Dedicated to

"The National Mathematical Year 2012"

PREFACE

Éclat, with its roots in French, means brilliance. This journal aims at providing a platform for the students who wish to publish their ideas and understanding for the various concepts of mathematics which they might have come across. To present diverse concepts, the journal has been divided into four sections - History of Mathematics, Rigour in Mathematics, Extension of Course Contents and Interdisciplinary Aspects of Mathematics. The work contained here is not original but consists of the review articles contributed by both faculty and students.

Feeling enriched and encouraged from the publication of last two volumes of Éclat, this year, the idea of the journal has also been propagated to other departments of our college and mathematics department of other colleges of the University of Delhi. Consequently, we received an overwhelming response. We hope such participation continues in future as well.

The entire department of mathematics of our college has been instrumental in the publication of this journal. Specially, we express our sincere thanks to the faculty advisors involved in, without whose guidance such an effort would not have been possible to come to its result. Its compilation has evolved after continuous research and discussion. We hope this journal to be a regular annual feature of the department of mathematics and would keep on encouraging the students to hone their skills in doing individual research, and writing academic papers. It is an opportunity to go beyond the prescribed limits of the text and to expand and share the knowledge of the subject.

We would like to thank all the authors who have contributed their articles for this volume of *Éclat*.

Editorial Team

Sakshi Arya
Shefali Sharma
Manushi Gupta
Prashansha Kaushik
Jaya Dutta

Contents

History of Mathematics

The history behind various mathematical concepts and great mathematicians is intriguing. Knowing the history can lead to better development of concepts and enables us to understand the motivation behind ideas. Also, it shows us what inspired eminent mathematicians and highlights the problems they came across during their research.

RAMANUJAN- THE INDIAN WIZARD OF MATHEMATICS

SAKSHI ARYA AND ROOPAM JINDAL

ABSTRACT. Although Ramanujan is unknown throughout the West, except in mathematical circles, the recognition of Ramanujan's name in India can perhaps only be compared with the recognition that the names of Newton and Einstein have in West. In India today, Ramanujan is revered more than any other Indian scientific figure. So who was this mathematician, who inspired two movies about his life from India, a third from England, and a fourth now in the planning stages in Hollywood? Srinivasa Iyengar Ramanujan was a mathematician so great that his name transcend jealousies, the one superlatively mathematician whom India has produced. In this article, we are giving a brief from the Ramanujan's personal life, and his contribution to the world of mathematics.

Early Life

Ramanujan was born on December 22,1887 to Komalattmal and Srinivasa in Erode, 150 kms from Chennai. His father, K. Srinivasa Iyengar worked as a clerk in a sari shop and used to earn only about twenty rupees per month. His mother was a shrewed and cultural lady with deep belief in God, she used to sing bhajans or devotional songs at a nearby temple. After Ramanujan's birth she gave birth to five more children but only two of his brothers survived. Ramanujan was deeply influenced by his mother, he was a man who grew up praying to stone deities, who for most of his life took counsel from a family goddess Namagiri, declaring it was she to whom his mathematical insights were owed.

Ramanujan- The Man

The image of Ramanujan's personality that comes from the reminiscences of those who knew him is one of a genial, shy and modest person who was devoutly religious as well as being devoted to mathematics. He had a somewhat shy and quiet disposition, a dignified bearing and pleasant manners.

Mystic. Ramanujan knew and was interested in Astrology, Vedanta philosophy, theories like the theory of relativity. He used to give occult explanations of subjects such as life after death and Psychic Phenomena. One day he was explaining a mathematical relation to Mr. R. Srinivasan(Retired professor of mathematics, Trivandram); then suddenly turned around and said "Sir, an equation has no meaning for me unless it expresses a thought of God". In that statement he saw the real Ramanujan, the philosopher-mystic-mathematician.

FIGURE 1. Ramanujan

Modesty. Ramanujan seems to have been an embodiment of *Hri*- the quality of modesty. According to Hardy, when Trinity College Fellowship of Rs.250 a year was announced to him, Ramanujan felt embarrased and worried. In effect he told Hardy, "How do I deserve it? In this health how can I justify accepting it?" Hardy replied "What you have done deserves much more than this. This is given to you to enable you to live in comfort. There is no obligation of any kind either by the way of teaching or by the way of doing any further research".

EDUCATION

Ramanujan entered the town high school first at the age of ten. He used to impress his teachers and classmates with his extraordinary intuition and astounding proficiency in several branches of mathematics. During school days he became something of a minor celebrity.

Let us observe this instance: One day in a small high school in southern India, a teacher was explaining division and said that "If you divide any number by itself, you get 1". The teacher turned around to see a tiny hand trying to reach the ceiling. "Oh by the gods, him again!" That Iyengar boy with his horribly difficult questions. "Yes Ramanujan?" The small boy with shining eyes stood up. He spoke slowly, with the calm confidence of one who did not need to be told he was the best in the class. "Is zero divided by zero also equal to one?" This was perhaps an indication of Ramanujan's unusual insight into the behaviour of numbers.

Through college students boarding with his family he was introduced to more advanced mathematics and began learning on his own. By age 11, he had exhausted the mathematical knowledge of two college students, who were lodgers at his home. By 14, he was achieving merit certificates and academic awards. When Ramanujan was sixteen and still in high school an elderly friend who knew of his precocious mathematical talent gave him *George Carr's Synopsis* of elementary results in pure and applied mathematics. This two-volume encyclopedic tome contained six thousand theorems on all fields of mathematics. As Ramanujan read and worked his way through these theorems he discovered that he could derive results that were not in Carr. This was the beginnings of Ramanujan's mathematical productions and set the tone for his mathematical career.

When he completed high school in 1904 he took a competitive examination in which he earned such high marks that he was given a scholarship to a local college, the Government College at Kumbakonam. There his mathematical development proceeded well but he was not much interested in the other subjects. In part this was a matter of him being so fascinated with mathematics that he did not want to spend his time thinking about other academic subjects. Also some of the subjects were positively distasteful to Ramanujan, in particular psychology, which was probably really physiology. His course in psychology involved the dissection of frogs. Ramanujan was a devout Hindu and was appalled at what he perceived as senseless and immoral cruelty. When the examinations came at the end of the year he did very well in mathematics but he failed the other subjects and therefore lost his scholarship.

Despite his academic failure Ramanujan threw himself into the pursuit of new results in mathematics. He worked long and hard. A friend of Ramanujan known as Sandow related the following conversation with Ramanujan:

Sandow: Ramanju, they all call you a genius.
Ramanujan: What! Me, a genius! Look at my elbow, it will tell you the story.
Sandow: What is all this, Ramanju? Why is it so rough and black?
Ramanujan: Night and day I do my calculations on slate. It is too slow to look for a rag to wipe it with. I wipe the slate almost every few minutes with my elbow.
Sandow: So, you are a mountain of industry. Why use a slate when you have to do so much calculation? Why not use paper?
Ramanujan: When food itself is a problem, how can I find money for paper? I may require four reams of paper every month.
Sandow: Tell me honestly what do you do for your food. Do you work anywhere?
Ramanujan: Our professor Seshu Ayyar introduced me to Dewan Bahadur R. Ramachandra Rao, the Collector of Nellore. That great man has been providing me with money every month.
Sandow: Then why do you worry yourself?
Ramanujan: How long am I to depend on others? The humiliation of it has gone deep into me. Therefore I did not take the money from last month.
Sandow: What a rash thing to do? What are you going to do now?
Ramanujan: I joined the Madras Port Trust Office as a clerk on the 9th of this month. Pay Rs. 25 a month.

He was the most strangest man in all of mathematics, probably in the entire history of science. He has been compared to a bursting supernova, illuminating the darkest, most profound corners of mathematics.

Married Life

Ramanujan got married to Janaki on July 14,1909. At the time of marriage, Janaki was only 9 years old and joined her husband after 3 years and nursed him till his death.

FIGURE 2. Janaki Ramanujan (March 21, 1899–April 13,1994)

Association with Hardy

The role of Hardy in the life and career of Ramanujan is beyond praise. Their association started with the historic letter of Ramanujan to Hardy in 1913. He has been very generous in doing his very best in propagating the mathematics of Ramanujan, through lectures and articles. Hardy contracted the Indian office in London to bring Ramanujan to England. He reached England on April 1914 abreast and ahead of contemporary knowledge and he succeeded in recreating in his field his own unaided powers, a rich half century of European mathematicians.

FIGURE 3. Hardy and Ramanujan

On January of 1913 Ramanujan sent a letter to G. H. Hardy at Cambridge. Included with Ramanujan's letter were nine pages of Ramanujan's advanced mathematical work. Hardy, the pre-eminent mathematician of his time, would forever be changed by this humble but self-assured letter which began:

"*I beg to introduce myself to you as a clerk in the Accounts Department of the Port Trust Office at Madras on a salary of 20 pounds per annum. I am now about 26 years of age. I have had no university education... being inexperienced I would very highly value any advice you give me...*" Included in his letter were 100 theorems that Ramanujan had found in various parts of mathematics. Hardy, no stranger to letters from the unimpressive and suspect, was cautiously impressed. Upon closer examination, Hardy became convinced that he was reading from pages authored by a true mathematical genius.

Hardy, together with Littlewood, studied the long list of unproved theorems which Ramanujan enclosed with his letter. On 8 February he replied to Ramanujan, the letter beginning:- "I was exceedingly interested by your letter and by the theorems which you stated. You will however understand that, before I can judge properly of the value of what you have done, it is essential that I should see proofs of some of your assertions. Your results seem to me to fall into roughly three classes: (1) there are a number of results that are already known, or easily deducible from known theorems; (2) there are results which, so far as I know, are new and interesting, but interesting rather from their curiosity and apparent difficulty than their importance; (3) there are results which appear to be new and important...". Ramanujan was delighted with Hardy's reply and when he wrote again he said: "I have found a friend in you who views my labours sympathetically. I am already a half starving man. To preserve my brains I want food and this is my first consideration. Any sympathetic letter from you will be helpful to me here to get a scholarship either from the university or from the government."

Hardy asked Ramanujan to come to Cambridge. At first Ramanujan was unsure; moving conflicted with Ramanujan's Brahmin background. But finally he consented to the move and was admitted to bachelor programme in Trinity College in 1914. Ramanujan collaborated with Hardy on seven papers, as well as publishing many of his own works.

According to Hardy "The limitations of his knowledge were as startling as its profundity. Here was a man who could work out modular equations and the theorems of complex multiplications to orders unheard of, whose mastery of Continued Fractions was on the formal side at any rate beyond that of any mathematician in the world, who had found himself the Functional Equations, the Zeta Function, and the Dominant terms of many of the most famous problems in the Analytic Theory of Numbers; and he had never heard of Doubly Periodic Functions or of Cauchy's Theorems, and had indeed the vaguest idea of what the Function of a Complex Variable was."

Ramanujan was awarded the B.A. degree in 1916 for his work on highly composite numbers. He was the second Indian and the youngest person in the world to become a fellow of the Royal society.

His Works

To properly appreciate the legend of Ramanujan, it is important to assess his legacy to mathematics. For this task, we recall some of his works.
Once a teacher was teaching square root. She said that $9 = 3^2$. And $3 = \sqrt{9}$. But this boy wrote something else. He wrote :

$$3 = \sqrt{1+8}$$

$$3 = \sqrt{1+(2*4)}$$

$$3 = \sqrt{1+(2*\sqrt{16})}$$

$$3 = \sqrt{1 + (2 * \sqrt{1 + 15})}$$
$$3 = \sqrt{1 + (2 * \sqrt{1 + (3 * 5)})}$$

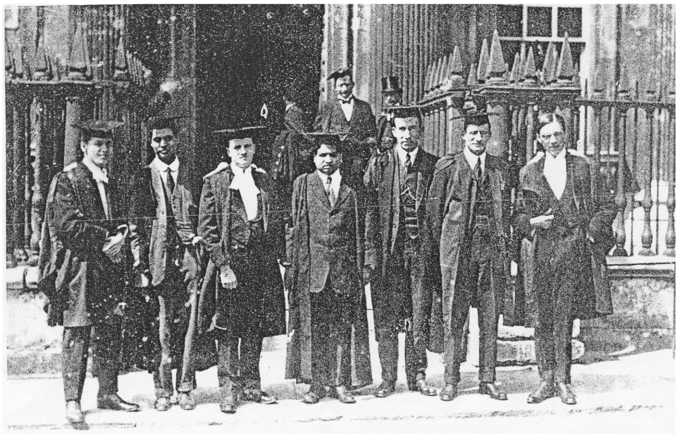


FIGURE 4. Ramanujan with other scientists at Trinity College

Ramanujan, with the help of Ramaswamy Aiyer who was the Deputy Collector in Tirtukkoileir in 1910, had his work published in the Journal of Indian Mathematical Society. One of the first problems he posed in the journal was the one mentioned below:

Evaluate

$$\sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + \cdots}}}.$$

He waited for a solution to be offered in three issues, over six months, but failed to receive any. At the end, Ramanujan supplied the solution to the problem himself. On page 105 of his first notebook, he formulated an equation that could be used to solve the infinitely nested radicals problem.

$$x + n + a = \sqrt{ax + (n + a)^2 + x\sqrt{a(x + n) + (n + a)^2 + (x + n)\sqrt{\cdots}}} \ .$$

Using this equation, the answer to the question posed in the Journal was simply 3.

Fractional differentiation. The differential coefficient of x^n is nx^{n-1}. The coefficient n can be written as $\frac{n!}{(n-1)!}$. If we differentiate x^n twice, we get $n(n-1)x^{n-2}$. The coefficient may be written as $\frac{n!}{(n-2)!}$. In this approach we can have only integral differentiation. Ramanujan suggested that we have the function $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}\,dx$ and when n is a positive integer, $\Gamma(n+1) = n\Gamma(n)$ and note that $\Gamma(n+1) = n!$.

Now, if D= differential coefficient; D^2= second differential coefficient and in general D^m= mth differential coefficient, then we may write ,

$$Dx^n = \frac{\Gamma(n+1)}{\Gamma(n)}.x^{n-1}\,;$$
$$D^2x^n = \frac{\Gamma(n+1)}{\Gamma(n-1)}.x^{n-2}\,;$$
$$\vdots$$
$$D^mx^n = \frac{\Gamma(n+1)}{\Gamma(n-m+1)}.x^{n-m}\,;$$

and on putting $m = 1/2$, we have

$$D^{1/2}x^n = \frac{\Gamma(n+1)}{\Gamma(n+1/2)}.x^{n-1/2}.$$

This fact enables us to have fractional differentiaton.

Every Integer a Friend of Ramanujan. According to Hardy, Ramanujan could remember the idiosyncracies of numbers in an almost uncanny way. According to Littlewood, every integer was one of the personal friends of Ramanujan. Ramanujan was in a sanatorium in Putney, London, United Kingdom. Hardy went to see him.

Hardy: I came in the taxi-cab 1729. It is rather a dull number. I hope it is not an unfavourable omen.
Ramanujan: No it is a very interesting number.
Hardy: How?
Ramanujan: It is the smallest number expressible as the sum of two cubes in 2 ways. $(1729 = 1^3 + 12^3 = 9^3 + 10^3)$.
Hardy: What is the smallest number expressible as the sum of fourth powers?
Ramanujan: (After thinking for a moment) I see no obvious example. I think the number must be very large.
After narrating this anectode, Hardy refers to a number by Euler
$N = 158^4 + 59^4 = 134^4 + 134^4$.

Ramanujan prime. The prime counting function $\pi(x)$, $x \in \mathbb{N}$ is the function giving the number of primes less than or equal to a given number x. For example, there are no primes less than or equal to one, so $\pi(1) = 0$. There is only one prime ≤ 2, which is 2 itself, so $\pi(2) = 1$. There are two primes (2 and 3) ≤ 3, so $\pi(3) = 2$. And so on.

The case for $\pi(x) - \pi(x/2) \geq 1$ for all $x \geq 2$ is **Bertrand's postulate**. It was proved in 1850 by Chebyshev. Ramanujan (1919) gave a new proof of Bertrand's postulate. Then he proved the generalization that $\pi(x) - \pi(x/2) \geq 1, 2, 3, 4, 5, \ldots$ if $x \geq 2, 11, 17, 29, 41, \ldots$, respectively.
The converse of this result gives us the definition of Ramanujan primes:

The nth Ramanujan prime is the least integer R_n for which $\pi(x) - \pi(x/2) \geq n$, for all $x \geq R_n$. Equivalently, Ramanujan primes are the least integers R_n for which there are at least n primes between x and $x/2$ for all $x \geq R_n$.

His research marched on undeterred by any environmental factors – physical, personal, economic or social. Magic squares, continued fractions, hyper-geometric series, properties and partition of numbers – prime as well as composite, elliptic integrals, and several other such regions of mathematics engaged his thought. Today Ramanujan's work has some applications in particle physics or in the calculation of π up to a very large number of decimal places. His work on Riemann's Zeta Function has been applied to the pyrometry, the investigations of the temperature of furnaces. His work on the Partition Numbers resulted in two applications, new fuels and fabrics like nylons. He recognized the multiplicative properties of the coefficients of modular forms that we know as cusp forms and his conjectures are still attracting the attention of quite a large group of mathematicians of our time.

The lost Notebook

Since the age of 16, Ramanujan had been steadily filling up notebooks with mathematical formulas. But the "Lost Notebook" refers specifically to the notebook that he wrote in the final year of his life. Despite its name, the "Lost Notebook" actually is not a notebook at all. It is a hodgepodge of papers, on which Ramanujan wrote mathematical formulas. Berndt (Bruce Carl Berndt is an analytic number theorist who is probably best known for his work explicating the discoveries of Srinivasa Ramanujan. He is a coordinating editor of The Ramanujan Journal and, in 1996, received an expository Steele Prize from the American Mathematical Society for his work editing Ramanujan's Notebooks) says,"The Lost Notebook represents possibly his deepest work. Thus, even though he worked in great pain, with his physical strength ebbing away, Ramanujan's creativity did not diminish but rather gained in strength". After his death, Ramanujan's materials followed a circuitous path, which is when the "Lost Notebook" became lost. The papers were first shipped to the University of Madras library, which later sent many of the materials to Hardy for publication. Then sometime between the late 1930s and late 1940s, Hardy passed on the "Lost Notebook" to the English mathematician G.N. Watson. The "Lost Notebook" sat in Watson's office for years until he died in 1965. That's when a fellow mathematician found the "Lost Notebook" among Watson's papers, which covered the floor of his large office to the depth of a foot. However, not realizing the significance of Ramanujan's papers, he simply sent them to Trinity College at Cambridge. The "Lost Notebook" finally surfaced eight years later, when Berndt's co-writer, Andrews, discovered the "Lost Notebook" in 1976 while sifting through papers left to the Trinity College Library at Cambridge University. As Berndt puts it, "The discovery of this Lost Notebook caused roughly as much stir in the mathematical world as the discovery of Beethoven's tenth symphony would cause in the musical world".

End Years

Ramanujan had to face considerable physical and emotional strain during the years of his stay in England. While the strangeness of cultural environment in the English society caused emotional strain, factors like unavailability of vegetarian food due to war conditions, his food habbits, the climatic change became a cause of his physical strain. He being a strict vegetarian used to cook his own food. Ramanujan developed a tubercular tendency on his return to India. He was subjected to fits of depression, and had a premonition of his death and was a difficult patient. He was brought to Madras for expert medical treatment in 1920 but ultimately he died on 26th April 1920 at a very young age of 32.

Reminiscences of Mrs. Janaki Ramanujan. His Last Days-" He returned from England only to die", as the saying goes. He lived for less than a year. Throughout this period I lived with him without break. He was only skin and bones. He often complained of severe pain. Inspite of it he was always busy doing his Mathematics. That evidently, helped him to forget his pain. He was uniformly kind to me. In his conversation he was full of wit and humour. Even when mortally ill he used to crack jokes. One day he confided in me that he might not live beyond 35 and asked me to meet the event with courage and fortitude. We did not have children. I have only two momentos with me. One consists of the two vessels in which I used to warm water for fomentation whenever he complained of acute pain. The second consists of two portraits of himself.

A Closing Note

The fact that Ramanujan's early years were spent in a scientifically sterile atmosphere, that his life in India was not without hardships that under circumstances that appeared to most Indians as nothing short of miraculous, he had gone to Cambridge, supported by eminent mathematicians, and had returned to India with very assurance that he would be considered, in time as one of the most original mathematicians of the century these facts were enough, more than enough, for aspiring young Indian students to break their bands of intellectual confinement and perhaps soar the way what Ramanujan had.

FIGURE 5. Ramanujan's Stamp produced in 1962 on his 75th Birth Anniversary

References

1. Robert Kanigel, *The Man Who Knew Infinity: a Life of the Genius Ramanujan*, Washington Square Press, 5 edition, 1992.
2. George E. Andrews, Bruce C. Berndt, *Ramanujan's Lost Notebook Part I*, Springer-Verlag, Berlin, New York, 2005.
3. S.R.Ranganathan, *Ramanujan-The Man and the Mathematician*, Asia Pub. House, London, 1967.

Sakshi Arya, B.Sc.(H)Mathematics, 3rd year, Lady Shri Ram College For Women
E-mail address: arya.sakshi44@gmail.com

Roopam Jindal, B.Sc.(H)Mathematics, 2nd year, Lady Shri Ram College For Women
E-mail address: roopam.jindal3@gmail.com

CONVERGENCE OF THE DISTINCT

DR. SMITA SAHGAL
AND
MS. BHAVNEET KAUR

ABSTRACT. History and Mathematics appear to be two ends of a pole. However, a closer study of the two disciplines helps us to locate areas of convergence that break the myth of their incompatibility. The paper intends to explore such avenues where their symbiosis stands the test of time. Renaissance in Europe was one such time which witnessed a cross disciplinary interaction for the benefit of humanity and the paper is an attempt to trace the trajectory of the coming together seemingly diverse themes over time and space.

It feels strange to write about the intersection of two extremely diverse disciplines; mathematics which can give me goose pimples till this date and history, the passion of my life. Yet a long innings at academics and ensuing research work has taught me that there are points of convergence in all disciplines. Both studying and teaching history has opened up vistas where people of cross disciplines can tread together. In fact one of the greatest scholars of ancient Indian stream, D.D.Kosambi was professionally trained as a mathematician and made effective use of this science in researching various fields of history. He was the first to use interdisciplinary methods in historical investigation, and to deploy his mathematical and scientific genius to the study of issues as wide ranging as the reading of ancient texts to their contexts, numismatics, philology, religion, archaeology, anthropology and historical reconstruction. He truly enriched the discipline of Indology and endeavored to make it a scientific discipline.

Inspired by D.D.Kosambi's methodology, Bhavneet and I* decided to investigate one such intersection in the worlds of mathematics and history of art. This is not from the world of Indology. Rather it revolves around the concept of Perspective. My interest in the notion was triggered off through the study of renaissance art. I teach a paper to II year History Honors titled The Rise of Modern West and Italian Renaissance is among the first topics to be unraveled.

Renaissance in Western Europe is generally associated with revival of arts and letters between the 14th and 16th centuries that marked a break with the immediate past and heralded an era of modernity. It was characterized by a change in human attitude towards

*References to I/me are to Dr. Smita Sahgal of History Department.

the problem of existence of humanity. In the middle ages painters, philosophers, and writers used their talents for a single purpose- to praise god and make His purpose plain. But during the Renaissance each branch of intellectual activity became distinct from the other and each was justified in terms of its means rather than its end. That is, a painting succeeded in terms of its excellence as a painting, quite apart from the purpose from which it was painted.

For art and science to evolve, this disassociation from a common purpose was essential. There was a new interest in aesthetics and art became a medium of praise of human beauty. But now the painter had to become conscious of technique before he could choose it to become a medium to convey his/her ideas. This implied the evolution of visual arts in a form of science with a sound theoretical grounding. The subjects may be drawn from the Antiquity, i.e., ancient Greece and Rome but the execution was distinctly novel. In the world of visual art the break with the immediate [i.e., medieval] past was visible in more than mere rejection of thematic inspirations. There was a clear effort to bring technique centre stage. No other epoch has brought together so many great painters, sculptors and architects who also duplicated as scientists of tall order. In the realm of visual art radical changes in technique appeared and one such was the concept of Perspective which was in some ways an amalgam of mathematical observations of the remote past and current interventions in the discipline of geometry.

What exactly is perspective? My definition is not confined for the benefit of students of mathematics but is intended to provide explanations to student body across board. The problem for the artist was to represent a three dimensional world on a two dimensional surface. Perspective is the use of lines and angles to create an illusion of three dimensional shapes. It lets the artist control the spatial elements of his drawing- what makes visual realism, or illusionism, look ‘right’. This is especially true when objects are drawn on a flat surface. One needs to be familiar with the fact that the distant objects appear smaller than those in the foreground. Following are some terms used in perspective with which we need to become a little familiar.

(1) **Horizon line:** The horizon line exists in all pictures and sometimes corresponds to actual horizon. Sometimes it is imaginary for the purposes of drawing. The location of the Horizon line is always at eye level whether you are sitting or standing.
(2) **Centre of vision:** The centre of vision is at the same level as horizon.
(3) **Vanishing Point:** The vanishing point is the apparent meeting point of all lines that are parallel to the ground. The point becomes clearer with the following demonstration of a rail road track [Figure1]. The rail road track appears to get closer as they recede in space towards horizon. The point where they appear to meet is the vanishing point.

Linear perspective works by representing the light that passes from a scene through an imaginary rectangle (the painting), to the viewer’s eye. It is similar to a viewer looking

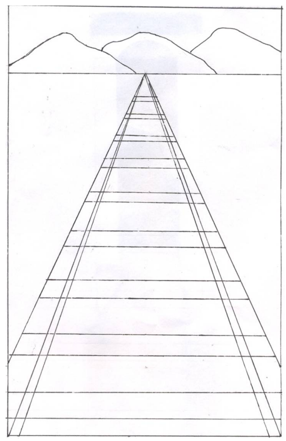


FIGURE 1.

through a window and painting what is seen directly onto the windowpane. If viewed from the same spot as the windowpane was painted, the painted image would be identical to what was seen through the unpainted window. Each painted object in the scene is a flat, scaled down version of the object on the other side of the window. As each portion of the painted object lies on the straight line from the viewer's eye to the equivalent portion of the real object it represents, the viewer cannot perceive any difference between the painted scene on the windowpane and the view of the real scene. All perspective drawings assume the viewer is a certain distance away from the drawing. Objects are scaled relative to that viewer. Additionally, an object is often not scaled evenly: a circle often appears as an ellipse and a square can appear as a trapezoid. This distortion is referred to as foreshortening [Figure 2]. Let's now discuss some basic ideas about perspective.

