing machine or cellular automata cannot solve, given sufficient time and space.

3. Public Key Cryptography
A public-key cryptosystem is a cryptographic protocol that relies on two keys an enciphering key E, which is made public, and a deciphering key D, which is kept private. Cellular automaton rules that are invertible are prime candidates for the invertible function we require to construct a public-key cryptosystem. The crucial factor that enables such a cellular automaton cryptosystem to work is the fact that it is extremely computationally expensive to unravel the original state from the cipher state without prior knowledge of the individual rules used in the composite enciphering function. This guarantees the security of the system. All the sender has to do is to apply the secret decryption key D to her name and then include that coded signature at the end of her message. This way, when the receiver receives the message, he can simply use the public key E to decipher the coded signature and see if the name matches the presumed sender.

6   Conclusion

We have described a model to generalise Conway's "Game of Life" to a continuous range of values and a continuous domain. The 8 pixel neighbourhood and 1 pixel cell of Game of Life have been generalised to a ring shaped neighbourhood and a disk shaped cell. The rule set has been generalised to four real numbers: the boundaries of the birth and death intervals. The goal of finding a glider that can move in arbitrary directions has been achieved. Of the original Game of Life it resembles both the glider and the spaceship at the same time. So we think we have found the generic, generalised glider, and call it the "smooth glider". Despite the hurdles above we have yet to think of a reason why life cannot occur within the Game of Life. Searching for artificial life with processors may be like searching for intelligent life on other planets with radio telescopes; it seems likely that it exists somewhere but finding it in the vast expanse of space, or vast numbers of possible patterns within a grid, may be impossible.

References

[1] "Continuous Spatial Automata", Bruce J.MacLennan,1990

[2] Game of Life Status page, Jason Summers, retrieved 2012-02-23

[3] "Life Universal Computer" ,Paul Chapman (11 November 2002)

[4] Gardner, M. "The Game of Life, Parts I-III." Chs. 20-22 in Wheels, Life, and other Mathematical Amusements. New York: W. H. Freeman, 1983.

[5] Generalization of Conway's "Game of Life" to a continuous domain ??? SmoothLife, Stephan Rafler Nurnberg, Germany.

[6] Bays, Carter. "Candidates for the Game of Life in Three Dimensions." Complex Systems1 (1987): 373-400. Web. 9 Sept. 2015.

[7] Kun, Jeremy. "Turing Machines and Conway's Dreams." Web log post. Math ??? Programming. N.p., 30 June 2011. Web. 09 Sept. 2015.

[8] Rendell, Paul. Turing Machine Universality of the Game of Life. Thesis. University of the West of England, 2014. Print.

[9] Dynamics of Complex Systems", Yaneer Bar-Yam, 1997, (section 7.2: pattern formation)

[10] Glider Dynamics on the Sphere Exploring Cellular Automata on Geodesic Grids, Author: Jeffrey Ventrella (Ventrella.com), published in the Journal of Cellular Automata, Volume 6 Number 1, 2011

[11] http://www.math.com/students/wonders/life/life.html

[12] http://www.nathanieljohnston.com/index.php/tag/conways-gameof- life/

[13] https://www.youtube.com/watch?v=C2vgICfQawE

[14] http://mathworld.wolfram.com/GameofLife.html

[15] web.mit.edu/jb16/www/6170/gameoflife/gol.html

[16] web.stanford.edu/ cdebs/GameOfLife/

SIMRAN BHATIA, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
bhatiasimran221@gmail.com

RASHI, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
rashichawla2608@gmail.com

Mathematical Analysis of Rangekeeping

Namrata Lathi

Abstract

When the long range guns were introduced in naval gunnery, in order to avoid the attack, an enemy ship would move some distance once the shells were fired towards it. Thus, it became important to develop a method to find out the position of the target ship at the time of impact of shells. This process of keeping the track of position of target ship is called 'Rangekeeping'. This paper aims to analyse how a mathematical model can be used for Rangekeeping along with a hypothetical example.

1 Introduction

Throughout history, nations have developed their armed forces as a means of defence and to achieve their political, religious, or economic goals. To equip those armies for battle, complex and destructive weapons were engineered to exploit the weakness of opponents.

