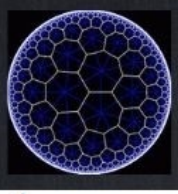
Blackboard

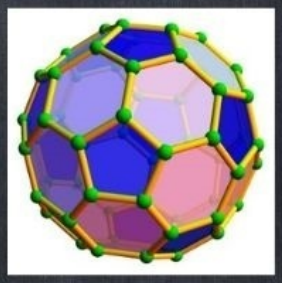


Issue 5

MTA (I)

$$\sum_n \frac{(-1)^n}{n^n} = \int_0^1 x^x\,dx$$

$$\sum_n \frac{1}{n^n} = \int_0^1 \frac{dx}{x^x}$$



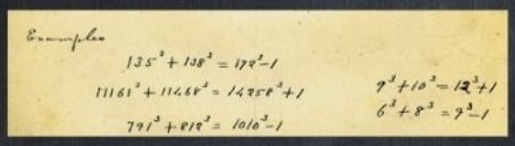
Here is the Ramanujan-Hardy formula for the calculation of the number of partitions:

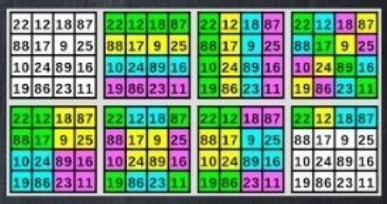
$$p(n) = \frac{1}{2\sqrt{2}} \sum_{k=1}^{v} \sqrt{k}\, A_k(n) \frac{d}{dn} \exp\left(\pi\sqrt{\frac{2}{3}}\frac{\sqrt{n-\frac{1}{24}}}{k}\right)$$

where

$$A_k(n) = \sum_{0 \le m < k;\ (m,k)=1} e^{\pi i\left[s(m,k) - \frac{1}{k}2nm\right]}$$

$$1/2+1/3+1/7+1/43+1/1807+\ \cdots = 1$$





$$e^{\pi\sqrt{163}} = 262537412640768743.99999999999992\ldots$$

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{6 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{8 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{10 + \cdots}}}}}}}}}}}}}$$

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathbb{Z}^2 & \xrightarrow{\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}} & 3\mathbb{Z} \oplus 2\mathbb{Z} & \longrightarrow & \\ \downarrow & & \downarrow{\scriptstyle \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 2 \end{pmatrix}} & & \downarrow & & \\ \mathbb{B}_d^a & \longrightarrow & \mathbb{Z}^4 & \xrightarrow{\begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}} & \mathbb{Z}^2 & \longrightarrow & \end{array}$$

The aim of *Blackboard*, the Bulletin of the Mathematics Teachers' Association (India), is to promote interest in mathematics at various levels and to facilitate teachers in providing a well-rounded mathematical education to their students, in curricular as well as extra-curricular aspects. The Bulletin also serves as an interface between MTA (I) and the broad mathematical community.

Registered Office

Homi Bhabha Centre for Science Education
Tata Institute of Fundamental Research
V. N. Purav Marg, Mankhurd
Mumbai, 400088 INDIA

https://www.mtai.org.in/bulletin

Blackboard

Bulletin of the Mathematics Teachers' Association (India)
Issue 5

January 2023

Contents

Editorial 3

1 Report on the Fourth Annual Conference of the MTA, by Jonaki Ghosh 5

2 Historical Roots of Calculus – 2, by Shailesh Shirali 9

3 M.C. Escher's Print Gallery, by Muskan Choudhary & Kartav Kesri 19

4 Fair Allocation Using an Unfair Randomizer, by Jyotirmoy Sarkar & Mamunur Rashid 31

5 Going Beyond Turing: A Fractal Approach Towards Decision Making, by Mayukh Mukhopadhyay 55

6 Two Consecutive Positive Integers Cannot Both Be Perfect, by Shailesh Shirali 65

7 The Moore-Penrose Inverse and Its Applications, by J. K. Verma 69

Editorial

During the publication of the previous issue, we had planned to bring out the next issue within six months but it did not come to fruition due to several reasons.

The present issue also carries a diverse assortment of articles that should cater to a wide readership.

Shailesh Shirali continues with the second article of a series on the historical development of calculus. In the previous issue, his first article had noted that several mathematicians prior to Newton and Leibniz had already laid the essential groundwork. In that article, Shirali looked particularly at the work of Roberval, Descartes and Fermat, and noted that Fermat's approach was the closest in spirit to the modern idea of the derivative as a limit. In this second article, Shirali considers the corresponding questions in relation to integration. Some readers may find it surprising to discover that the origins of integral calculus go back to Archimedes, and hence, lie further back in time than those of differential calculus.

Muskan Choudhary and Kartav Kesri narrate a fascinating account of the so-called 'Print Gallery' of Escher which is known by a tongue-twisting name 'Prentententoonstelling'. The beautiful pictures are a hallmark of Escher's work some of which can be understood in the language of hyperbolic geometry. It is all the more remarkable to realize that Escher was not a trained mathematician.

Following that, we have an extremely interesting article by Jyotirmoy Sarkar and Mamunur Rashid. This article studies a very natural problem – that of providing fair allocation through an unfair randomizer. More precisely, the task is to equitably determine a winner from among $b \geq 2$ participants by using a randomizer that generates $a \geq 2$ values with unknown probabilities. After all, a fair toss is a utopian dream as the authors say! This article has a very lucid discussion and has the additional advantage of introducing the usage of the program R in implementing their algorithms to readers who are not very familiar with it.

In an extremely remarkable piece, Mayukh Mukhopadhyay takes us through the fractal approach to decision-making vis-a-vis the traditional Turing approach. The prospective applications of this approach are mind-boggling, ranging as they do from mental healthcare to security and business analytics. The readers would surely find it fascinating

to learn about the Geometric Musical Language process in comparison to the Turing machine approach.

Jugal Verma writes about pseudo-inverses of matrices, which is a topic that is not as well known as it deserves to be. Its applications abound, including an important one to the least squares method (a method that is supposed to have helped Gauss predict accurately the next appearance of the asteroid Ceres).

For the number theory afficionados, Shailesh Shirali has a short and sweet morsel on the nonexistence of consecutive perfect numbers.

Activities of the MTA include conducting the first level mathematical olympiad exam, and an annual conference. The fourth annual conference was conducted during September 9 to 11, 2022. It comprised of keynote addresses by eminent speakers and panel discussions as well as paper presentations by practising teachers. A nice report on this has been written by Jonaki Ghosh

MTA is also involved in other outreach programmes. One such was a 4-day training camp on mathematical problem-solving for mathematics teachers of classes 8,9,10,11,12 of schools in Karnataka. This was held at the International Centre for Theoretical Sciences (ICTS), Bengaluru from Nov 4 to Nov 7, 2022. The camp was attended by 22 teachers who hailed from different parts of Karnataka and there were 7 resource persons who discussed various topics from Algebra, Geometry, Number Theory and Combinatorics. The camp was enjoyed by both the teacher participants and the resource persons as most of the sessions were interactive.

Here is wishing all of the readers a very Happy New Year with beautiful Mathematical pursuits that elevate, inform and amuse!

— *B. Sury, Indian Statistical Institute Bangalore.*

1 Report on the Fourth Annual Conference of the MTA: Enhancing teacher preparedness for mathematics teaching and learning

Jonaki Ghosh

Department of Elementary Education
Lady Shri Ram College for Women
University of Delhi
New Delhi – 110024

Email: jonakibghosh@lsr.edu.in

The Fourth Annual Conference of the Mathematics Teachers' Association of India was held online during September 9 to 11, 2022.[1] The MTA annual conferences aim to provide a nationwide forum for mathematics teachers, mathematics educators, mathematicians, and individuals interested in mathematics education to come together and deliberate on various aspects of mathematics education. The discussions of the past three conferences have largely centered around policy, curriculum and pedagogic aspects at the school and college level. The second and third annual conferences, both held in the virtual mode, focused on the challenges in mathematics education unleashed by the pandemic and deliberated on ways to meet these challenges. Indeed, as educators we need to continue the discourse and attempt to develop a more nuanced understanding of what mathematics education is likely to be in the post pandemic times. Most of all, mathematics teachers at all levels (be it school or college) need to engage in a continuous dialogue to address these challenges and find feasible solutions. Undeniably, a higher level of engagement is expected from students after the pandemic, and it would be unrealistic not to scaffold this journey of learning recovery for teachers as well as students.

[1] The recordings of all the conference events are available on YouTube at https://youtube.com/playlist?list=PLIPKa8ExlqS9erOvraBo0frKkMPObJ2Y5.

Thus, acknowledging the need to reflect on several critical questions related to mathematics instruction and teacher preparedness for learning recovery, the overarching theme of the fourth annual conference was articulated as *Enhancing teacher preparedness for mathematics teaching and learning*. The following four sub-themes were identified to steer the discussions:

Theme 1: Scaffolding learning recovery in face-to-face mathematics classes.

