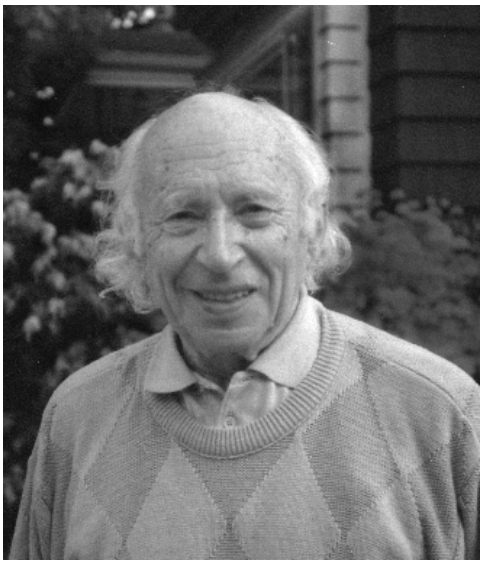
Èclat
Mathematics Journal

Lady Shri Ram College for Women

Volume –IV

2013



ISRAEL MOISEEVICH GELFAND
September 2, 1913 - October 5, 2009

This journal is dedicated to the memory of Israel Moiseevich Gelfand.

PREFACE

After witnessing the naissance, growth and success of our journal *Éclat*, we have immense pleasure in presenting its Volume IV. Our readers expect a high standard of exposition, so *Éclat* is a compilation of papers based on historical and new results, speculations and definitive treatments, broad developments and explorations by the students as well as the faculty members. Following a structured approach, we have divided the journal into four sections- History of Mathematics, Rigour in Mathematics, Extension of Course Contents and Interdisciplinary Aspects of Mathematics. This volume will stimulate further discussion and additional research in various realms of Mathematics, enabling a range of empirical studies on the topical concepts covered.

We hope in this way we keep our readers, as it were, on a level with the progressive state of Mathematics. Like the previous volumes, the idea of the journal has been promulgated to other departments as well as colleges. We hope such a response continues and hence, leads our readers to feel a greater interest in the study of Mathematics, helping them to learn and appreciate it.

The success and final outcome of *Éclat* involved a lot of guidance, research and efforts. The discursive compilation has evolved after rigorous perusing. As the editors of this journal, we owe profound gratitude to the entire department, which has been so supportive and involved throughout. And our heartfelt thanks to our faculty advisors for their valuable support and continued encouragement and discussion at all stages of the gestation of the journal.

We hope that our readers deem the journal beneficial and worthwhile, and find it instrumental in honing their skills in doing individual research.

Editorial Team:

Sruthi Sekar
Kritibha Rai
Megha Baid

Contents

History of Mathematics

Mathematics is 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, an understanding of the motivation behind the ideas is needed. This 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.

ISRAEL MOISEEVICH GELFAND
(1913-2009)

EDITORIAL TEAM

Israel Gelfand was a major figure in mathematics for seven decades. His research ranged over most of pure mathematics, including algebra, analysis, and geometry. He also worked in mathematical biology, opening up the field of integral geometry, a topic that is fundamental to medical scanners. He was an incomparable teacher and made significant advances in every field that he touched.

Gelfand was born on 2 September 1913, to Jewish parents in the small town of Okny (now Krasni Okny) to the north of Odessa in southern Ukraine, which was then a part of the Russian empire. In 1930 he moved to Moscow to complete his secondary education. However, he was not permitted to enroll as an undergraduate, having (according to some sources) been expelled from school because his father, a miller, was considered to be a capitalist. Israel took a part-time job as doorkeeper at the Lenin Library and taught evening classes on mathematics. The work made it possible for him to attend mathematics courses at Moscow State University.

He showed such talent that Andrei Kolmogorov, the leading Soviet mathematician of the period, took him on as a postgraduate student. His 1935 PhD thesis was in the relatively new area of functional analysis, where the ideas of calculus are extended from finitely many variables to infinitely many. One practical application is to partial differential equations, the mathematical physicist's favourite tool for describing the natural world. Another is the mathematical formulation of quantum mechanics.

Gelfand was appointed to the Steklov Mathematical Institute and taught at the university, but lost both positions temporarily through antisemitism. He was elected a corresponding (low-status) member of the Soviet Academy of Sciences, but it was more than 30 years before he was made a full member. His seminar series, run independently of the university and open to anybody, ran for nearly 50 years and is famous throughout the mathematical world. He moved to America in 1989, first to Harvard University, Cambridge, Massachusetts, and then Massachusetts Institute of Technology, then settling at Rutgers University, New Jersey.

The heart of Gelfand's research was representation theory, a formal setting for symmetry, a concept of central importance in mathematics and physics. A symmetry of an object is a transformation that preserves its structure, and the collection of all such transformations is the object's symmetry group. The physical world, at subatomic level, is highly symmetric: if you change an electron's direction of spin, or its electric charge, the laws of physics still work the same way. Representation theory studies all the contexts in which a particular symmetry group can arise. Its applications include subatomic particles and pattern formation - why snowflakes are six-sided, and why tigers have stripes but leopards have spots.

The most important types of symmetry are the "classical groups", a typical example being the group of all rotations of space. Gelfand solved many fundamental questions about classical groups, using a mixture of algebraic and geometric methods. His interests went beyond mathematics into theoretical and experimental science. In 1958, when his son, Aleksandr, contracted leukaemia, he started applying mathematics to cell biology, setting up the Institute of Biological Physics of the Russian Academy of Sciences.

Some of his discoveries have applications that are important for everyone: medical scanners. Doctors routinely use several different kinds of scanner. CT scanners, for example, use beams of x-rays to obtain a three-dimensional image of the body's internal organs. This is a bit like holding a semi-transparent object up to the light and using the resulting shadows to work out its true shape. The first steps in this area were taken in 1917 by Johann Radon. Gelfand developed Radon's ideas extensively, founding an entire field of mathematics, now called integral geometry. His ideas are vital to today's medical imaging methods.

Gelfand received many awards. The Soviet Union awarded him the Order of Lenin three times. He won the Wolf prize (comparable to a Nobel) in 1978, and the Kyoto prize (for "significant contributions to the progress of science, the development of civilisation, and the enrichment and elevation of the human spirit") in 1989. He was elected to innumerable academic bodies, including the Royal Society and the US National Academy of Science. He was also a great teacher. He set up a distance-learning school for mathematics in the Soviet Union, and a similar one in the US in 1992. He considered teaching and research to be inseparable, and was equally comfortable talking to schoolchildren or his research colleagues. He supervised 22 PhD students, several of them now outstanding mathematicians in their own right.

The Russian Academy of Sciences together with the Moscow Mathematical Society and Independent University of Moscow is organizing an international conference dedicated to the centenary of Israel Gelfand and another conference dedicated to the centennial of Gelfand will be held in the US.

References

[1] Ian Stewart, *The Guardian*, Sunday 8 November 2009 18.13 GMT.
[2] http://www.ams.org/meetings/calendar/2013_jul22-25_moscow.html

Editorial Team, B.Sc.(H) Mathematics, 4th Semester, Lady Shri Ram College For Women

AMALIE "EMMY" NOETHER

AASTHA BHATIA AND JASMINE BHULLAR

Abstract. Traditionally, mathematicians are considered to be men. This, however, is erroneous. Women have contributed to mathematics as much as their male counterparts. One such unmerited women mathematicians is German-born Amalie "Emmy" Noether. Emmy Noether revolutionized the areas of abstract algebra and theoretical physics. Even Einstein eulogized her in a letter to New York Times as "the most significant creative mathematical genius." Her groundbreaking work in abstract algebra and physics undergird much of todays vanguard research in physics, including the search for 'Higgs Boson.' Its time to bring this brilliant theorist out of anonymity and celebrate life and achievements of this great mathematician who with her sheer will and unshakable love for numbers overcame severe handicaps.

EARLY LIFE AND FAMILY BACKGROUND

Noether's siblings

Emmy Noether Circa 1910

Amalie "Emmy" Noether was born in Erlangen Bavaria, Germany. She was the daughter of Max Noether, a distinguished mathematics professor at University of Erlangen and Ida Amalie Kaufmann, who belonged to a wealthy Cologne family. Both Emmy's parents were of Jewish origin and Emmy was the eldest of their four children. Her younger siblings were all

boys named Alfred, Fritz, and Gustav Robert. Fritz became a quite respected mathematics professor like his father. Some might suspect her mathematical talent was in her genes.

EDUCATION

Emmy had a conventional upbringing, attending the *Hohere Tochter Schule* in Erlangen till she turned 18, when she was certified as a teacher in French and English at the Institute for Education and Instruction of Females. But her interests soon turned towards mathematics. Being a woman, she could not enroll in the university formally, but was allowed to audit at University of Erlangen where her father taught. She took her final exams and was granted an equivalent of Bachelor's degree. She then went to graduate school at University of Gottingen. Gottingen was at that time the centre of the mathematical universe. In 1904, she returned to University of Eralengen, where she prepared her dissertation under Paul Gordon and was awarded her doctorate (*summa cum laude*) in 1907.

PROFFESIONAL LIFE

The University of Erlangen did not hire Emmy as they had a policy against women professors. She helped her father at the Mathematics Institute in Erlangen and meanwhile did her own research. Soon, she began to publish papers on her work. In 1908, she was elected to the *Circolo Matematico di Palermo*, then in 1909 she was invited to become a member of the *Deutsche Mathematiker Vereinigung*. Till 1915, Noether worked at the Mathematical Institute of Erlangen without pay or title. In 1915, she joined the Mathematics department at the University of Gottingen and started working with Klein and Hilbert on Einstein's general relativity theory. Noether was only allowed to lecture under Hilbert's name, as his assistant, although still without a salary. Much of her work appears in papers written by colleagues and students, rather than under her own name. She became an "associate professor without tenure" in 1922 and began to receive a small salary. She was invited to address the International Mathematical Congress twice. In 1933, Emmy was fired from the University of Gottingen with the beginning of the Nazi rule in Germany. She continued to meet her students for some time but after receving a grant to become a guest professor at Bryn Mawr College she moved to Pennsylvania, U.S.A. According to Van der Waerden's obituary of Emmy Noether, she did not follow a lesson plan for her lectures. Rather, she used her lectures as a spontaneous discussion time with her students. A promising start to a new career was cut short by her untimely death due to complications after a uterine surgery in 1935.

ASSOCIATION WITH EINSTEIN

Emmy did work on the theory of invariants, whose main importance was seen in the framework of Einstein's theory of relativity. Her paper appeared in 1918. Einstein wrote to Hilbert in a letter of May 24, 1918: "Yesterday I received from Miss Noether a very interesting paper on the formation of invariants. I am impressed that one can handle those things from such a general viewpoint. She seems to understand her job."

FIGURE 1. Emmy Noether circa 1932

In 1933, when Emmy was fired by the Nazis from the University of Gottingen, her brother, Fritz Noether, now working in Soviet Union, urged her to join him. But, Albert Einstein convinced the Rockefeller Foundation to match a grant from the Emergency Committee to aid displaced German Scholars and Emmy was granted a one year instructor position in Bryn Mawr College in Pennsylvania.
Two years later, when Emmy died, the New York Times printed a letter that designated her as "The most significant creative mathematical genius thus far produced since the higher education of women began." The letter was signed by Albert Einstein.

HER CHARACTER AND PERSONA

Sociable: Her friends and colleagues remember her as a happy woman with a robust sense of humour. Her heart knew no malice and she did not believe in evil. Emmy Noether never married. She cared deeply about her students and considered her students as family. In his memorial speech at Emmy's funeral, Hermann Weyl said: "She was warm like a loaf of bread, there radiated from her a broad, comforting, vital warmth. A tea at her apartment was always pleasurable."
Dedication towards mathematics: French Mathematician Poisson once remarked: "Life is good for only two things, discovering mathematics and teaching mathematics." Emmy would have seemed to agree. She lived for mathematics and cared nothing for housework or possessions. As Hermann Weyl recalled in his speech at her funeral "She had a particular handkerchief which she had a way of jerking around, very energetically while explaining something." Appearance-conscious students cringed as she retrieved the handkerchief from her blouse and ignored the increasing disarray of her hair during a lecture.
A distinguished algebraist, Olga Taussky-Todd described a luncheon during which Noether, wholly engrossed in a discussion of mathematics, "gesticulated wildly" as she ate and "spilled her food constantly and wiped it off from her dress, completely unperturbed."
Perseverence and never say die attitude: The most noticeable aspect of her personality was her persistence in the face of the tremendous obstacles and barriers in her path to become one of the greatest algebraists.
Noether had two strikes against her: First, she was a woman at a time when education for women was severely limited. Most German universities did not enroll female students or

hire female professors. Second, she was Jew. The Jews in Nazi Germany suffered appallingly after January 1933. Nevertheless with her spirit and will she made great contributions to Mathematics.

CONTRIBUTION TO MATHEMATICS

Emmy Noethers scientific production fell into three clearly distinct epochs:

- The period of relative dependence, 1907-1919;
- The investigations grouped around the general theory of ideals 1920-1926;
- The study of the non-commutative algebras, their representations by linear transformations, and their application to the study of commutative number fields and their arithmetics.

Emmy Noether's mathematics was abstract, original and deep. Her contributions to the theory of algebraic invariants and the theory of ideals in rings are very significant. She laid down the broad foundations of the modern abstract theory of ideals and also a great deal more in modern algebra.
Emmy Noether's first piece of work when she arrived in Gottingen in 1915 is a result in theoretical physics sometimes referred to as *Noether's Theorem*, which proves a relationship between symmetries in physics and conservation principles. It was her work in the theory of invariants which led to formulations for several concepts of Einstein's general theory of relativity.

CLOSING NOTE

Among mathematicians, Emmy Noether is consistently ranked among the most famous mathematicians of the 20th century and recognized along with Newton, Gauss, Fourier, Leibnitz, as one of the greatest of all time for her work in theoretical physics and mathematics, which both contribute significantly to studies today. The outstanding theme of Emmy's life is that she pursued her goals with single-minded determination and not much fuss. The approach as developed by Emmy Noether and her pupils has come to be known as the 'Noether School'. It has been rightly said that "her mathematical originality was absolute beyond comparison" (by B. L. van der Waerden) and "Noether changed the face of algebra" (by Hermann Weyl). Noether's death was somewhat sudden. Her legacy in mathematics, however, remains.

References

[1] Ioan James, *Remarkable Mathematicians: From Euler to von Neumann (Spectrum Series)*
[2] M.B.W. Tent, *Emmy Noether: The Mother of Modern Algebra*
[3] Dwight E. Neuenschwander *Emmy Noether's Wonderful Theorem*

Aastha Bhatia, B.Sc.(H)Mathematics, 1st Year, Lady Shri Ram College for women
E-mail address: aasthabhatia94@gmail.com

Jasmine Bhullar, B.Sc.(H)Mathematics, 1st Year, Lady Shri Ram College for women
E-mail address: samsimmi@yahoo.com

PARACONSISTENT MATHEMATICS

HIRANMAYEE RAJAN, SANJANA GUPTA, TANISHA

Abstract. There are two schools of thoughts regarding mathematics, classical mathematics and paraconsistent mathematics. Classical mathematics believes that contradictions cannot exist in mathematics, and every theory is both complete and consistent. Paraconsistent mathematics, on the other hand, considers the possibility of existence of contradictions in mathematics. It is this part of mathematics that we will address in our article. We will discuss inconsistencies in three fields of mathematics - Set theory, Calculus and Geometry, by discussing paradoxes and giving relevant examples.

Introduction

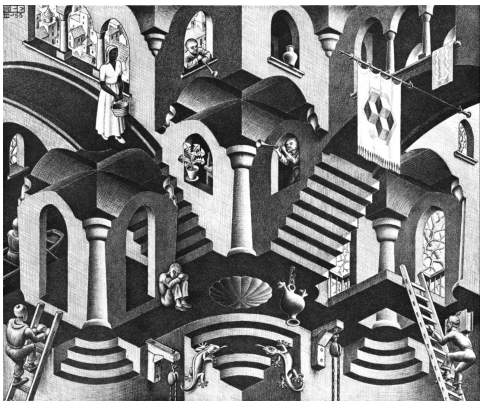
Paraconsistent mathematics is a system of mathematics in which it is possible for contradictions to exist. A statement 'A' and its negation 'not A' can both be true. Basically, paraconsistent mathematics accepts inconsistent theories (theories which contain contradictions). Formal logic is used to contain these contradictions and allow theories to remain coherent. This system of mathematics arose in response to paradoxes like Russell's Paradox and Liar's Paradox.

Hand with Reflecting Sphere, M.C. Escher

Motivation and Foundation. In 1920, David Hilbert tried to develop a set of axioms (that were both consistent and complete) which could be used to prove any true statement in mathematics. However, Gödel's incompleteness theorems showed that these axioms were impossible to develop and laid the foundation for work in the field of paraconsistent mathematics. Following are Gödel's Incompleteness Theorems:

- Gödel's first incompleteness theorem states that in any given system, there will always be some true statements that cannot be proved.
- Gödel's second incompleteness theorem states that a system can prove its consistency if and only if it is inconsistent.

For example, let 'T' be a formal theory i.e., a system of mathematics based on a collection of axioms. Let 'G' be a statement such that 'G' cannot be proved in the theory 'T'. If this statement is true, then there is at least one sentence in 'T' which cannot be proved (namely 'G'), making 'T' incomplete. However, if 'G' can be proved in 'T', we get a contradiction as 'G' is provable, but by virtue of its content, it also cannot be proven.
Therefore, one has to choose between incompleteness and inconsistency. Gödel showed that a sentence such as 'G' can be created in every theory. So, mathematics must always either be incomplete or inconsistent.

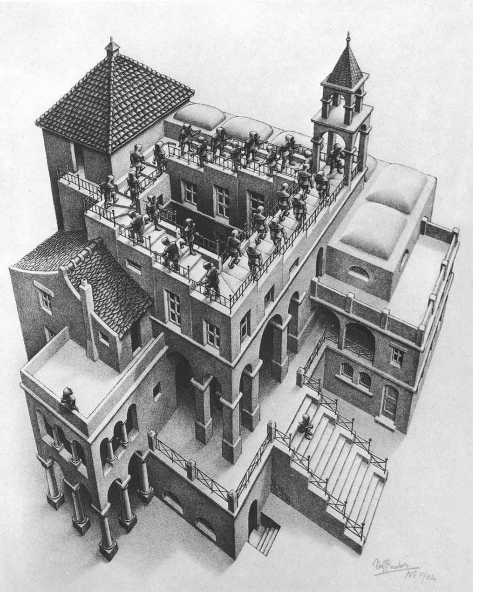


Convex and Concave, M.C. Escher

The inconsistency of mathematics is best seen in paradoxes. A paradox is a statement or proposition that, despite correct reasoning, leads to a conclusion that is contradictory. It is an argument that produces an inconsistency. Some of the famous paradoxes are the Liar's Paradox and the Barber's Paradox.

Liar's Paradox:
"This statement is false." If the statement is true, then the statement must also be false. Similarly, if the statement is false, then it implies that the statement is true.

Barber's Paradox:
"The Barber only shaves those people in town who do not shave themselves." Then who shaves the barber?
If the barber shaves himself, then he must also not shave himself. However, if he does not shave himself, then he ought to shave himself – a contradiction.

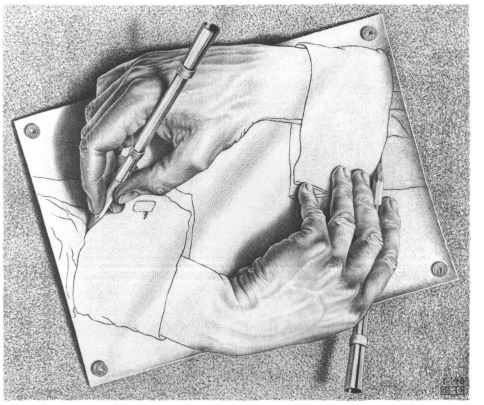


Ascending, Descending, M.C.Escher

Inconsistency in Set Theory

Set theory, often called the foundational system of mathematics, is a branch of mathematics that studies sets. An example of set theory defined informally i.e., without the use

of formal logic, is Naïve Set Theory. Naïve set theory is filled with inconsistencies. This is best highlighted by Russell's Paradox. Russell's Paradox was put forth by Bertrand Russell, a famous British philosopher, logician, mathematician and historian. Before discussing Russell's Paradox, we'll briefly outline Naïve Set Theory.