For maritime nations, ships were designed as weapons to inflict great damage to an enemy's fleet. The largest of those ships became known as 'Capital Ships'. Naval strength of a nation was measured by the type and quantity of "capital ships" that it possessed. Therefore, controlling the seas and protecting national interests became expensive so a large monetary commitment was required.

When the long range gunnery was introduced, the time of flight of the projectile increased, giving the target ship an opportunity to avoid impact by moving some distance once the shells were fired towards it. But as the nations had so much money invested in their capital ships, it became necessary to develop a method to find the exact position of the target at the time of impact.

At first, the guns were aimed using the technique of "Artillery Spotting" which involved firing a gun at the target, observing the projectile's point of impact (fall of shot), and correcting the aim based on where the shell was observed to land. This became difficult as the range of the gun increased. So, there was a need to develop a method which could calculate the exact position of the target ship without being restricted by the range of the gun. This was done by developing a mathematical model called RANGEKEEPING which used analog computing. It was widely used by US Navy during World War I and Imperial Japanese Navy for launching of torpedoes and for torpedo fire control system.

2 Theoretical Concepts

2.1 Definitions

1. **Line of Sight:** It is an imaginary straight line that stretches between observer's eye and the object that he is looking at. It is established using radar or optical instruments.

61

2. **Speed of the Target:** It is the distance traveled by the target over an interval of time using radar divided by time taken to cover that distance.

3. **Present Range:** It is the observed distance from ship to target, measured along the line of sight determined by rangefinder, fire control radar or estimate.

4. **Relative Bearing:** The angle between the vertical plane through the fore-and-aft axis of own ship and the vertical plane through the line of sight, measured in a horizontal plane clockwise from the bow, Figure 1. So, the motion of the target is a result of change in target's present range and relative target bearing.

Figure 1: Relative Bearing

Figure 2: Coordinate System

2.2 Coordinate System

Here, a moving coordinate system is used based on the line of sight (LOS) between the ship firing its gun (known as the "own ship") and the target. As is shown in Figure 2, we define the "y axis" as the LOS and the "x axis" as a perpendicular to the LOS with the origin of the two axes centered on the target. The rate of bearing change is positive in the clockwise direction. The rate of range is positive for increasing target range.

2.3 Course of the Target

Due to difference in the speed of the target and the projectile, the course of the projectile is set where the target will be when the projectile reaches it. This can be done by measuring a related quantity called 'Angle on the bow' which is the angle made by the ship's course and the line of sight, Figure 3. It is usually estimated based on observational experience of the observer. Some devices used for assistance include Radar and *Sokutekiban* which was first used by the Imperial Japanese Navy. The angle on the bow of the target can be computed using the following equation-

$$\cos(90° - \theta_{AOB}) = \frac{l_{observed}}{l_{actual}} = \frac{R_T . \theta_{Ship}}{l_{actual}}$$

$$\theta_{AOB} = \sin^{-1}\left(\frac{R_T . \theta_{Ship}}{l_{actual}}\right)$$

where l_{actual} is the true length of the ship.

$l_{observed}$ is the observed length of the ship.

θ_{Ship} is the angle subtended by the target ship from the own ship.
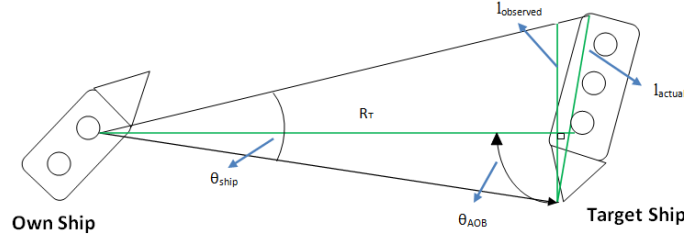
θ_{AOB} is the angle on the bow.

Figure 3: Determining the Angle on the Bow

2.4   Target Tracking

If the ship and the target both are stationary, then we can use the present range and relative bearing calculated using radar for launching the projectile. But as in our problem, when ship and target both are moving with respect to each other, gunfire, to be accurate, must compensate for the error caused by their relative motion during the time of flight of the projectile. For this, we need to find rate of change of range and rate of change of relative target bearing which can be applied to the original measured quantity to produce continuously accurate values of the coordinates of present target position.