Theme 2: Mathematical knowledge for teaching and teacher preparation.

Theme 3: Mathematics and computational thinking.

Theme 4: Mathematics laboratory activities: Opportunity for inquiry-based learning

The conference comprised keynote addresses by eminent speakers, panel discussions and paper presentations by practicing teachers. In the inaugural session, R. Ramanujam, President MTA(I), welcomed the participants and set the tone of the conference with his opening remarks. This was followed by an introduction to the overall theme of the conference by the program committee co-chairs.

The first sub-theme on *Scaffolding learning recovery in face-to-face mathematics classes* attempted to understand the coping mechanisms adopted by teachers and students as they moved from the online to offline classes and entered the recovery mode. A panel discussion on *Recovery and Resumption of Mathematics Learning in a Post-Pandemic Classroom* was organized where teachers and educators from across the country, teaching in schools and colleges, expressed their views. The deliberations focused on the methods adopted by teachers to help students cope with the challenges of offline classes after an extended period in online mode. The larger focus was on envisioning mathematics classrooms in the near future, the nature of resources that will be needed by teachers to ensure equitable access for learners and the role of digital technology as we go back to offline instruction.

The second sub-theme of the conference was *Mathematical knowledge for teaching and teacher preparation*. Much of the practice of teaching mathematics and how students develop the fundamental concepts depends on teachers' conceptual and pedagogical knowledge. This brings us to questions such as: What is the nature of mathematical and pedagogical knowledge that the teacher must be equipped with in times to come, especially after the pandemic? How can teacher preparedness be enhanced with a focus on developing learners' mathematical thinking?

Prof. Deborah Ball, University of Michigan, in her talk on *The work of teaching and the challenge of teacher education* demonstrated that teaching of mathematics is a common yet highly complex task. There is tremendous power in the professional work of teaching

and if it is not conducted properly it can perpetuate grave injustice. She explained, through examples of her own classroom videos, that teaching is dense with a multitude of discretionary spaces and reiterated the need to harness the power of this space. In fact her talk helped to unpack the complexity of mathematics teaching and left the audience with much to think about.

The third sub-theme of the conference was *Mathematics and computational thinking*. Computational Thinking (CT) has been identified as one of the key abilities to be acquired by children during their school years. While CT encompasses a broad skill-set applicable across contexts and domains, it is also intimately connected with mathematical thinking (MT). Thus, mathematics as a core school subject becomes a natural choice for integration of CT. Further the NEP 2020 in section 4.25 mentions that "mathematics and computational thinking will be given increased emphasis throughout the school years, starting with the foundational stage…" which posits questions such as: How can the integration of MT and CT be realised in the classroom? What kind of tasks should be designed which enable students to develop and employ CT practices and also foster MT at the same time?

The second keynote address on *Computation as a big idea in mathematics* by Dr. Weng Kin Ho, National Institute of Education, Singapore, was indeed an eye opener. He defined the *Big Ideas in Mathematics* and made a case for adding computation to the list of Big Ideas. According to him, the 21 big ideas have an overarching commonality, namely computation. He also elaborated on what it means to have a computational mindset and demonstrated through examples how teachers can encourage and develop CT in students. He further discussed the role of technology in developing CT and emphasised the role of spreadsheets in enabling students to create meaningful mathematical products.

One of the primary drawbacks of our present curriculum is the lack of opportunities for explorations and mathematical modelling activities. Setting up mathematics labs, both at school and undergraduate levels, where students engage in projects and meaningful explorations needs considered efforts and renewed emphasis. This aspect was highlighted in the fourth sub-theme of the conference – *Mathematics laboratory activities: Opportunity for inquiry-based learning*. Four presentations illustrating a range of activities that may be conducted with primary school students right up to the UG level were discussed. Nagesh Waiker demonstrated how readily available concrete materials can be used as manipulatives to help learners visualize concepts in algebra and the topic of integers. Jeenath Rahman focused on activities related to area and perimeter for middle school students. Swati Sircar shared a range of interesting activities for high school students from topics such as mensuration, square root spiral and fractals based on her work with students in the mathematics lab at Azim Premji university. Shantha Bhushan illustrated the power of games as math lab activities and also discussed a variety of exploratory tasks which she has integrated in the UG math curriculum. Jonaki Ghosh reiterated the importance of including explorations and projects as a part of mathematics lab activities and cited a few examples of projects done by school students.

It has been a tradition of the MTA conferences to invite a renowned mathematician, who has been a pioneer in a specialized mathematical field, to deliver a talk. Prof. Ram Murty, Queen's University, Canada, enlightened the participants on the topic *Brain networks and graph theory*. Prof. Ram Murty spoke on visualizing the brain as a neuronal network, as a directed graph with the nodes being either neurons or regions of the brain and the edge connectivity matrix representing communication pathways between them. He presented a method to identify "brain hubs" that are often implicated in various brain disorders such as epilepsy. Using elementary methods from graph theory, linear algebra and Markov processes, he showed exciting new applications of graph theory in neuroscience.

The conference could not be complete without a discussion on the National Education Policy and its implications for mathematics education. This was the focus of the panel discussion on *Change in the Education Policy Landscape*. Prof. Saumen Chattopadhyay, in his overview of NEP 2020, suggested that it was a market-based approach to education reform. He brought out the approach that NEP is formulating its aims from governance structures planned and the many questions that arise for us to deal with as educators. Prof. Farida Khan made the point that the NEP does not seem rooted at the ground level and focused on changes in school education as a result of the NEP. She emphasized that the teaching fraternity could and should use NEP to demand that various promises in terms of investment in infrastructure, space and education be kept. Prof. Amber Habib talked about the inconsistencies in many of the education policies. He pointed out various aspects of the CBCS, and later to NEP 2020 and their implications in mathematics education. He also presented an overview of the draft 4 year undergraduate brought out by UGC in response to NEP and expressed concern about the dilution of standards and lack of autonomy. However, the session ended in a positive note pointing to many initiatives at school and college level, outside the system, which can go a long way in addressing the challenges of education.

One of the primary objectives of the MTA conferences is to hear the voice of the teacher, learn about her classroom practices, the pedagogical approaches adopted by her and how she addresses the many challenges of everyday teaching. Twenty-three paper presentations by teachers from all across the country provided a glimpse into the mathematics classroom during the pandemic, the challenges they faced during online classes, how they adapted to different virtual platforms and developed innovative strategies to enhance learner engagement.

Overall, the participants' feedback was encouraging and the two-day conference ended on a positive note. In the valedictory session participants shared their views and suggested topics and themes for future MTA conferences. The conference co-chairs summarised the discussions held in the various sessions of the conference and proposed the vote of thanks acknowledging the contributions of the members of the Program Committee, the Local Organising Committee and the technical team.

2 Historical Roots of Calculus – 2

Shailesh Shirali

Sahyadri School KFI
Rajgurunagar, Khed
Pune – 410513

Email: shailesh.shirali@gmail.com

In the first part of this article [1], we had noted that Newton and Leibniz were not the first mathematicians to be interested in finding tangents to curves; that there were several mathematicians before them who may be regarded as having laid the essential groundwork for the development of the differential calculus. We had particularly looked at the work of Roberval, Descartes and Fermat, and noted that it is Fermat's approach that is closest in spirit to the modern idea of the derivative as a limit. We now look at the corresponding question concerning integration. It comes as a surprise to discover that the origins of integral calculus lie very much further back in time than those of differential calculus. Indeed, they go all the way back to Archimedes.

Archimedes and the method of exhaustion

We start by demonstrating how Archimedes 'squared' the parabola; specifically, his derivation of the formula for area of a parabolic segment. It is a remarkable piece of work, demonstrating an amazing dexterity of reasoning. (Just as remarkable is the story of how we came to know about this work. More about this fascinating story in a later article.) As we did earlier, we shall use modern notation to describe the approach followed by Archimedes.