(1) The perspective image P′Q′ of a line segment PQ is also a line segment. We observe that EPQ lies in a plane and the intersection of that plane with the picture plane is the line containing the line segment P′Q′.
(2) The line segment PQ that lies in a plane parallel to the picture plane has a perspective image P′Q′ that is parallel to PQ. The lines containing PQ and P′Q′ cannot intersect. Furthermore, these two lines are coplanar, since they lie in the plane containing EPQ. Since non intersecting coplanar lines are parallel, PQ and P′Q′ are parallel [Figure 3].
(3) The image of the line L in a space not parallel to the picture plane appear to vanish at a point V in the picture plane. Figure 4 shows the sketch of a street plan made by artist Perruzi in 15th century that indicates common vanishing point for all straight lines other than the one parallel to the horizon line converge at the vanishing point.

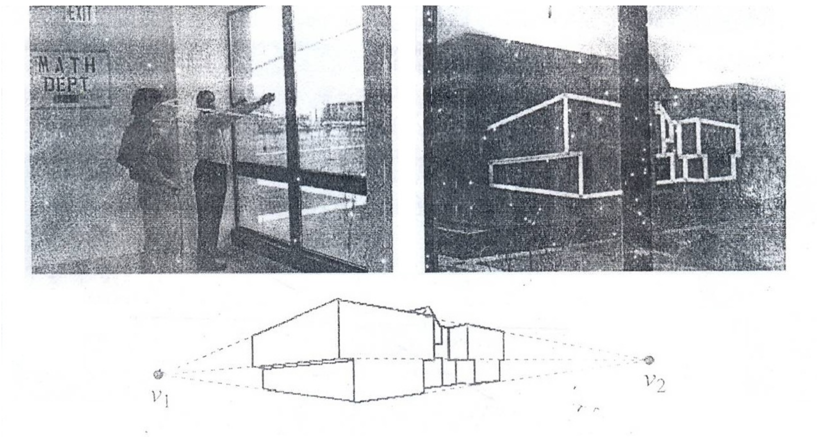


FIGURE 2.

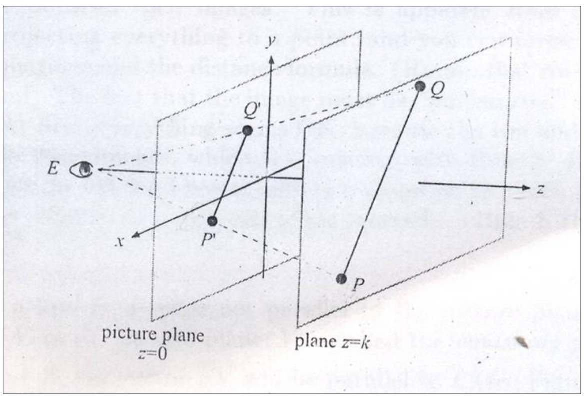


FIGURE 3.

Our intention here is not to give a lesson in a course on perspective but only to trace its history in the western world and to that an extent some basic comprehension of the concept is essential and therefore we decided to make the diagram and give a preview of what the talk would entail.

Before beginning the discussion of perspective in western art, it may be important to acknowledge the contribution made by Al-Haytham who around 1000 CE gave the first correct explanation of vision, showing that light is reflected from an object into the eye. He studied the complete science of vision, called *Perspectiva* in medieval times, and although he did not apply his ideas to painting, the renaissance artists later made use of Al-Haytham's optics.

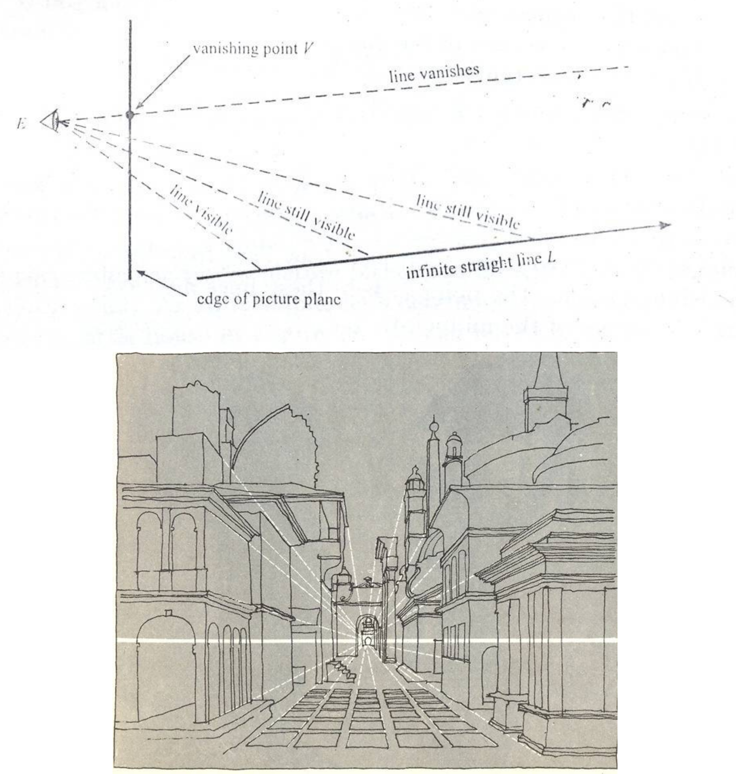


FIGURE 4.

Going back to Ancient Greek phase one has to concur that the artist of the day had some notion of this concept and that it was employed in visual arts as well in stage setting and architecture to provide an illusion of depth. However, there is no evidence that the artists understood the precise mathematical laws which govern correct representation. It was actually from the time of the renaissance that mathematics was used in a systematic way to make realistic paintings and more important to theorize these ideas in a scientific method for the benefit of posterity.

In the 13th century the celebrated painter Giotto was able to create the impression of depth by following certain rules. He inclined lines above eye level downwards as they moved away from the observer and lines below eye level upwards as they moved away from the observer. Similarly the lines to the left or right would be inclined towards the centre. Although not a precise mathematical formulation, Giotto clearly attempted to represent depth in space. His paintings were also different from the medieval ones as he threw in some action on the screen which was a contrast to lifeless medieval figures where perhaps the idea had been only to convey the spirit of the theme. In the medieval paintings (such as the one in Figure 5) one hardly gets a sense of space, depth or even human expression. But from the time Giotto it was almost like a live story being unfolded. Figure 6(a) narrates an action packed story of Christ's trial before his execution and Figure 6(b) refers to the meeting of Anna and Joschim at the Golden Gate, parents of Mother Mary, with onlookers in the background.

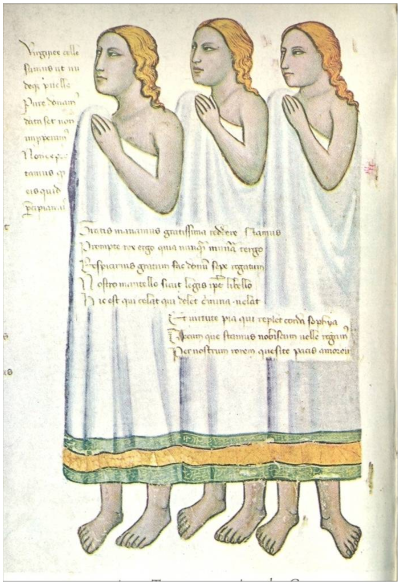


FIGURE 5.

The new realism heralded by Giotto was dramatically advanced in the 15th century by three great Florentines, Brunelleschi, Masaccio and Donatello and brought to its height by the Umbrian Piero della Francesca. Brunellschi scientifically plotted the laws of linear perspective for the first time, applying mathematics he learnt from Toscanelli, who counseled Christopher Columbus in plotting of maps. He appeared to have made the discovery around 1413. He understood that there should be a single vanishing point to which all parallel lines in a plane, other than the plane of the canvas converge. Also important was his understanding of the scale and he correctly computed the relation between the actual length of an object and its length in the picture depending on its distance behind the plane of the canvas. He demonstrated a geometrical method of perspective, used today by artists, by painting the outlines of various buildings onto a mirror. He set up a demonstration of his painting of the Baptistry. He had the viewer look through a small hole on the back of the painting, facing the Baptistry. He would then set up a mirror facing the viewer which reflected his painting. To the viewer, the painting of the Baptistry itself was nearly indistinguishable. As mentioned, Brunelleschi likely understood the mathematics behind

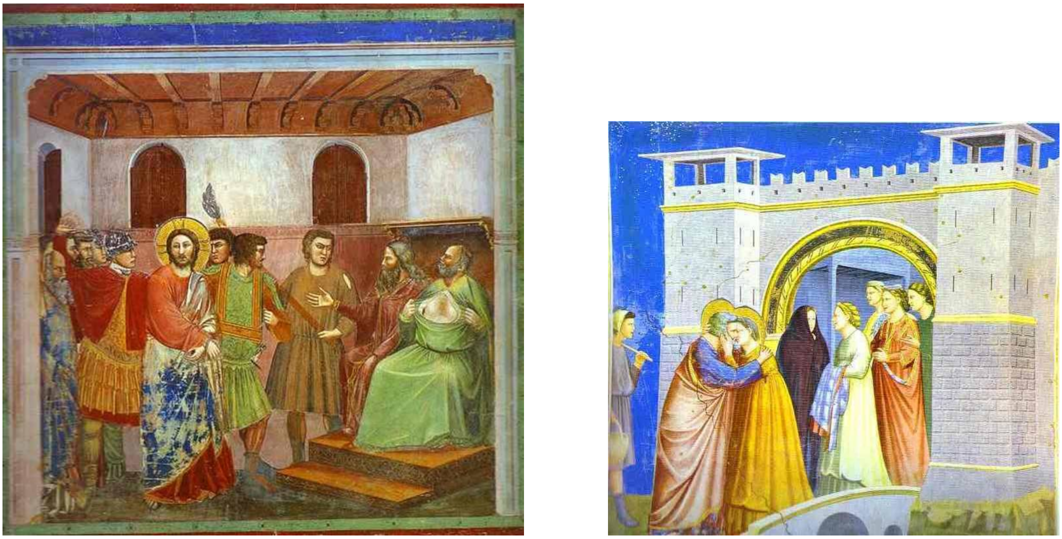


FIGURE 6. {a,b}

perspective but did not publish it. Alberti wrote "*Della Pittura*", a treatise on proper methods showing distance in painting. He was able to calculate the apparent height of a distant object using two similar triangles.

In viewing a wall, for instance, the first triangle had a vertex at the top and the bottom of the wall. The bottom of the triangle is the distance from the viewer to the wall. The second similar triangle, had a point at the viewer's eye and had a length equal to the viewer's eye from the painting. The height of the second triangle i.e., the height of the wall in the painting could be calculated due to the similarity of the triangles.

The scientific conception of art, which formed the basis of instruction in academies, begins with Leon Alberti. He was the first to express the idea that mathematics was the common ground of art and the sciences, as both the theory of proportions and of Perspective were mathematical disciplines.

Leon Battista Alberti was a major figure of renaissance, a philosopher, architect, musician, painter, sculptor and more important an elaborator of mathematical perspective and a theoretician of art. He wrote two treatises, the first was written in Latin in 1435 and titled *De Pictura* and the second dedicated to Brunelleschi, was an Italian work written in the following year and given the title *Della Pittura*. One should not confuse the latter as a translation of the other. Rather Alberti wrote these for different kinds of audience. The Latin work was technical and was for scholarly consumption while other was aimed at a general audience. In his *De Pictura*, Alberti defined painting from a scientific point of view which showed how fundamental he considered the notion of Perspective to be. In his words a painting was an intersection of a visual pyramid at a given distance, with a fixed centre and a defined position of light, represented by art with lines and colors on a given surface. Alberti worked on the principles of geometry and had a sense of the science of optics. He gave a precise concept of proportionality which determines the apparent size of an object in the picture relative to its actual size and distance from an observer.

One of the most famous examples used by Alberti was that of a floor covered with squared tiles called pavimento. Through this illustration what Alberti tried to do was to give a precise concept of proportionality which determines the apparent size of an object in the picture relative to its actual size and distance from the observer. We present a diagram with which Alberti tried to explain the perspective for the tiled floor [Figure 7]. The non horizontal floor lines are determined by spacing them equally along the base line and letting them converge to a vanishing point on the horizon. The horizontal floor lines are then determined by choosing one of them arbitrarily, thus determining one tile on the floor and then producing the diagonal of this tile to the horizon. The intersections of this diagonal with the non horizontal lines are the points through which the horizontal lines pass. This is certainly true on the actual floor and hence it remains true in the perspective view.

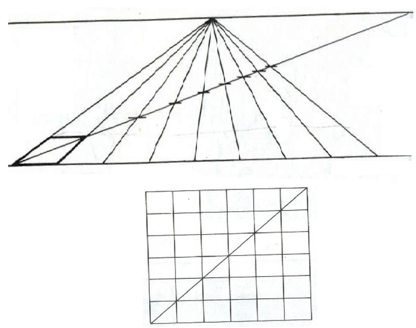


FIGURE 7.

The most mathematical of all the works on perspective written by Italian renaissance artist in the middle of the 15th century was by Piero Della Francesca. He went back to the Greek mathematicians [especially Euclid] for theoretical support, in order to write a treatise on perspective which helped even second rate painters to draw the relative size of objects correctly. He observed that measurement had become as important to art as drawing. In *Trattato d'abaco* which he probably wrote around 1450, Piero included material on arithmetic and algebra and a long section on geometry which was very unusual for the texts at that time. It also contained original mathematical results which again was atypical for the book written in the style of a teaching text. His book begins with a description of painting. He said, "Painting has three principal parts, which are drawing, proportion and coloring. Drawing we understand as meaning outlines and contours contained in thing.

Proportion we say is these outlines and contours positioned in proportion in their places. Coloring we mean as giving the colors as shown in things, light and dark according as the light makes them vary. Of the three parts I intend to deal only with proportion, which we call perspective, mixing in with it some part of drawing, because without this perspective cannot be shown in action. ..... We shall deal with that part which can be shown by means of lines, angles and proportion, speaking of points, lines, surfaces and bodies." Perhaps it is more accurate to say that he was studying the geometry of vision which he later made clearer when he observed, " first is the sight, that is to say the eye; second is the form of the thing seen; third is the distance from the eye to the thing seen; fourth are the lines which leave the boundaries of the object and come to the eye; fifth is the intersection which comes between the eye and the thing seen, and on which it is intended to record the object". In fact in the second of the three volumes Piero examined how to draw prisms in perspectives.

Artists of the time talked of points, lines and angles and described their subjects in terms of squares, cubes and tetragons. The goal of each artist whether interested in shadows, coloring, anatomical studies or perspectives was the same; the mastery of realism.

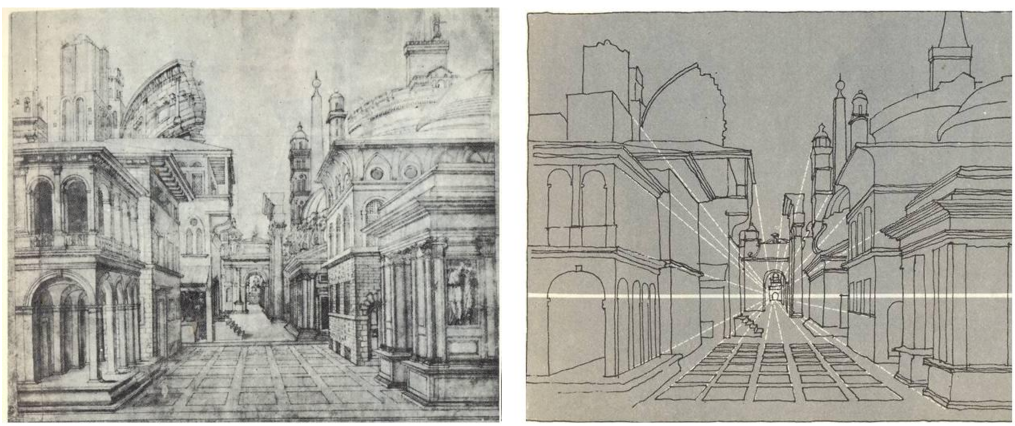


FIGURE 8. {a,b}

Before the picture would be finally executed, artists of renaissance carefully plotted converging lines in sketches made separately that would highlight the vanishing point. Following is one such illustration by Sienese painter and architect Baldassare Peruzi [Figure 8{a,b}]. The street scene shows how perspective conveys an illusion of depth on a flat surface. At the bottom in the simplified version of the sketch, superimposed white lines follow Peruzi's principal lines, converging at a single 'vanishing point'. The concept was used by other Venetian painters like Tinterreto to highlight the sense of space [Figure 9{a,b}].

There is little doubt that over a century from the time of Giotto the level of realism in painting had advanced considerably given the fact that the systems of perspective had been

FIGURE 9. {a,b}

worked out and the human body had been examined, dead and live at close range. By the time Leonardo Da Vinci arrived on the screen the system of linear perspective had been nearly perfected. In his early writings we find him echoing the precise theory of perspective as set out by Alberti and Piero. He wrote, "Perspective is a rational demonstration by which experience confirms that the images of all things are transmitted to the eye by pyramidal lines. Those bodies of size would make greater or lesser angles in their pyramids according to the different distances between the one and the other. By a pyramid of lines we imply those which depart from the superficial edges of bodies and converge over a distance to be drawn together in a single point". He developed mathematical formulae to compute the relationship between the distance from the eye to the object and its science on the intersecting plane, that is the canvas on which the picture would be painted. "If you place the intersection of one meter from the eye, the first object, being four meters from the eye, will diminish by three quarters of its height on the intersection; and if it is eight meters from the eye it will diminish by seven-eights and if it is sixteen meters away it will diminish by fifteen-sixteenths, and so on. As the distance doubles the diminution will double."

By around 1490 Vinci was convinced that perspective system required to be more nuanced. He realized that picture painted in correct linear perspective only looked if viewed from on exact location. A painting on the wall would not be viewed from one correct position only as in some cases the painting may well be placed well above the view's head. Leonardo then arrived at new conclusion that there should be different types of perspectives; artificial perspective which was the way a painter projected on to a plane and which may be seen foreshortened by an observer viewing at an angle; and natural perspective which reproduces faithfully the relative size of objects depending on their distance. In natural perspective, Leonardo claimed quite correctly, that objects would be of the same size if they lie on the circle centered on the observer. Moreover Leonardo Da Vinci looked at compound perspective where the natural perspective would be combined with a perspective viewed produced by viewing at an angle. There is little doubt that it was Vinci more than any other artist of the renaissance who successfully fused mathematics and art into a single concept. Leonardo Da Vinci, distrusted Brunelleschi's formulation of perspective as it failed to take into account the appearance of objects held very close to the eye. Leonardo called Brunelleschi's method artificial perspective projection. Projections closer to the image be held by the human eye he named natural perspective. Moreover, he looked at compound perspective where the natural perspective would be combined with a perspective produced by viewing at an angle. We, now, discuss variety of perspective drawings:

(1) *One point perspective* is typically used for roads, railway tracks, hallways, or buildings viewed so that the front is directly facing the viewer. Any objects that are made up of lines either directly parallel with the viewer's line of sight or directly perpendicular can be represented with one-point perspective. One-point perspective exists when the picture plane is parallel to two axes of a Cartesian scene. If one axis is parallel with the picture plane, then all elements are either parallel (either horizontally or vertically) or perpendicular to it. All elements that are parallel to the picture plane are drawn as parallel lines. All elements that are perpendicular to the picture plane converge at a single point (a vanishing point) on the horizon.
(2) *Two-point perspective* can be used to draw the same objects as one-point perspective, rotated: looking at the corner of a house, or looking at two forked roads shrink into the distance, for example. One point represents one set of parallel lines; the other point represents the other.
(3) *Three-point perspective* is usually used for buildings seen from above (or below). In addition to the two vanishing points from before, one for each wall, there is now one for how those walls recede into the ground. Three-point perspective exists when the perspective is a view of a Cartesian scene where the picture plane is not parallel to any of the scene's three axes. Each of the three vanishing points corresponds with one of the three axes of the scene.

By inserting into a Cartesian scene a set of parallel lines that are not parallel to any of the three axes of the scene, a new distinct vanishing point is created. Therefore, it is possible to have an infinite-point perspective if the scene being viewed is not a Cartesian scene but instead consists of infinite pairs of parallel lines, where each pair is not parallel to any other pair. In linear perspective, the ratio at which more distant objects decrease in size is constant. It is conceivable to have non-linear perspective.

But the sense of the painting's reality was not just a result of scientific innovation. The paintings were realistic copies but they were also platforms for inventive experiments. Perspective, for instance, was a rational and mathematical method for representation of space on a two dimensional medium, but it also tempted creative artist to play with it. Uccello, in his battle pieces, played with it in the interest of decorative pattern making. Leonardo Da Vinci, in his 'Mona Lisa' made use of compound perspective systems to evoke an atmosphere of mystery [Figure 11]. Even the 'Last Supper' where he made use of compound perspective and created lifelike figures with mere half strokes, nuanced shading and color and blurred outlines, the blend of imagination and technique triumphed [Figure 10(b)]. One can compare it to Giotto's painting on the same theme where the depth is not so obvious but realism starts penetrating and the picture shows a dinner taking place and people seated in the round with their backs showing [Figure 10(a)].

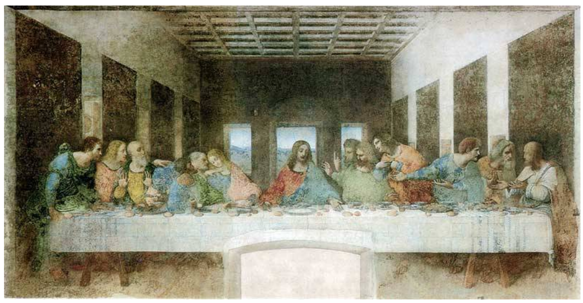



FIGURE 10. {a,b}

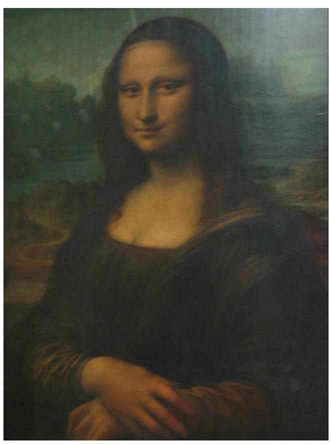


FIGURE 11.

The concept of perspectives may have originated and shaped up in renaissance Italy, but it did make progress in other European states as well. By 1500, Abrecht Durer, took the development of the topic in Germany. He did so only after learning much from trips to Italy where he learnt at first hand from mathematicians such as Pacioli. In 1525, he published Unterweisung der Messung mit dem Zirkel und Richtscheit, the fourth book of which contains his theory of shadows and perspective. Geometrically his theory was similar to that of Piero but he made an important addition stressing the importance of light and shade in depicting correct perspective. Among the other sixteen century mathematician- artists who worked on the problem of perspectives we can mention Daniele Barbaro, Egnatio Danti and Giovanni Battistta Benedetti. Daniele Barbaro, in his La Practica Della Perspectiva, published in 1569, complained that painters have stopped using perspective. Taken at face value that was not true, but what he said implied that painters were not painting architectural scenes. In fact another style of painting that was becoming a rage in the second half of sixteenth century was Mannerism that dramatized the stress of artist's emotions. This required manipulation of anatomy and dissolution of space and often turned perspective- once the artist pride into a jarring puzzle [Figure 12]. But it would be incorrect to suggest that the interest in perspective was entirely on wane. There were those Danti who traced a brief history of perspective in the introduction to Vignola's book on Perspective that was published in 1583 or Beneditti who wrote a short treatise on perspective titled De Rationibus Operationum Perpectivae where he was concerned not just with the rules for artists working on two dimensions but with the underlying three dimensional reasons for the rules. There is little doubt that thereafter perspective was gradually becoming more of a theoretical science and with the decline of the renaissance its art execution was also slipping.

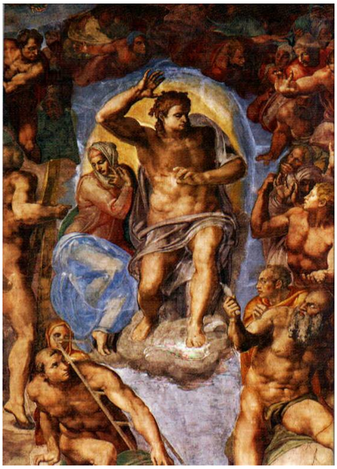


Figure 12.

Over all one cannot trivialize the mammoth contribution the artist-mathematicians of the renaissance had made to the concept of perspective. They had sought to comprehend the world empirically and derive rational laws from this experience as well as endeavored to know and control nature. But more important if the technician and the natural scientist now had the claim to be considered an intellectual on the basis of his mathematical knowledge, the artist, who was often identical with the technician and the scientist, could also expect to be distinguished from the craftsman and to have the medium in which he expressed himself be regarded as one of the 'free arts'. Painting, Leonardo maintained, was on the one hand, a kind of exact natural science; on the other, it was superior to the sciences, for these were 'imitable'. Leonardo, therefore, justified the claim of painting to be considered as one of the 'free arts' on the basis of both the artist's mathematical knowledge as well as on account of his talent which, according to Leonardo was equal to that of the poetic genius.

In other words the fusion of mathematics and fine arts that did come about in some measure through the emergence of the perspective system also had a social corollary. Renaissance artists began to be regarded as thinkers as well as decorators. Cennino Cennini, in his Book on Art, appealed to his fellow painters to seek social advancement and respect, and detailed for them the steps by which this change could be accomplished: 'Your lives should always be regarded as if you were studying theology, philosophy or other sciences' From Cennani onwards, i.e., from mid fifteenth century there was an attempt to associate the practice of painting, sculpture and architecture with the practice of skills such as poetry and mathematics- to represent them as liberal arts, as well as manual ones. Gradually painting, sculpture and architecture came to be accepted as gentlemanly professions. But somewhere there was a realization that social respect came only when their forms of art were associated with a theoretical rigor. Since the disciplines of poetry and mathematics proceeded from a basis of theory, so too, should art. Leonardo stated that those who devoted themselves to practice without science were like sailors who put to sea without rudder or compass and who could never be certain where they were going. 'Practice' he maintained, 'must always be founded on sound theory'.

We sincerely hope that a short study on convergence of two disciplines would make way for many more of such interactions. Whatever may have been the reason, we feel enriched by this exercise. The initial reticence of historians to even cast a glance into the direction of a discipline like Mathematics, has eventually given way to an opportunity to absorb some aspects of mathematical science. But more important it has assured us that a divorce between the disciplines of History and Mathematics would leave students on both the ends intellectually poorer.

References

1. Joseph D'Amelio, *Perspective Drawing Handbook*, p.19, Dover Publications, New York.
2. Website: http://www.math.iupui.edu/m290
3. John. R. Hale, *Renaissance*, Great Ages of Man, Time Life series, 1979, (Reprint).
4. Charles G. Nauert, Jr, *Humanism and Culture of Renaissance Europe*, Cambridge University Press, 1995.
5. *The Art of Renaissance Europe*, A Resource For Educators, The Metropolitan Museum of Art, 2000.