2.5   How to Analyse the Motion?

In order to analyse the motion of the target, we will study the components of relative target motion-

The apparent motion of the target as seen from the deck of own ship is the resultant of the motion of both vessels. To take a simple case, if both ships are steaming down the line of sight toward each other, each at the rate of 30 knots, it is apparent that the range is closing at the rate of 60 knots. Since this is true, the component of relative target motion affecting range can be obtained by combining the range component of own-ship motion with the range component of target motion. The deflection component can be similarly handled. We use vectors to represent the motion of own ship and target ship where length represents speed and direction indicates course. These components are:

VO_y: The horizontal range component of own-ship velocity along the line of sight.

VT_y: The horizontal range component of target velocity along the LOS.

VO_x: The horizontal lateral component of own-ship velocity across the LOS.

VT_x: The horizontal lateral component of target velocity across the LOS.

These components can be calculated as follows-

$$VO_x = S_O.\sin(180° - \theta_T) = S_O.\sin\theta_T = -S_O.\sin\theta_T$$
$$VT_x = S_T.\cos(90° - \theta_{AOB}) = S_T.sin\theta_{AOB}$$
$$VO_y = S_O.\cos(180° - \theta_T) = -S_O.\cos\theta_T$$
$$VT_y = S_T.\sin(90° - \theta_{AOB}) = S_T.\cos\theta_{AOB}$$

where S_O is the speed of own ship, S_T is the speed of the target ship, θ_T is the relative bearing, θ_{AOB} is the angle on the bow. Here, θ_T is measured in clockwise direction and θ_{AOB} is measured in anticlockwise direction.

The direction of the effect of the components on range or bearing can best be ascertained by observation. Consider the target stationary; then examine the direction of VO_y and VO_x. If VO_y causes range to increase, it is considered positive. VO_x, the bearing component, is negative if, as in the figure 2, it causes the line of sight to move to the left (or the numerical value of bearing to decrease). Then by considering own ship to be motionless, we can assign algebraic signs to target components by the same reasoning. Now, these components can be combined to provide range and bearing rates.

2.6   Range rate:

It is the time rate of change of range. It is the component of relative target motion along the LOS which can be obtained by adding the range components algebraically.

$$\text{Range rate (dR)} = VO_y + VT_y$$

The sign of range rate will depend upon the sign of its components which will decide whether range is increasing or decreasing. Range rate must be converted from knots into yards per second, since range is measured in yards. The conversion constant is the result of the following equation:

$1knot = 2,027$ yards per hour/$3,600$ seconds per hour=0.563 yards per second.

2.7   Target Bearing Rate

It is the time rate of change of target bearing. It is the component of relative target motion across the LOS obtained by adding the bearing components algebraically.

$$\text{Linear Target bearing rate (RdBs)} = VO_x + VT_x$$

The sign of target bearing rate will depend upon the sign of its components which will decide whether target bearing is increasing or decreasing.

Since bearing is an angle, rate of change of bearing must be converted to an angular rate. RdBs is linear. The conversion is accomplished as follows: 1 yard per second = $1/Range(yards)/1000 = 1000/$Range (yards) mils per second.

(Here, mil refers to milliradian, an angular measurement which is defined as a thousandth of the radian)

$$\text{1 mil} = 3.438 \text{ minutes of arc}$$

$$\text{1knot at Range (yards)} = 0.563 * 1000/\text{Range (yards)} * 3.438$$
$$= 1936/\text{Range (yards) minutes of arc per second}$$

Therefore, Angular target bearing rate (dBs) (min. of arc per sec.) $= 1936 * RdBs$ (kts)/Range (yards). It also implies that

Linear target bearing rate(RdBs) = Angular bearing rate(dBs) * Range/1936

Now, range and relative bearing of the target at any point of time can be computed by applying the rates derived above on the initial range and relative target bearing calculated above.

3   Position Prediction

The prediction of the target ship's position at the time of projectile impact involves prediction of range and azimuth.

3.1   Predicted Range

The target range at the time of projectile impact can be estimated using following equation:

$$R_{TP} \quad = \quad R_T + \frac{dR_T}{dt}.(t_{TOF} + t_{Delay}) \tag{1}$$

where R_{TP} is the range to the target at the time of projectile impact.

R_T is the range to the target at the time of gun firing.

t_{TOF} is the projectile time of flight(t_{TOF})plus system firing delays(t_{Delay}).