Consider the parabola $y = x^2$ and the region R bounded by the curve and the chord joining two points on the curve, $A = (a, a^2)$ and $B = (b, b^2)$; see Figure 1. The slope of AB is $a + b$, so its equation is $y - a^2 = (a + b)(x - a)$, i.e., $y = (a + b)x - ab$. The

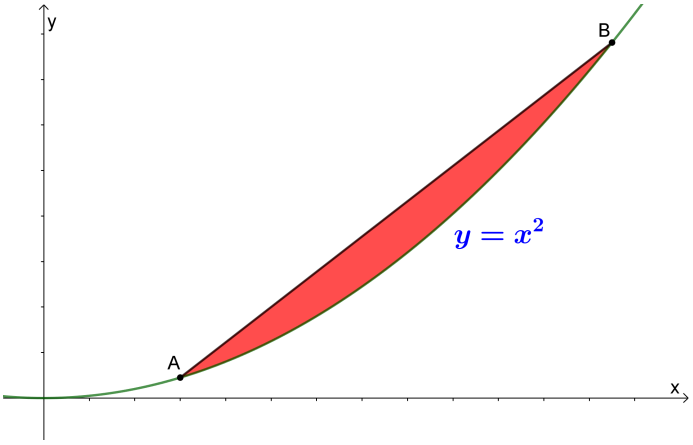


Figure 1

modern evaluation of the area proceeds thus:

$$\begin{aligned} \text{Area of R} &= \int_a^b \left((a+b)x - ab - x^2\right)\, dx \\ &= \frac{(b-a)^3}{6}, \quad \text{on simplification.} \end{aligned} \tag{1}$$

Let us now see how Archimedes finds this area. His strategy is to fill the region with infinitely many triangles; he 'exhausts' the space within the segment using these triangles and then computes the sum of the areas of these triangles. This is actually the sum of an infinite series, so we see here an early encounter with limits. (*Comment.* This is the celebrated 'method of exhaustion' applied here to find the area of a parabolic segment. The idea of using such an approach to compute areas goes back to an earlier mathematician called Eudoxus, but Archimedes was surely its finest exponent.)

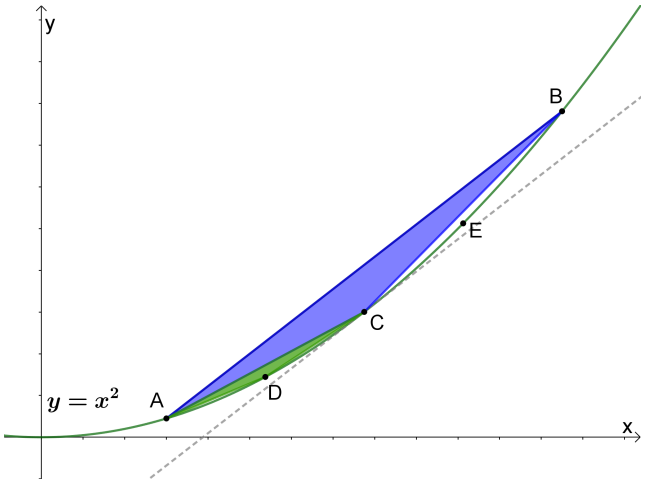


Figure 2

Let C be the point on the curve with x-coordinate halfway between that of A and B (Figure 2):

$$C = \left(\frac{a+b}{2}, \frac{(a+b)^2}{4} \right). \tag{2}$$

Consider now the area of $\triangle ABC$, which we will refer to as the 'central triangle' of segment R. The modern evaluation would proceed thus:

$$\begin{aligned} [\triangle ABC] &= \frac{1}{2} \begin{vmatrix} b^2 - a^2 & \dfrac{(a+b)^2}{4} - a^2 \\ b - a & \dfrac{a+b}{2} - a \end{vmatrix} \\ &= \frac{(b-a)^3}{8}, \quad \text{on simplification.} \end{aligned} \tag{3}$$

We express this in a more convenient form as follows:

$$[\triangle ABC] = \frac{(\text{Difference between } x\text{-coordinates of } A \text{ and } B)^3}{8}. \tag{4}$$

This form enables us to reuse the formula repeatedly. Let D and E be the points on the curve with x-coordinates halfway between those of A and C and halfway between those of C and B, respectively. Then:

$$\begin{aligned} [\triangle ADC] &= \frac{((a+b)/2 - a)^3}{8} \\ &= \frac{(b-a)^3}{8^2} = \frac{1}{8} \times [\triangle ABC]. \end{aligned} \tag{5}$$

and $\triangle BCE$ has the same area. We may similarly define the points F, G, H, I on the curve with x-coordinates halfway between A and D, halfway between D and C, halfway between C and E, and halfway between E and B (these points are not marked in Figure 2). Now we obtain:

$$[\triangle AFD] = \frac{(b-a)^3}{8^3} = \frac{1}{8^2} \times [\triangle ABC], \tag{6}$$

and likewise for three other such triangles. Observing the progression we see that

$$\begin{aligned} \text{Area of } \mathsf{R} &= \frac{(b-a)^3}{8} \left(1 + \frac{2}{8} + \frac{2^2}{8^2} + \frac{2^3}{8^3} + \cdots \right) \\ &= \frac{(b-a)^3}{8} \left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots \right) \\ &= \frac{(b-a)^3}{8} \cdot \frac{4}{3} = \frac{(b-a)^3}{6}, \end{aligned} \tag{7}$$

after summing the infinite series. Observe that we have obtained the same result as (1).

This is essentially the approach taken by Archimedes, but expressed in a form more understandable to us today, using algebraic notation. Note that we have made use of coordinate geometry in our derivation. The obvious question is: how did Archimedes establish the various in-between steps, in an era when algebra and coordinate geometry in their modern form did not exist? We now describe the steps involved, invoking properties of parabolas that may be relatively unfamiliar to us today.

Archimedes does not derive anything like (4) or (5). Indeed, such statements would not have made much sense to anyone from that era. Rather, what he proves is the following central relation:

$$[\triangle ABC] = 8 \cdot [\triangle ADC]. \tag{8}$$

If we accept this result, then by following the construction described above (exactly the same steps as earlier, with points $C, D, E, F, \ldots$), we immediately obtain the following:

$$\begin{aligned} \text{Area of } \mathsf{R} &= [\triangle ABC] \cdot \left(1 + \frac{2}{8} + \frac{2^2}{8^2} + \frac{2^3}{8^3} + \cdots\right) \\ &= [\triangle ABC] \cdot \left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots\right) = [\triangle ABC] \cdot \frac{4}{3}. \end{aligned} \tag{9}$$

This is the result established by Archimedes:

$$\text{Area of a parabolic segment} = \frac{4}{3} \times \text{Area of central triangle inscribed in segment.} \tag{10}$$

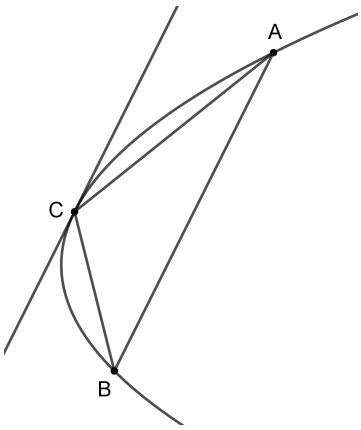


Figure 3: Vertex and central triangle of a parabolic segment

Recalling the definition of vertex and central triangle of a parabolic segment

We have referred to the *central triangle of a parabolic segment* earlier but we define it afresh here. Let AB be the bounding chord of such a segment (see Figure 3). Locate the point C on the arc that is *furthest from the chord*. This is equivalent to saying: *locate the point C on the arc at which the tangent to the arc is parallel to AB*. Then C is called the 'vertex' of the segment, and $\triangle ABC$ is the 'central triangle' of the segment.

The key questions we now address are the following.

1. How does Archimedes establish result (8)? —

$$[\triangle ABC] = 8 \cdot \triangle ADC].$$

2. How does Archimedes establish result (9)? —

$$1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{4}{3}.$$

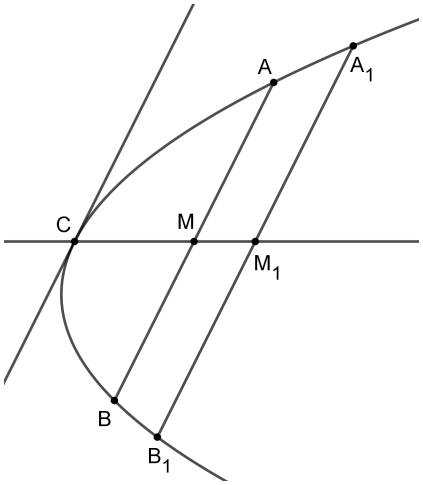


Figure 4: Chords parallel to the tangent at C are bisected by CM (where CM is parallel to the axis of the parabola), and AM^2/CM is a constant

Recalling some key properties of a parabola

With reference to Figure 4:

(P1) The line through C parallel to the axis of the parabola intersects AB at its midpoint M.

(P2) Every chord A_1B_1 parallel to AB is bisected by CM.

(P3) $AM^2 : {A_1M_1}^2 = CM : CM_1$, i.e., AM^2/CM is a constant.

We shall justify (P1)–(P3) later (using modern methods of co-ordinate geometry).