Dr. Smita Sahgal, Associate Professor, Department of History, Lady Shri Ram College for Women

E-mail address: smitasahgal16@yahoo.com

Ms.Bhavneet Kaur, Assistant Professor, Department of Mathematics, Lady Shri Ram College For Women

E-mail address: bhavneet.lsr@gmail.com

GÖDELS INCOMPLETENESS THEOREM: THE DISCOVERY OF AN INCOMPLETE UNIVERSE /THE INCOMPLETENESS OF EVERYTHING

ANISHA BANERJEE

Abstract. Gödels Incompleteness Theorems, proved by Kurt Gödel in 1931, focuses on the concept of mathematical logic. Gödel essentially demonstrated that all theories which consist of Number Theory follow the same pattern where **a consistent system is always incomplete, and a complete system is always inconsistent.** Alongside, a short biography on Gödel has also been included in order to elucidate the readers about a Mathematician whose contributions have been shadowed by the tides of time.

Background: How it all started?

Not long ago during the beginning of 20th Century, a popular German Mathematician named **David Hilbert** started off on a journey to prove the existence of a Complete Universe. If explained with a Mathematical twist, he started off on a journey to **set up a finite and complete system of axioms and provide a proof that these axioms were consistent**. That is, the axioms contained an exact and complete description of the relations existing between the basic elements of a given system.
He wanted to create a complete system free of paradoxes and proposed that the consistency of more complicated systems, such as Real analysis, could be proven in terms of simpler systems. Ultimately, the consistency of all of mathematics could be reduced to a simple branch, namely Number theory, also known as Peano's Arithmetic.

Hilbert's programme targeted the following properties

(1) **Formalization** of all mathematics; in other words all mathematical statements should be written in a precise formal language, and manipulated according to well defined rules.
(2) **Completeness:** a proof that all true mathematical statements can be proved in the formalism.
(3) **Consistency:** a proof that no contradiction can be obtained in the formalism of mathematics. This consistency proof should preferably use only "finitistic" reasoning about finite mathematical objects.
(4) **Conservation:** a proof that any result about "real objects" obtained using reasoning about "ideal objects" (such as uncountable sets) can be proved without using ideal objects.

(5) **Decidability:** there should be an algorithm for deciding the truth or falsity of any mathematical statement.

However, when Hilbert was really close to proving his theory, a young Mathematician named **Kurt Gödel** published a paper called *"On Formally Undecidable Propositions of "Principia Mathematica" and Related Systems"* which stated the **Theorems of Incompleteness**. His theorems were universally accepted, bringing an end to Hilbert's firm attempts to find a set of axioms sufficient for all mathematics.

Logical Completeness

In order to understand the statements better, one must understand the essential meaning and concept of Completeness.
The definition of **Completeness** varies from one field of study to the other. The two broad categories are essentially Logical Completeness and Mathematical Completeness, and we'll follow the former in details.

The following definitions are important for a deeper understanding of the concept:

(1) **Formal System:** A formal system is a combination of a Language (which consists of symbols, expressions, formulae and a set of rules) and a Deductive System (a proof system consisting of a set of axioms and theorems) which helps study a mathematical structure.
E.g., Euclid's geometry is a formal system.

(2) **Recursive System:** It is a method of defining functions in which the function being defined is applied within its own definition. However, no loop or infinite chain is formed. It consists of a base case such that all other functional values can be reduced to that base case.
E.g., Fibonacci sequence.

With the above mentioned definitions, we may now define **Completeness** as:
A Formal system 'S' is said to be Complete if and only if for each formula 'ϕ' of the language of the system, either ϕ or ϕ^{-1} (negation or inverse of ϕ) is a theorem of S. For instance, Gödel's theorem essentially shows that any strong recursive system cannot be both consistent and complete.

Consistency of a formal system

In simple words, every theorem, when interpreted, becomes a true statement. Inconsistency occurs when there is at least one false statement among the interpreted theorems. **This definition represents the scenario in the external world. In case of the internal inconsistencies, presumably, a system would be internally inconsistent if it contained two or more theorems whose interpretations were incompatible with one another, and internally consistent if all interpreted theorems were compatible with one another.**
For example, consider three individuals A, B and C, and let a relation R be defined between each pair of them such that R= "x beats y in Chess always". The three statements-

(1) A beats B in Chess always.
(2) B beats C in Chess always.
(3) C beats A in Chess always.

are not compatible together. This formal system follows an unusual loop with Chess players, because of which the system appears internally consistent, even though on the basis of their lack of compatibility (in terms of logic)* with one another, none of them appear to be true!

*According to the criteria of logic, we may apply Transitivity to the system to check the relation between a, b and c for relation R. If aRb and bRc, then logically aRc. That is, we get "a beats c in Chess always". But here, we get the result "c beats a in Chess always". Hence, this statement is not compatible on logical grounds. Therefore, this system is internally inconsistent.

Gödel's Incompleteness Theorem

Gödel stated that for any computable axiomatic system that is powerful enough to describe the arithmetic of the natural numbers, two properties hold true:

(1) If the system is **consistent**, it cannot be **complete**.
(2) The consistency of the axioms cannot be proven within the system; i.e. a complete system can never be consistent.

Further Insight

One of the simplest examples on the basis of a familiar formal system can be extracted from the concept of Isomorphism. Isomorphism can be studied under two situations :-

(1) **Rule governed systems:** A system which is artificially prepared by us by making certain assumptions. It is relatively less complex due to the lack of more variables.
(2) **Things in the real world:** A practical and naturally created system in which all the variables and constants are to be considered as they are.

The more complex the isomorphism, the more are the required tools and concepts to extract the meaning out of the symbols. Hence, several times we generalize meanings out of simple examples of systems, ultimately committing an error while using the same tools for a complex system.

For example, you might compare it to the naive belief that noise is a necessary side effect of any collision of two objects. This is a false belief; if two objects collide in a vacuum, there will be no noise at all. Here again, the error stems from attributing the noise exclusively to the collision, and not recognizing the role of the medium which carries it from the objects to the ear.

In short, human language attributes all the meaning to the object rather than to the link between that object and the real world.

The Liar's Paradox (Another example of a simple formal system)

The Liar's paradox (originally diagnosed by Alfred Tarski) was a simple example taken up by Gödel to demonstrate the **Unprovable nature** of certain complete systems. Whenever

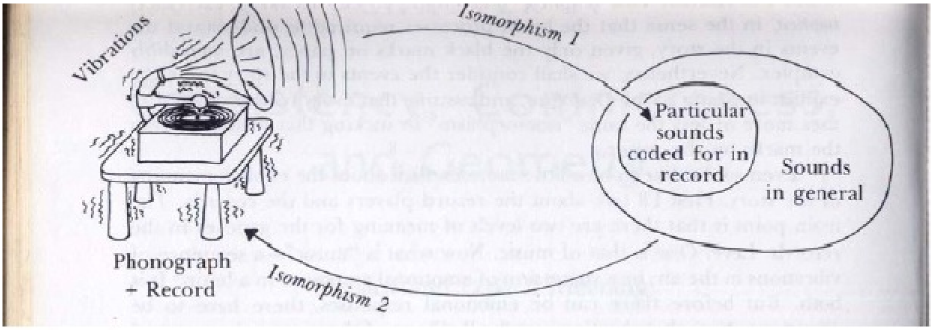


FIGURE 1. Diagram and working taken up from Gödel Escher Bach
Visual rendition of the principle underlying Gödel's Theorem: two back-to-back mappings which have an unexpected boomeranging effect. The first is from groove patterns to sounds, carried out by a phonograph. The second-familiar, but usually ignored – is from sounds to vibrations of the phonograph. Note that the second mapping exists independently of the first one, for any sound in the vicinity, not just ones produced by the phonograph itself, will cause such vibrations. The paraphrase of Gödel's Theorem says that for any record player, there are records which it cannot play because they will cause its indirect self-destruction.

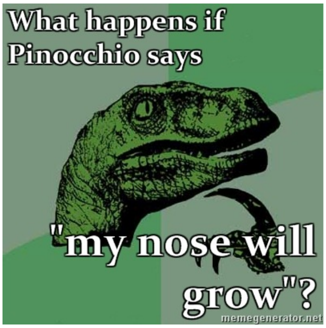
a binary truth value is assigned to the system 'S' - "This statement is false", the following behaviour is observed:-

We assume that the given statement 'S' is TRUE	If 'S' is said to be True, then 'S' itself claims its Falsity since (Read: S= This 'sentence is false' is CORRECT). Which implies that'S' is FALSE. Therefore, we get contradiction because a statement cannot be both TRUE and FALSE at the same time.
FALSE	If 'S' is said to be False, then the claim made by 'S' is said to be False (Read: S= 'This sentence is false' is INCORRECT). Which implies that the negation of 'S' should be true. This implies that 'S' is TRUE. Therefore, we get contradiction because a statement cannot be both TRUE and FALSE at the same time.

Therefore, both the truth values apply to the statement simultaneously, which isn't logically possible.

Gödel's Take on Hilbert's Formalization

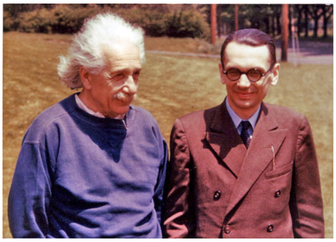
Kurt Gödel concluded that most of the targets of Hilbert's program were impossible to achieve. His **Second Incompleteness Theorem** stated that *any consistent theory powerful enough to encode addition and multiplication of integers cannot prove its own consistency.* This wipes out most of Hilbert's programme as follows:



(1) It is not possible to formalize **all** of Mathematics, as such an attempt towards a formalism will omit some true mathematical statements.
(2) An easy consequence of Gödel's incompleteness theorem is that **there is no complete consistent extension of even Peano arithmetic with a recursively enumerable set of axioms**, so in particular most interesting mathematical theories are not complete.
(3) A theory such as Peano arithmetic cannot even prove its own consistency, so a restricted "finitistic" subset of it certainly cannot prove the consistency of more powerful theories such as set theory.
(4) There is no algorithm to decide the truth (or provability) of statements in any consistent extension of Peano arithmetic. (Please Note that the respective result was inferred post Gödel's theorem.) However the consistency and completeness of Peano airthmetic is still a debate which is being worked upon.

A Brief Glance at Gödel's Life

In 1942, Kurt Gödel and Einstein met in Princeton and had become rather close friends. They walked to and from their offices every day, exchanging ideas about science, philosophy, politics, and the lost world of German science. By 1949, Gödel had produced a remarkable proof: *In any universe described by the Theory of Relativity, time cannot exist.* Einstein endorsed this result reluctantly but he could find no way to refute it, since then, neither has anyone else.



However, Gödel died a painful death due to mental instability and illness. He died because of starvation shortly after his wife's death in 1978. Anyhow, besides his tragic demise, he left a deep impression along with an even deeper meaning and mystery in the sphere of Physics and Mathematics, and continues to be a strong inspiration still.

References

1. Douglas R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*(1999), (A Pulitzer Prize Winner), Penguin Group, 2000, Online Edition.
   http://www.scribd.com/doc/6457786/Godel-Escher-Bach-by-Douglas-R-Hofstadter-
2. V.A. Uspensky, translated by Neal Koblitz, *Gödels Incompleteness Theorems*, MIR Publishers, Moscow, 1982.
3. Palle Yourgrau, *A World Without Time: The Forgotten Legacy of Gödel and Einstein* (Online Extract), Perseus Publishing, 2005.
4. Gregory J.Chaitin, *International Journal of Theoretical Physics 21*, 1982.
   http://www.cs.auckland.ac.nz/CDMTCS/chaitin/georgia.html
5. Dale Myers,*Godels Incompleteness Theorem.*
   http://www.math.hawaii.edu/ dale/godel/godel.html

Anisha Banerjee, B.Sc.(H) Mathematics, 3rd Year, Lady Shri Ram College For Women
E-mail address: anisha.banerjee9d@gmail.com

Rigour in Mathematics

This section aims to introduce advanced topics in mathematics to students. It serves to stimulate interest in different branches of mathematics and lay the foundation for further study.

NORM CONTINUITY OF C_0-SEMIGROUPS

MAHESH KUMAR

Abstract. In this article, we study the answer to a question (both in, particular as well as general context, which A.Pazy) asked nearly two decades ago: Given a C_0-semigroup $(T(t))_{t\geq 0}$ on a Banach space X, with generator A, the problem is to determine whether the resolvent decay condition

$$\lim_{|\beta|\to\infty} \|R(\omega + i\beta, A)\| = 0 \quad \text{for some } \omega \in \mathbb{R},$$

implies that the semigroup is immediately norm continuous, that is, norm continuous for $t > 0$. Author's MPhil thesis [1] is devoted to studying this problem in detail.

Some Basic Concepts and Definitions

Definition 1. Let X be a Banach space. A one parameter family $(T(t))_{t\geq 0}$ of bounded linear operators from X into X is a C_0-semigroup (or strongly continuous semigroup) of bounded linear operator on X if

(i) $T(0) = I$, (I is the identity operator on X);
(ii) $T(t+s) = T(t)T(s)$ for every $t, s \geq 0$ (the semigroup property);
(iii) $\lim_{t\to 0^+} T(t)x = x$ for every $x \in X$.

Definition 2. The linear operator A defined by

$$Ax = \lim_{t\to 0^+} \frac{T(t)x - x}{t} \quad \text{for } x \in D(A)$$

is the infinitesimal generator of the semigroup, $(T(t))_{t\geq 0}$, where $D(A)$ – the domain of A, is taken as

$$D(A) = \left\{ x \in X : \lim_{t\to 0^+} \frac{T(t)x - x}{t} \text{ exists} \right\}.$$

Definition 3. Let $(T(t))_{t\geq 0}$ be a C_0-semigroup on a Banach space X with generator A. We say that a strongly continuous semigroup $(T(t))_{t\geq 0}$ is eventually norm continuous if there exists $t_0 \geq 0$ such that the function $t \mapsto T(t)$ is norm continuous from (t_0, ∞) into $\mathcal{L}(X)$. The semigroup is called immediately norm continuous if t_0 can be chosen to be $t_0 = 0$. The semigroup is called uniformly continuous (or norm continuous) if $t \mapsto T(t)$ is norm continuous from $[0, \infty)$ into $\mathcal{L}(X)$.

It is pointed out that every uniformly continuous semigroup $(T(t))_{t\geq 0}$ on a Banach space X is of the form

$$T(t) = e^{tA}, \quad t \geq 0$$

for some bounded operator $A \in \mathcal{L}(X)$.

Clearly, uniform continuity implies immediate norm continuity and immediate norm continuity implies eventual norm continuity of a C_0-semigroup, but the converse is not true, as shown by the following example.

Example 1. Let X be the Banach space of continuous functions on $[0,1]$ which are equal to zero at $x=1$, with the supremum norm. Define

$$(T(t)f)(s) = \begin{cases} f(s+t), & \text{if } s+t \le 1 \\ 0, & \text{if } s+t > 1 \end{cases}$$

Its infinitesimal generator A is given by

$$D(A) = \{f : f \in C^1[0,1] \cap X, f' \in X\}$$

and

$$Af = f' \quad \text{for} \quad f \in D(A).$$

This semigroup is known as nilpotent semigroup. Since $\|T(t) - T(t_0)\| = 0,\ t, t_0 > 1$, it follows that it is eventually norm continuous for $t > 1$. But it is not immediately norm continuous. For that we show that the function $t \mapsto T(t)$ is not norm continuous at any $t < 1$. Let (t_n) be a sequence in $\mathbb{R}_+$ such that $t_n > t$ and $t_n \to t$. Let $g \in X$ with $\|g\| \le 1$ and $|g(s_0 + t)| > 0$, where $s_0 + t < 1$ and $s_0 \le 1 - t$. Then

$$\begin{aligned} \|T(t_n) - T(t)\| &= \sup_{\|f\|\le 1} \sup_{s\in[0,1]} |T(t_n)f(s) - T(t)f(s)| \\ &= \sup_{\|f\|\le 1} \sup_{s\in[0,1-t]} |T(t_n)f(s) - T(t)f(s)| \\ &= \sup_{\|f\|\le 1} \sup_{s\in[0,1-t]} |T(t_n)f(s) - f(s+t)| \\ &\ge \sup_{s\in[0,1-t]} |T(t_n)g(s) - g(s+t)| \ge |g(s_0+t)| > 0, \end{aligned}$$

since for some $s_0 \in [0, 1-t]$, $s_0 + t_n = 1$.

It has been one of the main issues in the study of norm continuity to characterize the classes of immediately norm continuous semigroups and of eventually norm continuous semigroups in terms of the resolvent of the semigroup generator. In particular, the so called norm continuity problem for C_0-semigroups attributed to A. Pazy was a focus for relevant research during the last two decades.

Given a C_0-semigroup $(T(t))_{t\ge 0}$ on a Banach space X, with generator A, the problem is to determine whether the resolvent decay condition

$$\lim_{|\beta|\to\infty} \|R(\omega + i\beta, A)\| = 0 \quad \text{for some } \omega \in \mathbb{R}, \tag{1}$$

implies that the semigroup is immediately norm continuous, that is, norm continuous for $t > 0$.

The decay condition (1) is certainly necessary for immediate norm continuity, by the fact that the resolvent of the generator is the Laplace transform of the semigroup, and by an application of the lemma of Riemann-Lebesgue. Hence, the question is whether condition (1) characterizes immediate norm continuity.

In the year 1992, P. You in [4] proved that (1) characterizes immediate norm continuity of the semigroup if X is a Hilbert space. Later his proof was analyzed, simplified and extended to eventual norm continuity by G. Q. Xu in [8], O. El-Mennaoui and K.J. Engel in [6] and [7], and by O. Blasco and J. Martinez in [5] in the years 1994 and 1996.

Then in the year 1999, a remarkable result of V.Goersmeyer and L.Weis, [3], shows that (1) also characterizes immediate norm continuity of the semigroup for positive semigroups in L^p spaces $(1 < p < \infty)$.

Very recently in 2008, T. Matrai [9] constructed a counterexample showing that the answer to the norm continuity problem is negative in general. The generator in his example is an infinite direct sum of Jordan blocks on finite dimensional spaces. The infinite sum is equipped with an appropriate norm and the resulting Banach space is reflexive. More precisely, let $1 \leq p < \infty$ be arbitrary but fixed, and set

$$(\mathcal{X}, \|.\|) = \bigoplus_{n=e^{61}}^{\infty} (X_n, \|.\|_n)$$

as an l^p-sum of Banach spaces, that is, for $x \in \mathcal{X}$,

$$\|x\| = \left(\sum_{n=e^{61}}^{\infty} \|x(n)\|_n^p \right)^{1/p},$$

$X_n = \mathbb{C}^{n+1}$, $\|.\|_n$ is the coordinate supremum norm As usual, $\|.\|$ stands for the norm of operators and functionals on $\mathcal{X}$. Consider the operators

$$\mathcal{T}(t) = \bigoplus_{n=e^{61}}^{\infty} T_n(t/\sqrt{\gamma_n}) \quad (t \geq 0),$$

where the sequence $(\gamma_n)_{n\in\mathbb{N}}$ satisfies

(1) $\gamma_n \geq 1 \quad (n \in \mathbb{N})$;
(2) $2(1 + 2\log(\gamma_{n+d}))e^{1+2\log(\gamma_{n+d})} \leq \log(n) \quad (0 \leq d \leq \sqrt{2n},\ e^6 \leq n < \infty)$;
(3) $\gamma_n \to \infty$ as $n \to \infty$;

and let $\mathcal{A} : \mathcal{X} \to \mathcal{X}$ be defined as

$$\mathcal{A} = \bigoplus_{n=e^{61}}^{\infty} \frac{1}{\sqrt{\gamma_n}} A_n$$

with natural domain $D(\mathcal{A}) = \{x \in \mathcal{X} : \mathcal{A}x \in \mathcal{X}\}$ and each A_n is a Jordan block with eigen value $-n$ respectively. We then have the Matrai's main result:

Proposition 1. *With the notation introduced above we have the following.*

1. $(\mathcal{X}, \|.\|)$ *is a Banach space which is reflexive for* $1 < p < \infty$*;*
2. $(\mathcal{T}(t))_{t\geq 0}$ *is a strongly continuous semigroup of bounded operators satisfying* $\|\mathcal{T}(t)\| \leq 1$ $(t \geq 0)$*;*
3. $(\mathcal{A}, D(\mathcal{A}))$ *is the generator of* $(\mathcal{T}(t))_{t\geq 0}$*;*
4. $R(\lambda, \mathcal{A})$ *exists for every* $\lambda \in \mathbb{C} \setminus \{-k/\sqrt{\gamma_k} : e^{61} \leq k < \infty\}$ *and*

$$\lim_{\mu\in\mathbb{R},\ |\mu|\to\infty} \|R(i\mu, \mathcal{A})\| = 0;$$

5. $(\mathcal{T}(t))_{t\geq 0}$ *is not eventually norm continuous.*

Algebra Homomorphism and The Norm Continuity Problem

Although the "Resolvent decay characterization of norm continuity" question is settled by Matrai's example above, many more issues remain. For example, why does such a characterization fails in general. Also, some nice characterization of immediate norm continuity is still lacking. In [2], Chill and Tomilov look at the norm continuity problem for semigroups from a different view point, in an attempt to throw light on some of the above points. They approach the problem of norm continuity via Banach algebra homomorphisms $L^1(\mathbb{R}_+) \to \mathcal{A}$ and have developed a new method for constructing C_0-semigroups for which properties of the resolvent of the generator and continuity properties of the semigroup in the operator norm topology are controlled simultaneously. In fact, Chill and Tomilov show that (a) there exists a C_0-semigroup which is continuous in the operator-norm topology for no $t \in [0,1]$ such that the resolvent of its generator has a logarithmic decay at infinity along vertical lines; (b) there exists a C_0-semigroup which is continuous in the operator-norm topology for no $t \in \mathbb{R}_+$ such that the resolvent of its generator has a decay along vertical lines arbitrarily close to a logarithmic one. These examples rule out any possibility of characterizing norm continuity of semigroups on arbitrary Banach spaces in terms of resolvent norm decay on vertical lines. The approach taken by Chill and Tomilov also opens up the possibilities as yet unexplored, of studying and perhaps settling, many other open problems in the theory of C_0-semigroups, particularly asymptotes. In order to understand, how Chill and Tomilov reformulated the norm continuity problem via Banach algebra homomorphisms $L^1(\mathbb{R}_+) \to \mathcal{A}$, we introduce the following

Definition 4. Let $T : L^1(\mathbb{R}_+) \to X$ be a bounded operator.

(1) T is called Riesz representable, or simply representable, if there exists a function $f \in L^\infty(\mathbb{R}_+; X)$ such that

$$Tg = \int_0^\infty g(s) f(s)\, ds \quad \text{for every } g \in L^1(\mathbb{R}_+).$$

(2) T is called Riemann-Lebesgue if

$$\lim_{|\beta|\to\infty} \|T(e_{i\beta} g)\| = 0 \quad \text{for every } g \in L^1(\mathbb{R}_+).$$

The following theorem gives a characterization of Riemann-Lebesgue operators using only exponential functions.

Theorem 1. *An operator $T : L^1(\mathbb{R}_+) \to X$ is a Riemann-Lebesgue operator if and only if*

$$\lim_{|\beta|\to\infty} \|Te_{\omega+i\beta}\| = 0 \quad \textit{for some/all } \omega > 0.$$

Definition 5. Two operators $T : L^1(\mathbb{R}_+) \to X$ and $S : L^1(\mathbb{R}_+) \to Y$ are said to be *equivalent* if there exist constants $c_1, c_2 \geq 0$ such that

$$\|Tg\|_X \leq c_1 \|Sg\|_Y \leq c_2 \|Tg\|_X \quad \text{for every } g \in L^1(\mathbb{R}_+).$$

The properties like Representability and Riemann Lebesgue are preserved under equivalence, that is,

Proposition 2. *Let $T : L^1(\mathbb{R}_+) \to X$ and $S : L^1(\mathbb{R}_+) \to Y$ be equivalent operators. Then T is representable if and only if S is representable.*

Proposition 3. *Let $T : L^1(\mathbb{R}_+) \to X$ and $S : L^1(\mathbb{R}_+) \to Y$ be equivalent operators. Then T is Riemann-Lebesgue if and only if S is Riemann-Lebesgue.*

Proposition 4. *Let $\mathcal{A}$ be a Banach algebra. If $(a(t))_{t>0} \subset \mathcal{A}$ is a uniformly bounded and continuous semigroup, then the operator $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ given by*

$$Tg = \int_0^\infty a(t)g(t)\,dt, \quad g \in L^1(\mathbb{R}_+) \quad \textit{(integral in the strong sense)}$$

is a bounded algebra homomorphism.

The converse of this result is also true, that is, if $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ is an algebra homomorphism, then T is represented as above, but $(a(t))_{t>0}$ is a semigroup of multipliers on $\mathcal{A}$ and the integral is to be understood in the sense of the strong topology of the multiplier algebra $\mathcal{M}(\mathcal{A})$. This result has been stated below in a slightly different form using the notion of equivalent operators.

Theorem 2. *For every algebra homomorphism $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ there exists a Banach space X_0, an equivalent algebra homomorphism $S : L^1(\mathbb{R}_+) \to \mathcal{L}(X_0)$ and a uniformly bounded C_0-semigroup $(S(t))_{t\geq 0} \subset \mathcal{L}(X_0)$ such that for every $g \in L^1(\mathbb{R}_+)$,*

$$Sg = \int_0^\infty S(t)g(t)\,dt \quad \textit{(integral in the strong sense).}$$

If $\mathcal{A} \subset \mathcal{L}(X)$ as a closed subspace, then X_0 can be chosen to be a closed subspace of X.

Remark 1. To give conditions on an algebra homomorphism $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ which imply that there exists an equivalent algebra homomorphism $S : L^1(\mathbb{R}_+) \to \mathcal{L}(X)$ on a Banach space X having additional properties, for example, being reflexive, being an L^p space, etc., is an open problem. We have seen in the proof of the above lemma that if T has dense range in $\mathcal{A}$ then $\mathcal{A}$ can be represented as a Banach subalgebra of $\mathcal{L}(\mathcal{A})$. So, if T has dense range in $\mathcal{A}$ then this is essentially the problem of representing $\mathcal{A}$ as a Banach subalgebra of $\mathcal{L}(X)$ for some Banach space X.

The next result relates the norm continuity problem with the problem of representability of homomorphisms $L^1(\mathbb{R}_+) \to \mathcal{A}$.

Theorem 3. *An algebra homomorphism $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ is representable if and only if there exists a uniformly bounded and continuous semigroup*

$$(a(t))_{t>0} \subset \mathcal{A} \quad \textit{(no continuity condition at 0)}$$

such that T is given by

$$Tg = \int_0^\infty a(t)g(t)\,dt \quad \textit{for every } g \in L^1(\mathbb{R}_+).$$

Let $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ be a Riemann-Lebesgue algebra homomorphism, $\mathcal{A}$ is a Banach algebra. Then, by Theorem 2, there exists a Banach space X, an equivalent algebra homomorphism $T_1 : L^1(\mathbb{R}_+) \to \mathcal{L}(X)$, which is Riemann-Lebesgue by Proposition 3, and a uniformly bounded C_0-semigroup $(T(t))_{t\geq 0} \subset \mathcal{L}(X)$ with generator A such that for every $g \in L^1(\mathbb{R}_+)$,

$$T_1 g = \int_0^\infty T(t)g(t)\, dt \quad \text{(integral in the strong sense)}.$$

By Theorem 3, $(T(t))_{t\geq 0}$ is immediately norm continuous if and only if T_1 is representable. But T_1 is representable if and only if T is representable, using Proposition 2.