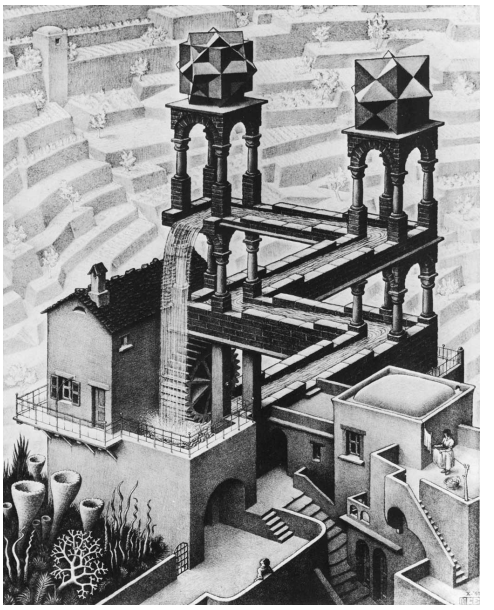


Drawing Hands, M.C.Escher

Naïve Set Theory. It consists of the following 3 axioms:

- **Axiom of Extensionality:** Given any set A and any set B, if for every element x, x is a member of A if and only if x is a member of B, then A is equal to B.
- **Axiom of Abstraction:** Any collection of objects that can either be listed or described by some predicate constitutes a set.
- **Axiom of Choice:** Let C be a collection of non-empty sets. Then we can choose a member from each set in that collection. In other words, there exists a function f defined on C with the property that, for each set S in the collection, f(S) is a member of S.

A set may contain other sets as its members. A set may also contain itself. For example, consider the set 'T' containing all things that are not triangles. Since 'T' is not a triangle, it contains itself.



Waterfall, M.C.Escher

Russell's Paradox. *"Let R (Russell set) be the set of all sets that are not members of themselves." Is R a member of R?*

Interpretation:
If it contains itself, it is by definition a set that does not contain itself – a contradiction. If it does not contain itself, it is a set that does not contain itself and so should contain itself – again a contradiction.

Alternate Theory:
To avoid contradictions and paradoxes, classical mathematicians adopted a different stance. They accepted a somewhat complex version of set theory called **Zermelo-Fraenkel Set Theory** (ZFT). ZFT discards the Principle of

Abstraction, and replaces it with eight other axioms. Pre-existing sets are used to construct new sets via the set building rules laid out in ZFT. The sets thus built are given ranks based on the number of times the set building rules are used.

For example, let's start with V_0.
Then, the set with the next rank $V_1 = \mathcal{P}(V_0)$, where $\mathcal{P}$ is the power set.
Similarly, $V_2 = \mathcal{P}(V_1)$.
In general, $V_{n+1} = \mathcal{P}(V_n)$.

The basic idea is that every set must be a member of a set V_α for some α. So, in ZFT it is impossible to create a set of all sets.

Reconciliation of Russell's Paradox:
In ZFT, Russell set cannot exist and thus Russell's paradox is avoided. **To create Russell set, Russell set is required, so building it using the axioms of ZFT is impossible.** In terms of rank, Russell set would need to be of some rank n, but also of rank $n + 1$ (and $n + 2$ and $n + 3$ and so forth) because to be created it needs to be of a higher rank than itself. As this is not possible, Russell set cannot exist in ZFT.

Belvedere, M.C.Escher

Interestingly, ZFT based on ad hoc axioms was later proved to be highly inconsistent.

Incompleteness In Set Theory

Cantor's Continuum Hypothesis highlights the incompleteness in set theory. It shows us that Naïve Set Theory cannot answer some fundamental questions about infinity. Cantor's Continuum Hypothesis can be stated as follows:

"There is no infinite set with a cardinal number between that of the "small" infinite set of integers $\aleph_0$ and the "large" infinite set of real numbers $\mathfrak{c}$ (the 'Continuum'), Symbolically, the Continuum Hypothesis is that $\aleph_0 < \mathfrak{c}$."

Hence, we see that Naïve Set Theory is both incomplete and inconsistent.

Inconsistency in Calculus

Infinitesimals are an integral part of calculus. Yet, it was pointed out that infinitesimals were being used inconsistently in equations. For example, suppose we are differentiating the polynomial $f(x) = ax^2 + bx + c$. In order to compute its derivative, we have

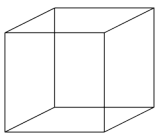
$$\begin{aligned}\frac{f(x+\varepsilon) - f(x)}{\varepsilon} &= \frac{a(x+\varepsilon)^2 + b(x+\varepsilon) + c - (ax^2 + bx + c)}{\varepsilon} \\ &= \frac{2ax\varepsilon + a\varepsilon^2 + b\varepsilon}{\varepsilon} \\ &= 2ax + b + a\varepsilon.\end{aligned}$$

So, $f'(x) = 2ax + b$, since ε is an infinitesimal. It is pointed out that ε marks a small but non-trivial neighbourhood around x, **and as $f(x+\varepsilon) - f(x)$ can be divided by ε, so ε is not zero.** Nevertheless, by the end ε simply disappears. Hence, we see that Calculus in its original form was outright inconsistent.

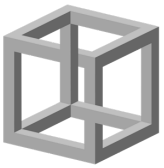
Inconsistency in Geometry

Impossible Objects. These are objects that can be drawn but can never be created in 3D. They are of 4 types:

Impossible Cube:

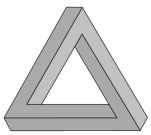


This diagram shows the most common representation of a cube. However, another interpretation of this cube is the impossible cube.



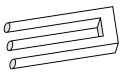
In an impossible cube, as one can see, the edges cross each other in an impossible manner.

Penrose Traingle:



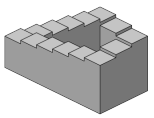
The Penrose triangle is an impossible equilateral triangle. It consists of three perpendicular sides.

Blivet:



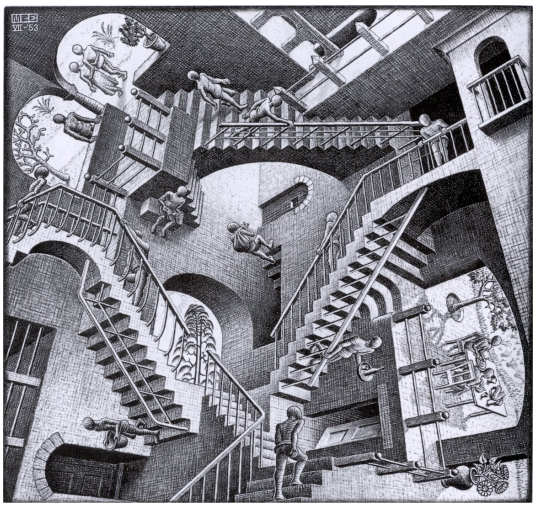
A blivet, also called Devil's Fork is an impossible object which appears to have two rectangular prongs on the right and three cylindrical prongs on the left.

Penrose Stairs:



Penrose stairs is an impossible object in which the stairs turn at an angle of 90° four times, yet the staircase forms a continuous loop. It is neither ascending, nor descending.
This impossible object is widely represented in pop culture, the most famous example being its use in the movie 'Inception'.

Depiction in Art. Inconsistencies in geometry have been represented several times in art. Some of the best examples are the works of M.C. Escher. M.C. Escher was a Dutch artist who is known for his mathematically inspired art and portrayal of impossible reality. His artworks have been depicted alongside this article.



Relativity, M.C.Escher

Conclusion

As one can see, mathematics is filled with inconsistencies. These inconsistencies span all branches of mathematics from calculus to geometry to set theory. In conclusion, mathematics is not perfect.

References

[1] McKubre-Jordens, M. *This is not a carrot: Paraconsistent Mathematics*, +PLUS Magazine.
[2] Weber, Z. *Inconsistent Mathematics*, The Internet Encyclopedia of Philosophy.
[3] Mortensen, C. *Inconsistent Mathematics*, The Stanford Encyclopedia of Philosophy.
[4] Suppes, P. *Axiomatic Set Theory*, Dover Publications.
[5] Sider T. *Set Theory*, Cornell University.
[6] LaFleur K. *Russell's Paradox*, Expository Paper, University of Nebraska-Lincoln.
[7] Szudzik, M. and Weisstein, E.W. *Continuum Hypothesis*, MathWorld–A Wolfram Web Resource.
[8] www.wikipedia.org

Hiranmayee Rajan, Sanjana Gupta, Tanisha, B.Sc.(H) Mathematics, 2nd Semester, Lady Shri Ram College For Women
E-mail address: hiranmayee.rajan@gmail.com,sanjana848@gmail.com,tanisha095@gnail.com

FOUR DIMENSIONAL SPACE

SRUTHI SEKAR

Abstract. This paper intends to induce the thought of the existence of a fourth spatial dimension and helps in visualizing it. There is an analogy between the way we move from 2nd to 3rd dimension, and the transition from 3rd to 4th dimension. This analogy is discussed briefly. One particular four dimensional object, tesseract, has also been discussed in detail.

Introduction

In mathematics, four-dimensional space ("4D") is an abstract concept derived by generalizing the rules of three-dimensional space. Algebraically it is generated by applying the rules of vectors and coordinate geometry to a space with four dimensions. In the 4th dimension, there would virtually be no limits. Through mathematics we know a great deal about the 4th dimension. Still, we cannot imagine it. Although we do have rational access to 4th dimension, it is non-existent for our senses and for our consciousness.

History

The possibility of spaces with dimensions higher than three was first studied by mathematicians in the 19th century. In 1827, **Mobius** realized that a 4th dimension would allow a 3D form to be rotated onto its mirror-image (just like we need the 3rd dimension to **rotate** a 2D object like a square onto its mirror image. This can be visualized by looking at a tessract), and by 1853 **Ludwig Schlafli** had discovered many polytopes in higher dimensions.

Higher dimensions were soon put on firm footing by **Bernhard Riemann's** 1854 **Habilitationsschrift**, in which he considered a "point" to be any sequence of coordinates $(x_1, x_2, ..., x_n)$. The possibility of geometry in higher dimensions, including four dimensions in particular, was thus established.

One of the first major expositors of the 4th dimension was **Charles Howard Hinton**, starting in 1880 with his essay "What is the Fourth Dimension?" published in the Dublin University magazine. He coined the terms **tesseract, ana** and **kata** in his book **A New Era of Thought**, and introduced a method for visualising the 4th dimension using cubes in the book **Fourth Dimension**.

Another important contribution to this field was made by **Alicia Boole Stott**, a woman mathematician, in the 19th century. The main idea behind Boole Stott's method was to transform a 4D problem into a 3D one, making it possible to use our intuition on the 3D

space to solve a 4D problem. Her results on this topic are collected in her publication of 1900, "On certain series of sections of the regular four-dimensional hypersolids."

In 1908, **Hermann Minkowski** presented a paper consolidating the role of time as the 4th dimension of **spacetime**, the basis for **Einstein's theories of special and general relativity**.

Vectors

Mathematically, a 4D space is simply a space with four spatial dimensions, i.e., a space that needs four parameters to specify a point in it. For example a general point might have position vector a, where, $a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}$.

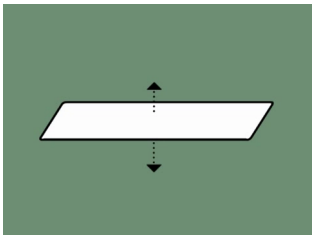
This can be written in terms of the four standard basis vectors $\{e_1, e_2, e_3, e_4\}$, given by:

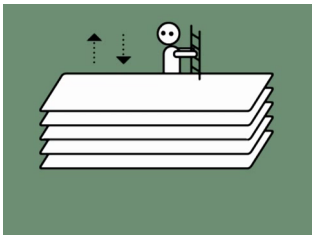
$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$; $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$; $e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$; $e_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$.

So, general vector a is $a = a_1e_1 + a_2e_2 + a_3e_3 + a_4e_4$.

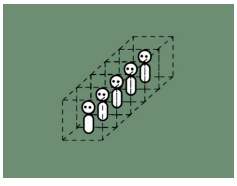
Visualizing the fourth dimension

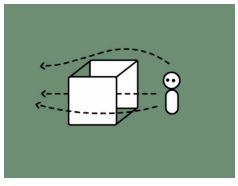
Firstly, if we consider a 2D world, it is impossible to go up or down. A 3D world implies that there are infinite planes adjacent to this 2D world. A 3D entity can see across a multitude of 2D planes and move up and down with ease. In a 3D world, we can move





horizontally and vertically but not in the 4D direction. A 4D world implies that there are infinite 3D spaces adjacent to our 3D world. We can only see one, the one we are in. If we could move 4D+ for a while and go back 4D− for the same while, we would end up in the same place, we didn't go up, down, left, right, forward or backward, **we went across**. We



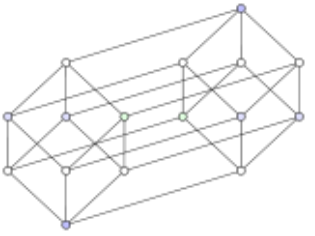


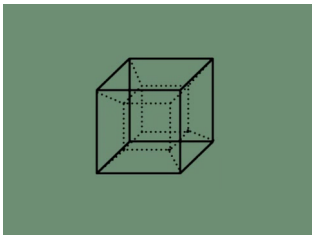
can imagine moving across several 3D rooms similar to ours, but each of them is actually a different 3D space. A 4D entity is able to see across a multitude of 3D spaces.

To understand easily, we can imagine a society of 2D beings, which are conscious of only 2 dimensions. Let us further imagine that these "flatlanders" live on the surface of a large sphere. Can they know this? If they are mathematically trained, they can measure that the sum of the interior angles of large triangles clearly exceeds the 180°,which is never expected in a flat Euclidean plane. They would have to conclude that their 2D space is not flat, but curved. From that they could deduce that their perceived 2D world is embedded in a reality of at least 3 dimensions. This does not imply that they can really imagine the 3rd dimension, but they would at any rate have proved its existence.

We ourselves are in an analogous situation: the space that we can habitually imagine has 3 dimensions. As we are sure that it is curved, we are also certain that our perceived 3D-space is embedded in a reality of at least 4 dimensions.

An object in 4th dimension consists of four units. For example, a hypercube has a length, width, height and a 4th dimension that is perpendicular to all three of the other units. If we can visualize stacking cubes into this 4th dimension, we create a hypercube.





Notations and Terms

In the familiar 3D space that we live in there are three coordinate axes usually labeled **x**, **y**, and **z** with each axis orthogonal (i.e., perpendicular) to the other two. The six cardinal directions in this space can be called *up*, *down*, *east*, *west*, *north* and *south*.

Positions along these axes can be called *altitude, longitude*, and *latitude*. Lengths measured along these axes can be called **height, width**, and **depth**.

Comparatively, 4D space has an extra coordinate axis, orthogonal to the other three, which is usually labeled **w**. To describe the two additional cardinal directions, **Charles Howard Hinton** coined the terms **ana** and **kata**, from the Greek words meaning **"up toward"** and **"down from"**, respectively. A length measured along the w axis can be called **spissitude**, as coined by **Henry More**.

Geometry

The geometry of a 4D space is much more complex than that of a 3D space, due to an extra degree of freedom.

- In 3-dimensions, there are polyhedra made of 2D polygons.
  In 4-dimensions, there are polychora (4-polytopes) made of polyhedra.


Tetrahedron
(4 faces)

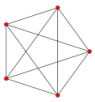

Cube
(6 faces)

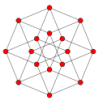

Octahedron
(8 faces)


Dodecahedron
(12 faces)


Icosahedron
(20 faces)

- In 3-dimensions there are five regular poyhedra known as the Platonic Solids.
  In 4-dimensions, there are 6 convex regular polychora, the analogues of the Platonic Solids.

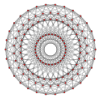

5-cell


8-cell

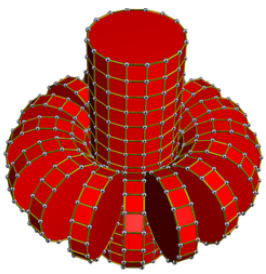

16-cell

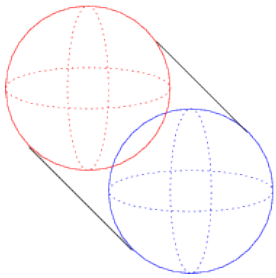

24-cell


120-cell

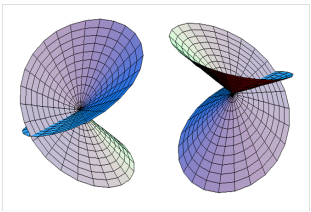

600-cell

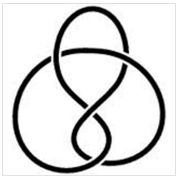
- In 3-dimensions, a circle may be extruded to form a cylinder.
  In 4-dimensions, a sphere may be extruded to obtain a spherical cylinder (a cylinder with spherical caps), and a cylinder may be extruded to obtain a cylindrical prism.

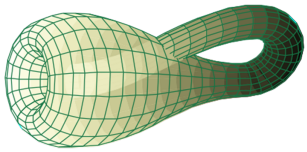




- In 3-dimensions, curves can form knots but surfaces cannot, unless they are self-intersecting. In 4 dimensions however, knots made using curves can be trivially


A Self Intersecting disk


A mathematical knot

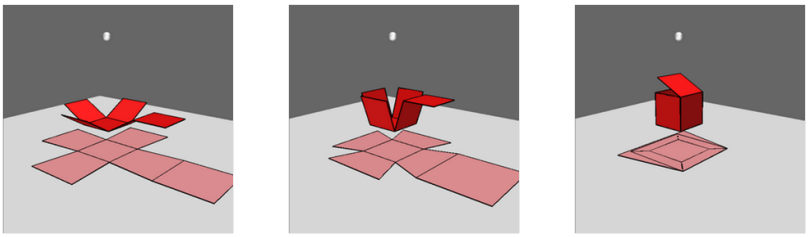

Klein Bottle: A knotted surface

untied by displacing them in the 4th direction, but 2D surfaces can form non-trivial, non-self-intersecting knots in the 4D.
A picture of Klien Bottle is shown on previous page. It is an example of a knotted surface in 4D. By adding a 4th dimension to the 3D space, the self intersection can be eliminated. To understand this, we can consider an analogy. If we consider a self-intersecting curve on the plane, we can eliminate the self-intersections by lifting one strand off the plane.

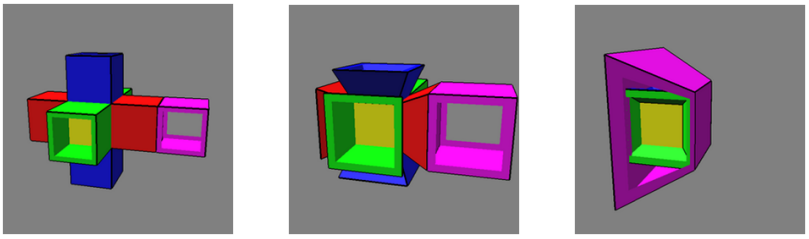
Tesseract- A four dimensional analog of a cube

The tessaract, also called an **8-cell** or **regular octachoron** or **cubic prism**, is the four-dimensional analog of the cube. A generalization of the cube to dimensions greater than three is called a "hypercube". The tesseract is the **four-dimensional hypercube**, or **4-cube**.
The word tesseract was coined and first used in 1888 by Charles Howard Hinton in his book "*A New Era of Thought*", from the Greek $\tau\varepsilon\sigma\sigma\varepsilon\rho\varepsilon\iota\varsigma\ \alpha\kappa\tau\iota\nu\varepsilon\zeta$ ("four rays"), referring to the four lines from each vertex to other vertices.

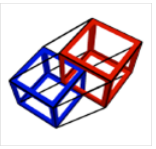
Geometry: Just like we obtain a cube by folding its 6 faces, which are all squares, we obtain a hypercube (tesseract) by folding its 8 cells, which are cubes.



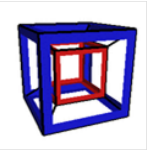
We begin with eight cubes forming a cross-like shape. Some faces are partially removed to make the interior structure easier to see. The central (yellow) cube will be the bottom of the hypercube, and the purple one will be the top. The remaining six cubes form the faces of the hypercube that join the bottom to the top. As these begin to fold up in the fourth dimension, we see their shadows become distorted in three dimensions (as one face of the cubes moves closer to the light source, its shadow get larger). Eventually, the faces of the cubes come together and are joined, just as the edges of the squares that form a cube are glued when they are folded together. This leaves just the top remaining to fold into place.