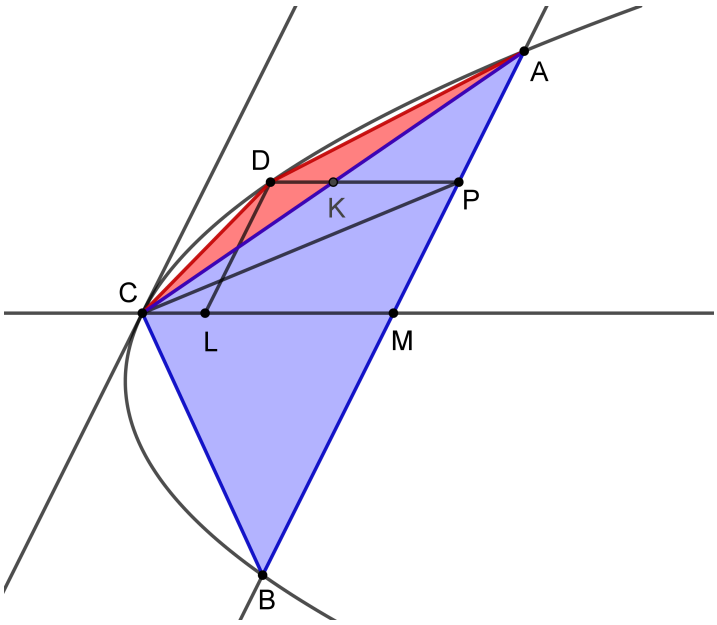


Figure 5: Proving that $[\triangle ABC] = 8 \cdot [\triangle ADC]$

Proof by Archimedes of (8)

Figure 5 shows a segment of a parabola $\mathcal{P}$, bounded by chord AB; its vertex is C (so the tangent to $\mathcal{P}$ at C is parallel to AB). The line through C parallel to the axis of $\mathcal{P}$ meets AB at its midpoint M. The midpoint of AM is P; next, PD is drawn parallel to MC (with D on the parabola), and DL is drawn parallel to AB (with L on CM). The point of intersection of AC and DP is K. Note that D is the vertex of the segment bounded by chord AC.

We now show that $[\triangle ABC] = 8 \cdot [\triangle ADC]$. Using (P3),

$$\frac{CL}{CM} = \frac{DL^2}{AM^2} = \frac{PM^2}{AM^2} = \frac{1}{4}, \tag{11}$$

hence $MC = \frac{4}{3} \cdot ML$, and so $MC = \frac{4}{3} \cdot PD$.

Since $MC = 2 \cdot PK$, we get $PK = \frac{2}{3} \cdot PD$, and so $PK = 2 \cdot KD$.

Next, consider $\triangle APC$ and $\triangle ADC$.

Since $AK = KC$ and $PK = 2 \cdot KD$, it follows that $[\triangle APC] = 2 \cdot [\triangle ADC]$.

Next, $[\triangle AMC] = 2 \cdot [\triangle APC]$, since $AM = 2 \cdot AP$, and $[\triangle ABC] = 2 \cdot [\triangle AMC]$, since $AB = 2 \cdot AM$.

It follows that $[\triangle ABC] = 8 \cdot [\triangle ADC]$. We have obtained the required relation. ■

Proof by Archimedes of (9)

Next, we study how Archimedes showed that

$$1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{4}{3},$$

or, expressed in another way,

$$\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{1}{3}. \tag{12}$$

Archimedes starts by deriving the following identity ([2]):

$$\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots + \frac{1}{4^n} + \frac{1}{3 \cdot 4^n} = \frac{1}{3}, \tag{13}$$

which it is true for all $n \in \mathbb{N}$. Archimedes most likely proved (13) recursively as follows. Writing $1/4$ as $4/4^2 = 3/4^2 + 1/4^2$, $1/4^2$ as $4/4^3 = 3/4^3 + 1/4^3$, and so on, we have:

$$\begin{aligned}
1 &= 3 \cdot \left(\frac{1}{4}\right) + \frac{1}{4} \\
&= 3 \cdot \left(\frac{1}{4} + \frac{1}{4^2}\right) + \frac{1}{4^2} \\
&= 3 \cdot \left(\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3}\right) + \frac{1}{4^3} \\
&= 3 \cdot \left(\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4^4}\right) + \frac{1}{4^4} = \cdots \cdots \cdots .
\end{aligned}$$

This immediately yields (13).

From this relation Archimedes would surely have concluded that (12) must be true, as the quantity $1/(3 \cdot 4^n)$ can be made as small as one may want.

However, such an argument could not have been made by Archimedes. We must keep in mind that the language of infinite series and limits did not exist during that era; indeed, we must move forward by more than 1,500 years to encounter those terms!

Instead, Archimedes had to use the language of 'exhaustion': he had to argue that the sum

$$\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots$$

cannot be less than $\frac{1}{3}$, nor cannot it be greater than $\frac{1}{3}$. Therefore the sum must be equal to $\frac{1}{3}$.

The reasoning used above can be depicted in an attractive form as a "proof without words" ([6]):

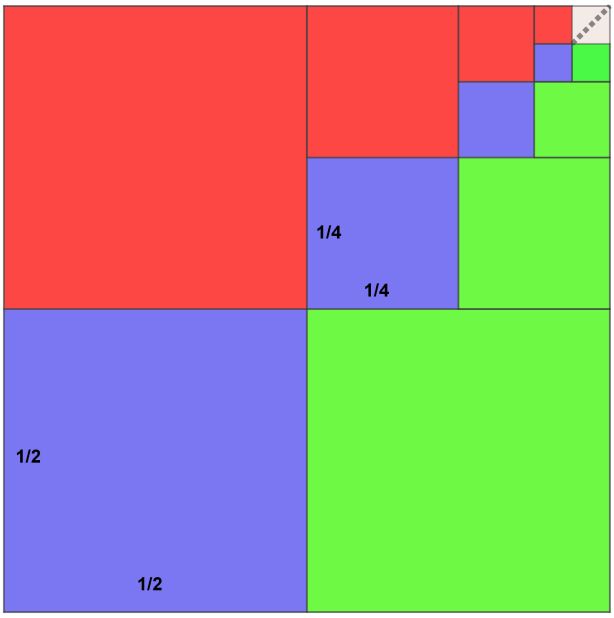


Figure 6: Proving that $\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{1}{3}$

Closing remarks

We see the remarkable ingenuity shown by Archimedes and his extraordinary dexterity in using methods of elementary algebra and geometry to prove results which are very far from elementary, as they depend implicitly on the idea of a limit. We see also how Archimedes was far, far ahead of his time in anticipating how area and volume can be computed using concepts that are equivalent to the modern concept of a limit. Using

much the same ideas, he was able to derive formulas for the surface area and volume of a sphere, and for the volume of a right circular cone.

In later parts of this series of articles, we shall study how mathematicians of the Renaissance era and after — Descartes, Fermat, Cavalieri, Wallis, Barrow, Mercator — approached the problem of area.

Appendix: Proofs of some properties of the parabola

We now give the proofs of (P1), (P2), and (P3).

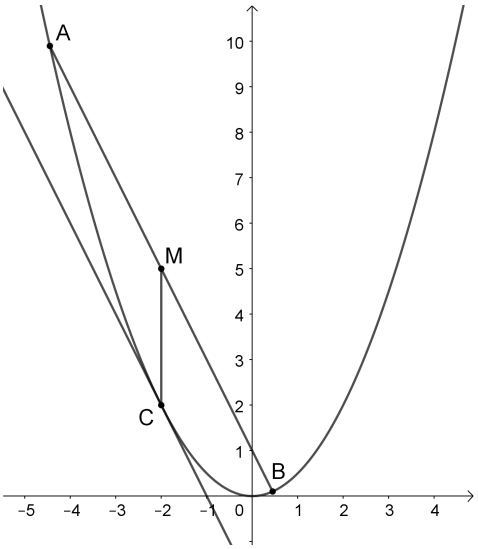


Figure 7: Chords parallel to the tangent at C are bisected by CM, and CM/AM^2 is a constant

With reference to Figure 7, we must prove: (P1 & P2) Chords parallel to the tangent at C are bisected by CM (where CM is parallel to the axis of the parabola). (P3) AM^2/CM is a constant. We shall approach the matter the way a college student would today — and not how Archimedes or Apollonius or their contemporaries would have done!