Hence, $(T(t))_{t\geq 0}$ is immediately norm continuous if and only if T is representable.

As

$$T_1(e_{\omega+i\beta}) = \int_0^\infty e^{-(\omega+i\beta)t}T(t)\, dt = R(\omega + i\beta, A),$$

the resolvent of A satisfies the resolvent decay condition (1) if and only if T_1 is Riemann-Lebesgue, by Theorem 1, which is, if and only if T is Riemann-Lebesgue, by Proposition 3.

Thus, the resolvent of A satisfies the resolvent decay condition (1) if and only if T is Riemann-Lebesgue.

Hence, the norm continuity problem can be reformulated in the following way.

Problem 1. (Norm continuity problem reformulated). *If $\mathcal{A}$ is a Banach algebra and if $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ is a Riemann-Lebesgue algebra homomorphism, is T representable?*

From Matrai's example [9], it follows that the answer to Problem 1 is negative, in general. In [2], suitable Riemann-Lebesgue homomorphisms has been constructed and from this different counterexamples to Problem 1 has been deduced for which it is possible to control the resolvent decay along vertical lines. The main result in this regard is –

Theorem 4. *Let $f \in L^\infty(\mathbb{R}_+)$ be a function such that its Laplace transform $\hat{f}$ extends to a bounded analytic function in some domain Σ_φ, where $\varphi \in C(\mathbb{R}_+)$ satisfies* $\inf \varphi > 0$.

Then there exists a Banach space X which embeds continuously into $L^\infty(\mathbb{R}_+)$ and which is left-shift invariant such that

(i) *the left-shift semigroup $(T(t))_{t\geq 0}$ on X is bounded and strongly continuous,*
(ii) *the resolvent of the generator A satisfies the decay estimate*

$$\|R(\lambda_\beta, A)\|_{\mathcal{L}(X)} \leq C \frac{\log d_\beta}{d_\beta} \quad \textit{for every } \beta \in \mathbb{R},$$

(iii) *if $T : L^1(\mathbb{R}_+) \to \mathcal{L}(X)$ is the algebra homomorphism which is represented (in a strong sense) by $(T(t))_{t\geq 0}$, then $f \in \operatorname{range} T^*$, and*
(iv) *the following inclusion holds:*

$$X \subset \overline{L^1(\mathbb{R}_+) \circledast f}^{(L^\infty, weak^*)}.$$

If, in addition, the function

$$\mathbb{C}_+ \to L^\infty(\mathbb{R}_+),$$
$$\lambda \mapsto e_\lambda \circledast f,$$

extends analytically to Σ_φ *and if there exists some* $r \in (0,1)$ *such that*

$$\sup_{\lambda \in B(\lambda_\beta, r d_\beta)} \|e_\lambda \circledast f\|_\infty \leq C \frac{1}{d_\beta} \quad \textit{for every } \beta \in \mathbb{R},$$

then the space X *can be chosen in such a way that the resolvent satisfies the stronger estimate*

$$\|R(\lambda_\beta, A)\|_{\mathcal{L}(X)} \leq C \frac{1}{d_\beta} \quad \textit{for every } \beta \in \mathbb{R}.$$

We need the following Proposition to make a remark on the above theorem.

Proposition 5. *Let* $T : L^1(\mathbb{R}_+) \to \mathcal{A}$ *be an algebra homomorphism. Then the following are true:*

(1) *If* T *is representable, then* $\operatorname{range} T^* \subset C(0,\infty)$.
(2) *If* T *is represented (in the strong sense) by a bounded* C_0*-semigroup which is norm continuous for* $t > t_0$*, then every function in* $\operatorname{range} T^*$ *is continuous on* (t_0, ∞).

Remark 2. It is important to note from the above theorem that the resolvent decay condition (1) is satisfied as soon as $\lim_{\beta\to\infty} \varphi(\beta) = \infty$, and that at the same time $f \in \operatorname{range} T^*$. Thus, if one is able to find a function $f \in L^\infty(\mathbb{R}_+)$ such that its Laplace transform $\hat{f}$ extends to a bounded analytic function on Σ_φ, where $\varphi \in C(\mathbb{R}_+)$ satisfies $\lim_{\beta\to\infty} \varphi(\beta) = \infty$, and such that f is not continuous on $(0,\infty)$, then the Riemann-Lebesgue operator from Theorem 4 is not representable by Proposition 5(1), that is, the semigroup $(T(t))_{t\geq 0}$ is not immediately norm continuous [see Theorem 3]. In other words, the existence of such a function f solves the norm continuity problem.

The following is the first counterexample to the norm continuity problem, where one takes $f = \chi_{[0,1]}$.

Theorem 5. *There exists a Banach space* X *and a uniformly bounded* C_0*-semigroup* $(T(t))_{t\geq 0} \subset \mathcal{L}(X)$ *with generator* A *such that:*

(i) *the resolvent satisfies the estimate*

$$\|R(2+i\beta)\| = O\left(\frac{1}{\log|\beta|}\right) \quad \textit{as } |\beta| \to \infty,$$

and in particular the resolvent satisfies the resolvent decay condition

$$\lim_{|\beta|\to\infty} \|R(\omega + i\beta, A)\| = 0 \quad \textit{for some } \omega \in \mathbb{R},$$

(ii) $T(1) = 0$*, that is, the semigroup* $(T(t))_{t\geq 0}$ *is nilpotent, and*
(iii) *whenever* $t_0 \in [0,1)$*, then the semigroup* $(T(t))_{t\geq 0}$ *is not norm continuous for* $t > t_0$.

The next counterexample shows that there are also C_0-semigroups which are never norm continuous, whose generator satisfies the resolvent decay condition, and the decay of the resolvent along vertical lines is even arbitrarily close to a logarithmic decay.

Theorem 6. *Let $h \in C(\mathbb{R}_+)$ be a positive, increasing and unbounded function such that also the function $\log^+ /h$ is increasing and unbounded. Then there exists a Banach space X and a uniformly bounded C_0-semigroup*

$$(T(t))_{t\geq 0} \subset \mathcal{L}(X)$$

with generator A such that:

(i) *the resolvent satisfies the estimate*

$$\|R(2+i\beta, A)\| = O\left(\frac{h(|\beta|)}{\log|\beta|}\right) \quad \text{as } |\beta| \to \infty,$$

and in particular the resolvent satisfies the resolvent decay condition, and

(ii) *the semigroup $(T(t))_{t\geq 0}$ is not eventually norm continuous.*

References

1. Mahesh Kumar, *Norm Continuity of C_0-Semigroups*, M.Phil. Thesis, University of Delhi, 2011.
2. Ralph Chill, Yuri Tomilov, *Operators $L^1(\mathbb{R}_+) \to X$ and the norm continuity problem for semigroups*, Journal of Functional Analysis 256 (2009), 352-384.
3. V.Goersmeyer, L. Weis, *Norm continuity of c_0-semigroups*, Studia Math. 134 (1999), 169-178.
4. P. You, *Characteristic conditions for C_0-semigroups with continuity in the uniform operator topology for $t > 0$*, Proc. Amer. Math. Soc. 116 (1992), 991-997.
5. O. Blasco, J. Martinez, *Norm continuity and related notions for semigroups on Banach spaces*, Arch. Math. 66 (1996), 470-478.
6. O. El-Mennaoui, K.J. Engel, *On the characterization of eventually norm continuous semigroups in Hilbert space*, Arch. Math. 63 (1994), 437-440.
7. O. El-Mennaoui, K.J. Engel, *Towards a characterization of eventually norm continuous semigroups on Banach spaces*, Quaest. Math. 19 (1996), 183-190.
8. G. Q. Xu, *Eventually norm continuous semigroups on Hilbert space and perturbations*, Journal of Mathematical Analysis and Applications 289 (2004), 493-504.
9. T. Matrai, *Resolvent norm decay does not characterize norm continuity*, Israel J. Math.(2008).

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

E-mail address: mahekumar81@gmail.com

WHAT IS REPRESENTATION THEORY?

GAUTAM BORISAGAR

ABSTRACT. We will define in this article basic concepts in representation theory. Assumption is that reader knows fundamentals of group theory and linear algebra. We also relate representation theory with analysis. Some advance topics are touched in the last section.

INTRODUCTION

First correct statement about representation theory is that it is initially a game between Groups and Vector spaces. In mathematics most of the times we find ourselves working with sets and functions together with more structures or conditions on them. Through out the article G stands for a group. What are possible ways to relate G with a vector space V? Relate corresponds to function. There are not many ways actually if we take structures also into consideration. One way to get a "good" map is homomorphism $\rho : G \to V$. But this is too poor and we cannot do much with this. Now we have two options. Either we produce a group G' from vector space and relate that group with G some way or we produce vector space W from G and study connections of two vector spaces V and W. Representation theory goes with first option. Now ask yourself, "what are most natural groups associated with a vector space?" Other than V itself under vector addition, important group associated with V is set $Aut(V)$ of all automorphism on V under composition of linear maps. "Natural" map from G to $Aut(V)$ is called representation of G in vector space V!

Definition 1. A representation of G is a pair (ρ, V) where V is a vector space and ρ is a homomorphism of G to $Aut(V)$.

Equivalently group G acting on a vector space V by automorphisms is called a representation. Here we note that group G is fixed. We talk of representation **of** a group. What now? The moment we think of a vector space, we have concept of dimension and underlying field. If the vector space V involved here is finite dimensional we say that the representation is finite dimensional. Degree or dimension of representation is the dimension of the vector space. When it is clear from the context we just say ρ is a representation or just that V is a representation. If the vector space is $n-$dimensional over field of complex numbers $\mathbb{C}$ then there is no exaggeration in saying that representation theory is linear algebra! In such case we can use the fact that the automorphisms are $n \times n$ invertible matrices.

Examples.

(1) **Trivial representation:** Consider any field $\mathbb{F}$ as a vector space of dimension one over itself. Trivial homomorphism $\rho : G \to \mathbb{F}$ defined by $\rho(x) = 1$ for all $x \in G$ is called trivial representation.
(2) Consider group S_3 of permutations in three symbols $\{1, 2, 3\}$. Take three dimensional vector space V with basis $\{v_1, v_2, v_3\}$. A map defined on basis can be linearly extended to whole vector space. The map $\rho : G \to GL(V)$ is defined by $\rho(\sigma)(v_i) = v_{\sigma(i)}$ gives remaining requirement for representation.
(3) **Permutation representation:** Above example can be generalized as follows: Let a group G be given which acts on a set X. Take vector space V which has basis $\{v_x\}_{x\in X}$ parametrized by X. Define $\rho : G \to GL(V)$ using $\rho(g)v_x = v_{g\cdot x}$. This is called permutation representation associated to given action.
(4) In particular if H is a subgroup of G (not necessarily normal subgroup) then G acts on the set of left cosets $\{gH \mid g \in G\}$. This gives a representation of G.
(5) **Character:** Since any non-zero complex number corresponds to automorphism of $\mathbb{C}$, a homomorphism $\chi : G \to \mathbb{C}^*$ defines a one dimensional representation. Such representations are also known as characters. For example S_3 has two characters $\chi_0 \equiv 1$ and χ_1 defined by
$$\chi_1 = \begin{cases} 1 & \text{if } \sigma \text{ is even permutation} \\ -1 & \text{if } \sigma \text{ is odd permutation.} \end{cases}$$
What are all characters of a finite cyclic group?

Subrepresentation and Intertwining Maps

The moment we study some new object, we are interested in sub-object also. For example, set-subset, group-subgroup, vector space - vector subspace, ring-subring etc. What is analogous to sub-representation? You might think of a sub group once, but note that we are given a fixed group G and talking about its representations. We are already given a representation, say, (V, ρ) and we want to talk of its sub-representation. So sub-representation must be some(W, ρ') satisfying some "natural " condition. Now we define

Definition 2. Let G be a group. Let (V, ρ) be a representation of G. A subrepresentation of V is a subspace W of V such that the restriction $\rho\mid_W$ defines an automorphism of W for each $g \in G$.

In other words, a subrepresentation is subspace which itself is a representation of the given group. Equivalently the subspace, invariant under $G-$action, is subrepresentation.

Examples.

(1) In Example 2 above, take W as span of $\{v_1 + v_2 + v_3\}$ and it is easily checked that W is a subrepresentation. Note that W is trivial representation when seen as representation of S_3.
(2) Whole space V and zero space $\{0\}$ are always subrepresentations of V. For one dimensional representation V these are the only subrepresentations..

Definition 3. If the only subrepresentations of $V \neq 0$ are V itself and $\{0\}$ space, V is said to be an **irreducible representation.**

Can we know all the representations of a given group? (We should be clear about word "all".) What are all finite abelian groups? Such questions are known as classification problems. We try to get answer up to some equality. So we now define "equality" of two representations.

Definition 4. Let (ρ, V) and (ρ', W) be two representations of a group G. A linear map $T : V \to W$ satisfying $T(\rho(g))(v) = \rho'(g)(T(v))$ for all $g \in G$ and $v \in V$ is called $G-$equivariant or intertwining map. The same meaning holds in saying the following diagram is commutative.

$$\begin{array}{ccc} V & \xrightarrow{\rho(g)} & V \\ \Big\downarrow T & & \Big\downarrow T \\ W & \xrightarrow{\rho'(g)} & W \end{array}$$

In addition to above conditions if T is one-one and onto then we say the representations are **isomorphic** or **equivalent**.

Theorem 1. Shur's lemma: *If V and W are non-isomorphic irreducible representations of G, then the only intertwining map is zero map.*

Proof. Let $T : V \to W$ be intertwining map. We will show if T is nonzero then T is isomorphism. In this situation the image $T(V)$ is W because it is a subrepresentation and W is irreducible. So remains to show T is one one. Kernel is also a sub representation of V which is being irreducible must be $\{0\}$ or V. Now $T(V) = 0 \iff ker(T) = V$. This forces $ker(T) = 0$ and consequently the required claim. □

Definition 5. If (ρ_1, V_1) and (ρ_2, V_2) are two representations of G their vector space direct sum $V_1 \oplus V_2$ can be given a natural representation structure defined by $\rho(g)(v_1, v_2) = (\rho_1(g)v_1, \rho_2(g)v_2)$. This is called direct sum of the representations.

Theorem 2. *Corresponding to every subrepresentation W of V, there exists a complementary sub space W' which is also a subrepresentation.*

Proof. Refer [4], Chapter 1. □

Theorem 3. *Any finite dimensional representation is direct sum of irreducible subrepresentations.*

Proof. With repeated application of Theorem 2 apply induction on dimension. □

With above theorem the classification of finite dimensional representation of finite group is reduced to know all irreducible representations only. For example S_3 has two characters and one $2-$dimensional irreducible representation. Any finite dimensional representation is isomorphic to a finite direct sum of these representations. An interested reader is advised to read the book [4].

Character Theory

Here we consider vector spaces over complex numbers only. Many times practically it is not easy to find all irreducible representations of a given group. Still we can talk about the dimensions of all its irreducible representation. This part of representation theory is known as character theory. Remember we have already defined one dimensional representation as character, but this is a different concept. This is a good example of how can a word be overused.

Definition 6. Let $\rho : G \to GL(V)$ define a finite dimensional complex representation. We can think of $GL(V)$ as invertible matrices. The map $\chi : G \to \mathbb{C}$ (defined using trace map) by $\chi(g) = Tr(\rho(g))$ is called the character of representation. Recall that trace of a matrix is sum of all diagonal entries of the matrix.

Facts:

(1) The number of irreducible representations of a finite group is the number of conjugacy classes in it.
(2) Character characterizes representation, i.e., two representations are isomorphic if and only if their characters are same.
(3) The value $\chi(e)$ gives the dimension of the representation where e is identity element of G.
(4) A character is a class function in the sense that $\chi(gag^{-1}) = \chi(a)$ for all $g, a \in G$.
(5) If for a finite group of order n, $\rho_1, \rho_2, \cdots, \rho_m$ are all the irreducible representations with dimensions $d_1, d_2, \cdots, d_m$ then $\sum_{i=1}^{m} d_i^2 = n$.

Some Advance Topics

In case of infinite group and infinite dimensional vector spaces one puts more structures on group and vector spaces to get useful results. For example, locally compact topological group admits Haar measure, which is used to construct a Hilbert space and one studies representation of the group in this space. Compact groups representation theory is just like the one of finite groups. Completely different flavour is in modular representation theory where the vector space is over filed of finite characteristic. This theory is used to classify finite simple groups. In this situation Theorems 2 and 3 are not true. There are possibility of indecomposable representations which are neither irreducible nor can be written as direct sum of irreducible sub representations! Brauer developed block theory to understand better way modular representations. If the vector space is over field of characteristic p, we call such representation as mod$-p$ representation. For general finite group we don't have classification of indecomposable mod$-p$ representations. In such cases one studies cokernel representation and socal filteration. The only irreducible mod$-p$ representation of a $p-$group is trivial representation!

Going to infinite groups, the one that are useful to number theory are $\mathrm{GL}_n(F)$ where F is local field([5]) of residual characteristic p. In these cases there are 4 types of irreducible mod$-p$ representations. In 1994-95 Barthel and Livne ([2], [1]) classified irreducible mod$-p$ representations of $\mathrm{GL}_2(F)$ where F is of residual characteristic p into four categories. Out of which one, namely, supersingular representation was left almost undone. Later in 2003, Breuil [3] classified supersingular representations only for $\mathrm{GL}_2(\mathbb{Q}_p)$. Classification of supersingular representations of $\mathrm{GL}_2(F)$ for most other local fields is still open.

References

1. L. Barthel and R. Livné, *Irreducible modular representations of* GL_2 *of a local field*, Duke Math. J. **75** (1994), no. 2, 261–292.
2. ———, *Modular representations of* GL_2 *of a local field: the ordinary, unramified case*, J. Number Theory **55** (1995), no. 1, 1–27.
3. Christophe Breuil, *Sur quelques représentations modulaires et p-adiques de* $\mathrm{GL}_2(\mathbf{Q}_p)$*. I*, Compositio Math. **138** (2003), no. 2, 165–188.
4. Jean-Pierre Serre, *Linear representations of finite groups*, Springer-Verlag, New York, 1977, Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.
5. ———, *Local fields*, Graduate Texts in Mathematics, vol. 67, Springer-Verlag, New York, 1979, Translated from the French by Marvin Jay Greenberg.

Gautam Borisagar, Research Scholar, Department of Mathematics, Indian Institute of Technology, Bombay

E-mail address: ghbsagar@gmail.com

HOLONOMY GROUP OF A SURFACE

SAHANA BALASUBRAMANYA

Abstract. Holonomy is the phenomenon that occurs when a tangent vector is parallely transported around a closed curve on a surface. The vector attained is a rotation of the original vector. The inability of the tangent vector to return to the original vector is due to the intrinsic properties of the surface. This article proves that the holonomy set of a surface forms a group; linking group theory and differential geometry.

One of the features that attracts people about co-ordinate or analytical geometry is that it is a synthesis of geometry and algebra. However, we are restricted in our vision, and the preliminary understanding of higher dimensions is not cultivated. We thus introduce differential geometry, which develops geometry in $(n+1)$ dimensions. The advantage of this approach is that one can then easily illustrate each concept in every lower dimension simultaneously. This approach is used in this article, and can be easily reduced to 2 or 3 dimensions.

Preliminaries

(1) **Level set of a function**: Let $f : U \to \mathbb{R}$ be a function; $U \subseteq \mathbb{R}^{n+1}$, where the space $\mathbb{R}^{n+1}$ consists of $(n+1)$ tuples. Let $c \in \mathbb{R}$. Then the level set of f, at a height 'c', is defined as:

$$f^{-1}(c) = \{(x_1, x_2, \ldots, x_{n+1}) \,|\, f(x_1, x_2, \ldots, x_{n+1}) = c\}.$$

(2) **Graph of a function**: Let $f : U \to \mathbb{R}$; $U \subseteq \mathbb{R}^{n+1}$ be a function. Then the graph of f is defined as

$$\text{graph}(f) = \{(x_1, x_2, \ldots, x_{n+1}, x_{n+2}) \,|\, x_{n+2} = f(x_1, x_2, \ldots, x_{n+1}),\ (x_1, x_2, \ldots, x_{n+1}) \in U\}.$$

The terms "level set" and "height" arise from the relation between level sets and the graph of a function. The level sets are just the projections of cross sections of the graph at various "heights".

(3) **Surfaces**: Given a smooth function(continuously differentiable) $f : U \to \mathbb{R}$; $U \subseteq \mathbb{R}^{n+1}$; we can define a function, called the gradient; denoted by $\overline{\bigtriangledown} f$; as

$$(\overline{\bigtriangledown} f)(p) = \left(p, \frac{\partial f}{\partial x_1}(p), \frac{\partial f}{\partial x_2}(p), \ldots, \frac{\partial f}{\partial x_{n+1}}(p)\right)$$

[NOTE: $(p, \overline{v})$ denote the vector $\overline{v}$ placed at the point p].
Let S be a non empty subset of $\mathbb{R}^{n+1}$. S is said to be a surface if $\exists$ a smooth function $f : U \to \mathbb{R}$, $U \subseteq \mathbb{R}^{n+1}$; U is open and some $c \in \mathbb{R}$ such that $\overline{\bigtriangledown} f(p) \neq 0$ for all $p \in S$ and $S = f^{-1}(c)$. S is called an n-surface in $\mathbb{R}^{n+1}$.

(4) **Vector fields on a surface**: A vector field $\overline{X}$ on an n-surface S is a function that assigns to each point p in S, a vector $\overline{X}(p)$ at p.
If S is a surface; then it can be seen that $\overline{\bigtriangledown} f$ defines a vector field on S, known as the gradient vector field.

(5) **Integral curve of a vector field**: A parametrized curve in $\mathbb{R}^{n+1}$ is a smooth function $\alpha : I \to \mathbb{R}^{n+1}$; I is an open interval in $\mathbb{R}$, given as

$$\alpha(t) = (x_1(t), x_2(t), \ldots, x_{n+1}(t))$$

where x_i is a smooth, real valued function on I for all i. The velocity vector of α is defined as:

$$\dot{\alpha}(t) = \left(\alpha(t), \frac{d\alpha(t)}{dt}\right) = \left(\alpha(t), \frac{dx_1(t)}{dt}, \frac{dx_2(t)}{dt}, \ldots, \frac{dx_{n+1}(t)}{dt}\right).$$

If $\overline{X}$ is a vector field on $U \subseteq \mathbb{R}^{n+1}$; U is open; then α is said to be an integral curve of $\overline{X}$ if $\alpha(t) \in U$ for all $t \in I$ and $\dot{\alpha}(t) = \overline{X}(\alpha(t))$ i.e., the vector field is tangent to α at every point of the curve.
For example: It can be checked that for the vector field

$$\overline{X}(p) = (p, -x_2, x_1) = ((x_1, x_2), (-x_2, x_1));$$

the integral curve through (0,1) is

$$\alpha(t) = (-\sin t, \cos t).$$

(6) **The tangent space**: Let $f : U \to \mathbb{R}$ be a smooth function; $U \subseteq \mathbb{R}^{n+1}$ is open. Let $c \in \mathbb{R}$ be such that $f^{-1}(c)$ is non empty, and let $p \in f^{-1}(c)$. A vector at p is said to be tangent to the level set $f^{-1}(c)$ if it is a velocity vector of a parametrized curve in $f^{-1}(c)$. The collection of all such velocity vectors forms the tangent space, denoted by T_p. It is pointed out that $T_p = [\bigtriangledown f(p)]^{\perp}$, where $\perp$ denotes the orthogonal complement i.e., the set of vectors that nullify $\bigtriangledown f(p)$.

Thus, if S is a surface, and $p \in S$; then the tangent space at p, denoted by S_p, is given by $S_p = [\bigtriangledown f(p)]^{\perp}$. Also, if $\overline{X}$ is a vector field on S such that $\overline{X}(p) \in S_p$ for all $p \in S$; then $\overline{X}$ is a *tangent vector field.* If $\overline{X}(p) \in S_p^{\perp}$ for all $p \in$S; $\overline{X}$ is said to be a *normal vector field* on S.
Example:

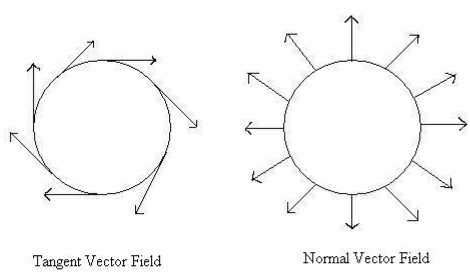


FIGURE 1.

Orientations

Let S be an n-surface in $\mathbb{R}^{n+1}$ i.e., there exists a smooth function $f : U \to \mathbb{R};\ U \subseteq \mathbb{R}^{n+1}$ is open such that $S = f^{-1}(c)$ and $\bigtriangledown f(p) \neq 0$ for all $p \in S$; for some $c \in \mathbb{R}$. Define a vector field on S as

$$\overline{N}(p) = \frac{\bigtriangledown f(p)}{\| \bigtriangledown f(p)\|}; \quad p \in S.$$

Then $\overline{N}$ is a smooth, unit, normal vector field on S. The existence of such a vector field on a surface is called an orientation. The surface along with an orientation is called an oriented surface. If S is a connected surface, there are exactly two orientations; $\overline{N}$ and $-\overline{N}$ (referred to as the outward and inner normals).

Derivative and Co-Variant Derivative of Vector Fields

Let $\overline{X}$ be a smooth vector field along α i.e., $\overline{X}(t)$ assigns a vector at $\alpha(t)$; given by

$$\overline{X}(t) = (\alpha(t), X_1(t), X_2(t), \ldots, X_{n+1}(t)),$$

X_i's are called the component functions. Then the *derivative* of $\overline{X}$ along α is defined to be the vector field $\dot{\overline{X}}$ given by

$$\dot{\overline{X}}(t) = \left(\alpha(t), \frac{dX_1(t)}{dt}, \frac{dX_2(t)}{dt}, \ldots, \frac{dX_{n+1}(t)}{dt}\right).$$

If the original vector field $\overline{X}$ is tangent to S along α i.e., $\overline{X}(t) \in S_{\alpha(t)}$ for all t, the derivative $\dot{\overline{X}}$ need not remain tangential to S. We thus define the *co-variant derivative* $\overline{X}'$ of $\overline{X}$, which is a smooth vector field given by

$$\overline{X}'(t) = \dot{\overline{X}}(t) - [\dot{\overline{X}}(t).\overline{N}(\alpha(t))].\overline{N}(\alpha(t))$$

where $\overline{N}$ is an orientation on S. This $\overline{X}'$ is also tangential to S along α, and $\overline{X}'$ is independent of the choice of orientation, $\overline{N}$ or $-\overline{N}$. Intuitively, $\overline{X}'$ measures the rate of change of $\overline{X}$ along α as seen from S.