As the top closes in to join the six other faces, we are left with the well-known "cube within a cube" view of the hypercube in perspective. A tesseract has 24 faces, 32 edges, 16 vertices

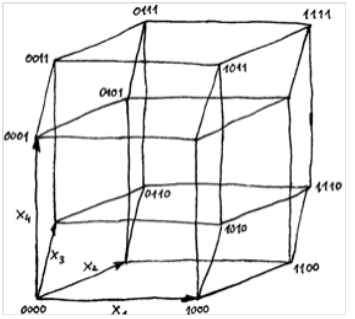


The Hypercube as linked cubes



Stereographic Projection of Hypercube Cube-first

and 8 cubical cells. Since each vertex of a tesseract is adjacent to four edges, the vertex figure of the tesseract is a regular tetrahedron. Three cubes and three squares intersect at each edge. There are four cubes, six squares, and four edges meeting at every vertex.
The coordinates of corner points of a tesseract are : (1,1,1,1), (1,1,1,0), (1,1,0,1), (1,1,0,0), (1,0,1,1), (1,0,1,0), (1,0,0,1), (1,0,0,0), (0,1,1,1), (0,1,1,0), (0,1,0,1), (0,1,0,0), (0,0,1,1), (0,0,1,0), (0,0,0,1), (0,0,0,0).



A Wire model of Tesseract in 3D space. Its coordinates are marked.

References

[1] T.Gosset, *On the Regular and Semi-Regular Figures in Space of n Dimensions*, Messenger of Mathematics, Macmillan, 1900.
[2] Coxeter, *Regular Polytopes*, London, Methuen and Co, 1948.
[3] John H. Conway, Heidi Burgiel, Chaim Goodman-Strass, *The Symmetries of Things*, 2008.
[4] Irene Polo-Blanco, Jon Gonzalez-Sanchez, *Four-Dimensional Polytopes: Alicia Boole Stott's Algorithm*, The Mathematical Intelligencer, September 2010, Volume 32, Issue 3, pp 1-6
[5] www.math.union.edu/ dpvc/math/4d/welcome.html
[6] www.wikipedia.org

SRUTHI SEKAR, B.Sc.(H) MATHEMATICS, 4th SEMESTER, LADY SHRI RAM COLLEGE FOR WOMEN
E-mail address: sruthi.sekar1@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.

REPRODUCING KERNEL HILBERT SPACES

MAHESH KUMAR AND H.B.SAHANA

Abstract. In functional analysis, a reproducing kernel Hilbert space (RKHS) is a Hilbert space of functions in which point-wise evaluation is a continuous (bounded) linear functional. This article explains the basic concepts and introduces to the theory of reproducing kernel Hilbert spaces and attempts to shed light on some of the applications of this field of mathematics.

Introduction

The topic of reproducing kernel Hilbert spaces (hereafter abbreviated as RKHS) was used for the first time by S.Zaremba in 1907, while working on boundary value problems. He introduced, in a particular case, the kernel associated with a class of functions and stated its reproducing property. But he did not develop any theory nor give any particular name to the kernels he introduced. In 1909, J.Mercer examined functions that satisfy the reproducing property in the theory of integral equations, and called such functions positive definite kernels. The main idea of reproducing kernels appeared in the dissertations of three Berlin mathematicians, G.Szego (1921), S.Bergman (1922) and S.Bochner (1922). In particular, Bergman introduced reproducing kernels in one and several variables for the class of harmonic and analytic functions, and called them kernel functions. In 1935, E.H.Moore examined the positive definite kernels under the name of positive Hermitian matrix. The credit for systemising the theory of reproducing kernels, however, goes to N.Aronszajn in 1948, whose classic paper on the subject is still referred to.

Prerequisites

(i) An inner product space (or pre-Hilbert space) is a vector space X with an inner product $\langle\,,\,\rangle$ defined on X. A **Hilbert space** is a complete inner product space; complete in the metric defined by the inner product

$$d(x,y) = \|x-y\| = \sqrt{\langle x-y,\, x-y\rangle}.$$

We will consider Hilbert spaces over either the field of real or complex numbers. We will use $\mathbb{F}$ to denote either $\mathbb{R}$ or $\mathbb{C}$, so that we may state results which are true for both simply by using one notation.

(ii) A **linear functional** f is a linear operator with domain in a vector space X and range in the scalar field $\mathbb{F}$ of X. A **bounded** linear functional f on a Hilbert space X is a linear functional for which there exists a real number c such that

$$|f(x)| \leq c\|x\| \quad \text{for all } x \in X.$$

(iii) Given a set $\mathbb{X}$, if we equip the set of functions from $\mathbb{X}$ to $\mathbb{F}$, denoted by $\mathcal{F}(\mathbb{X}, \mathbb{F})$, with the usual operations of addition

$$(f+g)(x) = f(x) + g(x),$$

and scalar multiplication

$$(\lambda.f)(x) = \lambda.f(x),$$

then $\mathcal{F}(\mathbb{X}, \mathbb{F})$ is a vector space over $\mathbb{F}$.

Definition

Given a set $\mathbb{X}$, we say $\mathcal{H}$ forms an RKHS on $\mathbb{X}$ over $\mathbb{F}$, provided that

(i) $\mathcal{H}$ is a vector subspace of $\mathcal{F}(\mathbb{X}, \mathbb{F})$,
(ii) $\mathcal{H}$ is endowed with an inner product, $\langle\ ,\ \rangle$, making it a Hilbert space,
(iii) For every y $\in \mathbb{X}$, the linear evaluation functional $E_y : \mathcal{H} \to \mathbb{F}$, defined by $E_y(f) = f(y)$ is bounded.

We know that every bounded linear functional on a Hilbert space can be given by the inner product with a unique vector in it. So, if $\mathcal{H}$ is an RKHS on $\mathbb{X}$, then we have that for every $y \in \mathbb{X}$, there exists a unique vector (function) $k_y \in \mathcal{H}$, such that for every $f \in \mathcal{H}$, $E_y(f) = \langle f, k_y\rangle = f(y)$. The function k_y is called the **reproducing kernel for the point y.** The 2-variable function $K : \mathbb{X} \times \mathbb{X} \to \mathbb{F}$, defined by

$$K(x, y) = k_y(x) = \langle k_y, k_x\rangle$$

is called the **reproducing kernel for** $\mathcal{H}$.

The term 'reproducing' is used in the sense that the Hilbert space $\mathcal{H}$ can be recovered from the functions k_y, as their linear span is dense in $\mathcal{H}$. Recall that, a subset of a Hilbert space is said to be dense if its closure is equal to the whole space. In fact, if the orthogonal complement of the closure of the linear span of a subset of a Hilbert space is the trivial space, then that linear span must be dense in that Hilbert space.

Note that a function $f \in \mathcal{H}$ is orthogonal to the closed linear span of the functions $\{k_y, y \in \mathbb{X}\}$ if and only if $\langle f, k_y\rangle = f(y) = 0$ for all $y \in \mathbb{X}$, which holds if and only if $f = 0$. This proves that the span of k_y's is dense in $\mathcal{H}$, and so the Hilbert space can be reproduced with the help of these k_y's.

It is also interesting to note that in the case that $\mathbb{F} = \mathbb{R}$, the Hilbert space can be **complexified**. Let $\mathcal{H}$ be an RKHS of real valued functions on the set $\mathbb{X}$ with reproducing kernel $K(x, y)$. Let $\mathcal{W} = \{f_1 + if_2 : f_1, f_2 \in \mathcal{H}\}$, which is the vector space of complex valued functions on $\mathbb{X}$. Set

$$\langle f_1 + if_2,\, g_1 + ig_2\rangle_{\mathcal{W}} = \langle f_1,\, g_1\rangle_{\mathcal{H}} + \langle f_2,\, g_2\rangle_{\mathcal{H}} + i\langle f_2,\, g_1\rangle_{\mathcal{H}} - i\langle f_1,\, g_2\rangle_{\mathcal{H}}.$$

Then this defines an inner product on $\mathcal{W}$, with norm $\|f_1 + if_2\|_w^2 = \|f_1\|^2 + \|f_2\|^2$. Thus, $\mathcal{W}$ forms a Hilbert space. Also, as

$$f_1(y) + if_2(y) = \langle f_1 + if_2, k_y\rangle = (f_1 + if_2)(y),$$

we have that $\mathcal{W}$ equipped with this inner product is an RKHS of complex-valued functions on $\mathbb{X}$, with kernel $K(x, y)$. The space $\mathcal{W}$ is called the **complexification** of $\mathcal{H}$, which still preserves the reproducing kernel. Henceforth, we shall consider only the case of complex-valued RKHS since every real-valued RKHS can be complexified in a way that still preserves the reproducing kernel.

Examples

Sometimes to fix ideas it helps to look at a non-example. Suppose that we take the continuous functions on $[0, 1]$, $C[0, 1]$, define the usual 2-norm on this space, i.e., $\|f\|^2 = \int_0^1 \|f\|^2\, dt$, and complete to get the Hilbert space $L^2[0, 1]$. Now, given any point $x_0 \in [0, 1]$, one can always construct a sequence $f_n \in C[0, 1]$, such that $\lim_n \|f_n\| = 0$, and $\lim_n f_n(x_0) = +\infty$, for example let g_n be the piecewise linear function defined on [0,1] with support on $[x_0 - 1/n^2,\, x_0 + 1/n^2]$ and $g_n(x_0) = n$, as shown in figure below for $n = 1\,, 2,\, 3$, etc, and define $f_n = \sqrt{g_n}$ for each $n \in \mathbb{N}$. Here, we have taken $x_0 = 1/2$. Note that $\|f_n\| = 1/n$ and $f_n(x_0) = \sqrt{n}$ for each $n \in \mathbb{N}$. Thus, E_{x_0} is not bounded for any x_0 in [0,1] and hence $L^2[0, 1]$

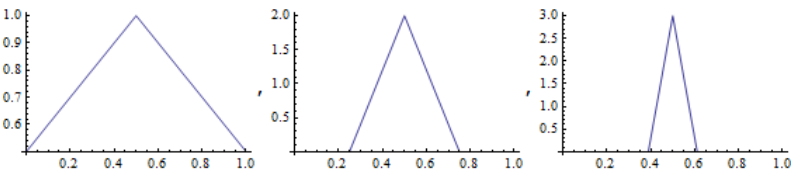


FIGURE 1. Graph of g_n for $n = 1\,, 2,\, 3$

is not an RKHS on $[0, 1]$. Thus, reproducing kernel Hilbert spaces are quite different from L^2-spaces.

We now look at a few key examples.

The Hardy Space of the Unit Disk, $H^2(\mathbb{D})$. This space is defined as

$$H^2(\mathbb{D}) = \left\{ f : \mathbb{D} \to \mathbb{C} \,\middle|\, f(z) = \sum_{n=0}^{\infty} a_n z^n, z \in \mathbb{D}, \sum_{n=0}^{\infty} |a_n|^2 < \infty \right\}.$$

Endow $H^2(\mathbb{D})$ with the inner product

$$\langle f, g \rangle = \sum_{n=0}^{\infty} a_n \overline{b_n},$$

where $f,\, g \in H^2(\mathbb{D})$ with $f(z) = \sum_{n=0}^{\infty} a_n z^n$, $g(z) = \sum_{n=0}^{\infty} b_n z^n$. Then the map $L : H^2(\mathbb{D}) \to l^2(\mathbb{Z}^+)$ defined as $L(\sum_{n=0}^{\infty} a_n z^n) = (a_0, a_1, \ldots, a_n, \ldots)$ is a linear, inner product preserving isomorphism, $\mathbb{Z}^+ = \mathbb{N} \cup \{0\}$. Thus, $H^2(\mathbb{D})$ can be identified with $l^2(\mathbb{Z}^+)$, which

is a Hilbert Space. Next, we see that for $z \in \mathbb{D}$,

$$|E_z(f)| = |f(z)| = \left| \sum_{n=0}^{\infty} a_n z^n \right| \leq \sum_{n=0}^{\infty} |a_n||z^n| \leq \left(\sum_{n=0}^{\infty} |a_n|^2 \right)^{1/2} \left(\sum_{n=0}^{\infty} |z^{2n}| \right)^{1/2}$$
$$= \|f\| \frac{1}{(1-|z|^2)^{1/2}},$$

using the Cauchy-Schwartz inequality. Thus, $\|E_z\| \leq \frac{1}{(1-|z|^2)^{1/2}}$, which shows that E_z is bounded for all $z \in \mathbb{D}$. This also shows that each power series defines a function on $\mathbb{D}$ and the usual vector space operations are well-defined. Thus, $H^2(\mathbb{D})$ defines an RKHS on $\mathbb{D}$.

To compute the kernel, let $w \in \mathbb{D}$. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$, $g(z) = \sum_{n=0}^{\infty} \overline{w}^n z^n$ Then, $\langle f, g \rangle = \sum_{n=0}^{\infty} a_n w^n = f(w)$. Thus g is the reproducing kernel for w and so

$$K(z, w) = k_w(z) = g(z) = \sum_{n=0}^{\infty} \overline{w}^n z^n = \frac{1}{1 - \overline{w}z}.$$

This function is called the **Szego kernel** on the disk.

Sobolev Spaces on $[0, 1]$**.** These are very simple examples of the types of Hilbert spaces that arise in differential equations. Let

$$\mathcal{H} = \left\{ f : [0, 1] \to \mathbb{R} : f \text{ is absolutely continuous, } f(0) = f(1) = 0, \ f' \in L^2[0, 1] \right\}$$

Recall that a function $f : [0, 1] \to \mathbb{R}$ is absolutely continuous if and only if it is differentiable almost everywhere, $f' \in L^1[0, 1]$ and is equal to the integral of its derivative modulo $f(0)$, i.e., $f(x) = f(0) + \int_0^x f'(t)\, dt$. In this case, $f(x) = \int_0^x f'(t)\, dt$, since $f(0) = 0$ [see [8, Cor. 20.18] for more details]. Clearly, $\mathcal{H}$ is a vector space of functions on [0,1]. Endow $\mathcal{H}$ with the non-negative, sesquilinear form,

$$\langle f, g \rangle = \int_0^1 f'(t) g'(t)\, dt.$$

Since f is absolutely continuous and $f\,(0) = 0$, for any $0 \leq x \leq 1$, we have that,

$$f(x) = \int_0^x f'(t)\, dt = \int_0^1 f'(t) \chi_{[0,x]}(x)\, dt.$$

Thus, by the Cauchy-Schwartz inequality,

$$|f(x)| \leq \left(\int_0^1 f'(t)^2\, dt \right)^{1/2} \left(\int_0^1 \chi_{[0,x]}(t)\, dt \right)^{1/2} = \|f\| \sqrt{x}.$$

This last inequality shows that $\|f\| = 0$ if and only if $f = 0$ .Thus $\langle , \rangle$ is an inner product on $\mathcal{H}$ and that for every $x \in [0, 1]$, E_x is bounded with $\|E_x\| \leq \sqrt{x}$.

All that remains to show that $\mathcal{H}$ is an RKHS, is to show that it is complete. If $\{f_n\}$ is a Cauchy sequence in this norm, then $\{f_n'\}$ is Cauchy in $L^2[0, 1]$ and hence there exists $g \in L^2[0, 1]$ that this sequence converges to. By the above inequality, $\{f_n\}$ must be pointwise Cauchy and hence we may define a function by setting $f(x) = \lim_n f_n(x)$. Since,

$$f(x) = \lim_n f_n(x) = \lim_n \int_0^x f_n'(t)\, dt = \int_0^x g(t)\, dt,$$

it follows that f is absolutely continuous and that $f' = g$, a.e. and hence, $f' \in L^2[0,1]$. Finally, $f(0) = \lim_n f_n(0) = 0 = \lim_n f_n(1) = f(1)$, and so $f \in \mathcal{H}$.

Thus, $\mathcal{H}$ is a RHKS on [0,1].

It remains to find the kernel function. To do this we first formally solve a differential equation and then show that the function we obtain by this formal solution, belongs to $\mathcal{H}$.

To find $k_y(t)$, we apply integration by parts to see that,

$$f(y) = \langle f, k_y \rangle = \int_0^1 f'(t)\, k_y'(t)\, dt = f(t)\, k_y'(t)\Big|_{t=0}^1 - \int_0^1 f(t)\, k_y''(t)\, dt = -\int_0^1 f(t)\, k_y''(t)\, dt.$$

If we let, δ_y denote the formal **Dirac-delta function**, then $f(y) = \int_0^1 f(t)\delta_y(t)\, dt$, where

$$\delta_y(t) = \delta(t-y) = \begin{cases} 1, & \text{if } t = y \\ 0, & \text{if } t \neq y \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(t) = 1.$$

[For more details about Dirac-delta function, see [7, sec 3.4]]. Thus, we need to solve the boundary-value problem,

$$-k_y''(x) = \delta_y(x), \quad k_y(0) = 0 = k_y(1).$$

The solution to this system of equations is called the **Green's Function** for the differential equation. Solving formally, by integrating twice and checking the boundary conditions we find that

$$K(x,y) = k_y(x) = \begin{cases} (1-y)x, & \text{if } x \leq y \\ (1-x)y, & \text{if } x \geq y. \end{cases}$$

The computations are as follows. First, notice that

$$\int_{-\infty}^{x} \delta(t)\, dt = H(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0. \end{cases}$$

So, integrating the equation $-k_y''(x) = \delta_y(x)$ w.r.t. x, implies

$$k_y'(x) = \begin{cases} A(y) - 1, & \text{if } y < x \\ A(y), & \text{if } x \leq y, \end{cases}$$

where $A(y)$ is an arbitrary function, which again, on integrating implies,

$$k_y(x) = \begin{cases} (A(y)-1)x + B_1(y), & \text{if } y \leq x \\ A(y)x + B_2(y), & \text{if } x \leq y, \end{cases}$$

where $B_1(y), B_2(y)$ are arbitrary functions. As $k_y(0) = 0$, we have $B_2(y) = 0$. Also,

$$k_y(y) = A(y)y + B_1(y) - y = A(y)y, \quad \text{which implies } B_1(y) = y.$$

So,

$$k_y(x) = \begin{cases} A(y)x + y - x, & \text{if } y \leq x \\ A(y)x, & \text{if } y \geq x \end{cases}$$

As $k_y(1) = 0$, $A(y) + y - 1 = 0$, which implies $A(y) = 1 - y$ and we have the required kernel.

Bergman Spaces on Complex Domains. Let G be an open and connected subset of $\mathbb{C}$. Set

$$B^2(G) = \left\{ f : G \to \mathbb{C} \,|\, f \text{ is analytic on } G \text{ and } \iint_G |f(x+iy)|^2 dxdy < +\infty \right\}$$

Define a sesquilinear form on $B^2(G)$ as

$$\langle f, g \rangle = \iint_G f(x+iy)\overline{g(x+iy)}\, dxdy.$$

Then this defines an inner product on $B^2(G)$, so that it becomes an inner product space. In fact, $B^2(G)$ forms an RKHS on G [see [1] for details]. The reproducing kernel for $B^2(G)$ is called the **Bergman kernel** for G.

The result above extends to open connected subsets of $\mathbb{C}^n$ and to many complex manifolds. Knowledge of the Bergman kernel of such such domains has many applications in complex analysis and the study of this kernel is still an active area of research. It is remarked that $B^2(\mathbb{C}) = (0)$. Thus, the only analytic function defined on the whole complex plane that is square-integrable is the 0 function. Also, for $G = \mathbb{D}$, the Bergman kernel for $B^2(\mathbb{D})$ is given by $K(z, w) = \frac{1}{(1-z\overline{w})^2}$.

Moore-Aronszajn Theorem

This theorem first appeared in Aronszajn's *Theory of Reproducing Kernels*, Transactions of the American Mathematical Society 68(3) in 1950, [2]; although he attributed it to E.H.Moore. The theorem gives a necessary and sufficient condition for a function $K(x, y)$ to be the reproducing kernel for some RKHS, thereby providing a characterization of reproducing kernels. We state some preliminaries and make some of the remarks before stating the theorem.

(i) Let $A = (a_{ij})$ be an $n \times n$ complex matrix. Then A is said to be **positive**, denoted by $A \geq 0$, if and only if for every $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$, we have that $\sum_{i,j=1}^{n} \overline{\alpha_i}\alpha_j a_{ij} \geq 0$.

(ii) If $\langle\ ,\ \rangle$ is the usual inner product on $\mathbb{C}^n$, given by

$$\langle x,\, y \rangle = y^* x = \sum_{i=1}^{n} x_i \overline{y_i}$$

then in terms of the inner product, $A \geq 0$ if and only if $\langle Ax,\, x \rangle \geq 0$ for all $x \in \mathbb{C}^n$. It is remarked that, $A \geq 0$ if and only if $A = A^*$ and every eigenvalue, λ, of A, satisfies $\lambda \geq 0$. For this reason, some authors might prefer to call such matrices positive semi-definite or non-negative. In the case that $A = A^*$ and every eigenvalue, λ, of A, satisfies $\lambda > 0$, then we call A **strictly positive**, denoted by $A > 0$. Since A is a matrix, we see that $A > 0$ is equivalent to $A \geq 0$ and invertible.