Let the equation of the parabola (relative to an appropriate co-ordinate system) be $y = x^2$. Here the axis of the parabola is the y-axis. Let $C = (c, c^2)$. The slope of the tangent to the parabola at C is $2c$. Let $M = (c, c^2 + k)$. Since AB is parallel to the tangent, its equation is $y - (c^2 + k) = 2c(x - c)$, i.e., $y = 2cx - c^2 + k$. Let $A = (a, a^2)$ and $B = (b, b^2)$. Then a, b are the roots of the quadratic equation $x^2 - 2cx + c^2 - k = 0$, therefore $a + b = 2c$, so $(a + b)/2 = c$. This implies that M is the midpoint of AB, as

claimed. This justifies (P1), and (P2) follows similarly. For (P3), we compute the value of AB^2. We have:

$$\begin{aligned} AB^2 &= (a-b)^2 + (a^2-b^2)^2 \\ &= (a-b)^2 \cdot \big(1 + (a+b)^2\big) \\ &= \big((a+b)^2 - 4ab\big) \cdot \big(1 + (a+b)^2\big) \\ &= 4k \cdot (1 + 4c^2), \text{ after simplification.} \end{aligned}$$

Hence $AM^2/CM = 1 + 4c^2$, which is independent of k. This means that AM^2/CM is a constant for all chords parallel to the tangent at C, as claimed. ■

Bibliography

[1] Shailesh Shirali, "Historical Roots of Calculus – I." *Blackboard*, Jan. 2022, Issue 4, pp. 9–17, https://www.mtai.org.in/wp-content/uploads/2022/01/blackboard-issue4.pdf

[2] John Abbott College, "Archimedes' quadrature of the parabola and the method of exhaustion." https://www.math.mcgill.ca/rags/JAC/NYB/exhaustion2.pdf

[3] Gordon Swain and Thomas Dence, "Archimedes' Quadrature of the Parabola Revisited." *Mathematics Magazine*, Apr., 1998, Vol. 71, No. 2 (Apr., 1998), pp. 123-130, Mathematical Association of America, https://www.jstor.org/stable/2691014

[4] John B. Little, "Archimedes' Quadrature of the Parabola." https://mathcs.holycross.edu/~little/Archimedes.pdf

[5] A. Kursat Erbas, "An Explanatory Approach to Archimedes's Quadrature of the Parabola." http://jwilson.coe.uga.edu/EMT668/EMAT6680.F99/Erbas/emat6690/essay1/essay1.html

[6] Wikipedia, the free encyclopedia, "Quadrature of the Parabola." https://en.wikipedia.org/wiki/Quadrature_of_the_Parabola

3 M.C. Escher's Print Gallery

Muskan Choudhary & Kartav Kesri

IISER Bhopal
Bhopal Bypass Rd, Bhauri
Madhya Pradesh 462066

Email: muskan20@iiserb.ac.in, kartav20@iiserb.ac.in

1 Introduction

Have you ever noticed the variety of buildings? Old architecture? New-age architecture? Some drawings of artists? Computer simulations? They all have one thing in common. They all have repeating patterns. For more than a hundred years, the act of repeating basic shapes has stimulated the neurons of both artists and mathematicians. The Dutch graphic artist *M. C. Escher* (1898–1972) was most likely the first to be highly inspired. He came up with many artworks like that. However, this article will specifically discuss the "*Prentententoonstelling*", popularly known as "*The Print Gallery*", including the hyperbolic patterns by Escher. Also, we will try to understand the mathematics behind it in brief.

2 Hyperbolic Geometry

2.1 Origin

Mathematicians classify geometry into two types, *Euclidean* and *non-Euclidean*. Euclidean geometry can be related to the Greek mathematician *Euclid*. It is the geometry with which most of us are familiar because it is the geometry we were taught in school. On the other hand, Non-Euclidean Geometry is more prevalent among physicists and artists. (For example, *spherical geometry*, *hyperbolic geometry*, *elliptic geometry* etc.)

The significant difference between Euclidean and Non-Euclidean geometry is the property of parallel lines.

We all are aware of the five axioms of Euclid. However, we should look closely at the fifth one for our topic of interest. It is as follows:

> "If a straight line falls on two straight lines so that the interior angles on the same side are together less than two right angles, the straight lines if produced indefinitely, meet on that side on which the angles are less than the two right angles."

Many mathematicians tried to find a proof of this axiom by contradiction. They investigated the consequences of either of two assumptions; at least two lines parallel or no lines parallel to the original line. However, instead of finding a contradiction, their work gave rise to new geometries. This was the origin of hyperbolic geometry.

2.2 Studies and models

In the nineteenth century, hyperbolic geometry was intensely investigated by *János Bolyai* and *Nikolai Ivanovich Lobachevsky*. Euclidean geometry can be modelled by a "flat plane". The most straightforward model for elliptical geometry is a sphere, but developing a model for hyperbolic geometry proved to be problematic. In 1868, the mathematician *Eugenio Beltrami* showed that the pseudosphere had the proper curvature to model a part of hyperbolic space.

There are four famous models that define a real hyperbolic space which meets the axioms of hyperbolic geometry.[1]

- *The Klein model* takes the *hyperbolic plane* as the interior of a circle. Hyperbolic lines are the chords of the given circle. It is the simplest; however, it deforms the hyperbolic angles.

- *The Poincaré half-plane model* assumes half of the *Euclidean plane*, as specified by a Euclidean line l, to be its *hyperbolic plane*. Lines are either semicircles which are orthogonal to l or rays perpendicular to l.

- *The Lorentz model* utilises a two-dimensional hyperboloid of revolution in a three-dimensional space. (It looks like a Pringles potato chip.) It is the most difficult to understand of the four models known.

[1] All the references of this section are taken from [4] and from pages 6-7 of [1].

- *The Poincaré disc model* employs the interior of a circle as a plane. Lines are represented by circular arcs that are orthogonal to the boundary circle as well as the diameters of the boundary circle.

2.3 Poincaré Disk Model

The *Poincaré disk model*[2] is the most popular of all the four models. Also, it is our model of interest. It is a 2-dimensional disk with *hyperbolic geometry* implemented via the hyperbolic metric:

$$ds^2 = \frac{(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}$$

In the Poincaré disk a line is modelled by the arc of a circle whose ends are perpendicular to the given disk's boundary. Diameters are also counted as arcs for this purpose.

Two angles, θ_1 and θ_2, around the disk can specify the endpoints of an arc. Let us define θ_1 and θ_2 as follows:

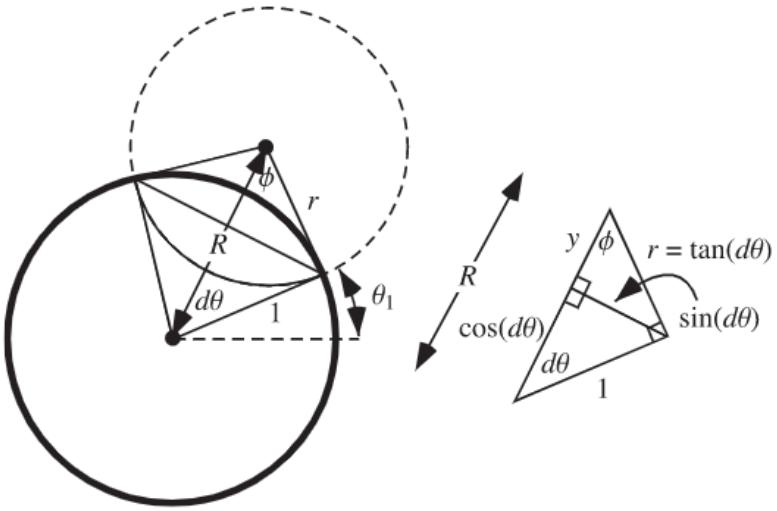


Figure 1

$$\theta \equiv \frac{1}{2}(\theta_1 + \theta_2), \qquad d\theta \equiv \frac{1}{2}|\theta_1 - \theta_2|$$

Using the above diagram and basic trigonometry, we get,

$$r = \tan(d\theta), \qquad y = \sin(d\theta)\tan(d\theta)$$

[2]All the references of this section are taken from [5] and from pages 7-10 of [1].

Thus, the radius of the circle forming the arc is r, and the centre is located at $(R\cos\theta, R\sin\theta)$, where R is defined by

$$R = \cos(d\theta) + y = \sec(d\theta)$$

The half-angle subtended by the arc is,

$$\sin\phi = \frac{\sin(d\theta)}{\tan(d\theta)} = \cos(d\theta)$$
$$\phi = \sin^{-1}[\cos(d\theta)]$$

The Poincaré disk exhibits a conformal mapping for directly measuring the angles between the arcs. Mathematicians and artists have widely used this property to fit shapes, mainly triangles, to get repeating patterns.

2.4 Poincaré disk and patterns

Patterns have constantly stimulated the human brain. Patterns add symmetry and additional beauty to objects. Classically, artists would have used tools like the compass and ruler to create their geometric patterns. However, presently, we would use computer graphics instead. Coming back to *M. C. Escher*, he too was amazed by patterns, and especially by infinite patterns. The Poincaré disk showed him how to accomplish his long-held desire to break the two-dimensional circular limit.[3]

In 1958, *H. S. M. Coxeter* sent Escher a copy of his paper "*Crystal symmetry and its generalisations*". In his reply, Escher wrote, "*…some of the text-illustrations, especially figure 7, page 11, gave me quite a shock*". Coxeter described the basics of these methods in the return letter to Escher. However, by that time, Escher had figured out most of this, as evidenced by *Circle Limit I*. Like Escher, mathematicians have traditionally drawn triangle tessellations in the Poincaré disk model using straightedge and compass techniques, sometimes showing the structure. This technique was something of a geometric 'folk art' until the recent paper by *Chaim Goodman-Strauss* [5], in which the construction methods were finally written down.