A smooth, tangent vector field is said to be *Levi-Civita* parallel, or simply parallel if $\overline{X}' = 0$ i.e., $\overline{X}$ is constant along α, as seen from S.

Parallel transport

Let S be an n-surface in $\mathbb{R}^{n+1}$ and $\alpha : I \to S$ be a parametrized curve in S and $t_0 \in I$. Let $\overline{V} \in S_{\alpha(t_0)}$ i.e., $\overline{V}$ is tangent to S at $p = \alpha(t_0)$. Then we can always find a unique vector field $\overline{V}$, tangent to S along α and parallel ($\overline{V}' = 0$) such that $\overline{V}(t_0) = \overline{V}(\alpha(t_0)) = \overline{V}$.

So if p and q are two points on an n-surface S, we can always find a "path" from p to q i.e., an $\alpha : [a, b] \to S$ such that $\alpha(a) = p;\ \alpha(b) = q;$ and α is smooth. Then we can find a unique, parallel vector field $\overline{V}$, tangent to S along α such that $\overline{V}(a) = \overline{V}$; for a given $\overline{V} \in S_p$. We then define a map $P_\alpha : S_p \to S_q$ as

$$P_\alpha(\overline{V}) = \overline{V}(b) = \overline{V}(\alpha(b)).$$

This $P_\alpha(\overline{V})$ is called the *parallel transport* of $\overline{V}$ along α, from p to q. The function P_α is a vector space isomorphism preserving dot product i.e.,

- P_α is linear.
- P_α is one-one and onto.
- $P_\alpha(\overline{V}).P_\alpha(\overline{W}) = \overline{V}.\overline{W}$; for all $\overline{V}, \overline{W} \in S_p$.

Holonomy Group

Let S be an n-surface and $p \in S$. Let G_p denote the group of non-singular, linear transformations from S_p to itself. Define

$$H_p = \{T \in G_p \,|\, T = P_\alpha \text{ for some piecewise smooth } \alpha : [a,b] \to S \text{ with } \alpha(a) = \alpha(b) = p\}.$$

Then H_p is a subgroup of G_p, called *the holonomy group of S at p*. We now prove that H_p is indeed a group. For this, we will use the following result, which can be easily proved.

"Let S be an n-surface and let $\alpha : I \to S$ be a parametrized curve in S. Let $\beta : \tilde{I} \to S$ be defined as $\beta = \alpha \circ h$, $h : \tilde{I} \to I$ is a smooth function with $h'(t) \neq 0$ for all $t \in \tilde{I}$. Then a vector field $\overline{X}$, tangent to S along α, is parallel if and only if $\overline{X} \circ h$ is parallel along β."

Thus, parallel transport from p to q in S is unique upto reparametrizations of the curve α. We now note that:

(1) $H_p \subseteq G_p$ (by definition).

(2) Let $p, q \in S$ and $\alpha : [a,b] \to S$ be a curve such that $\alpha(a) = p$; $\alpha(b) = q$. Let $\overline{V} \in S_p$. Then $P_\alpha : S_p \to S_q$ is defined as $P_\alpha(\overline{V}) = \overline{V}(b)$ where $\overline{V}$is the parallel, tangent vector field such that $\overline{V}(a) = \overline{V}$, along α. Define

$$h(t) = -t;\ h : [-b, -a] \to [a, b].$$

Then $h'(t) \neq 0$ for all t ; $h(-b) = b$; $h(-a) = a$. Define

$$\beta(t) = \alpha \circ h(t) = \alpha(-t);\ \beta : [-b, -a] \to S.$$

Then $\beta(-b) = q$; $\beta(-a) = p$. Also $\overline{V}(b) \in S_q$.
Let $\overline{Y}$ be a vector field given by $\overline{Y}(t) = \overline{V}(h(t))$. Then $\overline{Y}(-b) = \overline{V}(b)$ and $\overline{Y}$ is parallel along β (since $\overline{V}$ is parallel along α, by the result given above). Thus $P_\beta : S_q \to S_p$ is defined as

$$P_\beta(\overline{V}(b)) = \overline{Y}(-a) = \overline{V}(h(-a)) = \overline{V}(a) = \overline{V}.$$

Therefore, $P_\beta \circ P_\alpha(\overline{V}) = P_\beta(\overline{V}(b)) = \overline{V} = I(\overline{V})$. Thus, for each path α in S from p to p, there is a path $\beta = \alpha \circ h$; $h(t) = -t$ in S such that $P_\beta = {P_\alpha}^{-1}$. Thus, existence of an *inverse* is guaranteed.

(3) Let $\alpha : [a,b] \to S$ and $\beta : [c,d] \to S$ are two piecewise smooth curves such that $\alpha(a) = \alpha(b) = p = \beta(c) = \beta(d)$ and $\overline{V} \in S_p$. Let the parallel transport of $\overline{V}$ along α result in vector $\overline{w}$, and that of $\overline{w}$ along β result in vector $\overline{z}$.

By rescaling, we can construct $\hat{\beta} : [b, e] \to S$; $\hat{\beta}(b) = \hat{\beta}(e) = p$; so that $\hat{\beta}$ is a reparametrization of β. Then, by the result given above, the parallel transport of $\overline{w}$ along $\hat{\beta}$ also results in $\overline{z}$.

We can now define a curve δ as the path covered by moving along α first and then along $\hat{\beta}$; $\delta : [a, e] \to S$. Further, the parallel transport of $\overline{V} \in S_p$ along δ will result in $\overline{z}$, as is obvious by the construction. Thus $P_\delta = P_\beta \circ P_\alpha$; $\delta(a) = \alpha(a) = p$ and $\delta(e) = \hat{\beta}(e) = p$. As α and β are piecewise smooth, so is δ.

Therefore, H_p forms a subgroup of G_p, called the holonomy group of S at p.

Examples

Parallel transport shows us that the curvature of the surface measures the amount of rotation obtained by transporting vectors along curves. Holonomy is exactly the phenomenon that occurs when the curve is closed. The vector thus obtained is a rotation of the original vector and so corresponds to an element of the rotation group-$SO(n)$; the group of all positive rotations in n-dimensional space. We now give a few examples of the holonomy group of some surfaces.

EXAMPLE 1. Consider the space $\mathbb{R}^n$ (which can be regarded as a surface in $\mathbb{R}^{n+1}$; given by the equation $x_{n+1} = 0$). Then, the holonomy group is trivial, consisting of the identity element. This is because the parallel translation in the sense of the Riemannian metric coincide with the usual parallel translation. So the vector attained is the vector we started with.

EXAMPLE 2. Consider a 2-plane in 3-dimensions. As the normal of a plane is constant, the tangent space at every point is identical, co-inciding with the plane itself. Thus, if a vector is a tangent at one point of the plane, it is a tangent at every point. Thus parallel transport around a closed loop results in the same vector; so the holonomy group in this case is also trivial.

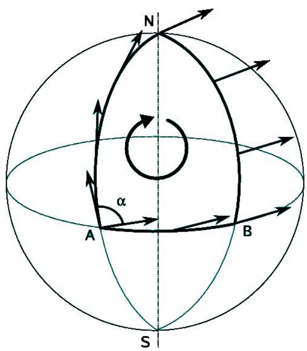


FIGURE 2.

EXAMPLE 3. Consider a 2-sphere in 3-dimensions. Here the holonomy group is not trivial. As seen in the Figure 2, transporting the vector from $A \to N \to B \to A$ yields a vector different from the initial vector; but at an angle to the original vector. This failure

to return to the original vector is measured by the holonomy group; which is $SO(2)$ at every point of this sphere. The same is true for a torus.
It is important to note that this is the holonomy of the loop; because the angle α does not depend on the choice of the starting vector. The angle α corresponds to an element from the rotation group $SO(2)$; an angle modulo 2π.

References

1. J.A.Thorpe, *Elementary Topics in Differential Geometry*, Springer-Verlag, New York Inc., 1979,First Indian Reprint, 2004.
2. Lecture notes on 'Holonomy Groups', University of Adelaide, 2010.
3. www.wikipedia.org
4. I.Singer and J.A.Thorpe, Lecture notes on 'Elementary Topology and Geometry', Scott Foreman & Co. Reprint 1976.

Sahana Balasubramanya, M.Sc. Mathematics, 4th Semester, Mathematics Department, South Campus, University Of Delhi
E-mail address: hbsahana@gmail.com

Extension of Course Contents

The paucity of time restricts us to our course content. Thus, we wish to present this section which goes beyond the scope of our text and introduces concepts which are intriguing and also strengthen our knowledge and understanding.

GOSSIPING SEQUENCES AND SERIES

V. P. SRIVASTAVA

ABSTRACT. 'Convergence' is a commonly used word. We hear: the city is converging to Ramlila Maidan. Everyone understands its meaning. Even convergence in Mathematics is understandably SEEN by all, when we say that the non-ending sequence 1, 2, 3 ,4, . . . converges to infinity or 1, 1/2 , 1/3 , 1/4, . . . converges to zero. Yet, *the delicate structure of real number system does not make it so obvious that increasing sequence does not always converge to infinity nor does decreasing sequence (of positive numbers) necessarily converge to zero.* The natural transition from commonly understood 'convergence' to its mathematical definition deserves greater respect than usually the text-books and the teachers in the classroom give. Outside the classroom, in the corridors, shall we continue the gossip.

MOTIVATION

Let me ask you a question:

If infinite number of positive numbers are added, what is the answer expected?

'Infinity' is the instant reply, I get in the class.

"Must it be infinity", I repeat!

Even the considered reply I get is infinity.

There are genuine objections to my question of the kind: an infinity of numbers cannot be described or written in a finite time space; so no question arises that of adding them up!

Leaving the question unanswered and unattended let me ask a second question:

Can you in some way distribute a unit piece into infinite number of people?

And the reply I often get is very very very small bits.

How small ..... can we numerically quantify it? To this, we don't find an answer.

In the classroom, a chalk-piece I find to be readily available object to be asked to be distributed among infinite number of people. And I get the suggestion to powder it fine, presumably to distribute it to an infinite number of people; each one getting a particle of the powder like the pious powder (Bhabhoot), the medicine for all ailments. Interesting, isn't it! Let me insist *that by distribution we mean to know the exact fraction an individual would get* as his or her share. And then a final hint: I suggest that in an iniquitous world, equal distribution should not be expected to yield result. If the answer is not coming yet, offering half the chalk-piece to the front-corner student, the attempt for the distribution I start. How much to offer to the next is the biggest question? Not to frustrate the third and the subsequent ones in the row leaving nothing for them to be shared, the second should not claim the entire of the remaining half, a quarter looks reasonable for the second. That sets the trend, and a consensus voice announces the share of the remaining in the queue (the queue is much more important mathematically than culturally or civilizationally, a non-queue crowd cannot be obliged, non-queue is a symbol of barbarism) as $1/4, 1/8, 1/16, \ldots$ and so on. It is important to emphasize that in the proposal each person's share in the queue is exactly determined. We may further observe that at each stage of the distribution, the left over is as big as the fraction last given away. And the proposed distribution process can be continued endlessly. It is great discovery to know our ability to distribute one unit or one-tenth, or one-hundredth of a unit to an infinite number of queueable persons. It is important to emphasize that in school-days an infinite geometric series like $1/2 + 1/2^2 + 1/2^3 + \ldots$ is not treated with the respect it deserves. Summing infinite number of members is a GREAT STEP – forward, worthy of jubilations and celebrations. The equality sign used below:

$$1/2 + 1/2^2 + 1/2^3 \ldots = 1$$

has a meaning (to be assigned) different from the usually known meaning of the equality sign used below:

$$5 + 4 + 1 = 10, \quad 6 + 5 + 4 = 15, \quad 5 - 1 = 4.$$

Giving a meaning to this equality is the beginning of interpreting infinite processes (you may note that at no stage of the summing of $1/2 + 1/2^2 + 1/2^3 + \ldots$, the sum reaches 1; as is in the cases like $1/3 + 2/3$ or $1/10 + 1/10 + 1/5 + 3/5$. This entry into the study of infinite processes takes us into a new realm of Mathematics - Analysis.

Sequences : Fast and Slow

Observe the following sequences:

- $1,\ 2,\ 3,\ 4,\ 5,\ 6, \ldots, n, \ \ldots;$
- $1^2, 2^2, 3^2, 4^2, 5^2, 6^2 \ldots, n^2, \ldots;$
- $1^3, 2^3, 3^3, 4^3, 5^3, 6^3 \ldots, n^3, \ldots;$

Clearly each of the above sequences are approaching to infinity. What is the difference? The successive following ones were in greater hurry to go to infinity. Can we get sequences going to infinity yet faster? Yes, for example

$$1^{10}, 2^{10}, 3^{10}, 4^{10}, \ldots, n^{10}, \ldots;$$

and more faster is

$$1^{100}, 2^{100}, 3^{100}, \ldots, n^{100}, \ldots;$$

Can we think of sequences faster than all of the above $\{n^p\}$ types? Yes we can, for example:

$$2^1, 2^2, 2^3, \ldots, 2^n, \ldots;$$

and yet faster ones are

$$5^1, 5^2, 5^3, \ldots, 5^n, \ldots;$$

$$23^1, 23^2, 23^3, \ldots, 23^n, \ldots;$$

We also find that faster than $\{20^n\}$ is the sequence: $1^1, 2^2, 3^3, \ldots, n^n, \ldots$. And rocketing yet faster is $\{n^{n^n}\}$. You can guess, no fastest can be there.

Let us think of slower sequences than all those discussed above and yet going to infinity. A simple answer is:

$$1^{1/2}, 2^{1/2}, 3^{1/2}, \ldots, n^{1/2}, \ldots, \text{ and yet slower shall be:}$$
$$1^{1/3}, 2^{1/3}, 3^{1/3}, \ldots, n^{1/3}, \ldots;$$
$$1^{1/10}, 2^{1/10}, 3^{1/10}, \ldots, n^{1/10}, \ldots;$$

Shall we look for sequences slower than all of these $\{n^{1/p}\}$ type sequences ? Yes thinking the reverse of exponentials, we have the sequence $\log 2, \log 3, \log 4, \ldots, \log n, \ldots$; It is really worthwhile to compare the growth of sequences $\{n\}$ and $\{\log n\}$. Logarithm considered on base 10, the sequence has to take eight steps to attain the value 2, and 90 steps more to attain the value 3 and 900 more steps to reach the value 4 and so on. Let us explore for even more slower sequences and here they are:

$$\log\log 100, \log\log 101, \log\log 102, \ldots, \log\log n, \ldots;$$

$$\{\log\log\log n\},\ n \geq 10^{100} \text{ (to avoid negative terms).}$$

We observe that there is no slowest growing sequence approaching to infinity.

Having seen that $\{n^2\}$ goes towards infinity faster than $\{n\}$, without much effort we observe that $\{1/n^2\}$ approaches to zero faster than $\{1/n\}$; in general if $\{x_n\}$ goes to infinity faster than $\{y_n\}$; then $\{1/x_n\}$ goes to zero faster than $\{1/y_n\}$. We may further expect that $\{x_n\} \to 0$ faster than $\{y_n\} \to 0$ shall imply $\{x_n/y_n\} \to 0$, or equivalently, $\{y_n/x_n\} \to \infty$. We avoid to exactly define 'faster' or 'slower' sequences used above. We may instead assume or define $\{x_n\} \to 0$ faster than $\{y_n\} \to 0$ as $\{x_n/y_n\} \to 0$. It should be clear that $\{1/n^2\} \to 0$ neither faster nor slower than $\left\{\dfrac{1}{n^2+5n+7}\right\}$ as

$$\left\{\frac{1/n^2}{1/(n^2+5n+7)}\right\} = \left\{\frac{n^2+5n+7}{n^2}\right\} = \left\{1 + \frac{5}{n} + \frac{7}{n^2}\right\} \to 1.$$

Attending public functions, you might have found people looking towards M.P.'s and M.L.A.'s and on the arrival of the Chief Minister ignoring them altogether. I have seen heads of people and cameras of media turning away from the Cabinet Minister on the arrival of Madam Sonia Gandhi. Same is our approach towards infinity, you will soon learn to ignore $5n$ and 7 in the presence of n^2; and shall ignore $n^5 + 3n^3 - 40$ in the presence of 3^n. The dominance of a particular term is well-recognized in Mathematics as in everyday life.

Infinite Series

Deferring the exploration of the meaning and definition of convergence of series yet allows me to put a small weight on your head, very small not to offend you in the least. We may assume its weight to be 0.001 grams. But if allowed to put another weight of 0.001 grams and then one more.... can I go indefinitely? Obviously not, this add on will collapse you at some stage and even a stronger person than you may be at a later stage. An endless add on, even with an arbitrary small number, cannot result in a situation like

$$1/2 + 1/2^2 + 1/2^3 + \ldots = 1.$$

So to obtain a non-collapsing convergent situation, beginning with a small number is not enough. The 'fixed' shall have to be done away with. Small or big, a subsequent squeezing is unavoidable. Even a subsequent squeezing, if stopped, will lead to the collapsing situation again and so endless squeezing of terms is a must for convergence of series. So we find that for the series $a_1 + a_2 + a_3 + \cdots + a_n + \ldots$, with $a_n > 0$ for all n, to converge, the necessity is that $a_1 \geq a_2 \geq a_3 \geq a_4 \geq \ldots a_n \geq \ldots$. Shall I tell you, squeezing or giving away does not always lead to bankruptcy. To demonstrate this let us start with 1/100. Break

$$\begin{aligned}
1/100 &= 1/200 + 1/200 \\
&> 1/200 + 1/201 \\
&> 1/200 + 1/202 \\
&\vdots \\
&> 1/200 + 1/300 \\
&\vdots \\
&> 1/200 + 1/3000 \\
&\vdots
\end{aligned}$$

See how I have reserved 1/200 and squeezed the other 1/200. Is it not a miserly give away method? More miserly way is:

$$\begin{aligned}
1/100 &= 9/1000 + 1/1000 = a_1 \\
&> 9/1000 + 1/1001 = a_2 \\
&> 9/1000 + 1/1002 = a_3 \\
&\vdots \\
&> 9/1000 + 1/2000 = a_{1001} \\
&> 9/1000 + 1/5000 = a_{4001} \\
&\vdots
\end{aligned}$$

What needs to be observed is that a miserly squeezing ($a_1 > a_2 > a_3 > \ldots a_n > a_{n+1} > \ldots$) will give $a_1 + a_2 + a_3 + \ldots$ a collapsing situation. And so the convergence of $a_1 + a_2 + a_3 + \ldots$ necessitates a non–miserly (no reservation for oneself) squeezing that leads to bankruptcy. In other words the convergence of $a_1 + a_2 + a_3 + \ldots$ necessitates $a_1 > a_2 > a_3 > \ldots > a_n > \ldots$; and $a_n \to 0$. A landmark conclusion we arrive at is that convergence of Σa_n implies $\{a_n\}$ decreases to 0. However a careful consideration of the series :

$$1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6 + \ldots \geq 1 + 1/2 + 1/4 + 1/4 + 1/8 + 1/8 + 1/8 + 1/8 + \ldots$$

explains $a_n \to 0$ is not enough for convergence. We find that what is required is that $\{a_n\} \to 0$ faster than $\{1/n\}$ does. We find actually $a_n \to 0$ as fast as $\{1/n^2\}$, so hard squeezing of terms a_n gives convergence of Σa_n.

In reality, so hard squeezing is not required; even if $a_n \to 0$ as fast as $\left\{\frac{1}{n^{1+1/2}}\right\}$, even $\left\{\frac{1}{n^{1+1/200}}\right\}$ or even $\left\{\frac{1}{n^{1+1/2000}}\right\} \to 0$ guarantees convergence of Σa_n. But don't be tempted to squeeze a_n at the rate $\left\{\frac{1}{n^{1+1/n}}\right\}$ or even at $\left\{\frac{1}{n \log n}\right\}$ as that would give divergence. However a little harder squeezing at the rate of $\left\{\frac{1}{n(\log n)^2}\right\}$ or even $\left\{\frac{1}{n(\log n)^{200}}\right\}$ guarantees convergence of Σa_n.

Convergence of series defined

Let us explore, the meaning of this new equality sign in $1/2 + 1/2^2 + \ldots = 1$; or the definition we wish to assign to the statement: $a_1 + a_2 + a_3 + \ldots \text{ad inf} = 1$. In what sense, are we inclined to say that $1/2 + 1/2^2 + \ldots = 1$. One obvious property of the infinite addition (or infinite series) above is that at no finite stage the sum is 1; it is always less than 1. It is tending to 1; or is approaching to 1 can be seen. But we may as well say it is approaching to 2 or 1.1 or 33. (This crooked logic is of the type: suppose our car is approaching Delhi from South-direction, someone may say that we are approaching Chandigarh and yet someone else may say, we are approaching Amritsar, places further north). So what is the specialty or uniqueness about 1, which 2, 1.1 or 33 do not possess, when we say that the infinite series is approaching 1. We may for convenience denote S_n to be the sum of the first n terms of the series. We observe that the partial sum S_n is always at a distance greater than 1 from 2, it is always at a distance greater than 0.1 from 1.1, it is always at a distance greater than 32 from 33. Symbolically, we may write

$$2 - S_n > 1$$
$$1.1 - S_n > .1$$
$$33 - S_n > 32$$

for all n. But corresponding to 1, there is no such (as 1, .1, 32 above) positive p so that we may say that $1 - S_n > p$ for all n. The truth is that for arbitrary positive p, we find a corresponding stage N of the sequence where from the inequality gets reversed, that is

$$1 - S_n < p \quad \forall\, n \geq N.$$

The property referred above of 1 is geometrically much more appealing. The specialty of 1 is that the partial sums S_n come near and near, arbitrarily near to 1. That is to say whatever be the standard of nearness chosen, S_n is found close to 1. Let $1/10$ be a standard of nearness then we find that at the 4th stage of the series $1/2+1/2^2+1/2^3+1/2^4 = 1-1/2^4$. That is to say that the distance of the partial sum at the 4th stage from 1 is less than $1/10$, that is

$$1 - S_4 = 1 - (1/2 + 1/2^2 + 1/2^3 + 1/2^4) = 1/2^4 = 1/16 < 1/10.$$

Also

$$1 - S_5 = 1 - (1/2 + 1/2^2 + 1/2^3 + 1/2^4 + 1/2^5) = 1/2^5 = 1/32 < 1/10,$$
$$1 - S_6 = 1 - (1/2 + 1/2^2 + \ldots + 1/2^6) = 1/2^6 = 1/64 < 1/10.$$

Figure 1 below shows that $1 - S_n < 1/10$ for $n \geq 4$. If the standard of nearness is 1/100,

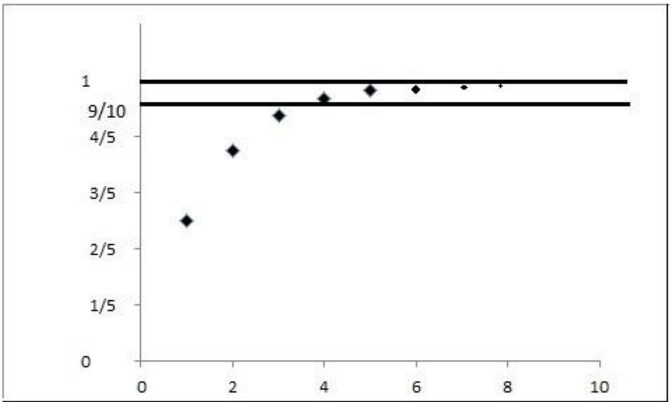


FIGURE 1.

we find that S_n's are in 1/100 closeness of 1, after the 7th stage. To be exact

$$1 - S_7 = 1 - (1/2 + 1/2^2 + \ldots + 1/2^7) = 1/128 < 1/100,$$
$$1 - S_8 = 1 - (1/2 + 1/2^2 + \ldots + 1/2^8) = 1/256 < 1/100,$$
$$1 - S_9 = 1 - (1/2 + 1/2^2 + \ldots + 1/2^9) = 1/512 < 1/100,$$
$$\text{or, } 1 - S_n < 1/100 \quad \text{for all } n \geq 7.$$

Now being at a distance of 1/10 be called near or being at 1/100 distance be called near.

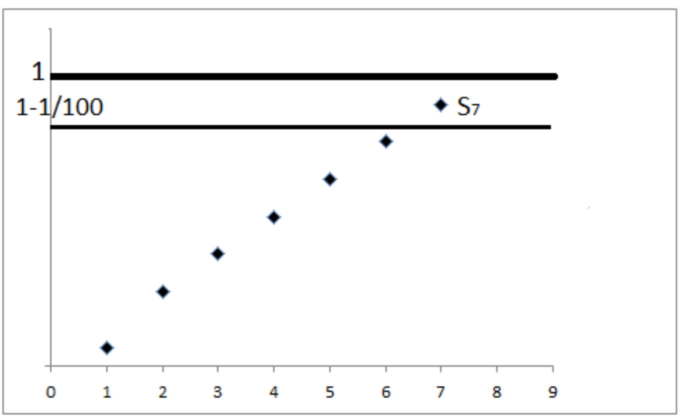


FIGURE 2.

It is relative. They are too big numbers and too large relative to $1/10^{10}$. And the criterion of arbitrarily near cannot be truly tested by 1/10, 1/100 or $1/10^{10}$, none being the smallest positive number. *Truly there is no smallest positive number.* And as such we are left with no choice than to take smaller and smaller positive numbers and demonstrate our ability to find a corresponding N. To fulfill our objective, we may choose arbitrary positive number p and show our ability to find the corresponding N in terms of p, so that we get

$$1 - S_n < p \quad \text{for all} \quad n \geq N.$$

For a given $p > 0$, to ensure $1 - S_n = 1/2^n < p$ for $n \geq N$; it is readily seen that the stage N be chosen as the integer greater than $\log_2(1/p)$.

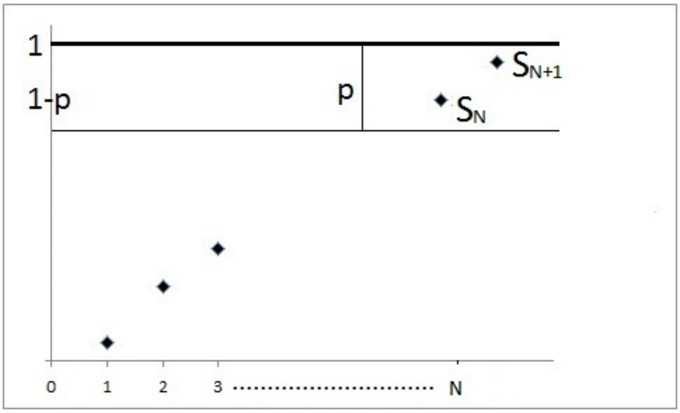


FIGURE 3.