(iii) Let $\mathbb{X}$ be a set and let $K : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$ be a function of two variables. Then K is called a **kernel function**, denoted by $K \geq 0$, if and only if for every n and every choice of n distinct points $\{x_1, x_2, \ldots, x_n\}$ in $\mathbb{X}$, the matrix $(K(x_i, x_j))_{n\times n} \geq 0$.

Please note that some authors refer to such functions as positive definite functions, while others call them positive semi-definite functions. This terminology is not standard, but we

have adopted it following Aronszajn's terminology. Also, we shall see that there is a one-to-one correspondence between kernel functions and reproducing kernels.

It is clear from the definition of the RKHS that the reproducing kernel defined is conjugate symmetric and a kernel function, as

$$\sum_{i,j=1}^{n} \overline{\alpha_i}\alpha_j K(x_i,x_j) = \sum_{i,j=1}^{n} \overline{\alpha_i}\alpha_j \left\langle k_{x_j}, k_{x_i} \right\rangle = \left\langle \sum_{j=i}^{n} \alpha_j k_{x_j}, \sum_{i=1}^{n} \alpha_i k_{x_i} \right\rangle = \left\| \sum_{j=1}^{n} \alpha_j k_{xj} \right\|^2 \geq 0$$

for every choice $\{x_1, x_2, \ldots, x_n\}$ in X and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$.

We would like to remark here that, generally for a reproducing kernel, $(K(x_j, x_j)) > 0$, for if not, then it follows from the above equation that there must exist some non zero vector such that $\| \sum_j \alpha_j k_{x_j} \| = 0$.

Hence, for every $f \in \mathcal{H}$ we have that $\sum_j \overline{\alpha_j} f(x_j) = \left\langle f, \sum_j \alpha_j k_{x_j} \right\rangle = 0$.

Thus, in this case there is an equation of linear dependence between the values of every function in $\mathcal{H}$ at some finite set of points.

Such examples do naturally exist. Recall that in the Sobolev spaces on $[0,1]$, we were interested in spaces with boundary conditions, like $f(0) = f(1)$, in which case $k_1(t) = k_0(t)$, since $f(0) = f(1)$ implies $\langle f, k_0 \rangle = \langle f, k_1 \rangle$, which imples $\langle f, k_0 - k_1 \rangle = 0$. Since $f \neq 0$, $k_0(t) = k_1(t)$.

Alternatively, many spaces of analytic functions, such as the Hardy or Bergman spaces, contains all polynomials. Note that there is no equation of the form $\sum_j \beta_j p(x_j) = 0$, that is satisfied by all polynomials (try with $1, x, x^2, \ldots, x^n$). Consequently, the reproducing kernels for these spaces always define matrices that are strictly positive and invertible!

Thus, for example, recalling the Szego kernel for the Hardy space, we see that for any choice of points in the disk, the matrix, $\left(\frac{1}{1-\overline{\lambda_i}\lambda_j}\right)$is invertible.

For one glimpse into how powerful the theory of RHKS can be, one can try to show this matrix to be invertible by standard linear algebraic methods.

Moore-Aronszajn Theorem. We have seen above that the reproducing kernel of an RKHS defines a kernel function. The **Moore-Aronszajn theorem** states the converse of this result:

Let $\mathbb{X}$ be a set and let $K : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$ be a function. If K is a kernel function, then there exists an RKHS $\mathcal{H}$ of functions on $\mathbb{X}$ such that K is the reproducing kernel of $\mathcal{H}$.

The proof of this theorem involves the construction of the required RKHS, by defining functions $k_y(x) = K(x,y)$ for each $y \in \mathbb{X}$, and taking W to be the space of functions spanned by these k_y's. We then complete W by taking equivalence classes of Cauchy sequences from W, to obtain a complete Hilbert space $\mathcal{H}$.

It is remarked that if $\mathcal{H}_i$, $i = 1, 2$ are RKHS's on $\mathbb{X}$ with kernels, $K_i(x,y)$, $i = 1, 2$, and if $K_1(x,y) = K_2(x,y)$ for all $x, y \in X$, then $\mathcal{H}_1 = \mathcal{H}_2$ and $\|f\|_1 = \|f\|_2$ for every f. This shows that there is a one-to-one correspondence between RKHS's on a set and kernel functions on that set. Thus, given a kernel function $K : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$, there exists a unique RKHS $\mathcal{H}$ on $\mathbb{X}$ with reproducing kernel K.

A Method for calculating reproducing kernels

Let $\mathcal{H}$ be an RKHS on $\mathbb{X}$ with reproducing kernel K and let $\{f_n\} \subseteq \mathcal{H}$. As

$$|f_n(x) - f(x)| = |\langle f_n - f, k_x\rangle| \leq \|f_n - f\| \, \|k_x\|,$$

we see that $\lim_n \|f_n - f\| = 0$, implies that $f(x) = \lim_n f_n(x)$ for every $x \in \mathbb{X}$.

This result although looks quite simple, but gives us another means of calculating the kernel for an RKHS that is very useful.

Recall that for given vectors $\{h_s : s \in S\}$ in a Hilbert space $\mathcal{H}$, indexed by an arbitrary set S, we say that $h = \sum_{s\in S} h_s$, provided that for every $\varepsilon > 0$, there exists a finite subset $F_0 \subseteq S$, such that for any finite set F, $F_0 \subseteq F \subseteq S$, we have that $\left\|h - \sum_{s\in F} h_s\right\| < \varepsilon$. Two examples of this type of convergence are given by the two **Parseval Identities**. When $\{e_s : s \in S\}$ is an orthonormal basis for a Hilbert space, $\mathcal{H}$, then for any $h \in \mathcal{H}$, we have that

$$\|h\|^2 = \sum_{s\in S} |\langle h, e_s\rangle|^2 \text{ and } h = \sum_{s\in S} \langle h, e_s\rangle e_s.$$

Note that these types of sums do not need S to be an ordered set to be defined. Perhaps, the key example to keep in mind is that if we set $a_n = (-1)^n/n$, $n \in \mathbb{N}$, then the series, $\sum_{n=1}^{\infty} a_n$ converges, but $\sum_{n\in\mathbb{N}} a_n$ does not converge. In fact, for complex numbers, one can show that $\sum_{n\in\mathbb{N}} z_n$ converges if and only if for every permutation π of $\mathbb{N}$, $\sum_{n=1}^{\infty} z_{\pi(n)}$ converges, which is if and only if $\sum_{n=1}^{\infty} |z_n|$ converges. Thus, this convergence is equivalent to absolute convergence in the complex case.

The following result gives us another means of calculating the kernel for an RKHS.

Theorem. *Let $\mathcal{H}$ be an RKHS on $\mathbb{X}$ with reproducing kernel, $K(x, y)$. If $\{e_s : s \in S\}$ is an orthonormal basis for $\mathcal{H}$, then $K(x, y) = \sum_{s\in S} \overline{e_s(y)} e_s(x)$ where this series converges pointwise.*

Proof. For any $y \in X$, we have that

$$\langle k_y, e_s\rangle = \overline{\langle e_s, k_y\rangle} = \overline{e_s(y)}.$$

Hence, $k_y = \sum_{s\in S} \overline{e_s(y)} e_s$, where these sums converge in the norm on $\mathcal{H}$. But since they converge in the norm, they converge at every point. Hence

$$K(x, y) = k_y(x) = \sum_{s\in S} \overline{e_s(y)} e_s(x). \qquad \square$$

For a quick example of this theorem, note that in the Hardy space, the functions $e_n(z) = z^n$, $n \in \mathbb{Z}^+$ form an orthonormal basis and hence, the reproducing kernel for the Hardy space is given by

$$\sum_{n\in\mathbb{Z}^+} e_n(z)\overline{e_n(w)} = \sum_{n=0}^{\infty} (z\overline{w})^n = 1/(1 - z\overline{w}).$$

Returning to our earlier example of Sobolev space on $[0, 1]$,

$$\mathcal{H} = \left\{f : [0, 1] \to \mathbb{R} : f \text{ is absolutely continuous, } f(0) = f(1) = 0, f' \in L^2[0, 1]\right\},$$

it can be checked that for $n \neq 0$, the functions $e_n(t) = (e^{2\pi int} - 1)/2\pi n$ belongs to $\mathcal{H}$ and are orthonormal.

If $f \in \mathcal{H}$ and $0 = \langle f, e_n \rangle = i \int_0^1 f'(t) e^{2\pi int}\, dt$, for all $n \in \mathbb{N}$, then since the function, $\{e^{2\pi int}\}_{n\neq 0}$ together with the constants span $L^2[0,1]$, we see that $f'(t)$ is constant and hence that $f(t)$ is a first degree polynomial. But the boundary conditions, $f(0) = f(1) = 0$, imply that this polynomial is 0.

Hence, we have shown that these functions are an orthonormal basis for $\mathcal{H}$. Applying the above theorem and our earlier calculation of the reproducing kernel, we have that,

$$\begin{aligned} \operatorname{Re}\left(\sum_{n\neq 0} \frac{(e^{2\pi inx} - 1)(e^{-2\pi iny} - 1)}{4\pi^2 n^2}\right) &= \sum_{n\neq 0} \frac{\cos 2\pi n(x-y) - \cos 2\pi nx - \cos 2\pi ny + 1}{4\pi^2 n^2} \\ &= \begin{cases} (1-y)x & \text{if } x \leq y \\ (1-x)y & \text{if } x \geq y \end{cases} \end{aligned}$$

The real part is considered because we know that the kernel function is real valued.

Applications

RKHS's are a very useful and powerful tool of functional analysis with application in many diverse paradigms. RKHS's arise in a number of areas, including approximation theory, statistics, machine learning theory, group representation theory, theory of one or several complex variables, study of invariant Riemannian metrics, interpolation problems, band-limited signal models/information control, multivariate statistics, fractal interpolation, boundary value problems, harmonic analysis, quantum mechanics(bound state problems), functional data analysis among others. They have been studied and used in the works of many mathematicians, such as A.N.Kolmogorov, E.Parzen, E.Hille, etc. We give one such application to illustrate the scope and importance of this field of analysis.

Band-limited Signal Models. In communication and information theory, band-limited signal models are used for analysis and representations. These models are used as they represent fairly well the signals encountered in practice. Basic properties of the abstract RKHS have been applied to study band-limited signals. In fact, the classes of finite energy Fourier-, Bessel-, Sine- and Cosine- transformed band-limited signals are specific realizations of RKHS.

In a given class of signals, there are certain properties of the signal that can be attained from extremal problems. If the given class of signals forms an RKHS, then the reproducing kernel plays a very important part in such problems. This in turn allows sampling expansions, along with specific truncation bound errors.

Now suppose K is the reproducing kernel of an RKHS $\mathcal{H}$. Consider a simple problem: Suppose t_0 is a fixed point in the set T of the abstract RKHS $\mathcal{H}$. What signal $f \in \mathcal{H}$, with specific value $f(t_0) = M$, where M is a real constant, has minimum energy $\|f\|^2$? On the other hand, what signal with energy $\|f\|^2 \leq E$ has a maximum value for $f(t_0)$?
It has been shown that the required signal for the first problem is given by

$$f(t) = M \frac{K(t,t_0)}{K(t_0,t_0)},$$

with minimum energy $\|f\|^2 = M^2$, in the subspace generated by the constraint $g(t_0) = M$. The solution to the second problem is given by the signal

$$f(t) = \pm E^{1/2} \frac{K(t,t_0)}{K(t_0,t_0)}.$$

We now extend the problem for n specified sampling instants, $f(t_i) = M_i, i = 1, 2, \ldots, n$, and consider the same problem of finding a signal $f \in H$ satisfying the above condition, with minimum energy. The solution to this slightly extended problem is given as a unique signal that interpolates over a finite number of points, and approximates any other specified signal with minimum energy. The solution is based on the Gram- Schmidt orthogonalization process used in interpolation theory, and the required signal is also generated by the reproducing kernel. This RKHS approach simplifies the solutions and offers a unified point of view. The interested readers are referred to [3] for more details.

References

1. Vern I. Paulsen, *An introduction to the theory of RKHS*, Lecture Notes, University of Houston, 2006.
2. N.Aronszajn, *Theory of Reproducing Kernels*, Transactions of the American mathematical Society, 68 (1950), 337- 404.
3. K.Yao, *Applications of RKHS- Bandlimited Signal Models*, UCLA, 1967.
4. P. Bouboulis, M. Mavroforakis, *Reproducing Kernel Hilbert Spaces and Fractal Interpolation*, Journal of Computational and Applied Mathematics, Volume 235, Issue 12, 2011.
5. *Reproducing Kernel Hilbert Spaces*, Lecture Notes, Bilkent University, Turkey, 2005.
6. E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley India Pvt. Ltd, 2007.
7. K. Sankara Rao, *Introduction to Partial differential equations*, Third edition, PHI learning private limited, 2012.
8. N.L. Carothers, *Real Analysis*, Cambridge University Press, 2000.

Mahesh Kumar, Assistant Professor, Department of Mathematics, Lady Shri Ram College For Women

E-mail address: mahekumar81@gmail.com

H.B.Sahana, M.A.Mathematics, Lady Shri Ram College For Women, University of Delhi

E-mail address: hbsahana@gmail.com

ALAN TURING'S REACTION-DIFFUSION SYSTEMS: UNDERSTANDING PATTERN FORMATION IN ANIMALS

IPSA BAJAJ, APURVA AGARWAL AND SRISHTI BANSAL

Abstract. This paper describes a model that provides a comprehensive explanation of Alan Turings reaction-diffusion systems. The reaction-diffusion model is widely studied in the field of mathematical biology and provides a basis for understanding pattern generation in several animals during their developmental stages. Further, we shall illustrate a method for reproducing numerical calculations with Microsoft Excel and discuss some examples of the patterns generated by the model.

Background - How it all began?

Haven't we often wondered how the cheetah got its spots or the zebra its stripes?

This very question was first addressed by the great mathematician **Alan Turing** in 1952 in his theoretical work **"The Chemical Basis of Morphogenesis."** It was observed that a small ball of cells (embryo), completely uniform (or homogeneous), during the early stages, gave rise to the dramatic patterns of a zebra, leopard, giraffe or butterfly.

Turing's Hypothesis

- Turings model hypothesized the existence of two molecules in the embryonic tissue- **activator** and **inhibitor**, collectively termed as **morphogens**.
- The "Inhibitor" is responsible for suppressing the production of both morphogens while, the "Activator" promotes their production.
- If the morphogens interact with each other in a specific manner, a periodic pattern is formed from a homogeneous initial distribution of activator and inhibitor.

Modeling Equations

Initial Conditions:

- For simplicity, let us consider a one-dimensional rod-like embryonic tissue.
- Define horizontal length as 1 and vertical length as $dy(<< 1)$.
- Cells inside this tissue produce two molecules - **activator** and **inhibitor**, and these control the production (or degradation) of both molecules and diffuse to neighbouring cells.

Discretization:

- Divide this rod-like structure into small pieces having horizontal length $dx(<< 1)$ and let the distribution of morphogens in these small pieces be homogeneous.
- Two factors affect the concentration of morphogens in these small pieces:
  1) Interaction of activator and inhibitor within each element.
  2) Transfer of activator and inhibitor between element and two nearest neighbors.
- Consider updating the system in discrete time steps dt. So, at time step m (m is a positive integer), a time of $m * dt$ has actually passed.
- Define concentration of activator molecule in the nth tissue element at time $m * dt$ as $p(n, m)$ and concentration of inhibitor as $q(n, m)$.

Reaction Term:

- Let rate of change of concentration of activator and inhibitor be $f(p, q)$ and $g(p, q)$, respectively.
- Then p (activator) and q (inhibitor) are called the 'reaction terms'.
- For simplicity we shall allow negative values of p and q and set their initial values to 0.

Diffusion Term:

- Consider the interaction between a tissue element and its two nearest neighbouring tissue elements during the time interval $(m * dt, (m + 1) * dt)$.
- Let concentration of activator in the nth tissue element be $p(n, m)$; then concentration of activator in the tissue element to the right will be $p(n + 1, m)$.
- The amount of activator transferred from element n to $n+1$ that is proportional to the concentration gradient $(p(n + 1, m) - p(n, m))/dx$ and transverse length of the element dy in time dt, is given as:

$$d_p * \frac{(p(n + 1, m) - p(n, m))}{dx} * dy * dt \tag{1}$$

where d_p represents the **diffusion coefficient** of the activator.
Then, the concentration change induced by this transfer is given as:

$$d_p * \frac{(p(n + 1, m) - p(n, m))}{dx^2} * dt \tag{2}$$

Similarly, the concentration change induced by the transfer of activator from the left neighbouring element is:

$$d_p * \frac{(p(n - 1, m) - p(n, m))}{dx^2} * dt \tag{3}$$

Taken together, the change of activator concentration in element n between time $m * dt$ and $(m + 1) * dt$ is:

$$d_p * \frac{(p(n + 1, m) + p(n - 1, m) - 2 * p(n, m))}{dx^2} * dt \tag{4}$$

Similarly, the concentration change of inhibitor is:

$$d_q * \frac{(q(n+1,m) + q(n-1,m) - 2 * q(n,m))}{dx^2} * dt \tag{5}$$

in which d_q is the **diffusion coefficient** of the inhibitor.

Governing Equations: Taking both reaction and diffusion into consideration, the concentrations of activator and inhibitor at time $(m+1) * dt$ are:

$$\begin{aligned} p(n,m+1) &= p(n,m) + \left(f\left(p(n,m), q(n,m)\right) + d_p \frac{p(n+1,m) + p(n-1,m) - 2 * p(n,m)}{dx^2} \right) dt \\ q(n,m+1) &= q(n,m) + \left(g\left(p(n,m), q(n,m)\right) + d_q \frac{q(n+1,m) + q(n-1,m) - 2 * q(n,m)}{dx^2} \right) dt \end{aligned} \tag{6}$$

Boundary Conditions: In the left-most and right-most tissue we define special conditions. If the total number of tissue elements is N_{total}, then $p(1,m)$ and $p(N_{total},m)$ have only one neighbour and cannot be treated as above; that is, Equations 6 only hold for integer n with n greater than 1 and less than N_{total}.

Periodic boundary condition: We define the values of $p(1,m)$ and $p(N_{total},m)$ to be equal for all m, so that the left-most tissue and right-most tissue are connected.

Zero-flux boundary condition: We assume that the boundary is impermeable. In this case, we have:

$$p(1,m+1) = p(1,m) + \left(f\left(p(1,m), q(1,m)\right) + d_p \frac{p(2,m) - p(1,m)}{dx^2} \right) dt \tag{7}$$

$$p(N_{total},m+1) = p(1,m) + \left(f\left(p(N_{total},m), q(N_{total},m)\right) + d_p \frac{p(N_{total}-1,m) - p(N_{total},m)}{dx^2} \right) dt \tag{8}$$

Fixed boundary condition: In this case, we have to set $p(1,m)$ and $p(N_{total},m)$ to specific values for all time. Hence:

$$p(1,m+1) = \alpha; \quad p(N_{total},m+1) = \beta, \tag{9}$$

where α and β are fixed (non-negative) numbers.

Transformation to continuous equation:

- Obtain the corresponding continuous differential equations by making dt and dx infinitely small.
- Define the concentration of activator and inhibitor as $u(x,t)$ and $v(x,t)$.
- Then, the above discrete governing equations become:

$$\frac{u(x,t+dt) - u(x,t)}{dt} = f(u(x,t), v(x,t)) + d_p \frac{\frac{u(x+dx,t) - u(x,t)}{dx} - \frac{u(x,t) - u(x-dx,t)}{dx}}{dx} \tag{10}$$

$$\frac{v(x,t+dt)-v(x,t)}{dt} = g(u(x,t),v(x,t)) + d_q \frac{\frac{v(x+dx,t)-v(x,t)}{dx} - \frac{v(x,t)-v(x-dx,t)}{dx}}{dx} \tag{11}$$

- If dt and dx tend to zero, then we have:

$$\begin{aligned} \partial u(x,t)/\partial t &= f(u(x,t),v(x,t)) + d_p \partial^2 u(x,t)/\partial x^2 \\ \partial v(x,t)/\partial t &= g(u(x,t),v(x,t)) + d_q \partial^2 v(x,t)/\partial x^2 \end{aligned} \tag{12}$$

Numerical calculation of the reaction-diffusion system: For instance, consider the following reaction-diffusion system:

$$\begin{aligned} \partial u/\partial t &= 0.6u - v - u^3 + 0.0002(\partial^2 u/\partial x^2) \\ \partial v/\partial t &= 1.5u - 2v + 0.01(\partial^2 v/\partial x^2) \end{aligned} \tag{13}$$

with domain size $[0, 1]$ and zero-flux boundary conditions.