Figure 2 illustrates an example. The points of the Poincaré disk model are the interior points of the bounding circle in the Euclidean plane. Hyperbolic lines are represented by diameters and circular arcs orthogonal to the bounding circle. We can see that the triangles on the disk become smaller as we approach the circular limit. Escher was interested in the pattern building implications of this shrinking.[4]

[3] All the references of this section are taken from pages 7-12 of [6].
[4] For a better understanding of this section, please go through [7].

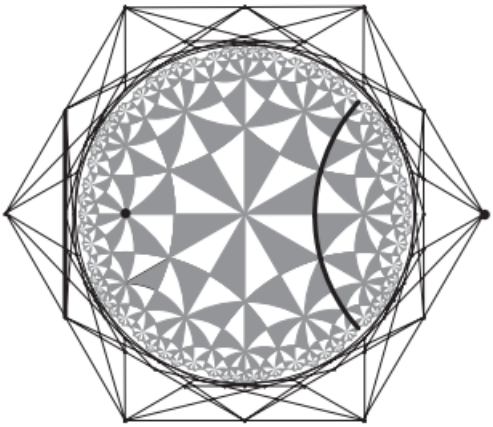


Figure 2

The Poincaré disk model appeals to artists because it is conformal, and it is displayed in a bounded part of the Euclidean plane. Escher exploited it to construct some incredible artworks, as shown in Figure 3. Take a minute to appreciate these works before moving to the next section.

(a)

(b)

Figure 3

3 The Print Gallery

Now, we are done with the preliminary discussions for this article. It is time to move toward our primary topic: *Prentententoonstelling*.[5] It shows a young man standing in a gallery, looking at the lithograph of a Mediterranean seaport. As his eyes track the quayside buildings displayed on the print from left to right and then down, he discovers among them the very same gallery in which he is standing. A circular patch in the middle of the print is left incomplete, containing Escher's signature. However, how did Escher draw such a masterpiece? What was going in his artistic mind when he drew this? Was he aware of the mathematics behind the lithograph?[6]

The best explanation of this can be found in *The Magic Mirror of MC Escher* by *Bruno Ernst*. We paraphrase his description of Escher's motivations: "*It should be possible to make an annular bulge and a cyclic expansion without a beginning and end.*"

This idea gave Escher headaches forever. He tried to put it into action by using straight lines. Later, he tried to use curved lines. Intuitively, he attempted to fix squares between those curved lines while small squares retain their appearance. After many iterations, Escher arrived at the grid shown in Figure 4. As one journeys from point A to point D, the squares expand four times in each direction. As one moves around the centre clockwise, the grid folds into itself by expanding by a factor of 256[7].

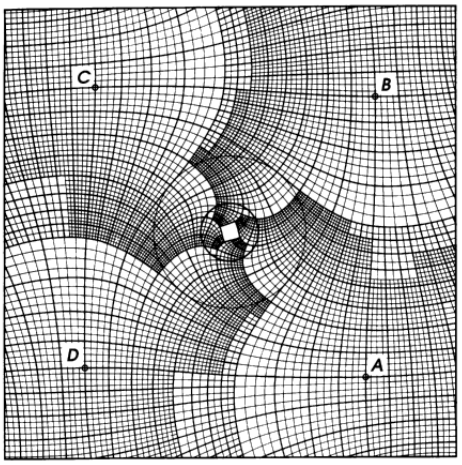


Figure 4: Escher's Grid

Escher did four location studies to achieve perfection, one for each grid corner. Mathematically we can view Escher's four studies as a single drawing invariant under scaling

[5] All the references, including the images for this section, are taken from the first half of [6].
[6] All images are taken from an open source website [3].
[7] See pages 1–3 of [6]

by 256. Square by square, Escher accommodated the straight square grid of his four studies onto the curved grid, and he created *Prententententoonstelling*. This is shown in Figure 5.

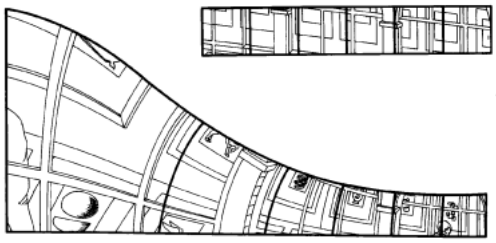


Figure 5: Fitting Squares in Curved Grid

Escher's ideas give us an exact way to travel back and forth between the straight and the curved world. Let us now try to study Escher's grid more mathematically. We will move along the curved grid and trace it in the straight world. Consider the path shown in Figure 6. As depicted in Figure 6, the corresponding path in the straight world takes three left turns, travelling four times as far as the last time before making the next turn. It is not a closed-loop; instead, the endpoint is 256 times the starting point if the origin is chosen correctly. With the same choice of origin, the same happens whenever one converts a single closed loop, counterclockwise around the centre, from the curved world to the straight world. It recollects the invariance of the straight picture under a blowup by a factor of 256. No such phenomenon happens when we move around the centre.

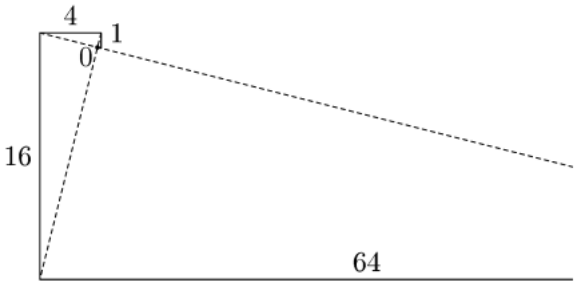


Figure 6: Square ABCD in the Straight World

3.1 Elliptic curves

We already know that the straight picture is periodic with a period of 256. An ideal version of the curved image has a complex period γ.

$$f(256z) = f(z)$$
$$g(\gamma w) = g(w)$$

Nevertheless, what is the connection between 256 and γ? Now, we reformulate what we know. Let us first remove 0 from $\mathbb{C}$. This leaves a hole too small to notice, unlike Escher's grid. Figure 7 illustrates the basic mathematical model for solving the problem.

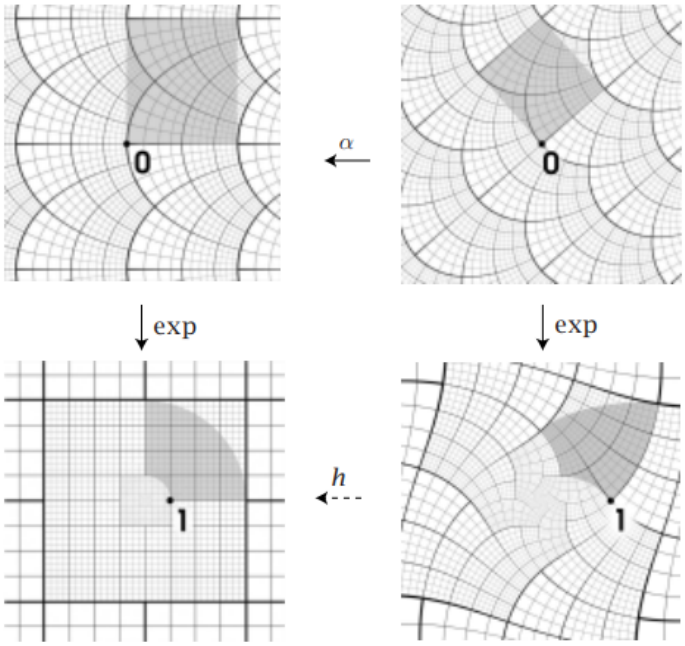


Figure 7

A standard result on complex tori implies that the map $\mathbb{C} \longrightarrow \mathbb{C}$ is a multiplication by a certain scalar $\alpha \in \mathbb{C}$ that satisfies $\alpha L_\gamma = L_{256}$. Now, we obtain an isomorphism between two short exact sequences:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & L_\gamma & \longrightarrow & \mathbb{C} & \xrightarrow{\exp} & \mathbb{C}^*/\langle\gamma\rangle & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & L_{256} & \longrightarrow & \mathbb{C} & \xrightarrow{\exp} & \mathbb{C}^*/\langle 256\rangle & \longrightarrow & 0. \end{array}$$

In order to compute α, we use the multiplication by α map $L_\gamma \longrightarrow L_{256}$ which may

be thought of as a map between fundamental groups. Indeed, it is nothing but the isomorphism between the fundamental groups of $\mathbb{C}^*/ < \gamma >$ and $\mathbb{C}^*/ < 256 >$.

The element $2\pi i$ in the fundamental group L_γ of $\mathbb{C}^*/ < \gamma >$ corresponds to a single counterclockwise loop around the origin in $\mathbb{C}^*$. It is the same as the path ABCDA along grid lines that we took earlier. As we saw, Escher's method transforms it into a path in $\mathbb{C}^*$ that goes once around the origin and at the same time multiplies by 256 in $\mathbb{C}^*/ < 256 >$. This path becomes a closed loop that represents the element 2πi + log 256 of L_{256}.