Having discussed the meaning of the sum of an infinite series with all positive terms, little is left to choice for the meaning to be assigned to the infinite series with same numbers with negative sign. Obviously, we are inclined to accept $-1/2 - 1/2^2 - 1/2^3 - \ldots = -1$, because for any given positive $p > 0$, we are able to work out analogously the stage $N > \log_2(1/p)$ such that $S_n - (-1) < p$ for all $n \geq N$ (Figure 4).

It is clear that members of a sequence, approaching a given number can be smaller or bigger than the number; and therefore it is natural that the positive difference should measure the nearness. There shall be situations indeed when members of the sequence approaching a number would be sometimes greater and some other time smaller than the number, coming however arbitrarily close to the number. Consider the series $1 - 1/2 + 1/3 - 1/4 + 1/5 - \ldots$. Let us plot the sequence of partial sums in this case (Figure 5):

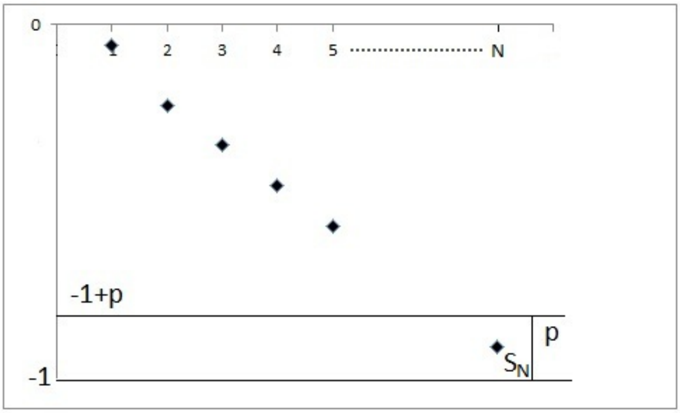


FIGURE 4.

$$
\begin{aligned}
&S_1 = 1\\
&S_2 = 1 - 1/2\\
&S_3 = 1 - 1/2 + 1/3\\
&\vdots\\
&S_n = 1 - 1/2 + 1/3 - \ldots + (-1)^{n-1}/n;
\end{aligned}
$$

The sequence $\{S_n\}$ in this case goes above and below (alternatively) the number L being

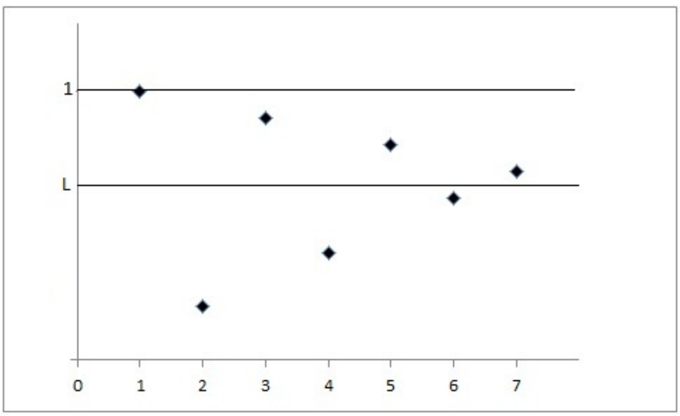


FIGURE 5.

approached. In this case, corresponding to the arbitrary positive number p, we must mark $L - p$ and $L + p$; and try to discover the stage N in the sequence, such that the N-th stage onwards, the sequence is inside the strip $L-p$ to $L+p$. In other words, N is to be found such that when $n \geq N$, we have

$$L - p < S_n < L + p.$$

Or equivalently,

$$-p < L - S_n < p \text{ for } n \geq N,$$

or,

$$|S_n - L| < p \quad \text{for all } \ n \geq N.$$

The modulus used here could be used to cover the earlier two cases also where the series had only positive terms or only negative terms. And without giving importance to the approach from upper (greater) or lower side; we may straight look for the stage in the sequence where onwards it falls in the strip $(L-p, L+p)$. It is not difficult to dispense away with the two–dimensional scene and look for the stage in the sequence from where onwards, the sequence falls in the given interval $(L-p, L+p)$ of the real line for an arbitrary positive p. We feel as much satisfied, whether this stage is 100 or 100^{100}. Indeed, success in getting a stage ensures that except a finite number (100 or 100^{100} or....) of members, the entire sequence lies in $(L-p, L+p)$, and so does it, for each positive p.

Sequences Revisited

It is easy to fabricate non-converging sequences, if you know the art of adulteration. Let us take sequences:

$$2/1, 3/2, 4/3, 5/4, 6/5, \ldots$$
$$1/3, 1/3, 1/3, 1/3, 1/3, 1/3, \ldots$$
$$1/5, 1/5, 1/5, 1/5, 1/5, 1/5, \ldots$$
$$1/4, 1/5, 1/6, 1/7, 1/8, 1/9, 1/10, \ldots$$

The sequences above are converging to $1, 1/3, 1/5$ and 0 respectively. We adulterate and form the sequence:

$$2, 1/3, 1/5, 1/4, 3/2, 1/3, 1/5, 1/5, 4/3, 1/3, 1/5, 1/6, 5/4, \ldots$$

Does this sequence behave like the earlier sequences? The adulteration has changed its character; the new sequence does not converge. It is natural to try to reverse the process of adulteration. Mixing of water with milk is easy - but the reverse process is not so. It is believed that a mythological bird on Himalayas called 'Hans' (similar to Swan) has the strange ability to separate the two. We invoke the ability of Hans to find subsequences of this adulterated sequence that converge. Mind it, there are not just four convergent subsequences. There are many many more, but essentially four; in the sense that the convergent ones have only four distinct limits and the two having the same limit differ only at finite number of terms.

Let us plot the adulterated sequence, and observe how there are basically four streams in the mixed sequence flowing four distinct ways: (Figure 6 and Figure 7).

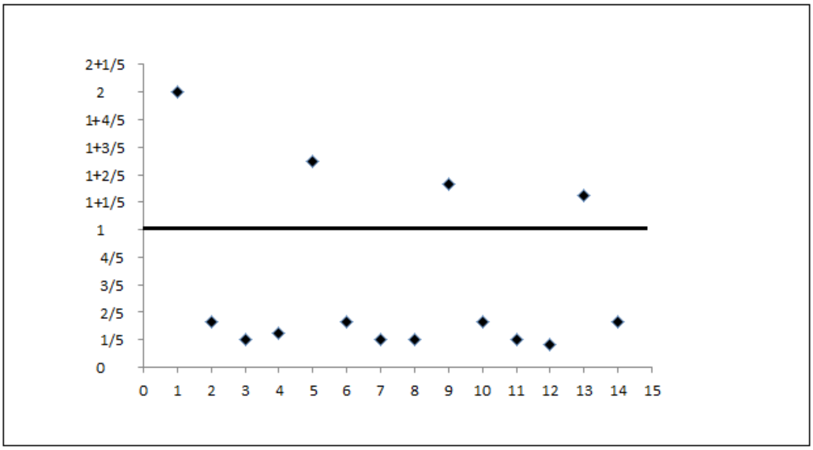


FIGURE 6.

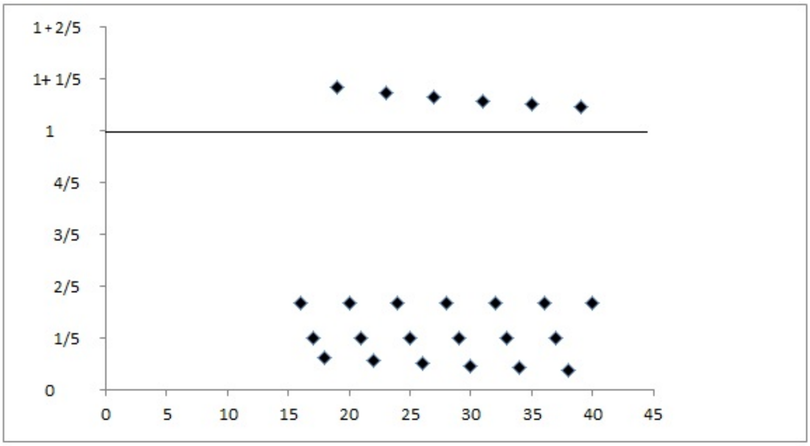


FIGURE 7. The sequence at a later stage

One of the four limits (obtained from a convergent subsequence) can be said to be one of the expectations, one of the ultimates of the adulterated sequence. We have in this case four expectations, the four ultimates of the mixed sequence. The biggest, we call the limit superior and the lowest as the limit inferior of the sequence. Thus the sequence has 1 and 0 as the limit superior and limit inferior respectively. We may think of them as the ultimate biggest and the ultimate smallest of the sequence also. In case of convergent sequences, obviously they are the same.

We can catch hold of the limit superior and the limit inferior of a sequence $\{a_n\}$ even without finding all the limits of the possible convergent subsequences. Let us now try to catch hold of the limit superior, that is to say the ultimate biggest of the sequence (approached by the sequence in the sense of the limit). First crude attempt to catch the ultimate biggest (supremum) can be to find the supremum of the entire sequence, that is $\sup\{a_n|n \geq 1\}$. Having taken the first step, we get a direction for finding a better estimate for the ultimate supremum as $\sup\{a_n|n \geq 2\}$, which purifies the earlier effort somewhat, in case $\sup\{a_n|n \geq 1\}$ was influenced by a_1 being the biggest. (We must not forget that the ultimate limits are not affected by first, second, or any finite number of terms of the sequence). Proceeding further, better and better declining estimates for the limit superior can be arranged as

$$\begin{aligned}
&\sup\{a_k|k \geq 1\} \\
&\geq \sup\{a_k|k \geq 2\} \\
&\geq \sup\{a_k|k \geq 3\} \\
&\vdots \\
&\geq \sup\{a_k|k \geq n\} \\
&\geq \dots \\
&\vdots
\end{aligned}$$

And to catch the best, the limit of the above or the smallest (infimum) of them (the ultimate among the suprema) is seen to be

$$\lim_{n\to\infty} \sup\{a_k|k \geq n\} = \inf_n \sup\{a_k|k \geq n\}.$$

We may well define (in notation)

$$\overline{\lim_{n\to\infty}} a_n = \limsup_{n\to\infty} a_n = \inf_{k\geq n} \sup a_k.$$

In our example above:

$$\begin{aligned}
\sup\{a_k|k \geq 1\} &= 2, \\
\sup\{a_k|k \geq 2\} &= 3/2, \\
\sup\{a_k|k \geq 3\} &= 3/2, \\
\sup\{a_k|k \geq 4\} &= 3/2, \\
\sup\{a_k|k \geq 5\} &= 3/2, \\
\sup\{a_k|k \geq 6\} &= 4/3.
\end{aligned}$$

Observe how from the biggest in the entire sequence, step by step, we are approaching to catch the ULTIMATE BIGGEST, as the

$$\begin{aligned}
\lim_{n\to\infty} \sup\{a_k|k \geq n\} &= \lim_{n\to\infty} \{\sup\{a_k|k \geq 1\}, \sup\{a_k|k \geq 2\}, \dots, \sup\{a_k|k \geq n\}, \dots\} \\
&= \lim\{2, 3/2, 3/2, 3/2, 3/2, 4/3, 4/3 \dots\} = 1.
\end{aligned}$$

We may as well take the earlier sequence with modification at three stages, so that the new sequence is (knowing well that the introduction of three members is not going to affect any of the limits): $2, 1/3, 1/5, 50, 1/4, 3/2, 60, 1/3, 1/5, 14, 1/5, 4/3, \ldots$ We find

$$\begin{aligned}&\sup\{a_k|k \geq 1\}\\&= \sup\{a_k|k \geq 2\}\\&= \ldots\\&= \sup\{a_k|k \geq 7\} = 60\end{aligned}$$

and

$$\begin{aligned}&\sup\{a_k|k \geq 8\}\\&= \sup\{a_k|k \geq 9\}\\&= \sup\{a_k|k \geq 10\} = 14\end{aligned}$$

and

$$\begin{aligned}&\sup\{a_k|k \geq 11\} = 4/3\\&= \sup\{a_k|k \geq 12\} = 4/3.\end{aligned}$$

Thereafter the earlier pattern repeats, leading finally to $\limsup a_n = \inf \sup\{a_k|k \geq n\} = 1$. We may remember that the limit superior (say M) being the infimum (the greatest lower bound) of $\{\sup\{a_k|k \geq n\}, n \in \mathbb{N}\}$, satisfies the following:

(1) For $\epsilon > 0, M + \epsilon$ is not a lower bound for $\left[\sup\{a_k|k \geq n\}\right]_n$; so $\sup\{a_k|k \geq n\} < M + \epsilon$ for some $n = n_0$, so, $a_n < M + \epsilon$ for $n \geq n_0$, that is, some n_0-th stage onwards every element of the sequence is less than $M + \epsilon$. In our example, for $\epsilon = 1/10$ we find that $a_n < 1 + 1/10$ for every $n > 37$. Looking graphically, after the 37-th stage, the sequence is below the horizontal line at $1 + 1/10$ (see Figure 8).

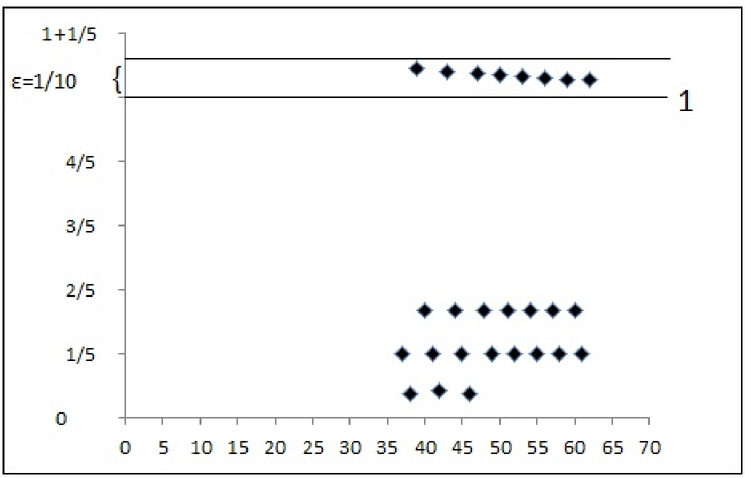


FIGURE 8.

(2) M is a lower bound for $\left[\sup\{a_k|k \geq n\}\right]_n$, so $M < \sup\{a_k|k \geq n\}$ for all n. For $\epsilon > 0$, $M - \epsilon$ is not an upper bound for $\{a_k|k \geq n\}$ for any n. So given integer n_0, we can find $n > n_0$ such that $a_n > M - \epsilon$. That is to say, we obtain $a_n > M - \epsilon$ for infinitely many arbitrarily large values of n. Taking $\epsilon = 1/10$ in our example we find that $a_n > 1 - 1/10$ for $n = 1, 5, 9, \ldots$. The missing members of our sequence $\{a_2, a_3, a_4, a_6, a_7, \ldots\}$ are essentially not part of the sequence converging to 1 (the limit superior), and so need not fall in the strip $(1 - \epsilon, 1 + \epsilon)$.

We may analogously catch hold of limit inferior and define it as $\underline{\lim}\{a_n\} = \sup_n \inf\{a_k|k \geq n\}$. The reader may try the above definition in finding the limit superior and limit inferior of the function $f(x)$, as x is tending to zero, where f is defined on $[0, 1]$ by:

$$f(x) = \begin{cases} 1, & \text{for } x \text{ a rational number in [0,1] except } 1/3, 1/5, 1/7 \\ 30, & \text{for } x = 1/3 \\ -50, & \text{for } x = 1/5 \\ -70, & \text{for } x = 1/7 \\ 2, & \text{for } x = \sqrt{2}/n \text{ for } n \in \mathbb{N},\ n \geq 2 \\ 3, & \text{for } x = \sqrt{3}/n \text{ for } n \in \mathbb{N},\ n \geq 2 \\ 5, & \text{for } x = \sqrt{5}/n \text{ for } n \in \mathbb{N},\ n \geq 3 \\ 3/2, & \text{elsewhere in [0,1].} \end{cases}$$

Let us close the discussion with the following two remarks:

We may observe that the factor $(-1)^n$ makes sequences $\{(-1)^n n\}$, $\{(-1)^n n^2\}$ oscillate wildly (unboundedly). A similar role is played by factors of the form $\sin n$ or $\cos n$ to make $\{n \sin n\}$ or $\{n^2 \cos n\}$ oscillate.

It is interesting to observe that the series $1+1/3+1/5+1/7+\ldots$ and $1/2+1/4+1/6+\ldots$ are both divergent, and yet the series $1-1/2+1/3-1/4+1/5-\ldots$ converges. It is like the case of two drunkards, not able to walk on their own yet managing to walk along balancing each other's step. Rearrangement on the steps can bring consequences unimagined. Both the drunkards leaning on the same side, making all terms of the series positive, is one such disaster.

V. P. Srivastava, Retired Professor and Head, Department of Mathematics, University of Delhi

E-mail address: vp.srivastava@rediffmail.com

URL: www.vpsrivastava.blogspot.com

RINGS OF REAL-VALUED UNIFORMLY CONTINUOUS FUNCTIONS

JASPREET KAUR

Abstract. The study of algebraic structure on function spaces such as the space of real-valued continuous functions has been beneficial in solving various problems in mathematics, and indeed it had a crucial role in the development of a branch of mathematical analysis, viz.‘ functional analysis’. In this article we present some recent results which deal with a basic problem related to ring structure of the set of all real-valued uniformly continuous functions defined on a metric space.

Introduction

The theory of mathematical analysis surrounds the notion of continuity of functions and the concept of uniform continuity occupies a special place in this theory. It is known that a uniformly continuous function is continuous at each point of the domain of the function. But there are several privileges a uniformly continuous function enjoys which may not be enjoyed by a continuous non-uniformly continuous function. For instance, a uniformly continuous function preserves Cauchy sequences. This property further gives the extension property of a uniformly continuous function. Facts like these make the class of all uniformly continuous functions a rich source of interesting problems. In this article we will discuss a simple yet important question concerning uniformly continuous functions. It will be assumed that reader is already familiar with the notions of groups, rings and metric spaces.

Let (X, d) be a metric space and $C(X)$ be the collection of all continuous functions from X into $\mathbb{R}$, the reals (with usual metric u, $u(x, y) = |x - y|$). It can be easily verified that $C(X)$ forms a commutative ring with unity under the operation of pointwise addition and pointwise multiplication. Now, if we denote the collection of all uniformly continuous functions from (X, d) into $(\mathbb{R}, u)$ by $U(X)$. Then the question arises whether $U(X)$ also forms a ring (under pointwise sum and pointwise product). Although $U(X)$ is an abelian group under the operation of pointwise addition but in general, $U(X)$ is not closed under pointwise multiplication i.e pointwise product of two elements of $U(X)$ may not be an element of $U(X)$. For example, the identity function on $(\mathbb{R}, u)$ is uniformly continuous but its pointwise product with itself is not uniformly continuous. This gives rise to the following characterization problem: **Determine intrinsic conditions that characterize those metric spaces for which $U(X)$ is closed under pointwise product.**

Beside compact metric spaces and discrete metric spaces there are other metric spaces as well for which $U(X)$ forms a ring. Such metric spaces have been studied by M. Atsuji[1], S. Nadler[6], [7] and others. In this article we will discuss some of the results from these research papers.

Preliminaries

In this section we cover background material which is needed to discuss the characterization problem as mentioned in the introduction. Unless otherwise stated, we will consider only the usual metric on $\mathbb{R}$ as well as on subsets of $\mathbb{R}$.

Proposition 1. *(Sequential Criterion for Non-Uniform Continuity) Let $A \subseteq \mathbb{R}$. Then a function $f : A \to \mathbb{R}$ is not uniformly continuous on A iff there exists an $\epsilon_0 > 0$ and two sequences (x_n) and (y_n) in A such that $\lim(x_n - y_n) = 0$ but $|f(x_n) - f(y_n)| \geq \epsilon_0$ for all $n \in \mathbb{N}$.*

Proof. See [12] □

Proposition 2.

(1) *Let $f : [a, \infty) \to \mathbb{R}$ be a continuous function such that $\lim_{x\to\infty} f(x)$ is finite. Then f is uniformly continuous on $[a, \infty)$.*
(2) *Let $f : (-\infty, a] \to \mathbb{R}$ be a continuous function such that $\lim_{x\to-\infty} f(x)$ is finite. Then f is uniformly continuous on $(-\infty, a]$.*
(3) *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that both the limits $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$ are finite. Then f is uniformly continuous on $\mathbb{R}$.*

Proof. [8] □

Remark 1. Characterizations of uniformly continuous functions on bounded intervals of $\mathbb{R}$ are well known but results related to uniform continuity on unbounded intervals (such as in Proposition 2) are often left out of classroom discussions and textbooks. More results like the one above can be found in the expository article [9].

Proposition 3. *Let f be a real-valued uniformly continuous function defined on a bounded subset A of $\mathbb{R}$. Then f is a bounded function.*

Remark 2. The proof of the above proposition is an easy exercise. Also we note that the above result cannot to extended to metric spaces in general. Consider the discrete metric d on the set of real numbers. The identity function $I : (\mathbb{R}, d) \to (\mathbb{R}, u)$ is clearly uniformly continuous. The set $\mathbb{N}$ of natural numbers is a bounded subset of $(\mathbb{R}, d)$ but its image under I is unbounded.

Definition 1. Let (X, d) be a metric space. A subset Y of X is called *uniformly discrete* or *uniformly isolated* provided there is an $\epsilon > 0$ such that $d(x, y) > \epsilon$ holds for all distinct points x and y in A.

Example 1. The set $\mathbb{N}$ of natural numbers is a uniformly discrete subset of the usual metric space $(\mathbb{R}, u)$ while the set $\{1/n : n \in \mathbb{N}\}$ is not uniformly discrete in $(\mathbb{R}, u)$.

Definition 2. Let (X, d) be a metric space and $\epsilon > 0$. An ϵ-chain in X from x to y of length m is a finite sequence $x = x_0, x_1, ..., x_m = y$ of points in X such that $d(x_{i-1}, x_i) < \epsilon$ for all $i = 1, 2, ..., m$. We say that (X, d) is *finitely chainable* provided that for each $\epsilon > 0$, there are finitely many points $p_1, ..., p_n \in X$ and a positive integer m such that there is an ϵ-chain in X of length m from any point x of X to one of the points $p_1, ..., p_n$.

Theorem 1. *(Extension Theorem) Let (X, d_X) be a metric space and (Y, d_Y) be a complete metric space, and let A be a dense subspace of X. If f is a uniformly continuous function of A into Y, then f can be extended uniquely to a uniformly continuous function g of X into Y.*

Proof. See [11]. □

Proposition 4. *(Pasting Lemma for Continuous Functions) Let (X, d_X) and (Y, d_Y) be two metric spaces. Suppose A and B are two closed subspaces of (X, d_X) and $f : A \to Y$ and $g : B \to Y$ are continuous functions such that $f(x) = g(x)$ for all $x \in A \cap B$. Then the function $h : A \cup B \to (Y, d_Y)$ given by*

$$h(x) = \begin{cases} f(x), & x \in A \\ g(x), & x \in B \end{cases}$$

is well-defined and continuous.

Proof. See [5]. □

Characterization Theorems

We first state and prove the result which gives the solution to the characterization problem for subsets of $(\mathbb{R}, u)$. In other words, the result characterizes those sets in the real line on which the pointwise product of any two real-valued uniformly continuous functions is uniformly continuous.

Theorem 2. *The pointwise product of any two uniformly continuous real-valued functions on a subset X of $\mathbb{R}$ is uniformly continuous if and only if X is the union of a bounded set and a uniformly discrete set.*

Proof. Let $X = B \cup I$, where B is a bounded subset and I is a uniformly discrete subset of $\mathbb{R}$. Since the set I is uniformly discrete, there exists $\delta_0 > 0$ such that $|x - y| > \delta_0$ whenever x and y are distinct points of I. Also, since the maps f and g are uniformly continuous on X, for $\epsilon = 1$ there exists $\delta_1, \delta_2 > 0$ such that

$$|f(x) - f(y)| < 1 \text{ whenever } |x - y| < \delta_1$$

and,

$$|g(x) - g(y)| < 1 \text{ whenever } |x - y| < \delta_2.$$

Let $\delta_3 = \min\{\delta_1, \delta_2\}$. Now, by Proposition 3 the restriction of uniformly continuous maps f and g on bounded set B are bounded maps. Therefore, there exists $M_1, M_2 > 0$ such that for each $x \in B$, $|f(x)| < M_1$ and $|g(x)| < M_2$. Take $M = \max\{M_1, M_2\}$. Using the relation $||f(x)| - |f(y)|| \leq |f(x) - f(y)|$ we get that whenever $x \in B$ and $y \in X \cap (x - \delta_3, x + \delta_3)$, $|f(y)| < 1 + |f(x)| < 1 + M$ and similarly $|g(y)| < 1 + M$. We now prove uniform continuity of the pointwise product function $f.g$ by ϵ-δ definition. Let $\epsilon > 0$ be arbitrary. Again by uniform continuity of the functions f and g, for $\epsilon_0 = \frac{\epsilon}{2(M+1)} > 0$, there exists $\delta_4, \delta_5 > 0$ such that

$$|f(x) - f(y)| < \epsilon_0 \text{ whenever } |x - y| < \delta_4$$

and,

$$|g(x) - g(y)| < \epsilon_0 \text{ whenever } |x - y| < \delta_5.$$

Let $\delta_6 = \min\{\delta_4, \delta_5\}$ and $\delta = \min\{\delta_0, \delta_3, \delta_6\}$. Now, if x and y are points of X such that $0 < |x - y| < \delta$ then we note that x and y both together cannot be members of I. So, without loss of generality we assume that $x \in B$. Thus, $|f(x)| < M$ and since, $y \in (x - \delta, x + \delta) \subseteq (x - \delta_3, x + \delta_3)$, $|g(y)| < M + 1$. Hence,

$$\begin{aligned} |f(x)g(x) - f(y)g(y)| &= |f(x)g(x) - f(x)g(y) + f(x)g(y) - f(y)g(y)| \\ &\leq |f(x)||g(x) - g(y)| + |g(y)||f(x) - f(y)| \\ &< M\frac{\epsilon}{2(M+1)} + (M+1)\frac{\epsilon}{2(M+1)} < \epsilon \end{aligned}$$

This shows that $|f.g(x) - f.g(y)| < \epsilon$ whenever $|x - y| < \delta$ i.e., the pointwise product of f and g is uniformly continuous on X.