The exact prediction of the target range at the time of projectile impact is difficult because it requires knowing the projectile time of flight, which is a function of the projected target position. Here, we **assume that time of flight is linearly proportional to range** as shown in the following equation-

$$R_{TP} \quad = \quad R_T + \left(\frac{dR_T}{dt}\right).(k_{TOF}.R_T + t_{Delay}) \tag{2}$$

where, k_{TOF} is the constant of proportionality between time of flight (TOF) and target range.

The assumption of TOF being linearly proportional to range can be improved through the use of more sophisticated means of function evaluation. Range prediction requires knowing the rate of range change whose method of calculation was shown above. Thus,

$$\frac{dR_T}{dt} = VO_y + VT_y$$

where V_{OY} is the horizontal component of own-ship velocity along the line of sight.

V_{TY} is the horizontal component of target ship velocity along the line of sight.

Now, as we know speed is the magnitude of velocity, so the above equation can also be written as follows-

$$\frac{dR_T}{dt} \quad = \quad SO_y + ST_y \tag{3}$$

where S_{OY} is the own ship speed along the LOS.

S_{TY} is the target ship speed along the LOS.

Now, putting values from equation 2 and 3 in equation 1, we get the following equation-

$$R_{TP} \quad = \quad R_T + (SO_y + ST_y).(k_{TOF}.R_T + t_{Delay}) \tag{4}$$

This is the complete equation for the predicted range.

3.2 Predicted Azimuth

The prediction of azimuth is performed similarly to the range prediction. The target bearing at the time of projectile impact can be estimated using following equation, which is illustrated in the figure 4.

$$\theta_{TP} = \theta_T + \tfrac{d_T}{dt}.t_{TOF}$$

$$\theta_{TP} \quad = \quad \theta_T + \frac{d_T}{dt}.(t_{TOF} + t_{Delay}) \tag{5}$$

where θ_{TP} is the azimuth to the target at the time of projectile impact.

θ_T is the azimuth to the target at the time of gun firing.

t_{TOF} is the projectile time of flight(t_{TOF})plus system firing delays(t_{Delay}).

Azimuth prediction requires knowing the rate of change of bearing whose method of calculation was shown above. Thus,

$$\frac{d\theta_T}{dt} \quad = \quad \frac{VO_x + VT_x}{R_T}$$

where VO_x is the horizontal component of own-ship velocity across the line of sight.

VT_x is the horizontal component of target ship velocity across the line of sight.

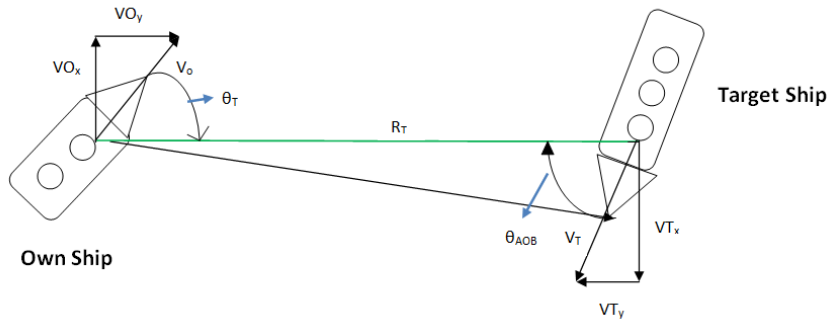R_T is the initial range to the target.

Figure 4: Determination of Angular Rate

Now, as we know speed is the magnitude of velocity, so the above equation can also be written as follows-

$$\frac{d\theta_T}{dt} = \frac{SO_x + ST_x}{R_T} \tag{6}$$

where SO_x is the own ship speed across the LOS.

ST_x is the target ship speed across the LOS.

Now, putting values from equation 6 in equation 5, we get the following equation-

$$\theta_{TP} = \theta_T + \frac{(SO_x + ST_x).(k_{TOF}.t_{TOF} + t_{Delay})}{R_T} \tag{7}$$

This is the complete equation for the predicted azimuth.