Thus, our isomorphism $L_\gamma \longrightarrow L_{256}$ maps $2\pi i$ to $2\pi i + \log(256)$, and therefore $\alpha = (2\pi i + \log(256)/(2\pi i$. The lattice L_γ is now given by $L_\gamma = \alpha^{-1} L_{256}$, and from $|\gamma| > 1$ we deduce

$$\begin{aligned} \gamma &= exp(2\pi i(\log(256)/(2\pi i + \log(256)) \\ &= exp(+3.1172277221 + 2.7510856371i) \end{aligned}$$

The new grid is given in Figure 8.The central hole is much smaller than Escher's grid because of the value $|\gamma| = 22.56$ (greater than Escher's grid) [8].

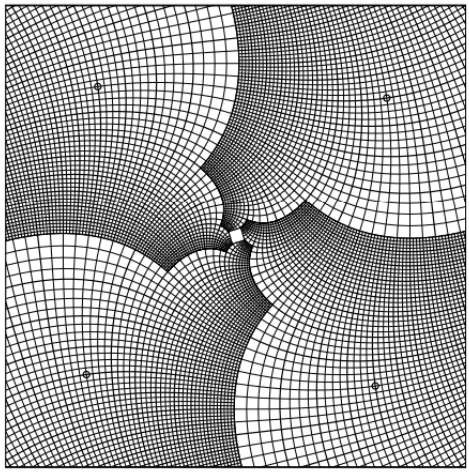


Figure 8

3.2 Filling the hole

First, the initial grid was reconstructed from Escher's grid to fill the hole. As illustrated in Figure 9, the blank spot in the middle gave rise to an empty spiral in the reconstructed

[8]For more detailed explanations, see [8]

studies, and there were other flaws. Next, the Dutch artists *Hans Richter* and *Jacqueline Hofstra* completed and adapted the pictures obtained; see Figure 10. However, making this grayscale, much unsteady resolution with changing line widths arose. A natural way to solve this problem was to make pixels uniform on the elliptical curve. The completed version of the *Prentententoonstelling* is shown in Figure 11.

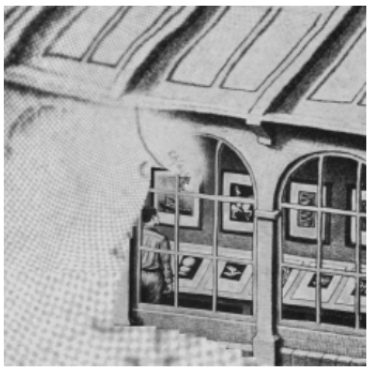


Figure 9

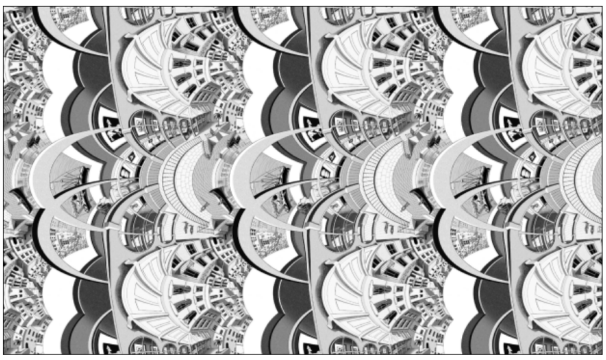


Figure 10: The straight drawing pulled back, by the complex exponential function, to a doubly periodic picture, with grayscale added. The horizontal period is log 256, the vertical period is $2\pi i$.

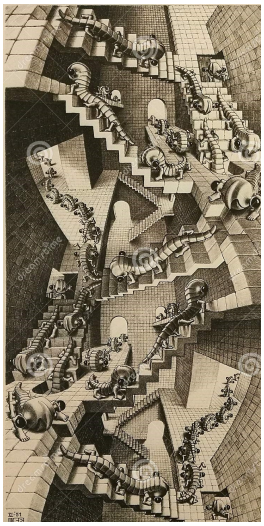
Figure 11: The completed version of Escher's lithograph with magnifications of the center by factors of 4 and 16.

4 Other Works

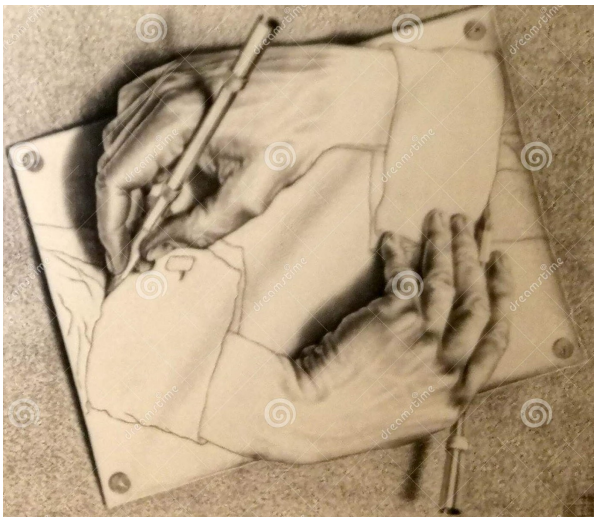
For an assessment of Escher's legacy in the art world, we quote Wikipedia [9]:

> His paintings feature mathematical objects and operations, including impossible things, explorations of infinity, reflection, symmetry, perspective, truncated and stellated polyhedra, hyperbolic geometry, and tessellations…
> Escher's work is inescapably mathematical. This has caused a disconnect between his full-on popular fame and the lack of esteem with which he has been viewed in the art world. His originality and mastery of graphic techniques are respected, but his works have been thought too intellectual and insufficiently lyrical…
> Although Escher did not have mathematical training – his understanding of mathematics was largely visual and intuitive – his art had a strong mathematical component, and several of the worlds he drew were built around impossible objects.

Escher did not get the recognition he deserved in the art world for quite a long time. Nevertheless, he became popular among scientists and mathematicians. Furthermore, his works are still inspiring and thought-provoking for upcoming mathematicians and art lovers. While Escher believed that he had no mathematical ability, he had an intuitive feel for mathematics and he has left mathematical works for all of us to appreciate long after he died. Here are some pieces by Escher that breach the limits of the human mind. (See [2].)

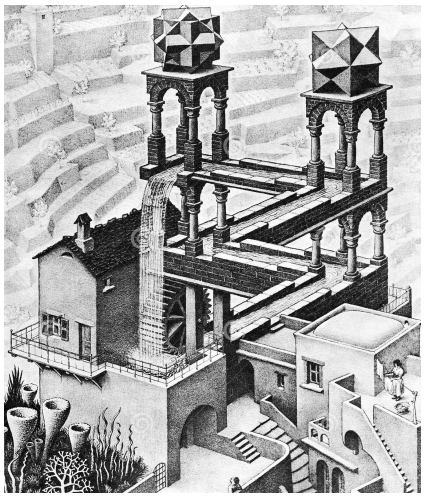


(a) Ascending and Descending

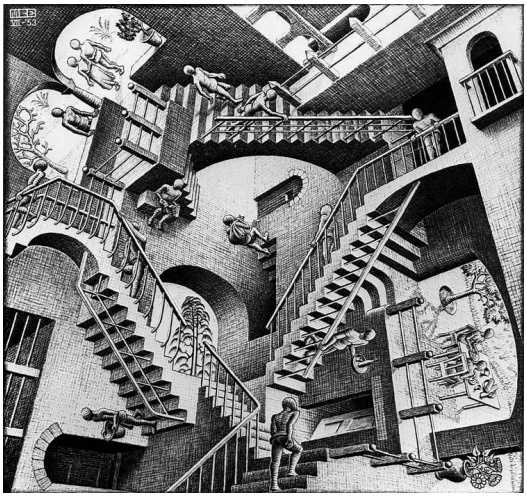


(b) Drawing Hands

Figure 12



(a) Waterfall



(b) Relativity

Figure 13

Bibliography

[1] James W Anderson, *Hyperbolic Geometry*. Springer Science & Business Media. 2006.

[2] John H. Conway, Heidi Burgiel and Chaim Goodman-Strauss, *The Symmetries of Things*, CRC Press, 2016.

[3] Dreamstime, *Escher Stock Photos, Images & Pictures*, https://www.dreamstime.com/photos-images/escher.html. Accessed: 2022-09-01.

[4] Douglas Dunham, "A Tale Both Shocking and Hyperbolic." *Math Horizons*, 2003, Vol. 10, No. 4, pp. 22–26.

[5] Chaim Goodman-Strauss, "Compass and Straightedge in the Poincaré Disk." *American Mathematical Monthly*, 2001, Vol. 108, No. 1, pp. 38–49, https://www.jstor.org/stable/2695674.

[6] B. de Smit and H. W. Lenstra Jr., "The Mathematical Structure of Escher's Print Gallery." *Notices of the AMS*, 2003, Vol. 50, No. 4, pp. 446–451, https://www.ams.org/notices/200304/fea-escher.pdf.