The proof of converse part proceeds by contradiction. Suppose, on the contrary X is not the union of a bounded set and a uniformly discrete set. This means that the set X is neither bounded nor uniformly discrete. Thus for each $n \in \mathbb{N}$ there exists an element x_n in X such that $x_n > n$. Also for each $n \in \mathbb{N}$, there exist two elements y_n, z_n in X such that $|y_n - z_n| \leq 1/n$. Hence, without loss of generality, there are sequences $\{a_n\}$ and $\{b_n\}$ in X such that $n < a_n < b_n < a_{n+1}$ and $\lim_{n\to\infty}(b_n - a_n) = 0$. Now define a function $f : \mathbb{R} \to \mathbb{R}$ as

$$f(x) = \begin{cases} \frac{x-a_n}{n(b_n-a_n)}, & a_n \leq x \leq b_n \\ \frac{x-a_{n+1}}{n(b_n-a_{n+1})}, & b_n \leq x \leq a_{n+1} \\ 0 & \text{otherwise} \end{cases}$$

By applying 'pasting lemma' (Proposition 4) it can be seen that f is a continuous function. Also, $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$. Therefore by Proposition 2, f is uniformly continuous on $\mathbb{R}$. Now, let h be the restriction of f to X and I_X be the identity function on X. Then h and I_X are uniformly continuous, but $g = h.I_X$ is not uniformly continuous because $\lim_{n\to\infty}(b_n - a_n) = 0$ and yet $|g(b_n) - g(a_n)| = |b_n.h(b_n) - a_n.h(a_n)| = |b_n/n - 0| > n/n = 1$. □

Next, we intend to discuss the characterization problem for metric spaces. But we first remark that the above result cannot be extended to metric spaces in general. The following example illustrates this fact.

Example 2. Consider the metric d given by $d(x, y) = \min\{|x - y|, 1\}$ on the set $\mathbb{R}$ of real numbers. Clearly, the space $(\mathbb{R}, d)$ is a bounded metric space. The identity function $I : (\mathbb{R}, d) \to (\mathbb{R}, u)$ is uniformly continuous because for $0 < \epsilon < 1$, we have $\delta = \epsilon$ such that $|I(x) - I(y)| = |x - y| < \epsilon$ whenever $d(x, y) < \delta$ and for $\epsilon \geq 1$ we choose ϵ_0, $0 < \epsilon_0 < 1 \leq \epsilon$ such that for $\delta_0 = \epsilon_0$ we have $|x - y| < \epsilon$ when $d(x, y) < \delta_0$. However, we show that the pointwise product of identity function I with itself is not uniformly continuous. We show that for $\epsilon = 1$ 'no δ works'. Let $\delta > 0$ be any real number. Then by Archimedean property of real numbers there exists $n \in \mathbb{N}$ such that $0 < \frac{1}{n} < \delta$. So for real numbers $x_\delta = n$ and $y_\delta = n + \frac{1}{n}$ we have $d(x_\delta, y_\delta) = \frac{1}{n} < \delta$ but $|{x_\delta}^2 - {y_\delta}^2| > 1$.

Proposition 5. *Let (X, d) be a finitely chainable metric space. Then every real-valued uniformly continuous function on X is bounded.*

Proof. Let $f : X \to \mathbb{R}$ be a uniformly continuous function. So, for $\epsilon = 1$ there exists $\delta > 0$ such that $d(x, y) < \delta$ implies $|f(x) - f(y)| < 1$. Now, since X is finitely chainable, for $\delta > 0$, there exists finitely many points $p_1, ..., p_n$ and a positive integer m such that there is an δ-chain in X of length m from any point $x \in X$ to p_j for some $j : 1 \leq j \leq n$. Let $x = y_0, y_1, ..., y_m = p_j$ be δ-chain joining x and p_j. Since $d(y_{i-1}, y_i) < \delta$ for all $i = 1, ..., m$, $|f(y_{i-1}) - f(y_i)| < 1$ for all $i = 1, ..., m$. Thus, $|f(x) - f(p_j)| \leq |f(x) - f(y_1)| + |f(y_1) - f(y_2)| + ... + |f(y_{m-1}) - f(p_j)| < m$. Let $K = \max\{f(p_j) : j = 1, .., n\}$ and $k = \min\{f(p_j) : j = 1, .., n\}$. Then for each $x \in X$ we have $k - m < f(x) < K + m$. This shows that f is bounded. □

Proposition 6. *Let f and g be two real-valued bounded uniformly continuous functions defined on a metric space (X, d). Then the pointwise product of f and g is uniformly continuous on X.*

Proof. Since f and g are bounded, there exists positive real numbers M_1 and M_2 such that for each $x \in X$, $|f(x)| < M_1$ and $|g(x)| < M_2$. Take $M = \max\{M_1, M_2\}$. Now, f and g are both uniformly continuous as well so for $\epsilon_0 = \frac{\epsilon}{2(M+1)} > 0$ there exists $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$|f(x) - f(y)| < \epsilon_0 \text{ whenever } d(x, y) < \delta_1$$

and,

$$|g(x) - g(y)| < \epsilon_0 \text{ whenever } d(x, y) < \delta_2.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then,

$$\begin{aligned} |f(x)g(x) - f(y)g(y)| &= |f(x)g(x) - f(x)g(y) + f(x)g(y) - f(y)g(y)| \\ &\leq |f(x)||g(x) - g(y)| + |g(y)||f(x) - f(y)| \\ &< M\frac{\epsilon}{2(M+1)} + M\frac{\epsilon}{2(M+1)} < \epsilon \text{ whenever } d(x, y) < \delta \end{aligned}$$ □

Combining Propositions 5 and 6 we get the following sufficient condition for the set $U(X)$ to be closed under pointwise product.

Theorem 3. *Let (X, d) be a finitely chainable metric space. Then $U(X)$ is closed under pointwise product.*

In fact, using Proposition 5 and arguments similar to that in Theorem 2 we can prove the following.

Theorem 4. *Let (X, d) be a metric space. If X is the union of a finitely chainable subspace F and a uniformly discrete subspace I, then $U(X)$ is closed under pointwise product.*

Remark 3. The converse of above theorem is also true in some cases:

(1) M.Atsuji [1] showed: If (X, d) is a connected metric space which is not finitely chainable, then $U(X)$ is not closed under pointwise product.

(2) S.Nadler [6] showed that the converse of above Theorem 4 holds for subspaces of those metric spaces in which all closed and bounded subsets are compact. Precisely, let (X, d) be a metric space in which all closed and bounded subsets are compact, and let $Y \subset X$. If $U(Y)$ is closed under pointwise product then Y is union of a finitely chainable set and a uniformly discrete set.

Exploiting the extension property of uniformly continuous functions one can find a relation between the ring structure of the space $U(Y)$ and that of $U(\overline{Y})$, where Y is a subspace of a metric space (X, d) and $\overline{Y}$ denotes the closure of Y in X.

Theorem 5. *Let (X, d) be a metric space and $Y \subset X$. Then $U(Y)$ is closed under pointwise product iff $U(\overline{Y})$ is closed under pointwise product.*

Proof. Let $U(Y)$ be closed under pointwise product and $f, g : \overline{Y} \to \mathbb{R}$ be two uniformly continuous functions. Then the restriction of functions f and g to Y are uniformly continuous. Since $U(Y)$ is closed under pointwise product, the pointwise product $h = f|Y.g|Y$ of the restriction functions is uniformly continuous. Now, the function $\overline{h} = f.g$ is extension of the function h on $\overline{Y}$. Therefore, by Theorem 1, $\overline{h}$ is also uniformly continuous.
Conversely, let $U(\overline{Y})$ be closed under pointwise product and $f, g : Y \to \mathbb{R}$ be uniformly continuous functions. Since $\mathbb{R}$ is a complete metric space, by Theorem 1, f and g can be extended uniquely to uniformly continuous functions $\overline{f}, \overline{g} : \overline{Y} \to \mathbb{R}$ respectively. Now, because $U(\overline{Y})$ is closed under pointwise product, the pointwise product $h = \overline{f}.\overline{g}$ is uniformly continuous. Hence the restriction of h to Y which is same as the pointwise product of f and g is also uniformly continuous. □

As a direct consequence of the above theorem we get the following.

Corollary 1. Let $(\hat{X}, \hat{d})$ be completion of the metric space (X, d). Then $U(\hat{X})$ is a ring iff $U(X)$ is a ring (under pointwise sum and pointwise product).

We conclude by some remarks.

Remark 4.

(1) In this article we addressed the question of determining those metric spaces for which the set $U(X)$ of all real-valued uniformly continuous functions forms a ring under pointwise sum and pointwise product. However, for any metric space (X, d) there are certain subsets of $U(X)$ consisting some special types of functions defined on X which always forms a ring. For example, the set $U^*(X)$ of all real-valued bounded uniformly continuous functions on any metric space (X, d) is closed under pointwise sum and pointwise product and thus forms a ring (see Proposition 6). More such subsets of $U(X)$ for a given metric space (X, d) have been determined in several research papers for eg. [3] and [4].

(2) We saw that pointwise product is not well behaved when it comes to uniformly continuous functions. But we also know that composition of two uniformly continuous functions is uniformly continuous. So a natural question arises as to whether the set $U(\mathbb{R})$ forms a rings with respect to addition operation as pointwise addition but

multiplication operation as composition of functions. The answer to this question is also negative. Because left distributive law fails to hold in $U(\mathbb{R})$ with these operations. For example consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \sin x$ and let I be the identity function on $\mathbb{R}$. Then $f \circ (I + I) \neq f \circ I + f \circ I$. However, one can find in literature the concept of 'near-rings' which is similar to that of rings but satisfy fewer axioms and it can be easily verified that $U(\mathbb{R})$ form a right near-ring. For sake of completeness we give here the definition of right near-ring.

Definition 3. A (right) near-ring is a set N together with two binary operations, denoted by $+$ and $\cdot$, such that

(a) $(a+b)+c = a+(b+c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all a, b, $c \in N$ (associativity of both operations).

(b) There exists an element $0 \in N$ such that $a + 0 = 0 + a = a$ for all $a \in N$ (additive identity).

(c) For each $a \in N$, there exists $b \in N$ such that $a + b = b + a = 0$ (additive inverse).

(d) $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$ for all $a, b, c \in N$ (right distributive law).

Also, one may note that the axioms of a near-ring differ from those of a ring primarily in the sense that they do not require addition to be commutative, and only require distributivity on one side.

References

1. Atsuji, M., *Uniform continuity of continuous functions of metric spaces*, Pacific Journal Math. 8 (1958), 11-16.
2. Bartle, R.G., Sherbert, D.R., *Introduction to Real Analysis*, Wiley India, (2007).
3. Elyash, E.S., Laush, G., Levine, G., *On the product of two uniformly continuous functions on the line*, American Math. Monthly, Vol 67, No.3 (1960).
4. Levine, N., Saber, N.J, *On the product of real valued uniformly continuous functions in metric spaces*, Amer. Math. Monthly, Vol 72, No. 1 (1965).
5. Munkres, J.R., *Topology*, Prentice-Hall of India, New Delhi, (2002).
6. Nadler, Sam B., *Pointwise product of uniformly continuous functions*, Sarajevo Journal of Math., Vol 1., No.3 (2005), 117-127.
7. Nadler, Sam B., Zitney, Donna M., *Pointwise product of uniformly continuous functions on sets in the real line*, American Math. Monthly, vol.114, No. 2 Feb. (2007), 160-163.
8. Parzynski, W.R., Zipse, P.W., *Introduction to Mathematical Analysis*, McGraw-Hill, New York, (1982).
9. Pouso, R.L.,*Uniform continuity on unbounded intervals: classroom notes*, Internationl J. of Math. Edu. in Sci. and Tech.,Vol 39, Issue 4 (2008).
10. Searcoíd, M., *Metric Spaces*, Springer-Verlag London Lim., (2007).
11. Simmons, G.F., *Introduction to Topology and Modern Analysis*, Tata McGraw-Hill, (2004).
12. Stephen, A., *Understanding Analysis*, Springer, (2001).

Jaspreet Kaur, Research Scholar, Department of Mathematics, University of Delhi
E-mail address: jasp.maths@gmail.com

SINGULAR VALUE DECOMPOSITION

DOOTIKA VATS

ABSTRACT. This paper introduces matrix decompositions, with the focus being on the Singular Value Decomposition of a matrix. The main emphasis of the paper is to outline the construction of the decomposition and its few important applications.

INTRODUCTION

One of the most fruitful ideas in the theory of matrices is that of matrix decomposition. The theoretical utility of matrix decompositions has long been appreciated. More recently, they have become the mainstay of numerical linear algebra, where they serve as computational platforms from which a variety of problems can be solved. Just like an integer might factor many ways, a matrix can also be expressed as the product of two or more matrices.

The most trivial decomposition is to write any square matrix A as AI or IA where I is the identity matrix. One of the famous and useful decompositions is the QR factorization obtained by the famous Gram-Schmidt Algorithm.

There are many other decompositions that bring a matrix to a canonical form. The one taught most commonly to linear algebra students is the LU Decomposition. I am going to touch upon this decomposition a bit, in order to get the reader familiarized with decompositions.

LU DECOMPOSITION

Let A be a square matrix. An LU Decomposition of A is of the form $A = LU$ where L is a lower triangular matrix and U is an upper triangular matrix.
An elementary method used in linear algebra to find solutions to a system of linear equations is the Gaussian Elimination. In this method, we try and reduce any given matrix A to an upper triangular matrix U by applying some row eliminations. Consider the example of a 3×3 matrix below:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

by applying the eliminations–
(2 times the first equation from the second),
(-1 times the first equation from the third), and
(-1 times the second equation from the third).

How these row eliminations work is that, each elimination is coded as a matrix which multiplies to A on the left to give a transformed matrix (for column eliminations, we multiply on the right). For example, the first elimination (2 times the first equation from the second) can be coded as:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow EA = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix}$$

This is exactly what we get when we apply the first elimination to A. Next we code the second elimination as F, and the third as G eventually reaching $GFEA = U$. The beauty of this decomposition is, GFE is **always** a lower triangular matrix with '1's on the diagonal. This property is indifferent to the number of eliminations applied and to the eliminations themselves. It is this property that enables the LU decomposition. Now

$$GFEA = U \quad \text{implies} \quad A = E^{-1}F^{-1}G^{-1}U,$$

so that

$$A = LU, \quad \text{where } L = E^{-1}F^{-1}G^{-1} = (GFE)^{-1} \text{ is also a lower triangular matrix.}$$

Thus we have, $A = LU$ where $L = (GFE)^{-1}$ is lower triangular with '1's on the diagonal and U is upper triangular. In this example, we have

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

The questions remains, what is the purpose of decomposing a matrix? What additional information does it give us? In the LU Decomposition, one main advantage is that since both L and U are triangular matrices, numerical calculations become much simpler and for large matrices it can reduce the running time of any algorithm.

Singular Value Decomposition

This article focuses mainly on the Singular Value Decomposition(SVD). Before we start describing it, we would need a few definitions to guide us to SVD (also refer [1]).

Orthogonal Matrix: A matrix Q is orthogonal if, $Q^TQ = I$. The columns of an orthogonal matrix are orthonormal to each other.
Diagonal Matrix: A diagonal matrix is a matrix with 0's everywhere but the diagonal.
Definition of SVD: Any m by n matrix A can be factored into $A = U\Sigma V^T$ where U and V are orthogonal matrices and Σ is a diagonal matrix.

- The columns of U (m by m) are eigenvectors of AA^T.
- The columns of V (n by n) are eigenvectors of A^TA.
- The σ singular values on the diagonal of Σ (m by n) are the square roots of the nonzero eigenvalues of both AA^T and A^TA.

Constructing the SVD of a Matrix

The proof of SVD involves a chain of lemmas and theorems that need to be introduced to the undergraduate audience. It is difficult to present the whole proof here, but I present the method for constructing the SVD for any given matrix.
For this purpose, assume that given any matrix A, its SVD is $U\Sigma V^T \Rightarrow AV = \Sigma U$, where U, $V$and Σ are appropriate. I present the following steps to achieve this:

- For simplicity lets consider a 2×3 matrix A. We find the symmetric matrix A^TA and its eigenvalues. Consider the following matrix:

  $$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

  This implies

  $$A^TA = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

  The eigenvalues of A^TA are $\lambda = 3, 1, 0$. Now as the diagonal entries in Σ are square roots of the nonzero eigenvalues of A^TA, $\sigma_1 = \sqrt{3}$, $\sigma_2 = \sqrt{1} = 1$. We get

  $$\Sigma = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

- Next we find the eigenvectors corresponding to the eigenvalues of A^TA. These eigenvectors will make up the columns of V. We have to keep in mind though, that V has to be an orthogonal matrix. Since A^TA is symmetric, we can use the following theorem to find such a V.

  Theorem. *It is always possible to find an orthogonal matrix of eigenvectors for a symmetric matrix, by using the Gram-Schmidt algorithm.*

  Using the Gram-Schmidt process, the eigenvectors corresponding to eigenvalues 3, 1 and 0 are

  $$v_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad v_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{respectively.}$$

  We get

  $$V^T = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

- Our only task now is to find U. Technically we can find U by finding a set of orthonormal eigenvectors of AA^T. This process however, can involve some computation. Instead we use the information we already have. Since

$$AV = \Sigma U \Rightarrow Av_j = \sigma_j u_j \Rightarrow \frac{Av_j}{\sigma_j} = u_j.$$

  In our example we get $u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, giving
  $U = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$.
- Finally we get

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

Applications of SVD

The applications of SVD are innumerous, and ever increasing. The following are some of its advantages that make the SVD so useful [2]:

- The fact that the decomposition is achieved by orthogonal matrices makes it an ideal vehicle for discussing the geometry of n-space.
- It is stable; small perturbations in A correspond to small perturbations in Σ, and conversely.
- The diagonality of Σ makes it easy to determine when A is near to a rank-degenerate matrix.
- Thanks to the pioneering efforts of Gene Golub, there exist efficient, stable algorithms to compute the singular value decomposition.

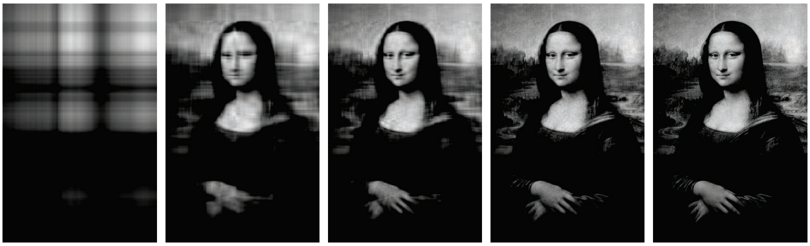
Image Processing. Suppose a satellite takes a picture, and wants to send it to Earth [1]. The picture contains (say) 1000 by 1000 pixels, each with a definite color. We can code the colors into numbers such that each pixel is a number and the whole picture is a 1000 by 1000 matrix. To send this picture to earth, we will need to store $1000 \times 1000 = 1$ million numbers in a matrix and send it. It is better to find the essential information inside the 1000 by 1000 matrix, and send only that. This is where SVD comes in.

Suppose we know the SVD of the matrix. The key is in the singular values (in Σ). Typically, some σ's are significant and others are extremely small. If we keep 20 and throw away 980, then we send only the corresponding 20 columns of U and V. The other 980 columns are multiplied in $U\Sigma V^T$ by the small σ's that are being ignored.

If only 20 terms are kept, we send 20×2000 numbers instead of a million (25 to 1 compression). The pictures are really striking, as more and more singular values are included.

At first you see nothing, and suddenly you recognize everything. The cost is in computing the SVD this has become much more efficient, but it is expensive for a big matrix.

The pictures below show the difference in sharpness with increasing singular values. The last one is the original 600×386 image.



The first one is with one singular value, the second with ten singular values, the third with twenty and the fourth with fifty singular values. Notice that there is hardly any difference between the original and fourth image. Even the one with twenty singular values is as good as the original.

References

[1] Gilbert Strang, *Linear Algebra and Its Applications*, Cengage Learning, India, 2011, 4th Edition.

[2] G.W. Stewart, *On the Early History of the Singular Value Decomposition*, Society for Industrial and Applied Mathematics Review, Vol 35, No.4, Dec. 1993, pp. 551-566.

Dootika Vats, Graduate Student, Department of Statistics,, Rutgers University, U.S.A.
E-mail address: dootika1990@gmail.com

Interdisciplinary Aspects of Mathematics

Mathematics is everywhere around us. The applications of mathematics are diverse and innumerable. This section highlights such concepts of mathematics that are used extensively in day to day life. The practical aspect of mathematics, often overlooked, is what we seek to show in the following section.

EXTREME VALUE DISTRIBUTIONS

DR. ANURADHA
AND
A. NUPUR, E. ANURADHA, G. ADITI, T. NUPUR

Abstract. The distinguishing feature of an extreme value analysis is the objective to quantify the stochastic behaviour of a process at unusually large or small levels. This paper delves into the study of Extreme Value distributions with major emphasis on the application of three distributions namely - Gumbel, Frechet and Weibull. These distributions have extensive applications in the field of Geology, Finance, Earth Sciences and related fields.

Introduction

Extreme value theory (EVT) is the study of the maximum of a random sample of observations, as the sample size grows. Historically, one of the important applications of EVT was in determining the height of the dikes in Netherlands. The government wanted the height to be such that the probability of a flood in a particular year is at most 10^{-4}, or in other words, in the next 10,000 years, it should be unlikely that the sea will overflow the dikes. However, the data available at that point was just over 100 years. Classical statistical methods, for example looking at the empirical distribution, were insufficient for the following simple reason. If one believes that the empirical distribution faithfully approximates the true distribution, then it follows that the maximum water level ever is just the maximum water level in the last 100 years, which clearly is misleading. Thus, for this particular problem, one needs a way to extrapolate out of the range of the data, and that is precisely where EVT helps us. What EVT essentially does is the following. It assumes a semiparametric model, and fits that into the data by estimating some tail parameters. Once the tail parameters are estimated, precise statistical predictions can be made about the maximum of a certain number of observations, even when that number is much larger than the available sample size.

Suppose that $X_1, X_2, \ldots$ are independent identically distributed (i.i.d.) random variables. The central question in EVT is whether or not there exist sequences of constants $a_n > 0$ and b_n such that

$$\frac{\max(X_1, X_2, \ldots, X_n) - b_n}{a_n} \tag{1}$$

has a non-degenerate limit as $n \to \infty$, and if yes, what is that limit? Fisher, Tippet and Gnedenko have characterized the class of distributions which can possibly be a limit. A cumulative distribution function (c.d.f.) is a possible limit if and only if it is of the form $G_\gamma(ax + b)$ with $a > 0$, b real, where

$$G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \ 1 + \gamma x > 0\,,$$

with γ real, the right hand side to be interpreted as $\exp(-\exp(-x))$ when $\gamma = 0$. The parameter γ is called the extreme value index. If the quantity in (1) converges to G_γ, then the distribution of $X_1, X_2, \ldots$ is said to belong to the maximal domain of attraction of G_γ. In EVT, one assumes that the distribution underlying a particular sample is in the maximal domain of attraction of some G_γ, and the statistical task is to estimate γ.

Applications of extreme value theory include predicting the probability distribution of catastrophe models, climatology, extreme floods, amounts of large insurance losses, equity risks, day to day market risk, operational risk management, mutational events during evolution.

Model Formulation

Let $X_1, ..., X_n$, be a sequence of independent random variables having a common distribution function F. Here X_i may denote values of the process measured on a regular time scale. For example hourly measurements of sea-level, daily mean temperatures etc. Then the model focuses on the statistical behavior of

$$M_n = \max\{X_1, X_2, \ldots, X_n\}.$$

Here M_n represents the maximum of the process over n time units of observation. Now,

$$P\{M_n \le z\} = P\{X_1 \le z, \ldots, X_n \le z\} = P\{X_1 \le z\} \cdots P\{X_n \le z\} = \{F(z)\}^n. \quad (2)$$

However, this is not immediately helpful in practice, since the distribution function F is unknown. One possibility is to use standard statistical techniques to estimate F from the observed data, and then to substitute this estimate into (2). Unfortunately, very small errors in the estimate of F can lead to substantial errors in F^n.

Alternatively, F^n can be approximated using families of distribution based on extreme data. This is similar to approximating the distribution of sample means by normal distribution using Central Limit Theorem.

We look at the behavior of F^n as $n \to \infty$, but this alone is not enough: for any $z < z_+$, where z_+ is the upper end-point of F, $F^n(z) \to 0$ as $n \to \infty$ so that the distribution of M_n degenerates to a point mass on z_+. This difficulty is avoided by allowing a linear renormalization of the variable M_n:

$$M_n^\star = \frac{M_n - b_n}{a_n}$$

for sequences of constants $\{a_n > 0\}$ and $\{b_n\}$. Appropriate choices of the $\{a_n\}$ and $\{b_n\}$ stabilize the location and scale of $M_n^\star$ as n increases, avoiding the difficulties that arise with the variable M_n.

Extreme Value Distributions

The limiting distribution of the normalised variable $M_n^\star$ is illustrated in the following theorem:

Theorem 1. *If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that*

$$P\left\{\frac{(M_n - b_n)}{a_n} \leq z\right\} \to G(z) \text{ as } n \to \infty,$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$G(z) = \exp\left\{\exp[-(\frac{(z-b)}{a})]\right\}, -\infty < z < \infty \tag{3}$$

$$G(z) = \begin{cases} 0 & \text{if } z \leq b \\ \exp[-(\frac{(z-b)}{a})^{-\alpha}] & \text{if } z > b \end{cases} \tag{4}$$

$$G(z) = \begin{cases} \exp[-(-(\frac{(z-b)}{a}))^{\alpha}] & \text{if } z < b \\ 1 & \text{if } z \geq b \end{cases} \tag{5}$$

for parameters $a > 0$, b and, in the case of families (4) and (5), $\alpha > 0$.

Collectively, these three classes of distribution are termed as the extreme value distributions, with types (3), (4) and (5) widely known as the Gumbel, Frechet and Weibull families respectively. Each family has a location and scale parameter, b and a respectively; additionally, the Frechet and Weibull families have a shape parameter α.

The probability distribution functions and cumulative distribution functions of the three types of extreme value distributions are shown by the following figures.

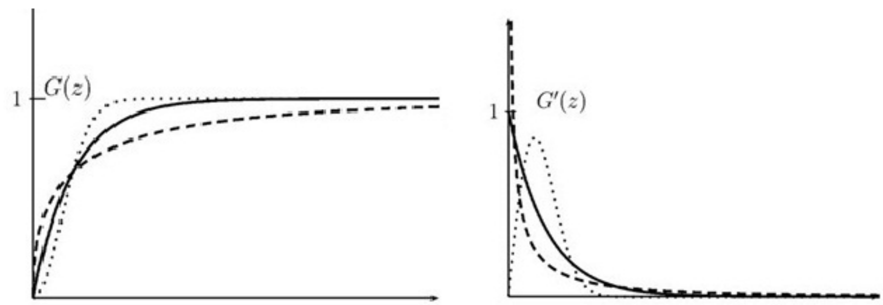


Figure 1. Weibull Distribution - cumulative distribution and probability density function

Generalised Extreme Value Distribution

A better analysis is offered by a reformulation of these models. The Gumbel, Frechet and Weibull families can be combined into a single family of models having distribution functions of the form $G(z)$ shown by Theorem 2.