4   Example

World War II is moving towards its end. A fierce battle is taking place between US Navy battleship USS Texas (BB-35) and Japanese Navy battleship Fubuki in the Pacific Ocean. But the Japanese Navy has employed several small ships before their main battleship. Missiles are placed on these small ships, but the guidance system of these missiles is on the main ship which is 5 miles away from the US ship and is moving farther. Suddenly, an officer informed the General of US Navy that the enemy attacked the ship on which all the long-range missiles were positioned and thus all the long-range weapons got destroyed. Now, only one long-range missile which was kept hidden on their main ship is left. US Navy can win only if this missile strikes the Japanese main ship and thus destroy their missiles guidance system. But as the chief Japanese ship is moving, the General must devise such a strategy to launch the missile, so that at the end of its time of flight, when it strikes; the target ship is at its striking point. What strategy the General should use?

Given the speed of missile is 3.72 miles/hr. The delay in firing is 10 minutes. The speed of the own ship is 65.2 Km/hr. The angle between the vertical plane through the fore-and-aft axis of own ship and the vertical plane through the line of sight is 55°. The length of the target ship is 175meters. Its speed is 75 Km/hr. The angular width of the target is 0.016°. Constant of proportionality between time of flight (TOF) and target range is 0.036.

Solution

The General should develop such a model which can predict the position of the target ship at the end of time of flight of missile. It is shown as follows:

$$\text{Given-} \quad R_T = 5\text{miles} = 10,000\text{yards}$$
$$t_{Delay} = 10\text{minutes}$$
$$\theta_{ship} = 0.016°$$
$$l_{actual} = 175\text{meters} = 192\text{yards}$$
$$S_O = 65.2Km/hr. = 71303.6yards/hr.$$
$$S_T = 75Km./hr. = 82,021yards/hr.$$
$$\theta_T = 55°$$

Step-1 Calculation of angle on the bow

$$\theta_{AOB} = \sin^{-1}\left(\frac{R_T.\theta_{Ship}}{l_{actual}}\right) = \sin^{-1}\left(\frac{10,000*0.016°}{192}\right) = 60°$$

Step -2 Analyzing the motion of the target ship

The components of motion of own and target ship are calculated as follows-

$$VO_x = -S_O.\sin\theta_T = -71,303.\sin 55° = -58,408 \text{yards}$$
$$VT_x = S_T.\sin\theta_{AOB} = 82,021\sin 60° = 71,032.27 \text{yards}$$
$$VO_y = -S_O.\cos\theta_T = -71,303.\cos 55° = -40,898 \text{yards}$$
$$VT_y = S_T.\cos\theta_{AOB} = 82,021.\cos 60° = 41,010.5 \text{yards}$$

Step-3 Calculation of Range rate and Target bearing rate

$$\text{Range rate (dR)} = VO_y + VT_y = -40898 + 41010.5 = 112.5 \text{yards}$$
$$\text{Linear Target bearing rate (RdBs)} = VO_x + VT_x = -58408 + 71032.27$$
$$= 12624.27 \text{yards}$$

(Positive sign indicates that range and target bearing rate is increasing.)
Angular target bearing rate (dBs) (min. of arc per sec.)
= 1936* RdBs (kts)/Range (yards) = 1936 * (12624.27*1/0.563)/10,000 = 42.86
Angular target bearing rate (dBs) (in degrees) = 42.86 * 0.0167 = 0.72

Step-4 Position prediction

Range prediction

$$R_{TP} = R_T + (SO_y + ST_y).(k_{TOF}.R_T + t_{Delay})$$
$$= 10,000 + (112.5).(0.036 * 10,000 + 0.17)$$
$$= 50,519.125 \text{yards}$$

Azimuth prediction

$$\theta_{TP} = \theta_T + \frac{(SO_x + ST_x).(k_{TOF}.t_{TOF} + t_{Delay})}{R_T}$$
$$= 450 + (0.72).(0.036 * 10000 + 0.17)$$
$$= 304.32°$$

Thus, the General should set the target of the missile at $50,519.125$yards and $304.32°$ to be able to hit the target ship.

References

[1] John L. Fleming "CAPITAL SHIPS: A HISTORICAL PERSPECTIVE"

[2] en.wikipedia.org/wiki/Mathematical discussion of rangekeeping

[3] http://www.eugeneleeslover.com/USNAVY/CHAPTER-19-B.html

NAMRATA LATHI, B.Sc.(H) MATHEMATICS, 2nd SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
namratalathi@gmail.com