Numerical calculation in Microsoft Excel:

- Use the system in Equations 13 with domain size 1 and take $dx = 0.05$ and $dt = 0.1$.
- So, we have 20 pieces of tissue ($1/0.05 = 20$) that contain both activator and inhibitor.
- In an Excel spreadsheet, let each column represent the concentration of activator or inhibitor in a specific tissue piece. This is expressed as a 20×2 matrix of numbers.
- Input the governing equations (Equation 6) to obtain concentration of activator (p) and inhibitor (q) at a certain tissue piece after time dt has passed.
- By repeating this 100 to 200 times, we observe that the concentrations gradually form a periodic structure as shown in Figure 1.

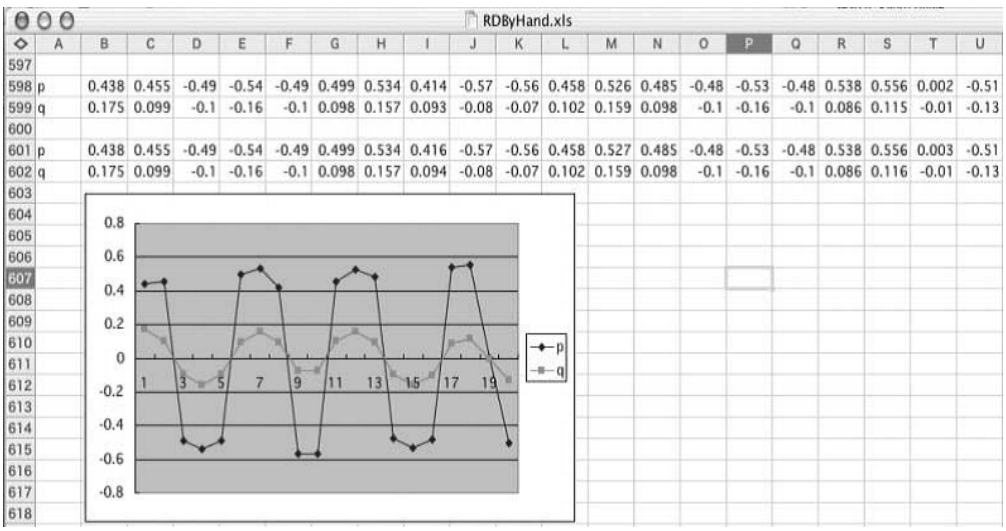


FIGURE 1. Numerical calculation results of the reaction-diffusion system (Equations 13) by Excel

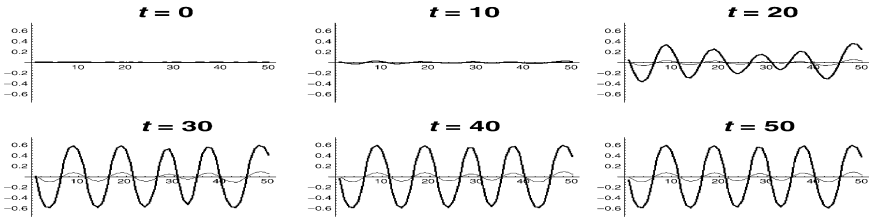


FIGURE 2. Time-course. **Thick line**-distribution of activator; **Thin line**-distribution of inhibitor.

PROPERTIES OF TURING REACTION–DIFFUSION SYSTEMS

Relationship between domain size and number of structures: If the domain size is changed, the number of structures changes, but the size of each structure remains the same, as can be seen in Figure 3.

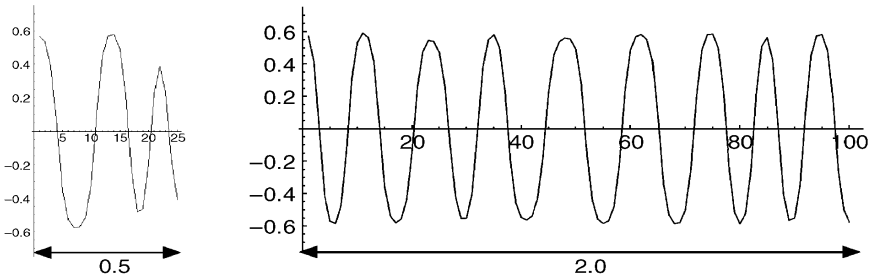


FIGURE 3. Each structure stays the same size, but number of structures increases when domain size is changed from 0.5 to 2.0.

Changing initial conditions: If the initial condition is not homogeneous, pattern formation can occur sequentially. However, the final periodic structure is more or less the same as shown in Figure 4.

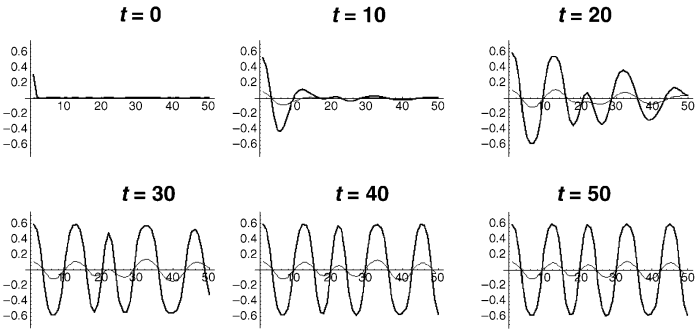
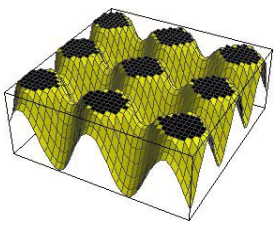


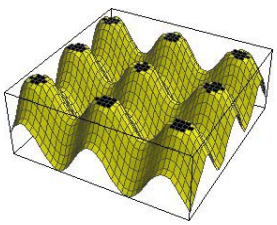
FIGURE 4. Initial value of p at the left-most point is increased.

REVEALING THE PATTERN - A COMPREHENSIVE CONCLUSION

- The Activator and Inhibitor are not color pigments themselves, just the morphogens that interact to create an underlying pattern.
- If the Activator also promotes the generation of a pigment in the skin of the animal then this pattern can be made visible.
- Skin cells could produce yellow pigment unless they detect high levels of Activator instructing them to produce black. This would yield a visible pattern similar to that of a 'CHEETAH'.
- The size of the spots depends on **threshold values** as a certain concentration of Activator is required to turn 'ON' the pigment. If threshold is high, then tiny spots are seen, but if the threshold is lowered,then the spots are larger. Such a mechanism may explain the difference in markings between two subspecies of 'GIRAFFE': the Rothschild's giraffe and the reticulated giraffe, the first of which has smaller, more widely-spaced spots than the other as shown in Figure 5a & 5b.
- Saturation can also be an important factor. If the concentration of Activator can reach a maximum value (i.e., it is produced as fast as it breaks down or diffuses away) then the spots may join up into stripes. This is believed to be what happens in the 'ZEBRA'.



(A) A low threshold for turning pigment ON



(B) A high threshold for turning pigment ON

FIGURE 5

REFERENCES

[1] Turing (1952) *The Chemical Basis of Morphogenesis*, Philosophical Transactions of the Royal Society of London.

[2] Takashi Miura and Philip K. Maini (2004), *Periodic pattern formation in reactiondiffusion systems: An introduction for numerical simulation*, Anatomical Science International.

[3] Lewis Dartnell (2004), *How the leopard got its spots?*, Plus magazine (Online Edition) http://plus.maths.org/content/how-leopard-got-its-spots

IPSA BAJAJ, B.SC.(H)MATHEMATICS, 3RD YEAR, LADY SHRI RAM COLLEGE FOR WOMEN
E-mail address: ipsabajaj@gmail.com

APURVA AGARWAL, B.SC.(H)MATHEMATICS, 3RD YEAR, LADY SHRI RAM COLLEGE FOR WOMEN
E-mail address: apurva.sag17@gmail.com

SRISHTI BANSAL, B.SC.(H)MATHEMATICS, 3RD YEAR, LADY SHRI RAM COLLEGE FOR WOMEN
E-mail address: srishti102@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 transformative learning.

A MATHEMATICA ENABLED EXPLORATION OF GIBBS PHENOMENON

JONAKI B GHOSH

Abstract. Fourier series can be used to model various types of problems in physics, engineering and biology. They arise in many practical applications such as modelling air flow in the lungs, electric sources that generate wave forms that are periodic and frequency analysis of signals. In contrast to Taylor series which can be used only to approximate functions that have many derivatives, Fourier series can be used to represent functions that are continuous as well as discontinuous. The partial sums of the series, approximates the function at each point and this approximation improves as the number of terms are increased. However, if the function to be approximated is discontinuous, the graph of the Fourier series partial sums exhibits oscillations whose value overshoots the value of the function. These oscillations do not disappear even as the terms are increased. This phenomenon is referred to as Gibbs phenomenon (Libii, 2005). Thus the approximation of functions by Fourier series, near the points of discontinuity, are inaccurate and this limits the use of Fourier Series in some cases. The topic of Fourier series is an integral part of mathematics courses at most undergraduate programmes in Engineering and Science. A computer algebra system like Mathematica, Maple or Matlab would enable students to visualize Fouries series of different functions and perform computations quite easily. This article presents an exploration of Gibbs phenomenon using Mathematica.

Study of the Basics of Fourier Analysis

The following are some preliminary results related to Fourier Series.

Result 1. Any function $g(t)$ defined and continuous on the interval $(0, \pi)$ can be written as

$$g(t) = b_1 \sin t + b_2 \sin 2t + b_3 \sin 3t + \ldots + b_n \sin nt + \ldots. \tag{1}$$

This is referred to as the *Fourier sine series.*

Result 2. The coefficients b_n in (1) are given by $b_n = \frac{2}{\pi} \int_0^\pi g(t) \sin nt \, dt$.

This is as follows. If we multiply both sides of (1) by $\sin nt$ and integrate from 0 to π we obtain

$$\int_0^\pi g(t) \sin nt \, dt = \int_0^\pi b_1 \sin t \sin nt \, dt + \int_0^\pi b_2 \sin 2t \sin nt \, dt + \ldots + \int_0^\pi b_n \sin^2 nt \, dt + \ldots.$$

Using the results

$$\int_0^\pi \sin mt \, \sin nt \, dt = 0, \;\; m \neq \pm n \quad \text{and} \quad \int_0^\pi \sin^2 nt \, dt = \frac{\pi}{2},$$

all the integrals vanish except the one with coefficient b_n. Thus

$$\int_0^\pi g(t)\sin nt\,dt = b_n\int_0^\pi \sin^2 nt\,dt = \frac{\pi}{2}b_n,$$

$$\text{or}\quad b_n = \frac{2}{\pi}\int_0^\pi g(t)\sin nt\,dt. \tag{2}$$

Result 3. If a periodic function $f(t)$ with period 2π is integrable on $[-\pi,\pi]$ then the *Fourier series* associated with the function $f(t)$ can be written as,

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n\cos nt + \sum_{n=1}^{\infty} b_n\sin nt, \tag{3}$$

where the Fourier coefficients a_0, a_n, b_n are given by

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\,dt,\quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt,\quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt \tag{4}$$

for $n = 1, 2, 3, \ldots$.

If $f(t)$ is piecewise continuous in $[-\pi,\pi]$ and has a left and right hand derivative at each point in that interval, then the corresponding Fourier series (3) with coefficients (4) is convergent. Its sum is $f(t)$, except, at a point x_0, at which $f(t)$ is discontinuous. Thus if the Fourier series corresponding to a function $f(t)$ converges with the sum $f(t)$, then the series will be called the Fourier series of $f(t)$.

Evaluating the Fourier Series of Some Elementary Functions.

The Fourier Series associated with some elementary functions such as $g(t) = 1, t$ and t^2 in the interval $(0,\pi)$ may be evaluated as follows:

For $g(t) = 1$, where $0 < t < \pi$, (2) gives

$$b_n = \frac{2}{\pi}\int_0^\pi g(t)\sin nt\,dt = \frac{2}{\pi}\int_0^\pi \sin nt\,dt = -\frac{2}{n\pi}\left[\cos n\pi - \cos 0\right] = \frac{2}{n\pi}\left[1-\cos n\pi\right].$$

Since $\cos n\pi = 1$ when n is even and -1 when n is odd,

$$b_n = \begin{cases} \frac{4}{n\pi} & \text{if } n \text{ is odd,}\\ 0 & \text{if } n \text{ is even.}\end{cases}$$

Substituting these in $g(t) = \sum_{n=1}^{\infty} b_n\sin nt = b_1\sin t + b_2\sin 2t + b_3\sin 3t + \ldots$, we get

$$1 = \frac{4}{\pi}\left[\sin t + \frac{\sin 3t}{3} + \frac{\sin 5t}{5} + \frac{\sin 7t}{7} + \ldots\right] = \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{\sin(2n+1)t}{2n+1}.$$

As an extension of this let us consider the *square wave function* which is defined as

$$g(t) = \begin{cases} -1 & \text{when } -\pi < t < 0,\\ 1 & \text{when } 0 < t < \pi,\end{cases}$$

and which is periodic with period 2π. Using (3) we can obtain its Fourier series. Using (4) it can be easily seen that a_0 and a_n are equal to 0. Computing b_n we get,

$$b_n = \frac{2}{n\pi}(1-\cos n\pi), \quad \text{which gives} \quad b_n = \begin{cases} \frac{4}{n\pi} & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even.} \end{cases}$$

In a similar manner, for $g(t) = t\,(0 < t < \pi)$, (2) gives

$$b_n = \begin{cases} \frac{2}{n} & \text{if } n \text{ is odd,} \\ -\frac{2}{n} & \text{if } n \text{ is even,} \end{cases}$$

and the Fourier Series for $g(t) = t$ for $0 < t < \pi$ is obtained as

$$t = 2\left[\sin t - \frac{\sin 2t}{2} + \frac{\sin 3t}{3} - \frac{\sin 4t}{4} + \ldots\right] = 2\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin nt}{n}.$$

The Fourier series for $g(t) = t^2$ for $0 < t < \pi$ may be respectively obtained as

$$t^2 = \frac{\pi^2}{3} - 4\left[\cos t - \frac{\cos 2t}{4} + \frac{\cos 3t}{9} - \frac{\cos 4t}{16} + \ldots\right] = \frac{\pi^2}{3} - 4\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\cos nt}{n^2}.$$

The reader is urged to do these computations on their own.

Plotting the Partial Sums of Fourier Series

In order to visualize how the Fourier series actually approximates a function, say $f(t) = 1$, we use a Mathematica code to plot the partial sums of the series.

```
fourierseries1[k_]:=
(4/Pi)*Sum[(Sin[(2n+1)t])/(2n+1), { n, 0, k } ];
Plot[{ 1, fourierseries1[10] }, { t, 0, Pi},
PlotRange− >{{ 0,Pi}, {.75, 1.25}}]
```

In line 2 of the above code the **Sum** command has been used to define the Fourier Series where k denotes the number of terms of the series. In lines 3 and 4 the **Plot** command is used to plot k terms of the series (here k = 10). Figure 1 shows the plot of the first 10 terms of the series.

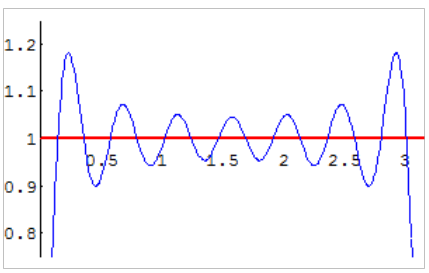


Figure 1. Plot of the Fourier series partial sum of $f(t) = 1$ for 10 terms.

The above code may be modified slightly by varying 'k' which represents the number of terms of the Fourier Series Partial Sum (FSPS). In the following code k is varied from 1 to 100 in steps of 10. Here the **Plot** command has been enclosed in the **Table** command in line 3 so that the output generates the plots of the partial sums as k varies from 1 to 100 in steps of 10. Also **PlotStyle** with **Thickness** and **RGBColor** options display the actual function $f(t) = 1$ in red and the Fourier series plot in blue. Figure 2 illustrates how the FSPS plot 'oscillates' around the actual function $f(t) = 1$.

```
fourierseries1[k_]:=
(4/Pi)*Sum[(Sin[(2n+1)t])/(2n+1), { n, 0, k } ];
Table[Plot[{ 1, fourierseries1[k] }, { t, 0, Pi},
PlotRange− >{{ 0,Pi}, {.75, 1.25}}],
PlotStyle− >{{ Thickness[0.009],
RGBColor[1,0,0] }, RGBColor[0,0,1]}],{k,1,100,10}]
```

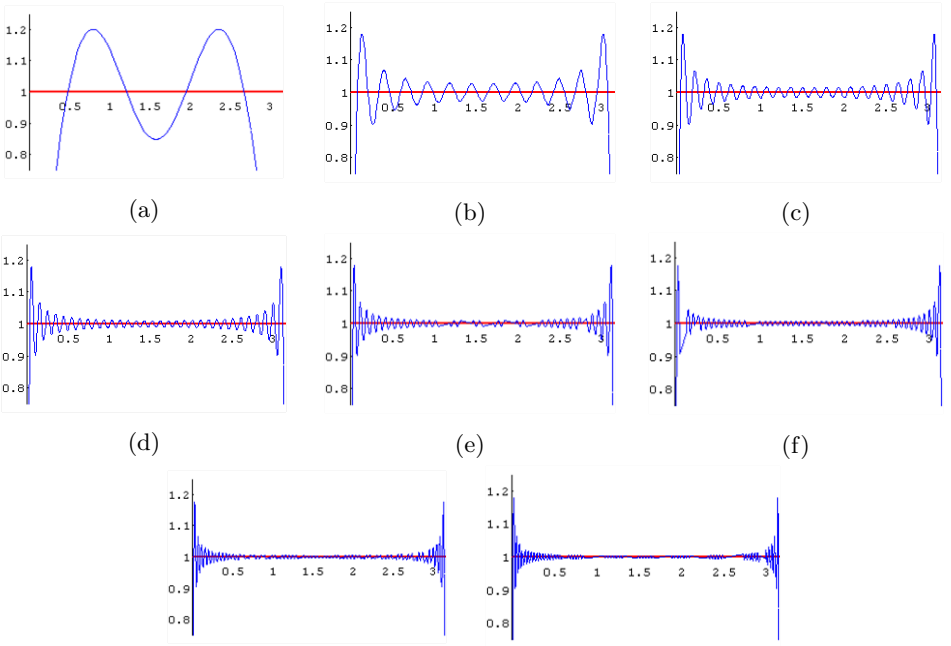


FIGURE 2. Plots of the FSPS of $f(t) = 1$ for 1, 11, 21, 31, 41, 51, 71 and 91 terms.

Graphical Analysis of the Fourier series Partial Sums

The FSPS plots show that the function $f(t) = 1$ can be approximated by the terms of its Fourier series for t lying in the interval $(0, \pi)$. The larger the number of terms, the better

is the approximation. The plots of the first $1, 11, 21, 31, 41, 51, 71$ and 91 terms between 0 and π shown in figure 2 (a) to (h) reveal that the approximation gets better as the number of terms are increased.

It is easy to observe that the Fourier series plots give a good approximation of the function within the interval $(0, \pi)$ but tend to oscillate towards the end points of the interval creating peaks which overshoot the function value. These oscillations persist and seem to approach the end points as the terms are increased. This is known as *Gibbs Phenomenon*.

Let us look at the square wave function which is periodic with period 2π and exhibits jump discontinuity at $x = -\pi$, $x = 0$ and $x = \pi$. The above Mathematica code may be used to obtain the FSPS plots for this function by changing the interval from $(0, \pi)$ to $(-2\pi, 2\pi)$. In figure 3 these plots have been generated for $1, 5, 10$ and 50 terms respectively in the interval $(-2\pi, 2\pi)$.