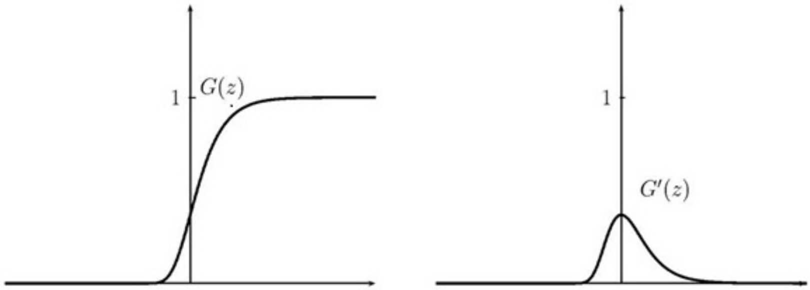


Figure 2. Gumbel Distribution - cumulative distribution and probability density function

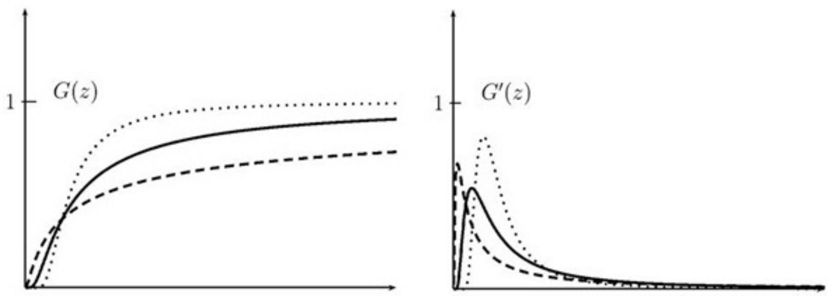


Figure 3. Frechet Distribution - cumulative distribution and probability density function

Theorem 2. *If there exist sequences of constants* $a_n > 0$ *and* b_n *such that*

$$P\{(M_n - b_n)/a_n \leq z\} \to G(z) \text{ as } n \to \infty$$

for a non-degenerate distribution function G, then the function G is a member of the GEV family:

$$G(z) = \exp\{-[1 + \xi(z-\mu)/\sigma\ ]^{-1/\xi}\} \tag{6}$$

defined on,

$$\{z : 1 + (z-\mu)/\sigma > 0\}, \text{ where } -\infty < \mu < \infty, \sigma > 0 \text{ and } -\infty < \xi < \infty.$$

This is known as the generalized extreme value (GEV) family of distributions. The Frechet and Weibull classes of extreme value distribution correspond respectively to the cases $\xi > 0$ and $\xi < 0$ in this parameterization. The subset of the GEV family with $\xi = 0$ is interpreted as the limit of GEV as $\xi \to 0$, leading to the Gumbel family.

By making suitable inferences on ξ, an appropriate type of tail behavior can be determined and there is no necessity to make apriori judgement about which individual extreme value family to adopt.

The above theorem leads to the following approach for modeling extremes of a series of independent observations $X_1, X_2, \ldots$. Data are blocked into sequences of observations of length n, for some large value of n, generating a series of block maxima, $M_{n,1}, \ldots, M_{n,m}$ say, to which the GEV distribution can be fitted. These blocks are chosen to correspond to a time period of length one year, in which case n is the number of observations in a year and the block maxima are annual maxima.

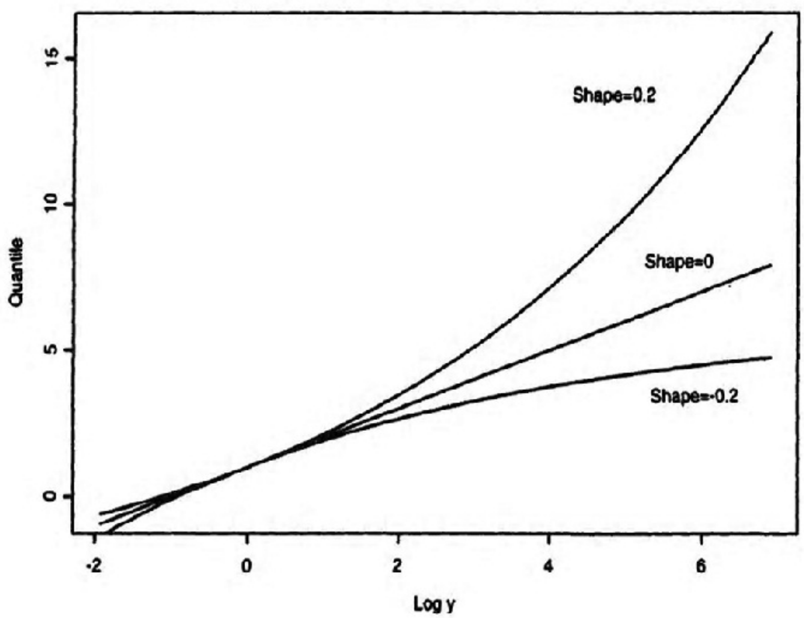
Return Level associated with Return period $1/p$

By inverting the equation (6), estimates of extreme quantiles of the annual maximum distribution are obtained as:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{for } \xi = 0. \end{cases}$$

Here, z_p is the return level associated with the return period $1/p$. It is a statistical measurement denoting the average recurrence interval over an extended period of time, and is usually required for risk analysis and also to dimension structures so that they are capable of withstanding an event of a certain return period (with its associated intensity).This imples that the return level z_p is expected to be exceeded on average once every $1/p$ years. More precisely, z_p is exceeded by the annual maximum in any particular year with probability p.

Let $y_p = -\log(1-p)$. Then, if z_p is plotted against y_p on a logarithmic scale - or equivalently, if z_p is plotted against $\log y_p$, then we obtain the return level plot. If $\xi = 0$, then the plot is linear. If $\xi > 0$ then the plot is concave and has no finite bound where as if $\xi < 0$ the plot is convex.



The return level plots are particularly helpful for both model presentation and validation thereby giving us an indication on the type of limiting distribution being followed.

Definition 1. A distribution G is said to be max-stable if, for every $n = 2, 3, \ldots$ there are constants $a_n > 0$ and β_n such that

$$G^n(a_n z + \beta_n) = G(z).$$

Since G_n is the distribution function of $M_n = \max\{X_1, X_2, \ldots, X_n\}$, where X_i's are independent variables each having distribution function G, max-stability is a property satisfied by distributions for which the operation of taking sample maxima leads to an identical distribution, apart from a change of scale and location.

The connection with the extreme value limit laws is made by the following theorem which states that

Theorem 3. *A distribution is max-stable if, and only if, it is a generalized extreme value distribution.*

Example 1.1 Suppose that $X_1, X_2, \ldots$ is an i.i.d. sequence of standard exponential random variables with parameter 1. Then, the cumulative distribution function is given by

$$F(x) = 1 - \exp(-x) \text{ for } x > 0.$$

In this case, letting $a_n = 1$ and $b_n = \log n$, we have

$$\begin{aligned} P\left\{(M_n - b_n)/a_n \le z\right\} &= F^n(z + \log n) = [1 - \exp(-(z + \log n))]^n \\ &= [1 - \exp(-z)/n]^n \to \exp(-\exp(-z)) \end{aligned}$$

as $n \to \infty$, for each fixed $z \in \mathbb{R}$. Hence, with the chosen a_n and b_n, the limit distribution of M_n as $n \to \infty$ is the Gumbel distribution, corresponding to $\xi = 0$ in the GEV family.

Example 1.2 Suppose that $X_1, X_2, \ldots$ is an i.i.d. sequence of standard Frechet random variables with pdf:

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \exp(-1/x) & \text{if } x > 0. \end{cases}$$

In this case, letting $a_n = n$ and $b_n = 0$, we have

$$P\left\{(M_n - b_n)/a_n \le z\right\} = F^n(nz) = [\exp\{-1/(nz)\}]^n \to \exp(-1/z)$$

as $n \to \infty$, for each fixed $z > 0$. Hence, the limit in this case which is an exact result for all n, because of the max-stability of F is also the standard Frechet distribution with $\xi = 1$ in the GEV family.

Example 1.3 Suppose that $X_1, X_2, \ldots$ is an i.i.d. sequence of uniform(0, 1) random variables with pdf:

$$F(x) = \max\{0, \min\{1, x\}\} \text{ for } x \in \mathbb{R}.$$

For fixed $z < 0$, suppose $n > -z$, let $a_n = 1/n$ and $b_n = 1$, we have

$$P\{(M_n - b_n)/a_n \le z\} = F^n\left[1 + \frac{z}{n}\right] = \left[1 + \frac{z}{n}\right]^n \to \exp(z)$$

as $n \to \infty$. Hence, the limit in this case is of Wiebull type with $\xi = -1$ in the GEV family.

Asymptotic Models For Minima

On the contrary, sometimes the strength of the system is dependent on the strength of its weakest component which leads us to the weakest link principle. Extreme value statistics can be employed to obtain an approximation to the statistical behavior of this weakest link, providing a plausible model for the statistical properties of system failure.

Some applications require models for extremely small, rather than extremely large observations. The overall system lifetime is then,

$$M'_n = \min\{X_1, X_2, \dots, X_n\}$$

where the X_i are independent and identically distributed random variables having a common distribution F. Analogous arguments apply to M'_n as were applied to M_n, leading to a limiting distribution of a suitably re-scaled variable.

The results are also immediate from the corresponding results for M_n. Letting Y_i = -X_i for $i = 1, 2, \dots, n$, the change of sign means that small values of X_i correspond to large values of Y_i. So if $M'_n = \min\{X_1, X_2, \dots, X_n\}$ and $M_n = \max\{Y_1, Y_2, \dots, Y_n\}$, then $M'_n =$ -M_n. Hence, for large n,

$$\begin{aligned} P\{M'_n \le z\} = P\{-M_n \le z\} &= P\{M_n \ge -z\} = 1 - Pr\{M_n \le -z\} \\ &\approx 1 - \exp\left\{-[1 + \xi(\frac{-z-\mu}{\sigma})]^{-1/\xi}\right\} \\ &\approx 1 - \exp\left\{-[1 + \xi(\frac{-(z-\mu')}{\sigma})]^{-1/\xi}\right\} \end{aligned}$$

on the set $\left\{z : 1 - \xi(\frac{z-\mu'}{\sigma}) > 0\right\}$ where μ' = -μ. This distribution is the GEV distribution for minima.

References

1. OHEN, J. P., *Fitting extreme value distributions to maxima, Sankyha*, Series A, 1988, 50, 74-97.
2. OLES, S. G., HEFFERNAN, J. and TAWN, J. A., *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001, 45-73.
3. UMBEL, E. J., *Statistics of Extreme*, Columbia University Press, New York, 1958.
4. UMBEL, E. J., *Distributions de valeurs extremes en plusieurs dimensions*, Publications of the Institute of Statistics of the University of Paris 9, 1960, 171-173.

Dr. Anuradha, Associate Professor, Department of Statistics, Lady Shri Ram College For Women

Aditi Garg, Anuradha Eswaran, Nupur Aggarwal and Nupur Thirani, B.Sc.(H) Statistics, Lady Shri Ram College For Women

GRAPH THEORY AND FOOTBALL

GAURANGNA MADAN AND SURABHI KHANNA

Abstract. Our paper aims to analyse a football match using graph theory. We see that a game of football can be viewed as a directed network with the players acting as vertices and the passes made between them serving as arcs. We aim to analyse a match and also comment on the importance of players in the network and the strategy of play.

Introduction

Football is one of the most popular sports played as well as followed worldwide. Two teams of eleven players compete with each other to score the maximum number of goals in the course of ninety minutes (plus extra time, the amount of which depends on the format), which is measured by the number of times the ball is kicked into the opposing teams net.

A football team consists of exactly 1 goalkeeper, 3-5 defenders, 3-5 midfielders and 1-3 strikers or forwards. The number of players in each position varies with different teams and their respective strategies. A football match lasts for a regulated period of ninety minutes, plus extra time to account for the time lost due the ball going out of play during the game. The three possible outcomes for a team involved in a football match are a win, a loss and a draw. In case of a competitive final, where the winner needs to be decided on the same day itself, a match that ends in a draw is followed by half an hour of extra time. In case both the teams are still drawing at the end of extra time, the match is followed by penalties. The world cup final of 2010, being played between the Netherlands and Spain ended in a draw after the regulatory period of 90 minutes were up, and so followed two halves of extra time. Spain defeated the Netherlands by a solitary goal in extra time, scored by their pragmatic midfielder, Andres Iniesta.

Our paper is based on the observation that a football match can be viewed as an analogue to a directed network in graph theory.

A **directed network** is a digraph with an integer weight attached to each arc. In the case of a football match we consider the players of the team as **VERTICES** and the passes exchanged between the players as **ARCS**. The direction of the arc will be in accordance with the direction of the pass, i.e., it is directed outwards if the player is passing to another player and inwards if the player receives a pass. We assume the weight of each arc to be 1.

It is important to note that the digraph which we are studying is a CONNECTED GRAPH, i.e., there exists at least one path between all vertices of the graph. This is due to the fact that during the course of the match, it is inevitable that each player would pass the ball to every one of his team members at least once.

Centrality

Centrality is used to measure the importance of a vertex to a digraph. We shall use this concept to assess the importance of a player to a team. Centrality helps us understand how removing a vertex impacts a network. Analogously, it helps us understand the implications of when a player is removed from a team. This scenario can arise for example, when a player gets a red card and is sent off from the game. We watched it happen in the 2006 Fifa World Cup Final, when the French captain Zinedine Zidane was sent off after a red card. Losing a player with high centrality impacted the game and the French ultimately ended up on the losing side.

The simplest of the centrality measures is the **degree centrality**, i.e., the number of arcs incident with a vertex. This concept can be extended to a player by counting the number of passes he is involved in. Logically, when a player has to pass to another player, he would wish to pass to a player who would help achieve the aim of the game better, i.e., to score a goal. So clearly a person crucial to game strategy will receive more passes and hence will have higher centrality.

Now, if a vertex is removed from a network, all the arcs incident to it also get removed from the network. Correspondingly, if a player gets removed from a team then all the passes associated with him also get removed. Thus, the number of paths that the ball can travel also reduces accordingly. So it follows that if a player who makes a large number of successful passes gets removed, the number of possible paths the ball can travel decreases significantly leading to a reduction in the probability of scoring a goal.

So, strategically if the team members with high centrality get eliminated from play, either through injury or substitution or the misfortune of a red card, it will impact the goal scoring abilities of the team and thus lower their chances of winning.

For our analysis, we are choosing the example of the 2010 Fifa World Cup Final held in Johannesburg, South Africa, that took place between Holland and Spain. We shall analyse the outcome of the match and try to determine the various factors that led to it.

Centrality Tables

Jersey Number	Player Name	Position	Centrality
1	Stekelenburg	Goalkeeper	48
2	Van Der Wiel	Defender	31
3	Heitinga	Defender	27
4	Mathijsen	Defender	33
5	Van Bronckhorst	Defender	29
6	Van Bommel	Midfielder	29
7	Dirk Kuyt	Forward	11
8	Nigel de Jong	Midfielder	32
9	Robin Van Persie	Forward	19
10	Wesley Sneijder	Midfielder	35
11	Arjen Robben	Forward	14
15	Braafheid(substitute)	Defender	4
17	Elia(substitute)	Forward	4
23	Rafael Van Der Vaart (substitute)	Midfielder	7
		MEAN	23.07142857

CENTRALITY TABLE FOR THE NETHERLANDS

Jersey Number	Player Name	Position	Centrality
1	Iker Casillas	Goalkeeper	31
3	Gerard Pique	Defender	51
5	Carlos Puyol	Defender	54
6	Andres Iniesta	Midfielder	43
7	David Villa	Forward	13
8	Xavi	Midfielder	95
11	Capdevila	Defender	53
14	Xabi Alonso	Midfielder	47
15	Sergio Ramos	Defender	54
16	Sergio Busquets	Midfielder	75
18	Pedro	Forward	13
9	Fernando Torres(substitute)	Forward	2
10	Cesc Fabregas(substitute)	Midfielder	34
22	Jesus Navas(substitute)	Forward	17
		MEAN	41.57142857

CENTRALITY TABLE FOR SPAIN

Observations from Centrality tables:

- Since the Spanish players have comparably higher degrees, i.e., there is a lot of passing between all the different players, so they have a balanced network.
- The Spanish team on an average has centrality 41 which is much higher than the Dutch team's 23. This means that the Dutch attack is straightforward whereas the Spanish play is more intricate.
- The higher number of passes amongst the Spanish team members suggests more fluid circulation of the ball.
- For the Dutch team, we observe that the centrality is concentrated to a few players. This means that the Dutch attack can be adversely affected by blocking/covering/marking these players.

Centre of the Graph

The **centre of a graph** is defined as the set of those vertices whose **eccentricity equals the graph radius** (in this case digraph). **Eccentricity** of a vertex is defined as the maximum distance of that vertex, say v, of the digraph from any other vertex, say w, of the digraph. **Graph radius** is defined as the minimum distance between any two vertices of the graph. Since we assumed the weight of each arc to be 1, the graph radius is 1. So the centre will be the vertex whose eccentricity is one. And it follows that the centre will include all those vertices whose distance to the other vertices in the digraph is 1.

For a game of football, the centre will be the player whose distance from the other players is minimum (here assumed as one, given the assumption that every player will pass to his teammate at least once during the game, thereby making the minimum number of passes required by him to reach the player as one). Here we have considered only the successful passes made by the players, i.e., the passes that were involved in a build-up for a goal or a move that leads to a goal-scoring chance. By observing the number of passes that each player of either side makes that is involved in an attack, we make the following conclusions:

(1) Centre for Spain: Xavi, Iniesta and Cesc Fabregas
(2) Centre for Holland: Sneijder, Robben

It is important to note that a high centrality need not translate into the vertex being in the centre. This is illustrated by the fact that the Dutch goalkeeper, Maarten Stekelenburg has a centrality of 48, but obviously he is not the centre as Holland's goal-scoring attempts do not always go through him. We observe that the centre shall appear more in goal scoring attempts.

Graphical Analysis of Goal-Scoring Opportunities

We define a successful pass as one that was involved in an attack. We selectively mapped all of these, i.e., the passes that resulted in shots on target to get a better understanding

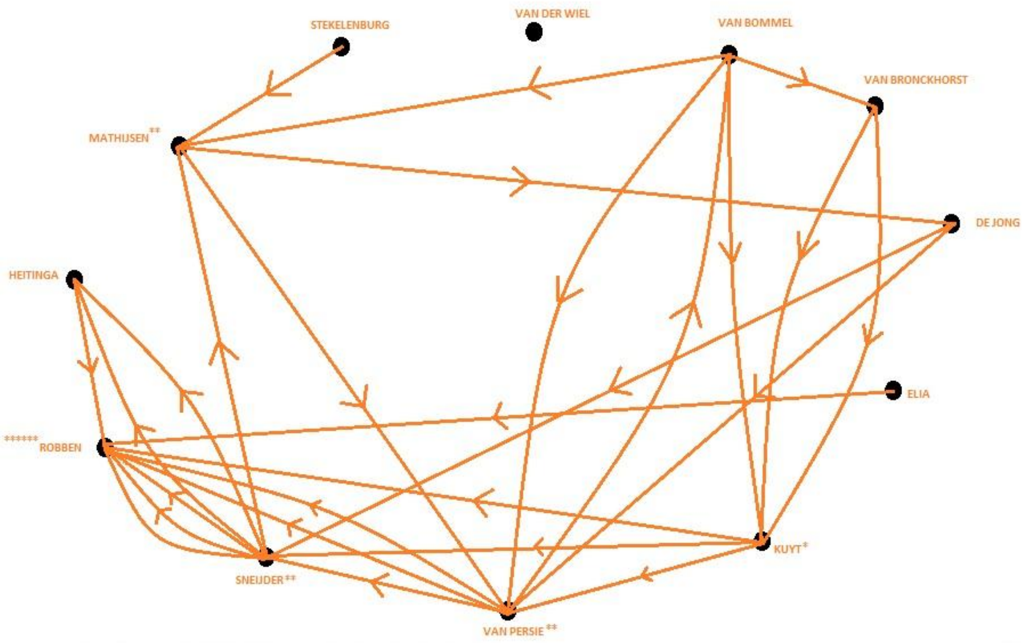


FIGURE 1. MAPPING THE DUTCH ATTACK: Graph illustrating the various moves of the Dutch team that resulted in goal-scoring oppurtunities

of goal scoring strategy. We also marked all the players who had shots on target with the symbol * and used the symbol @ for a goal. We have indicated the frequency of this occurrence. From the given graphs, we can clearly observe the different strategies being played out, which have been summarised as follows:

(1) Clearly, as can be observed from the graphs, the centres for Spain are Xavi, Iniesta and Fabregas; and the centres for Netherlands are Sneijder and Robben.
(2) The passing statistics reveal that the Spanish players make an extremely large number of passes during the game, much more than their Dutch counterparts. This is illustrated by the high network density in the Spanish team's graph, while the moderate network density in the Dutch team's graph.
(3) The Spanish attack is often unpredictable because of the huge number of passing outlets. For instance, we observe that the Spanish defender Sergio Ramos is at the heart of many attacks, and even has a few shots on goal to his name. This indicates the high mobility and volatility of the Spanish attack, which doesnot depend on just one creative outfit to help create goals.
(4) On the other hand, the Dutch attack seems to be much more traditional, with most attacks being carried out by the strikers. This is seen by the high density of passes

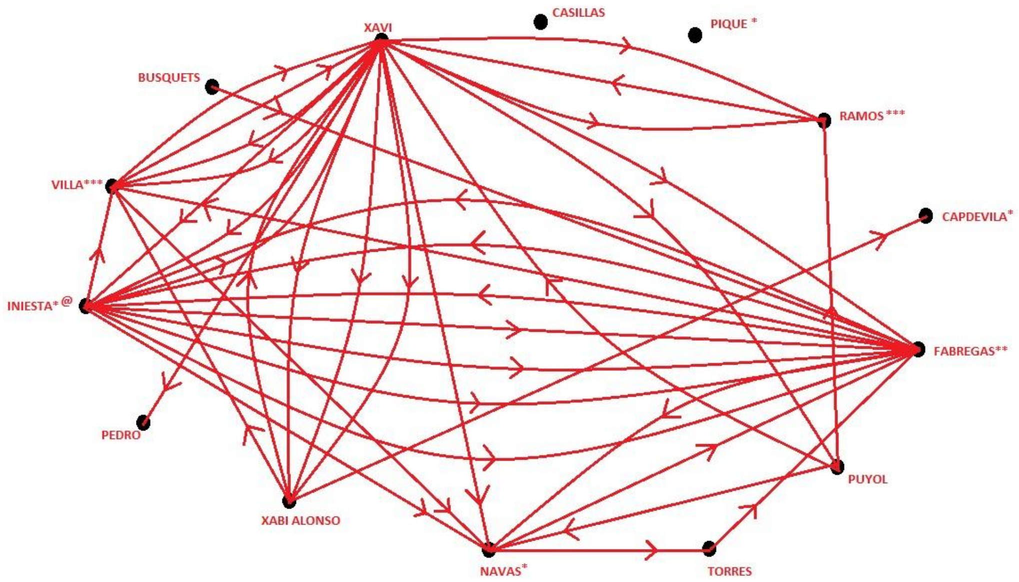


Figure 2. MAPPING THE SPANISH ATTACK: Graph illustrating the various moves of the Spanish team that resulted in goal-scoring oppurtunities

Due to the large amount of data involved, we only could map the number of passes that were involved in the build up of an attack. The Spanish team passed the ball around a total of 715 times, while the Dutch made a total of 475 passes. Mapping each one of those was proving to be impossible!

and shots on goal concentrated around their strike force. This is also supported by the fact that the players in the centre of the graph are all strikers.

(5) The Spanish attack relies on swift passes between the players, which are evenly distributed among the midfield. This gives them the element of surprise in attack, which leads to more shots on goal. This is supported by the fact that the midfielders Xavi and Iniesta are part of the centre.
(6) The low number of arcs in the Dutch team suggests a preference for quick attacks and counter attacks.

Graphical Analysis of the Goal

The world cup winning goal was scored by the Spanish midfielder, Andres Iniesta, who was set up by his midfield compatriot, Cesc Fabregas. The following graph shows the various stages involved in the build-up to the goal. We observe that the centres, namely Iniesta and Fabregas, were central to the winning goal. We see that the move was started by the

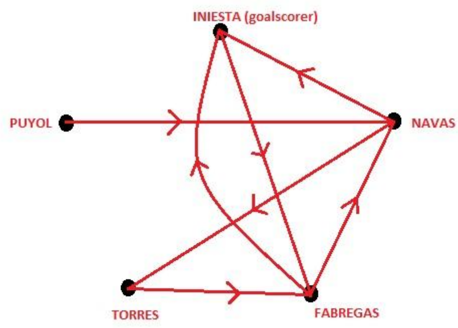


Figure 3. MAPPING THE GOAL: Puyol-Navas-Iniesta-Fabregas-Navas-Torres-Fabregas-Iniesta-World Cup Winning Goal

defender Puyol and it found its way in the back of the net after passing through several attacking players from the Spanish side. This reinforces our claim that the Spanish attack having several outlets, was able to out-manoeuvre it's opponent. Not depending on a lone attacking option and distributing the ball evenly among the entire team gave them the edge over the Dutch and enabled the Spanish to a victory.

Conclusion

By using graph theory to map the football world cup final, we were able to determine (as far we could!) the primary reasons for the Spanish triumph. Their well-oiled network of connected players made for a sound defence, a cohesive midfield, and an effective attack. The Dutch lost because their strategy was predictable and easily countered by the Spanish. They should have considered passing more and involving players other than their attacking midfielders and strikers in their goal scoring strategy. This would have led to comparable player centralities, and hence resulted in a balanced directed network. Similarly, the outcome of the game could have been different had they succeeded in blocking the players who belonged to the centre of the Spanish network, namely Xavi, Iniesta and Cesc Fabregas.

Extension

This study can be extended to sports such as Basketball, Hockey, American Football, Rugby, etc. which have similar playing strategies.

References

1. Edgar G. Goodaire and Micheal M. Parmenter, *Discrete Mathematics with Graph Theory*, Third Edition, Pearson Education, 2008.
2. plus.maths.org/content/os/latestnews/may-aug10/football/index
3. www.fifa.com/live/competitions/worldcup/matchday=25/day=1/math=300061509/index.html
4. www.fifa.com/worldcup/archive/southafrica2010/matches/round=249721/match=300061509/index.html

5. en.wikipedia.org/wiki/Association_football
6. www.answers.com/topic/soccer

Gaurangna Madan, B.Sc.(H) Mathematics, 3rd year, Lady Shri Ram College For Women
E-mail address: gaurangna.madan@gmail.com

Surabhi Khanna, B.Sc.(H) Mathematics, 3rd year, Lady Shri Ram College For Women
E-mail address: surabhi.khanna9@gmail.com