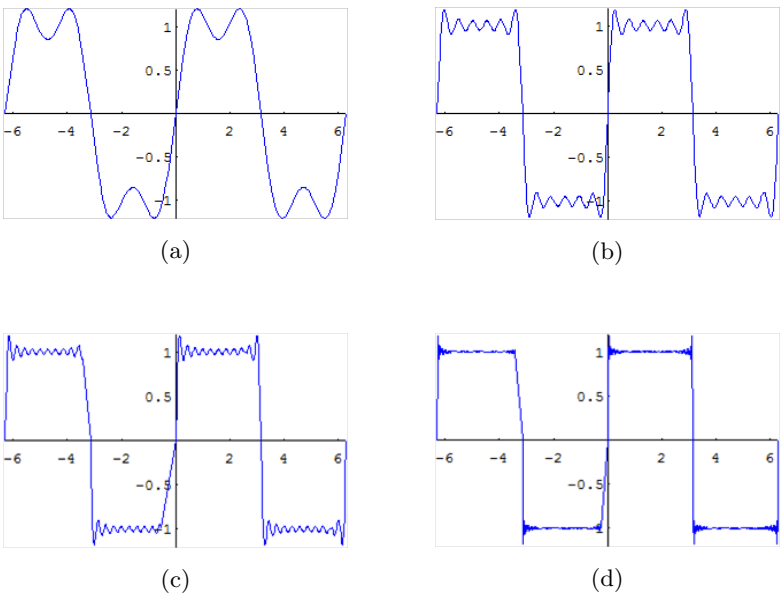
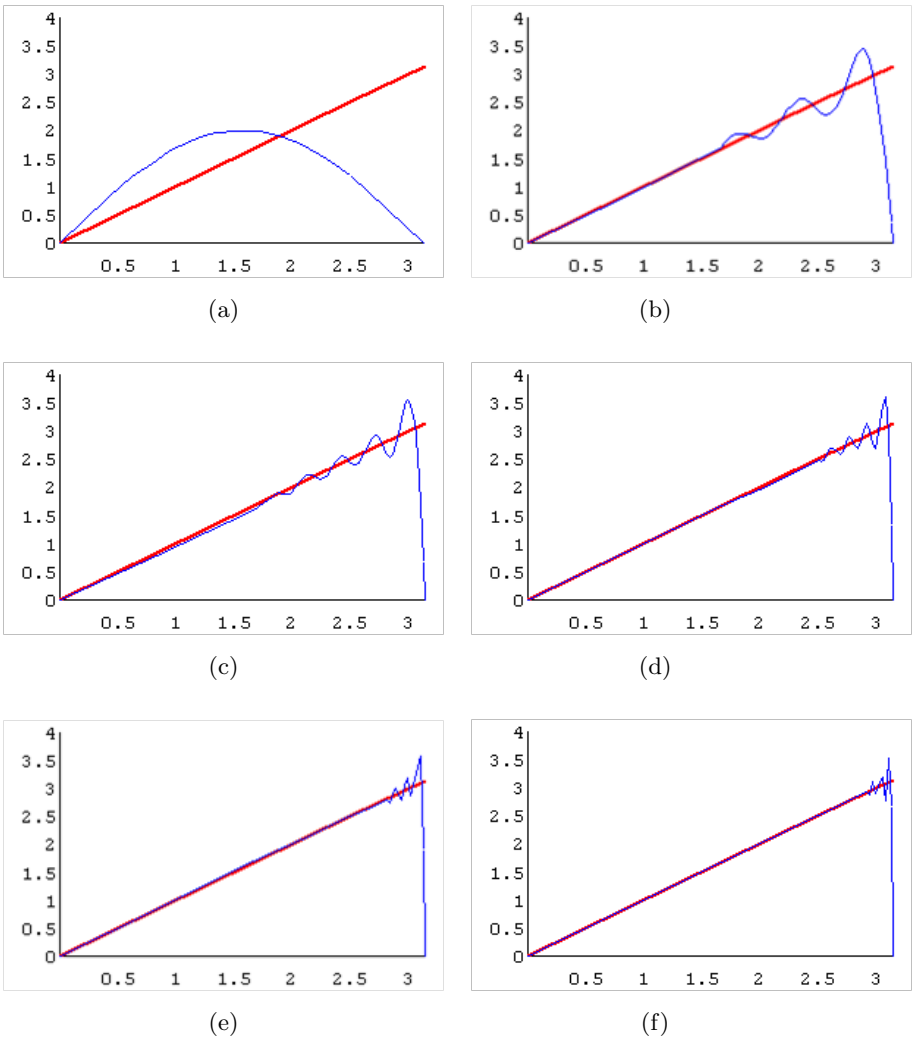


FIGURE 3. Plots of the FSPS of the square wave function for 1, 5, 10 and 50 terms.

We observe from the plots that peaks appear closer to the discontinuities followed by rapid oscillations. In fact as the number of terms of the partial sums are increased, these peaks get closer to the discontinuities but they do not disappear and the oscillations get smaller and more rapid.

Similarly the Mathematica code may be appropriately modified to obtain the plots for the FSPS of $f(t) = t$ for t in the interval $(0, \pi)$. Figure 4 shows that as the number of terms increase, the plot of the FSPS comes closer to the graph of the actual function.



(a) (b)

(c) (d)

(e) (f)

FIGURE 4. Plots of the FSPS of $f(t) = t$ for 1, 11, 21, 41, 71 and 91 terms.

The following Mathematica code may be used to plot the FSPS of $f(t) = 1$, closer to the endpoint 0 (Figure 5). The Table command (last line of the code) can be used to generate a table of values to measure the height of the highest peak, that is, the peak closest to the discontinuity.

```
fourierseries1[k_]:=
(4/Pi)*Sum[(Sin[(2n+1)t])/(2n+1), { n, 0, k } ];
plota[Plot[{ 1, fourierseries1[10] }, { t, 0, Pi},
PlotRange–>{{ 0,1}, {.75, 1.25},
PlotStyle–>{{ Thickness[0.009],
```

RGBColor[1,0,0]}, RGBColor[0,0,1]}]
Table[{t, N[fourierseries1[10]]},{t,0.1,0.2,0.001}]

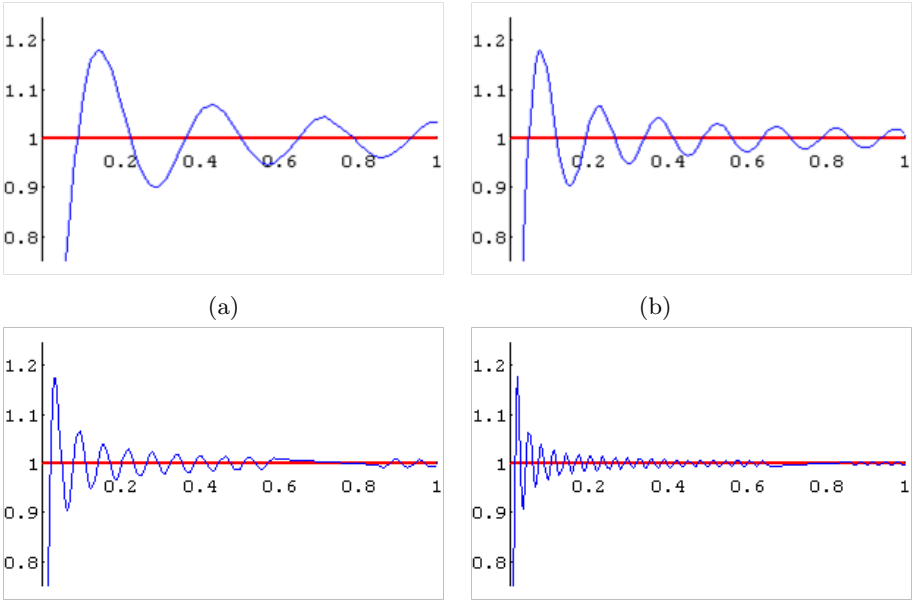


FIGURE 5. Plots of the FSPS of $f(t) = 1$ for 11, 21, 51 and 91 terms closer to the endpoint $t = 0$ of the interval $(0, \pi)$.

The Fourier Series partial sum values of the highest peaks of the function $f(t) = 1$ are summarized in the following table.

Table 1: Calculations of the FSPS for f(t)=1.

k (no of terms in Fourier Series)	FSPS value of highest peak	Value of t at which the highest peak is attained
10	1.1796	0.144
20	1.1791	0.075
30	1.1790	0.051
50	1.1789	0.031
100	1.1781	0.016

The output of the program as tabulated in Table 1 helps us to conclude that the overshoots approach the $y - axis$ as number of terms is increased but the value of the highest overshoot remains constant at 1.179. The successive peaks get smaller. A similar analysis can be done for the other functions as well.

Conclusion

This article presents a way of visualizing and exploring Fourier Series and Gibbs phenomenon using *Mathematica*. This can form an interesting lab activity as students can evaluate the Fourier series of different functions, plot their partial sums within an interval and also calculate the value of the overshoots. In this manner they can observe how *Gibbs phenomenon* occurs for different functions.

References

[1] Libii, J. N. (2005), *Gibbs Phenomenon and its applications in science and engineering*, Proceedings of the 2005 American Society for Engineering Education (ASEE) Annual Conference and Exposition, Portland.

[2] http://mathworld.wolfram.com/GibbsPhenomenon.html

Dr. Jonaki B Ghosh, Department of Elementary Education, Lady Shri Ram College
E-mail address: jonakibghosh@gmail.com

MARKOV CHAINS – THE KEY TO STOCK MARKET

AKANKSHA MITTAL AND EKTA SHARMA

Abstract. The aim of this research is to analyze the stock price movements of the shares of a company (in this case Reliance Industries Ltd.) and create a model using Markov Chains theory. The model has been created using past share prices of RIL and has been used to comment on their long and short term behavior. The results have then been verified using actual short term data.

Introduction

Before introducing Markov Chains, it is essential to define a Stochastic Process that is, a process in which the elements of a set are each classified as being in one of the several fixed states that can switch over time. Markov Chain is a special kind of Stochastic Process where the outcome of an experiment depends only on the outcome of a previous experiment. In other words, the next state of the system depends only on the current state and not on the previous states. Markov Chains have the property that probabilities involving how the process evolve in the future depend only on the present state of the process, and so are independent of events in the past.

Example 1. *To explain the above process with a simple example, we consider the population of a country in the year 2010. Based on the assumption that the population of the country in the given year remains constant and that there is a continual movement of the people between the city and the suburbs, we obtain the following matrix:*

	Currently living in the city	Currently living in the suburbs
Living next year in the city	0.9	0.02
Living next year in the suburbs	0.1	0.98

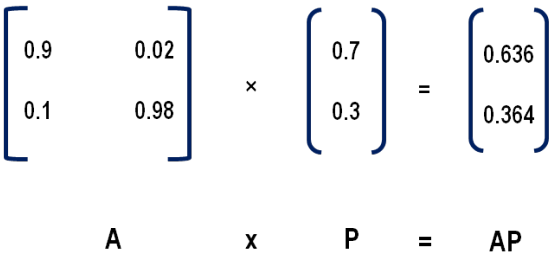
= A

Each entry of this matrix is non - negative and the sum of the entries in each column is 1. Such a matrix is called a ***transition matrix*** *or a* ***stochastic matrix****. For an arbitrary* $n \times n$ *transition matrix* M *the rows and the columns correspond to* ***n states*** *and the entry* M_{ij} *represents the probability of moving from state* j *to state* i *in one stage. So in this example the two states are living in the city and living in the suburbs.*

Each entry of this matrix A tells us the probability of a person moving from his current state to another state in the next year. For example, the entry A_{11} *tells us that the probability of a person, currently living in the city, continues to live in the city is* 0.90 *while the entry* A_{21} *tells that there is* 0.1 *probability for a person, currently living in the city, to move to the suburbs in the next year. Likewise other entries hold their respective meaning. Hence this transition matrix A gives us the probability of a person moving from one state to another in one year.*

Suppose 70% of the current population lives in the city while 30% lives in the suburbs. Expressing this initial distribution as a column vector, we get the initial probability vector P given by

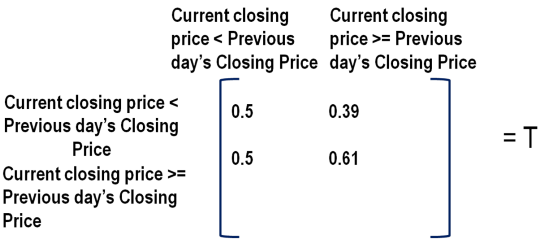
Now if we wish to determine the proportion of the current population that will reside in the city and the suburbs respectively in the next year, we simply multiply this initial probability vector P with the transition matrix A, to obtain the vector AP which gives the respective distribution i.e.,



This means that 63.6% of the current population will live in the city in the next year whereas 36.4% of the current population will be in the suburbs.
With sufficient knowledge in hand, we proceed onto the model that we have created to predict the future movement of the stock prices of Reliance Industries Limited, and comment on their profitability by analyzing their long term behavior.

METHODOLGY

The daily closing values of share prices of Reliance Industries Ltd. for the months of December 2012 and January 2013 were collected from Reuters. This data was interpreted to obtain the Transition Matrix T.



In the above matrix, the columns and rows signify the movement of current days share prices as compared to the previous days share prices. The entry T_{11} gives the probability that the current days price of the stock which was lower than the price of the previous day will decrease on the next day. The process used to calculate the probabilities can be better understood with the help of the following hypothetical example:
Suppose the following table gives the closing share prices of Reliance Industries Ltd. for a week:

Days	Prices
Monday	6
Tuesday	5
Wednesday	4
Thursday	3
Friday	5

To calculate the Total Outcomes for the first column, we need those current prices which are less than the previous days prices that is 5 (Tuesday), 4 (Wednesday) and 3 (Thursday). Hence, **Total Outcomes = 3**
Now, to calculate the Favorable Outcomes out of the Total Outcomes, that is those closing prices which are less than the current closing prices, which in turn were less than the previous days closing prices. In this case, the Favorable Cases are $(6 \to 5 \to 4)$ and $(5 \to 4 \to 3)$. Here, $(4 \to 3 \to 5)$ is not a Favorable Case as 5 which is the next days closing price is greater than 3 which is the current days closing value.
Hence, **No. of Favorable Cases = 2**
Therefore, the hypothetical probability in T_{11} must be = Favorable Outcomes/ Total Outcomes = 2/3

Predicting the Long Term Behavior of the Stock Prices: The following theorems were used as the basis for the formation of the transition matrix T and analyzing its long term behavior.

Theorem 1. *If A is an n×n matrix satisfying the following:*

(i) *Every eigen value λ of A satisfies either $\lambda = 1$ or $|\lambda| < 1$*
(ii) *A is diagonalizable.*

$$\textit{Then } \lim_{n\to\infty} A^n \quad \textit{exists.}$$

Theorem 2. *Every Transition matrix has 1 as an eigen value.*

Theorem 3. *If A is a regular transition matrix, then as n approaches infinity,*

$$A^n \rightarrow \quad S$$

where S is the matrix of the form $[v, v, ..., v]$ with v being a constant vector.

The eigen values of the transition matrix T was calculated to be 1 and 0.1087 . Computing the eigen vectors corresponding to each eigen value, we obtain an invertible matrix Q given by:

$$Q = \begin{bmatrix} 1 & 1 \\ 1.278 & -1 \end{bmatrix}$$

The columns of Q are the eigen vectors of T such that $T = QDQ^{-1}$ where D is a diagonal matrix given by:

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 0.1087 \end{bmatrix}$$

Hence T is diagonalizable.
Therefore Theorem 1 guarantees the existence of the limit of our transition matrix T. So

$$\lim_{n\to\infty} T^n = \lim_{n\to\infty} (QDQ^{-1})^n = \lim_{n\to\infty} QD^nQ^{-1} =$$

Probability of Current Closing Price < Previous Day's Closing Price	0.439	0.439
Probability of Current Closing Price >= Previous Day's Closing Price	0.561	0.561

= S

The columns of the matrix S are same and so the model goes well in accordance with Theorem 3.

Observation: In the limit matrix S, the probability of the current closing price being greater than or equal to the previous days closing price is more as compared to the probability of the current closing price being less than the previous days closing price.

Inference: In the long run, the probability of the closing prices of the stocks of Reliance Industries Ltd. to move up as compared to their previous days value is more. This clearly implies that the stocks of Reliance Industries Ltd. are profitable in the long run and hence validating the general perception of the stocks of RIL.

Predicting the Stock Price Movement in the next month (February): To predict the share price movement in the next month that is February 2013, we need an Initial Vector, as obtained in the Example 1. The Initial Vector obtained was as follows:

$$\begin{matrix}\text{Probability of Current Closing Price} \\ \text{< Previous Day's Closing Price} \\ \\ \text{Probability of Current Closing} \\ \text{Price >= Previous Day's Closing} \\ \text{Price}\end{matrix} \begin{pmatrix} 0.41 \\ \\ 0.59 \end{pmatrix} = V$$

The Initial Vector V was obtained from the Closing values of the stock price for the month of January. For the column vector V, the first row figure that is 0.41 indicates the probability that the stock prices of January decreased over their previous days value. And the second row indicates that the Closing Price of the Current day was greater than or equal to the closing price of the previous day.
As in Example1, we multiply the Transition Matrix T with the Initial Vector V to obtain TV:

$$\begin{pmatrix} 0.5 & 0.39 \\ 0.5 & 0.61 \end{pmatrix} \times \begin{pmatrix} 0.41 \\ 0.59 \end{pmatrix} = \begin{pmatrix} 0.435 \\ 0.565 \end{pmatrix}$$

Probability of Current Closing Price < Previous Day's Closing Price

Probability of Current Closing Price >= Previous Day's Closing Price

$$T \times V = TV$$

Observation: The first row of TV gives the probability that the stock price will decrease in February and the second row gives the probability that the stock prices will increase in the month of February, the latter being more than the former.

Inference: Since the probability in the latter case is more, this model predicts that the stock prices of Reliance Industries Ltd. will generally increase in the month of February.

Validation: The closing stock prices of Reliance Industries Ltd. were obtained from February 1, 2013 to February 14, 2013. The probability of the current days closing prices being less than or greater than equal to the previous days prices was calculated. The following column vector was obtained:

$$\begin{matrix}\text{Probability of Current Closing Price <} \\ \text{Previous Day's Closing Price} \\ \\ \text{Probability of Current Closing Price >=} \\ \text{Previous Day's Closing Price}\end{matrix} \begin{pmatrix} 0.333 \\ \\ 0.667 \end{pmatrix}$$

Comparing the corresponding rows of the above vector with the vector TV,

it is seen that the predictions of the model are in sync with the actual price movements. Both the vectors indicate a general increase in the stock prices of Reliance Industries Ltd. This validates the predictions of our model.

References

[1] Friedberg, Insel, Spence, *Linear Algebra.*
[2] S.Vasanthi, M.V.Subha, *An Empirical Study on Stock Index Trend Prediction using Markov Chain Analysis.*
[3] Suchi Mitra, *Predicting Stock Prices.*
[4] Kevin J. Doubleday and Julius N.Esunge, *Application of Markov Chains to Stock Trends.*
[5] http://aix1.uottawa.ca/ jkhoury/markov.htm
[6] http://in.reuters.com/finance/stocks/overview?symbol=RELI.BO$exchange=INB

Akanksha Mittal,B.Sc.(H) MATHEMATICS, 3rd Year, LADY SHRI RAM COLLEGE FOR WOMEN
E-mail address: akanksha.mittal1392@yahoo.com

Ekta Sharma, B.Sc.(H) MATHEMATICS, 3rd Year, LADY SHRI RAM COLLEGE FOR WOMEN
E-mail address: ekta.sharma2891@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.

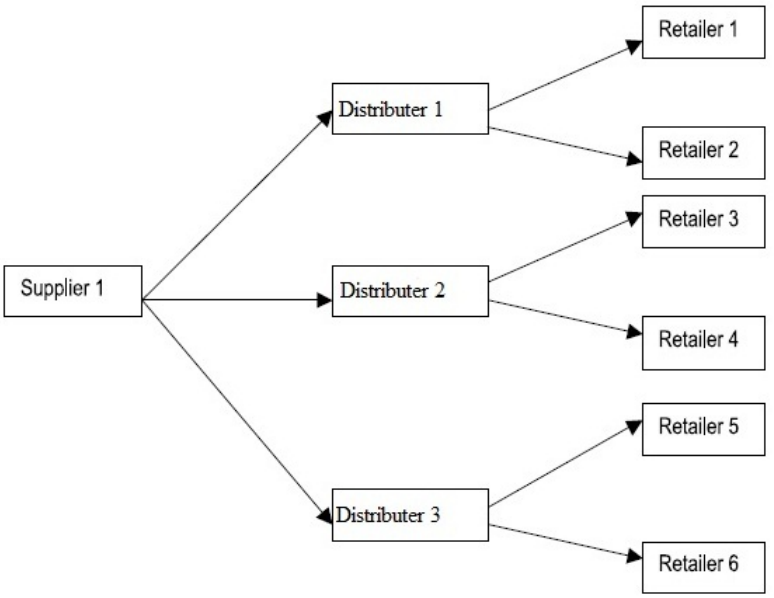
PROCUREMENT- DISTRIBUTION COORDINATION IN TWO STAGE SUPPLY CHAIN FOR MULTI PRODUCT

JYOTI DARBARI AND KIRAN GARG

Abstract. A supply chain is a global network of organizations that cooperate to improve the flows of material and information between suppliers and customers at the lowest cost and the highest speed. The focus of our current study is to formulate an optimization problem to specifically determine the optimal order quantities in a two stage supply chain along with the objective of minimizing the total cost that includes cost of purchasing, holding and transportation for the distributors and cost of holding and transportation for the retailers. We develop an integrated inventory-transportation two stage supply chain model incorporating the discounted policies on ordered goods and transportation network. The formulated model explains the flow of ordered quantity from single source to multiple distributors and from each distributor to its group of retailers.

Introduction

This paper deals with a two stage supply chain model that consists of a single supplier, multiple distribution centers, and multiple retailers where each distribution center supplies products to a group of retailers in each period. The distribution centers replenish their inventory from the supplier, and the retailers replenish inventory from the distribution centers in which group they belong. Further, we assume the supply chain is vertically integrated. At first stage, coordination is implemented between supplier and distribution centers in which the distributor avails all unit discount on purchased quantity as well as various freight policies on transportation. At second stage quantity ordered by the retailer is transported by the distributor in a single trip, by selecting the type of vehicle for a particular retailer on the basis of requirement of the retailer. The objective is to minimize the total cost of the two stage supply chain for a finite planning horizon taking into account the inventory holding costs and the transportation costs for entire chain. Mathematical models are formulated to determine optimal order quantities for an integrated inventory-transportation problem. The mode of transporting the goods from supplier to distributors takes place through two categories: truck load (TL) and less than truck load (LTL) transportation. In TL transportation, there is a fixed cost per load up to a given capacity. In LTL transportation, we assume a constant cost per unit. In the second stage of transportation, the goods are transported in a single trip by choosing a vehicle on the basis of the demand of the retailer and the capacity of the vehicle.



Sets

We will use the following Sets in our model:

- Product set with cardinality P and indexed by i.
- Period set with cardinality T and indexed by t.
- Item discount break point set with cardinality L and indexed by small l.
- Distributors set with cardinality M and indexed by m.
- Retailers set with cardinality J and indexed by j.
- $J_m \subseteq J$ set of retailers associated with m^{th} distributor.
- Vehicle set with cardinality F and indexed by f.

Decision variables

X_{imt} : Amount of item i ordered by m^{th} distributor in period t.
R_{ilmt} : If the ordered quantity falls in l^{th} price break then the variable takes value 1 otherwise zero,

$$R_{ilmt} = \begin{cases} 1, & \text{if } X_{imt} \text{ falls in } l^{th} \text{ pricebreak} \\ 0, & \text{if otherwise} \end{cases}$$

ID_{imt} : Inventory level at the end of period t of product i for m^{th} distributor.
IN_{im} : Inventory level at the beginning of planning horizon for product i for m^{th} distributor.
IR_{ijt} : Inventory level at the end of period t of product i for j^{th} retailer.
INR_{ij} : Inventory level at the beginning of planning horizon for product i for j^{th} retailer.

δ_{mt} : Total items ordered (transported in weights) in period t by m^{th} distributor.
α_{mt} : Total number of truck loads in period t to m^{th} distributor.
γ_{mt} : Amount in excess of truckload capacity (in weights) loads in period t to m^{th} distributor.
p_{jt} : Total items ordered (transported in weights) in period t to j^{th} retailer.
cap_f : capacity(in weight) of the f^{th} truck ($cap_0 = 0$).

$$u_{mt} = \begin{cases} 1, & \text{if considering TL \& LTL both policies,} \\ 0, & \text{if considering only TL policy.} \end{cases}$$

$$Z_{jft} = \begin{cases} 1, & cap_{f-1} < p_{jt} \leq cap_f, \\ 0 & \text{otherwise.} \end{cases}$$

Parameters

C : Total cost.
ID_{imt} : Demand for item i in period t at m^{th} distributor.
h_{im} : Inventory holding cost per unit of item i of m^{th} distributor.
D_{ijt} : Demand for item i in period t at j^{th} retailer.
CR_{ijt} : Consumption of item i in period t at j^{th} retailer.
q_{ij} : Inventory holding cost per unit of item i of j^{th} retailer.
w_i : Per unit weight of item i.
ϕ_i : Unit purchase cost for i^{th} item.
β_{mt} : Fixed freight cost for each truck load in period to m^{th} distributor.
df : Slab for discounts (price breaks).
ω : weight transported in each full truck (in kgs).
s : cost/kg of shipping in LTL policy.
a_{ilt} : Limit beyond which a price break becomes valid in period t for item i for l^{th} price break.
d_{ilt} : It reflects the fraction of regular price that the distributor pays for ordered items.
cap_f : capacity of the f^{th} truck (in weights).
c_f : Cost of transporting per ton of weight through f^{th} truck.

Objective Function

The objective function (minimization) is given by:
$\sum_m \sum_t [\sum_{i=1}^{P} \{h_{im} ID_{imt} + \sum_{l=1}^{L} R_{ilmt} d_{ilt} \phi_i X_{imt}\} + (s\gamma_{mt} + \alpha_{mt}\beta_{mt}) u_{mt} + (\alpha_{mt} + 1)\beta_{mt}(1 - u_{mt})] + \sum_j \sum_f \sum_t c_f p_{jt} Z_{jft} + \sum_t \sum_j \sum_i q_{it} IR_{ijt}$.

Constraints

Following are the constraints :
The inventory level of m^{th} distributor at the end of period 1:

$$ID_{im1} = IN_{im} + X_{iml} - \sum_{j=1}^{J_m} D_{ij1} \quad \forall i = 1, \ldots, P, \;\; m = 1, \ldots, M \tag{1}$$

The inventory level of the m^{th} distributor at the end of period t :

$$ID_{imt} = ID_{imt-1} + X_{imt} - \sum_{j=1}^{J_m} D_{ijt} \quad \forall i, m, t = 2, \ldots, T \tag{2}$$

Since there are no shortgaes, therefore,

$$\sum_{t=1}^{T} ID_{imt} + \sum_{t=1}^{T} X_{imt} \geq \sum_{j=1}^{J_m} \sum_{t=1}^{T} D_{ijt} \quad \forall i = 1, \ldots, P, \;\; m = 1, \ldots, M \tag{3}$$

The inventory level of the j_{th} retailer at the end of period 1:

$$IR_{ij1} = INR_{ij} + D_{ij1} - CR_{ij1} \quad \forall i = 1, \ldots, P, \;\; j = 1, \ldots, J \tag{4}$$

The inventory level of the j_{th} retailer at the end of period t:

$$IR_{ijt} = IR_{ijt-1} + D_{ijt} - CR_{ijt} \quad \forall i = 1, ..P, j = 1, \ldots, J, \;\; t = 2, \ldots, T \tag{5}$$

Since there are no shortages, therefore,

$$\sum_{t=1}^{T} IR_{ijt} + \sum_{t=1}^{T} D_{ijt} \geq \sum_{t=1}^{T} CR_{ijt} \quad \forall i = 1, \ldots, P, \;\; j = 1, \ldots, J \tag{6}$$

The above equations are balancing equations.
The m^{th} distributor will order minimum quantity to get discount *i.e.*

$$X_{imt} \geq \sum_{l=1}^{L} a_{ilt} R_{ilmt} \quad \forall i = 1, \ldots, P, \;\; m = 1, \ldots, M, \;\; t = 1, \ldots, T \tag{7}$$

In any period, exactly one level will be activated, therefore

$$\sum_{l=1}^{L} R_{ilmt} = 1 \quad \forall i = 1, \ldots, P, \;\; m = 1, \ldots, M, \;\; t = 1, \ldots, T \tag{8}$$

Transported quantity to the m^{th} distributor according to item weight is:

$$\delta_{mt} = \sum_{i=1}^{P} [w_i X_{imt} \sum_{l=1}^{L} R_{ilmt}] \quad \forall m = 1, \ldots, M, \;\; t = 1, \ldots, T \tag{9}$$

The minimum weighted transported quantity to the m^{th} distributor is equal to:

$$\delta_{mt} \leq (y_{mt} + \alpha_{mt}\omega)u_{mt} + (\alpha_{mt} + 1)\omega(1 - u_{mt}) \quad \forall m = 1, \ldots, M, \;\; t = 1, \ldots, T \tag{10}$$

truckload Overhead units from capacity in weights are:

$$\delta_{mt} = (y_{mt} + \alpha_{mt}\omega) \quad \forall m = 1, \ldots, M, \;\; t = 1, \ldots, T \tag{11}$$

Only one vehicle will be chosen as per the requirement of the j^{th} retailer in period t

$$p_{jt} \geq \sum_{f=1}^{F} c_{f-1} Z_{jft} \quad \forall j = 1, \ldots, J, \;\; t = 1, \ldots, T \tag{12}$$

$$\sum_{f} Z_{jft} = 1 \quad \forall j = 1, \ldots, J, \;\; t = 1, \ldots, T \tag{13}$$

Transported quantity to the j^{th} retailer in period t according to item weight is:

$$p_{jt} = \sum_{i=1}^{P} [w_i D_{ijt} \sum_{f} Z_{jft}] \quad \forall j = 1, \ldots, J, \;\; t = 1, \ldots, T \tag{14}$$

$$X_{imt}, ID_{imt}, IR_{ijt}, D_{ijt}, \delta_{mt}, \alpha_{mt}, \gamma_{mt}, p_{jt} \geq 0;$$

$$R_{ilmt}, u_{mt}, Z_{ift} \in \{0, 1\}, i = 1, \ldots, P, \;\; j = 1, \ldots, J, \;\; k = 1, \ldots, K,$$
$$t = 1, \ldots, T, \;\; l = 1, \ldots, L, \;\; m = 1, \ldots, M$$

Price breaks are defined as:

$$d_f = \begin{cases} d_{ilt}, & a_{ilt} \leq X_{it} \leq a_{il+1t} \\ d_{iLt}, & X_{it} \geq a_{iLt} \end{cases} i = 1, \ldots, P; \;\; t = 1, \ldots, T; \;\; l = 1, \ldots, L;$$

Lower limit of first price break in the models is zero and after the upper limit of first break buyer will get discount.

Case Study

One of the reputed company which supplies LED televisions has three distribution centers in three different cities each distribution center has two retailers each in its city. Supplies are transported to each distribution center and each center supplies to only its group of retailers once in the beginning of January, May and September. Same Transporter offers services to each of these distribution centers and retailers. He offers various discounts depending upon the mode of transport chosen. At first stage, TL/LTL mode for distribution and at second stage requirement based vehicle is used. Supplier also offers some quantity discounts on bulk purchase to the distributors Goal of the company is to minimize the total cost of the entire supply chain which includes Inventory holding cost and transportation cost of the distributors and retailers.. The relevant data is provided as under (Table $1 - 8$):

	J 1			J 2			
	1	**2**	**3**		**1**	**2**	**3**
J	18	18	11	**J**	19	18	17
My	22	21	17	**My**	23	15	17
S	27	21	17	**S**	25	15	15

	J 3			J4			
	1	**2**	**3**		**1**	**2**	**3**
J	14	19	14	**J**	20	19	15
My	16	24	19	**My**	25	23	20
S	18	25	19	**S**	29	27	14

	J 5			J 6			
	1	**2**	**3**		**1**	**2**	**3**
J	24	20	18	**J**	17	14	17
My	25	24	19	**My**	19	19	15
S	30	24	22	**S**	20	17	17

Table 1: Consumption of three styles of $LEDs$ in three periods at the retail stores (CR_{ijt})

	1	2	3
J_1	250	300	350
J_2	250	300	350
J_3	250	300	350
J_4	250	300	350
J_5	250	300	350
J_6	250	300	350

Table 2: Inventory carrying cost per item(in Rs) incurred by six retailers for three items (h_{im})

The initial inventory in hand for all retailers would be zero for all items
$IRN_{ij} = 0$ for $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5, 6$

$Wt.(Kg.)$	1	2	3
	7	9	11

Table 3: The weight(in $kg.$) of each item (w_i)

RS	1	2	3
	25000	30000	38000

Table 4: Purchase Cost for all types of $LEDs$ (ϕ_i)

	1	2	3
M_1	200	270	350
M_2	200	270	350
M_3	200	270	350

Table 5: Inventory carrying cost per item(in Rs) incurred by three distributors for three items (h_{im})

Quantity Thresholds (a_{1lt})	Discount factor(d_{1lt}) Period 1,2,3	Quantity Thresholds (a_{2lt})	Discount factor(d_{i2lt}) Period 1,2,3	Quantity Thresholds (a_{3lt})	Discount factor(d_{3lt}) Period 1,23
$0 \le X_{1mt} < 50$	1	$0 \le X_{2mt} < 45$	1	$0 \le X_{3mt} < 40$	1
$50 \le X_{1mt} < 75$	0.98	$45 \le X_{2mt} < 65$	0.96	$40 \le X_{3mt} < 60$	0.95
$75 \le X_{1mt} < 100$	0.96	$65 \le X_{2mt} < 85$	0.95	$60 \le X_{3mt} < 80$	0.93
$100 \le X_{1mt}$	0.94	$85 \le X_{2mt}$	0.92	$80 \le X_{3mt}$	0.90

Table 6: Figures given by Supplier (Cost to bear)

Weight each truck carries is (ω) 500 kg. Cost of each unit of shower by LTL mode is (s) Rs 25. Cost of truckload (in Rs) for the two distributors for the three periods is (β_{mt}):

	M_1	M_2	M_3
J	2500	2600	1500
M_y	2700	2500	1700
S	2950	1650	1900

Table 7: Figures given by Transporter for distributor

There are 4 types of vehicles $F1$, $F2$, $F3$, $F4$

Cost/Kg. capacity	F1	F2	F3	F4
	30	25	22	17
	100	150	175	200

Table 8 : Figures given by Transporter for Retailer

Solution

The above formulated optimization problem is programmed in Lingo 11.0 software. The required data is fed in the program to generate the solution. We have considered 3 periods and 3 products and obtained the solution given below. The Minimum cost incurred by the Company for the entire supply chain is: Rs $4,22,77380$

The Ordered Quantity by the retailers(D_{ijt}) of given products in the respective periods are

• LED1,LED2,LED3	• Period 1	• Period 2	• Period 3
• Retailer1	• 29, 18, 18	• 22, 21,21	• 27,25,21
• Retailer 2	• 21,20,26	• 24,25,29	• 25,26,32
• Retailer 3	• 30,20,24	• 16,24,29	• 18,25,29
• Retailer 4	• 20,16,25	• 25,23,24	• 29,27,24
• Retailer 5	• 36,21,28	• 25,25,29	• 30,29,32
• Retailer6	• 14,24,27	• 19,29,35	• 17,27,37

The Ordered Quantity by the distributors(X_{imt}) of given products in the respective periods are

LED1,LED2,LED3	• Period 1	• Period 2	• Period 3
• Distributor1	• 50,45,44	• 50,45,63	• 50,45,40
• Distributor2	• 50,45,76	• 50,45,40	• 50,45,40
• Distributor3	• 50,65,60	• 50,45,60	• 50,45,68

Conclusion

In this paper we have investigated the two stage supply chain optimization model that minimizes purchasing, holding and transportation cost. Using the formulated model, the optimal ordered quantity, holding inventory and weighted transported quantity are determined for both stages. Hence we can conclude from our present research that integration of various functions of different entities is possible, in order to minimize the aggregate cost of purchasing and transportation activities.

References

[1] Abraham Mendoza and Jos.A.Ventura, *Incorporating quantity discounts to the EOQ model with transportation costs*, International Journal of Production Economics. Amsterdam: 113(2), 754-765, (2008).

[2] B.Q. Riekstsa, JosA.Ventura, *Two-stage inventory models with a bi-modal transportation cost*, Computers and Operations Research: 37, 20 31, (2010).

[3] Ertogral K., *Multi-item single source ordering problem with transportation cost: A Lagrangian decomposition approach*, European Journal of Operational Research, 191, 154163, (2008).

[4] Hwang, Hark, Moon,Dug Hee, Shinn, Seong Whan, *An EOQ model with quantity discounts for both purchasing price and freight cost*, Computers and Operations Research. New York, 17(1), 73-78, (1990).

[5] Jae-Hun Kang, Yeong-Dae Kim, *Coordination of inventory and transportation managements in a two-level supply chain*, International Journal Production Economics 123 92010), 137-145.

[6] Thiriez,H.OR software LINGO, *European Journal of Operational Research*, 12: 655-656(2000)

Jyoti Darbari, Assistant Professor, Department of Mathematics, Lady Shri Ram College For Women

Kiran Garg, Research Scholar, Department of Mathematics, University of Delhi
E-mail address: jydbr@hotmail.com

A PEEK INTO THE LIFE OF A STATISTICIAN

DOOTIKA VATS

Abstract. This paper aims at introducing to the reader basic statistical analysis done by most applied statisticians. The paper introduces simple linear regression with the help of an example. The objective is to build a linear model between two variables and test the significance of the model.

Introduction

Statistics is generally understood to be a branch of mathematics concerned with collecting and interpreting data. This however, is largely untrue. Statistics, by itself should not be understood as a branch of mathematics. Mathematics is definitive, and in most cases deterministic. A problem in mathematics has one correct answer. In statistical analysis, there is no correct answer. This is what makes statistics the most flexible and sought after tool in understanding the workings of the world.

This is not to say that mathematics renders itself useless in the field of statistics. Indeed, understanding most topics in statistics requires a sound knowledge of mathematics, which is why most statisticians have an undergraduate degree in mathematics. Today, we live in a world where everything turns into data at the end of the day, and there are not enough qualified statisticians to make sense of the data. In 2009, an article in the New York Times [1] elaborated very articulately, on why being a statistician is one of the most sought after professions.

One of the most important aspect of statistics is to find out if certain things are related to each other. For example, the fact that smoking causes lung cancer was proved largely due to statistical analysis. Similarly, we might want to analyze if a certain variable, Y (response) depends on a set of variables $X_1, X_2, \ldots, X_p$ (predictors). And if there is such a dependence, then we want to find the f such that $Y = f(X_1, X_2, \ldots, X_p)$. This function f is found by using Regression.

This paper focuses on Simple Linear Regression, i.e., when we have only one predictor, X, and f is a linear function, giving the relation, $Y = \beta_0 + \beta_1 X$. The concept of simple linear regression will be explained step by step with the help of an example dataset.

Dataset

A dataset is a collection of data, usually presented in tabular form, where each column represents a variable of interest. In simple linear regression, there are only two columns, one for the response variable, Y and one for the predictor variable, X. This paper uses the example of one of the most basic and famous datasets.

Karl Pearson organized the collection of data of over 1100 families in England in the period 1893 - 1898. This particular data set gives the heights in inches of mothers and their daughters. All daughters are at least age 18, and all mothers are younger than 65. The objective is to find out whether there is a relation between the height of mothers and their daughters.

Notice how the background of the dataset is as important as the numbers in the dataset. For example, it is important to know that the daughters are atleast age 18, so we can assume that they have attained their full height.

Now, the original dataset has 1100 observations, but for the purpose of this paper, I have chosen a random subset of 200 observations from the dataset. This is just to ensure that the graphs produced are not messy. The numbers in the dataset are given in the table below.

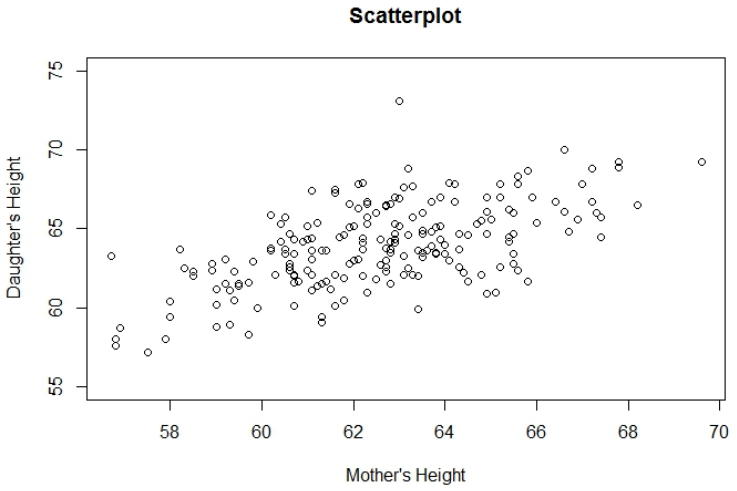
Of course, looking at the numbers does not really help us, specially when we have 200 such pairs. This is where graphical tools prove to be much more useful.

X	Y
Mother's Height	Daughter's Height
63.5	66.0
63.5	63.2
62.7	63.0
⋮	⋮

Each of the rows above, corresponds to a **datapoint**, which is to say that each point can be written as (x_i, y_i), and this would correspond to a point on a graph of Y vs X. When all the datapoints are plotted together, we get something called a **scatterplot**. This is essentially the first tool in understanding whether there is any relation between X and Y.

From the scatterplot above, we notice that as mother's height increases, we see some increase in the daughter's height. Note that this is not individually true, but the **trend** indicates that it is generally true.

As mentioned earlier, the objective is to find a linear relationship between X and Y.



Model

If there is a linear relationship between X and Y, all points (x_i, y_i) should lie on a common line. We also know that they do not exactly lie on a common line and there is some deviation(as demonstrated in the scatterplot above). This is represented by the **model** below

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where β_0 is the y-intercept, β_1 is the slope of the line, and ϵ_i is known as the **error**. When we write this model, we make the following very important **assumptions:**

- A linear model is appropriate.
- All the observations, $y_1, \ldots, y_n$ are independent of each other. In this dataset, we assumed that none of the daughters are related.
- The errors $\epsilon_i \sim N(0, \sigma^2)$. This means that the errors follow a normal distribution[3] with mean 0 and variance σ^2.
- All observations have a constant variance (σ^2, as opposed to σ_i^2).

Whenever a model is fit to the dataset, it is the duty of the statistician to ensure that the model assumptions stated above hold true.

One important point to note in the model, is that the unknown quantities are β_0 and β_1, and the points (x_i, y_i) are all known, since the line is fit only after we have the data. Thus, the term "linear" refers to the equation being linear in the βs, and not in the xs. If instead, we had the equation

$$y_i = \beta_0 + \beta_1 x_i^2 + \epsilon_i,$$

this would still be a linear regression model.

The next step is to try and estimate β_0 and β_1 from the data. Notice how we use the word "estimate", because every time the experiment is done, we get a different dataset, and every dataset gives a new "estimate" of these two parameters. These estimates are denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$, and are calculated by using a method known as Ordinary Least Squares (OLS).

OLS - Ordinary Least Squares. Clearly, we can not fit a line by joining all the points. This is where mathematics, gives way to statistics. We fit a line to the data, in such a way that the overall deviation of the datapoints from the line is minimized. This is done by a method known as Ordinary Least Squares, or OLS.

Notice that in the model, the error ϵ_i is nothing but the deviation of each point from the line. Now let us assume that we have already fit the line, and have found $\hat{\beta}_0, \hat{\beta}_1$ and for each x_i we have a $\hat{y_i}$. Thus each $(x_i, \hat{y_i})$ lies on the line:

$$\hat{y_i} = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

The estimated errors are known as **residuals** $= y_i - \hat{y_i}$, giving :

$$\hat{\epsilon_i} = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i).$$

The OLS method estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained by minimizing the sum of the squared residuals over all observations, i.e.,

$$(\hat{\beta}_0, \hat{\beta}_1) = \min_{(\beta_0, \beta_1)} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2.$$

This turns out to be an exercise in basic calculus, the proof for which can be found in most statistics books [2]. The OLS estimates we get are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

where $\bar{y}$ and $\bar{x}$ are the means of y_is and x_is. $\hat{\beta}_0$ and $\hat{\beta}_1$ represent the actual y-intercept and slope of the line.

In our dataset, we get $\hat{\beta}_0 = 25.05$ and $\hat{\beta}_1 = 0.58$ giving the equation

$$Daughter_{height} = 25.05 + 0.58 Mother_{Height}.$$

Interpretation:

$\hat{\beta}_0 = 25.05$ implies that when the mother's height is 0 inches, the daughters height on average is 25.05 inches. This of course does not make sense, and that is alright. In most cases β_0 need not be interpreted. We are more interested in β_1. $\hat{\beta}_1 = 0.58$ implies that

with one inch increase (decrease) in the mother's height, we expect the daughters height to increase(decrease) by 0.58 inches. This is what gives us the relation between mother's height and daughter's height. We need to analyze this more carefully.

Significance. The most important aspect of simple linear regression is to make sense of the $\hat{\beta}_1$. Since $\hat{\beta}_1$ is the slope of the line, a value of 0 would mean a horizontal line. If the line was horizontal, it would imply that there is no relationship between X and Y. Thus, our objective is to always check whether $\hat{\beta}_1 = 0$ or not.

The data that we collect is known as a **sample** and is a representation of the whole **population**. Clearly, it is impossible for us to collect the heights of mothers and their daughters all over the world. So we collect heights from a sample that represents the population. Every time, we collect data from a sample, we will get different estimates of β_0 and β_1.

Remember how in our dataset we had chosen 200 observations at random from 1100 observations so that the scatterplot was not messy. If we take different sets of 200 observations again, we will get different estimates of β_1.

No.	$\hat{\beta}_1$
1	0.61
2	0.59
3	0.51
4	0.54

Notice, from the table that the $\hat{\beta}_1$ values are close to 0.58, but not exactly 0.58. Thus, for each sample we get different estimates. And so we need to check if for our sample, the value of $\hat{\beta}_1$ is different enough from 0 for us to be confident that it is in fact, not 0.

This is done by using a method known as **Hypothesis Testing**. A step by step explanation on this can be found in the references [2], [3]. I present briefly, how we decide whether $\hat{\beta_1}$ is different enough from 0.

We first construct a **Null Hypothesis**, H_0, and an **Alternate Hypothesis**, H_a. H_0 is assumed to be true, and from the data, we want to gather enough evidence to reject the H_0 and accept H_a.

$H_0 : \beta_1 = 0$
$H_a : \beta_1 \neq 0$

This structure makes sense, because we want to be sure that there is in fact a relationship between X and Y, and for that to be true, we want the data to give us enough evidence to reject the null hypothesis.

Next, we calculate something called the **Test Statistic**, t which in this case is

$$t = \frac{\hat{\beta_1}}{se(\hat{\beta_1})} \quad \text{where } se(\hat{\beta_1}) = \sqrt{\frac{\hat{\sigma^2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

This t follows a t-distribution with $n-2$ degrees of freedom [3]. Intuitively, this t scales $\hat{\beta_1}$ down by its standard deviation, and gives us an idea on whether the $\hat{\beta_1}$ is different from 0.

If $|t| > 1.96$ (approx) this means, that we can reject the null hypothesis, H_0, and we say the the variable X is significant. That is to say, that we are confident that X and Y are related, and β_1 is in fact different from 0.

In our dataset, $\hat{\beta_1} = 0.58$, $se(\hat{\beta_1}) = 0.069$ and $t = 7.79$. Since $|t| = 7.79 > 1.96$, we can reject H_0, and claim that there is a significant relationship between mother's height and daughter's height.

Additional Comments

We have, at this point succeeded in analyzing the dataset. We have found the relationship between X and Y, and shown that this relationship is significant. This is what most statisticians have to do when they are given a dataset. However there are some roadblocks, and most datasets are trickier than this one. Following are some other important aspects of statistical analysis:

- Once the model has been fit, it is important to check the assumptions. In a lot of cases, the constant variance assumption is not valid, in which case we need to transform our data [2].
- In a lot of cases, a linear relationship is not adequate. Higher order regression models should then be tried [2].

- We generally have more than one predictor variable, $X_1, X_2, \ldots, X_p$. In that case, we fit the linear regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots x_{pi} + \epsilon_i.$$

  In our dataset for example, we could also introduce father's height, weight of the daughter, and time of menarche as potential predictor variables. However, in that case, we move from 2 dimensions to $p+1$ dimensions. The regression model is then built by using matrices [2].

- Sometimes, the data does not come from a Normal distribution, in which case we fit Generalized Linear Regression Models [4].

Another important aspect of statistical analysis is computer programming. Since most datasets are large, and computations are complicated, it is impractical to do regression on paper. Statistical programming languages make life a million times easier and also provide us with some excellent graphical tools. *R* and *SAS* are the two most famous languages used. *SAS* is used mostly in industry settings and biostatistics work. *R* is used extensively in academic settings. I used *R* to do the analysis in this paper.

Conclusion

Statistical analysis at its core is about quantifying uncertainty. It is about making inferences from raw data, figuring out trends and concluding with confidence that the results obtained are not coincidental.

In the end, the model obtained would probably not be the exact model. For example, if we were able to get data on all mothers and daughters in the world, then β_1 might be very different from 0.58. But, the idea is to get as much information as possible from the model. There is one statement most statisticians live by, "All models are wrong, but some are useful."

References

[1] http://www.nytimes.com/2009/08/06/technology/06stats.html?_r=0
[2] Sanford Weisberg, *Applied Linear Regression*, 3rd edition
[3] Jay L. Devore, *Probability and Statistics for Engineering and the Sciences*, 7th Edition
[4] Julian J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models*, 1st Edition

Dootika Vats, Research Scholar, Department of Statistics, Rutgers University, U.S.A.

E-mail address: dootika1990@gmail.com

PEAK LOAD MANAGEMENT

UMANG AGGARWAL, PRACHI SINGHAL AND RUCHI ARORA

Abstract. If you ask anyone living in Delhi what they dislike about the Metro, the chances are they will respond with "They're too crowded." As it happens, this problem has been studied using mathematical modeling techniques. This article looks at the above mentioned problem, a topic known as "Peak Load Management." The problem of predicting train loads (though not always reducing crowding!) has been successfully tackled by mathematicians working as consultants within the rail industry. This paper will give a feel for the techniques and approaches used by such consultants, which is highly representative of the way mathematics is used in real life.

Formal Definition Of Crowd

To start thinking about this problem, we will first need a clear definition, something better than "They're too crowded!" The metro has an average daily ridership of 1.8 million commuters. Each train can carry up to 600 passengers (4 coaches), with 50 seated and around 100 people standing passengers per coach. No wonder we complain of overcrowding. Getting to the first definition should be simple: "At certain times of the day, trains become very crowded, due to people wanting to travel to and from work." The diagram below shows how demand for a commuter train service typically varies during the day. Peak Load Management is about trying to cope with this changing demand, without trains becoming too overcrowded.

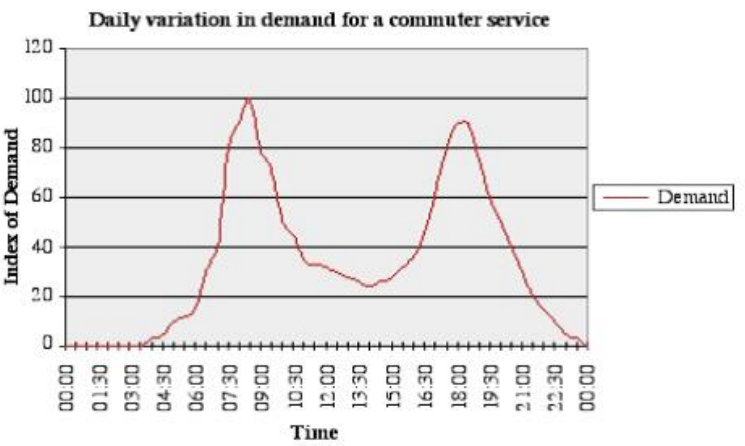


Figure 1

But first we need to understand exactly what the consultant is required to do about it! Here is a more thorough definition of the problem:

- Trains are crowded at peak times. This causes dissatisfaction amongst passengers, and may lead to fewer people using the trains.
- It is difficult to know if the right amount of rolling stock is currently being used for each route. Lengths of trains, frequency and stopping patterns can all be altered.
- It is even more difficult to know how people will behave when changes are made to the rolling stock. For example, is it better to introduce more frequent, smaller trains, or to make current trains longer?
- If more rolling stock is to be used, then a firm prediction must be made of how much is needed. Having too much wastes money, having too little drives customers away.

How Mathematics Can Help

A generic outline of how the mathematical model is built up to understand this problem, covering both the basic principles and some special cases from recent projects. There are three basic stages to tackle the problem. These are:

(1) Understanding the choices customers are making at present;
(2) Building models which can help understand future choices;
(3) Using the model to produce results meaningful to the client.

Understanding Current Choices: If we think simply about the choices passengers must make, then it is clear that each customer wishes to get from a point A to a point B. Customers want to do this in the shortest amount of time possible, and to arrive at the time that suits them. They may also have some preferences about the type of train they travel on; for example, regular travellers may prefer to avoid certain very crowded trains.
Since the ideal journey would be one that took no time at all, and every other journey of course falls short of this ideal, we can think of the problem as a minimisation problem. Each customer tries to minimise some combination of the inevitable real-world inconveniences of travel. We'll call this combination the attractiveness of a train to a customer who can choose between a number of different trains,

$$A_{C,T} = t_{C,T}$$

where, $A_{C,T}$ is the attractiveness of a train T to a costomer C and $t_{C,T}$, is the journey time on train T of the journey customer C wishes to make. This first attempt is very crude, but we can improve it by also taking into account the difference between the time the customer wanted to arrive at, and the actual arrival time of the train:

$$A_{C,T} = t_{C,T} - N(a_T - d_C),$$

where a_T is the actual arrival time of the metro, d_C is the time the passenger would have liked to arrive at, and N is a number telling us how much better or worse it is to be an extra minute early or late, compared with an extra minute spent traveling.

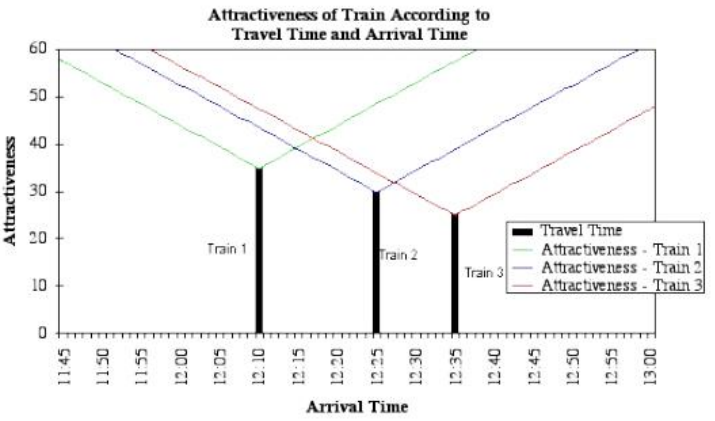


FIGURE 2

The final factor we will take into account is the crowding on the train. We modify the equation further, to get:

$$A_{C,T} = [t_{C,T} - N(a_T - d_C)] \times f(l_T/c_T),$$

where l_T is the trainload, c_T is the number of seats on the train, and f is a factor describing the importance of crowding.

As we can see, higher the attractiveness lower would be the preference for that particular train.

When this equation is used to describe the attractiveness of trains, the result looks something like the diagram above. Because of its appearance, this is called a "rooftop chart."

For a customer with a given arrival time, the attractiveness of a train is equal to the travel time, plus an allowance for the difference from their preferred arrival time. Using the graph, this means finding the train with the smallest attractiveness value at the customer's preferred arrival time. (Remember, as journey time increases, so does attractiveness. This means the trains become more "attractive" to customers as our attractiveness measure decreases.) For someone wanting to arrive at 12:15, Train 1 and Train 2 are equally attractive. Any earlier, and Train 1 becomes more attractive, any later and Train 2 wins.

Studying Data On Customer Behaviour: What we have described above is model-building largely by common sense and conjecture. The real effort goes into testing these hypotheses and calculating the true values of parameters such as N. We will not describe this process in detail, but here is an overview of how it works:

- Actual data on numbers of people travelling are collected. Cases where just one factor has changed in isolation are found (say, one train arrives 10 minutes later than previously, all the others stay the same).
- The numbers choosing that train before and after the change are analyzed to find a value for the factor N.

  Whole papers could be written on the approaches used to find different factors,

either in isolation or experimentation. The techniques might involve interviewing customers to assess their stated preferences for different options (for example, a fast, infrequent service, versus a slow frequent service), or analyzing very large amounts of historic data to assess whether the effects are truly linear (as assumed here) or whether more complex equations should be used.

Building a Model: Once equations have been arrived at to explain customer choices, we are ready to build a model. This will allow us to think about what might happen to customer behavior, and hence train loadings, if:

- Demand increases (or decreases!);
- The timetable changes (more trains, faster trains);
- The number of seats available changes (longer trains).

We'll look in detail at a simple case where there are only six trains, and only one origin and destination point say Rajiv Chowk. Obviously, in reality the situation would be much more complex, with trains calling at many points along the route. However, this simple example illustrates the general procedure.

The Demand Profile: First of all we need is to quantify, when people want to travel. We do this by splitting the time we want to study into time bands. In this case we are looking at a morning peak period, and the time is split into ten minute bands, as shown in the table below: We should ideally include the capacity of the trains and look at crowding, but for simplicity we'll leave that out for now.

Choosing The Best Train: Now, we simply apply the equation to see which train is most attractive in each time band. The table below shows the attractiveness values for each train and departure time, using a value of 0.5 for N. Remember that lower the attractiveness value, more attractive is the train.

Allocating Demand to trains: Since this is a simple model we'll assume that everyone boards the best train (in real models we assume some people board the second and third best trains, which does in fact happen in reality. This is sometimes referred to as a "fuzzy logic" approach). This gives the following train loads:

Train Number	Train1	Train2	Train3	Train4	Train5	Train 6
Train Load	551	559	807	499	487	397

This is the model in a nutshell. Now we can change the number of trains, the train times, the demand level or the demand profile and instantly see the effect.

Producing meaningful results for clients: Having the model isn't everything. We must understand the results it gives well enough to advise clients. Like, for example, if a client (DMRC here) was specifically interested in the number of passengers who would be standing on trains when a new change was introduced. When considering numbers standing,

Time	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6	Minimum	Best Train
7:00	30.0	42.5	65.0	70.0	80.0	90.0	30.0	Train 1
7:10	25.0	37.5	60.0	65.0	75.0	85.0	25.0	Train 1
7:20	20.0	32.5	55.0	60.0	70.0	80.0	20.0	Train 1
7:30	15.0	27.5	50.0	55.0	65.0	75.0	15.0	Train 1
7:40	20.0	22.5	45.0	50.0	60.0	70.0	20.0	Train 1
7:50	25.0	22.5	40.0	45.0	55.0	65.0	22.5	Train 2
8:00	30.0	27.5	35.0	40.0	50.0	60.0	27.5	Train 2
8:10	35.0	32.5	30.0	35.0	45.0	55.0	30.0	Train 3
8:20	40.0	37.5	25.0	30.0	40.0	50.0	25.0	Train 3
8:30	45.0	42.5	30.0	25.0	35.0	45.0	25.0	Train 4
8:40	50.0	47.5	35.0	30.0	30.0	40.0	30.0	Train 4
8:50	55.0	52.5	40.0	35.0	25.0	35.0	25.0	Train 5
9:00	60.0	57.5	45.0	40.0	20.0	30.0	20.0	Train 5
9:10	65.0	62.5	50.0	45.0	25.0	25.0	25.0	Train 5
9:20	70.0	67.5	55.0	50.0	30.0	20.0	20.0	Train 6
9:30	75.0	72.5	60.0	55.0	35.0	15.0	15.0	Train 6
9:40	80.0	77.5	65.0	60.0	40.0	20.0	20.0	Train 6
9:50	85.0	82.5	70.0	65.0	45.0	25.0	25.0	Train 6
10:00	90.0	87.5	75.0	70.0	50.0	30.0	30.0	Train 6

FIGURE 3

we found that statistical analysis of the model results was required to provide meaningful figures. To see why, consider the predictions of numbers standing produced in our example above:

Train number	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6
Capacity	250	200	300	200	200	250
Standing	301	359	507	299	287	147

So the model predicts a total of 1900 people standing.

CONCLUSION

The model used by us to predict the number of people traveling in a particular train or number of people standing in a train given a time frame is not a totally precise method to calculate the same. We are talking about people, their precise behavior is hard to predict. However, this model can help us predict how they "may" behave but we can never be certain. Also, all of the trains will have some people standing on certain days. This model has initially assumed that everyday is same and the actual number of people standing on any given day is same as average number of people standing on all days. As we have seen, random fluctuations in passenger numbers mean that this won't be so. A more accurate method for our calculation would be by considering the normal distribution, representing the average number of people standing on a given day and integrating the area above our acceptance region (capacity of train) to get a more accurate estimate. Over here, rather

than striving to understand everything about a subject we seek to know just enough to produce a solution which is good enough to bring about the changes required in the existing system.

References

[1] http://en.wikipedia.org/wiki/Mumbai_Suburban_Railway
[2] http://www.systemwide.com.au/train-load-predictor.php
[3] http://mumbailifeline.com/stats_and_facts.php

Umang Aggarwal,B.Sc.(H)Mathematics, 3rd Year, Lady Shri Ram College for women
E-mail address: umangagg@hotmail.com

Prachi Singhal, B.Sc.(H)Mathematics, 3rd Year, Lady Shri Ram College for women
E-mail address: prachi_singhal15@yahoo.co.in

Prachi Singhal, B.Sc.(H)Mathematics, 3rd Year, Lady Shri Ram College for women
E-mail address: prachi_singhal15@yahoo.co.in