[7] Christina L. Sheets, "Hyperbolic Geometry." *MAT Exam Expository Papers*, 2007, p. 39. https://digitalcommons.unl.edu/mathmidexppap/39/.

[8] Anood E. Alkatheeri and Waldo G. Arriagada, "An Algebraic Note On Print Gallery." *Applied Mathematics E-Notes*, 2021, Vol. 21, pp 669–677, https://www.emis.de/journals/AMEN/2021/AMEN-201022.pdf.

[9] Wikipedia, *M. C. Escher*. https://en.wikipedia.org/wiki/M._C._Escher. Accessed: 2022-09-01.

4 Fair Allocation Using an Unfair Randomizer

Jyotirmoy Sarkar

Indiana University-Purdue University Indianapolis
Department of Mathematical Sciences
402 N Blackford Street
Indianapolis, IN 46202-3216, USA

Email: jsarkar@iupui.edu

Mamunur Rashid

DePauw University
Department of Mathematical Sciences
2 E Hanna Street
Greencastle, IN 46135, USA

Email: mrashid@depauw.edu

Abstract: If you have access to a randomizer that generates a discrete random variable taking on $a \geq 2$ distinct values with unknown probabilities, how can you extract a sequence of independent discrete random variables taking on $b \geq 2$ equiprobable values? We provide R codes to implement algorithms that accomplish these tasks.

1 Motivating Problems

Three friends were hiking in the woods. They saw a sign staked in by the trailside pointing towards the mouth of a cave. It read:

> "See the cave? In you peep.
> There's a hole; please dig deep.
> What you find, you can keep.
> (Then) break this sign; dump in a heap."

The friends were intrigued. They saw the cave; they peeped in; there indeed was a hole; they dug; they found the treasure: It was a gold coin! Not to mislead other hikers, they broke the sign into pieces and tossed them into the hole. How obedient and how considerate!

But who will keep the coin? Having no other randomizer at hand, they agreed to toss the coin to determine the winner. The following two mathematical problems intrigued them:

1. Assuming the coin is fair, how to minimize the number of tosses necessary to choose the keeper, if each person must have a one-third chance of winning?

2. Not knowing that the coin is fair, how to minimize the number of tosses to generate an unbiased sequence of bits with which to choose the keeper, giving each person a one-third chance of winning?

A similar challenge and opportunity befell a group of ten scouts and their scout master who found a silver cube. That evening as they sat around the bonfire after dinner, they decided to roll the cube, after labeling its faces with numbers 1 through 6, to determine who will keep the cube, ensuring that each will have a one-eleventh chance of winning.

The objective of this paper is to equitably determine a winner from among $b \geq 2$ participants by using a randomizer that generates $a \geq 2$ values with unknown probabilities. We want to know (1) how to use a sequence of independent, unbiased values to determine the winner of one or more prizes, and (2) how to extract independent, unbiased values from a sequence of biased values. Section 2 uses independent, unbiased coin tosses to determine a prize winner from among several eligible candidates giving each candidate the same chance of winning. Section 3 does the same job using independent outcomes of an unbiased die with six faces. Section 4 discusses an efficient algorithm to extract independent, unbiased bits using a possibly biased coin. Section 5 extends the procedure to extract independent outputs which take one of 6 equiprobable outcomes using a possibly 6-faced biased die. Section 6 allows using a biased randomizer with $a \geq 3$ outcomes to extract independent outputs which take $b \geq 3$ equiprobable values. Section 7 discusses some variations and poses some open questions.

2 To Efficiently Determine the Winner Equitably Using an Unbiased Coin

Each toss of an unbiased coin results in a 0 (tail) or a 1 (head) with probability 1/2 each. If there are three friends eligible to win the prize, with only one toss, neither outcome can be allocated to any person to be the winner, for then that person will win with probability 1/2, which exceeds 1/3. With two tosses there are four possible outcomes — 00, 01, 10, 11 — each with probability 1/4. We may allocate one outcome to each person. For example, we can allocate 01 to A, 10 to B, 11 to C. In case the outcome is 00, then we restart the process: Toss twice more and apply the same rule to determine the winner.

Thus, after two tosses there is a 3/4 chance of a success (a winner is determined), and 1/4 chance of a failure resulting in a restart. Overall, $2N$ tosses are needed to determine a winner, where N, the number of iterations until a winner is determined, is a geometric(3/4) random variable. The average number of tosses μ_3 to determine the winner from among three friends is $\mu_3 = 2E[N] = 2(4/3) = 8/3$. Alternatively, by the renewal property of the allocation process, we have $\mu_3 = 2 + (1/4)\mu_3$, which is solved to obtain $\mu_3 = 2(4/3) = 2.67$. Also, each person's probability of winning is $(1/4)[1 + (1/4) + (1/4)^2 + (1/4)^3 + \ldots] = (1/4)/[1 - 1/4] = 1/3$, where we have used the familiar sum of a geometric sequence

$$1 + r + r^2 + r^3 + \ldots = \frac{1}{1-r}, \text{ for } |r| < 1. \tag{1}$$

Henceforth, we will use equation (1) whenever needed, completely unannounced.

What if the number of hikers was other than three? How would the allocation scheme work? Obviously, if there were two friends, the winner could be determined with a single toss of the unbiased coin; and had there been four friends, the winner could be determined in exactly two tosses. Likewise, for $n = 2^k$ friends exactly k tosses will suffice to determine a winner. Hence, $\mu_{2^k} = k$. What if n is not a power of two?

If there were $n = 5$ friends, then toss the coin three times; allocate one outcome to each friend and for the remaining three outcomes (not allocated to any person), toss the coin again so that there are now 6 outcomes. Allocate one such outcome to each friend. Should the only remaining outcome, unassigned to anyone, occur, restart the process. Therefore, using the renewal property, the average number of tosses needed to determine a winner is

$$\mu_5 = 3 + \frac{3}{8} * 1 + \frac{\mu_5}{16}, \text{ whence } \mu_5 = \frac{27}{8} * \frac{16}{15} = \frac{18}{5} = 3.6.$$

Each person's probability of winning is

$$\left(\frac{1}{8} + \frac{1}{16}\right)\left[1 + \frac{1}{16} + \left(\frac{1}{16}\right)^2 + \ldots\right] = \left(\frac{3}{16}\right)\ \frac{16}{15} = \frac{1}{5}.$$

If there were n=6 friends, form two teams of three friends in each team. With a single toss determine the winning team and then follow the recipe for the three-person game to determine which member of the winning team is the ultimate winner of the coin. Alternatively, form three pairs; determine the winning pair using the three-person game and then with one more toss determine the ultimate winner within the winning pair. Either way, on average $\mu_6 = \mu_2 + \mu_3 = 3.67$ tosses are needed to determine the winner from among six friends.

More generally, when there are a composite number of friends $n = p * 2^m$, then split them into 2^m teams of p friends each. Determine the winning team, using m tosses, and

then determine the winning member of the winning team, needing

$$\mu_{p*2^m} = m + \mu_p \tag{2}$$

tosses on average and giving each person a probability $(1/2^m)(1/p) = 1/(p * 2^m) = 1/n$ of winning. For example, using (2), $\mu_{10} = 1+\mu_5 = 1+3.6 = 4.6$ and $\mu_{12} = 2+\mu_3 = 4.67$. Therefore, it suffices to solve the problem for odd number of friends p.

If there are p friends (where p is an odd number), then the coin must be tossed a minimum of m times where $2^{m-1} < p < 2^m$. One outcome may be assigned to each friend and the remaining $(2^m - p)$ outcomes should be doubled, quadrupled, etc., by tossing the coin once, twice, etc., until for the smallest j, the number of outcomes $(2^m - p)\,2^j$ exceeds p. Again, one outcome may be assigned to each friend and the remaining outcomes should be doubled, quadrupled, etc. until the number of outcomes exceeds p again. And so on, until there remains exactly one outcome; thereafter, the process will repeat itself.

There is a beautiful number theoretic result (Fermat's little theorem) that guarantees the convergence of $\{c*2^j \pmod p : j \geq 1\}$ to 1, for any positive integer c. See Wikipedia [7]. The theorem is named after Pierre de Fermat, who stated in 1640 without proof that it holds for any prime p; it is also called the Fermat-Euler theorem because Leonhard Euler proved it in 1735; it is also called Euler's totient theorem because Euler generalized it in 1763 to hold for any n (not necessarily a prime).

For instance, when $n = 7$, toss the coin 3 times; assign one outcome to each friend and for the remaining single outcome let the process restart. The expected number of tosses until a winner is determined is $\mu_7 = 3+\mu_7/2^3 = 24/7 = 3.43$; and each person's chance of winning is $(1/8)\left[1 + (1/8) + (1/8)^2 + \ldots\right] = 1/7$.

When $n = 9$, toss the coin four times so that the number of outcomes exceeds 9; then assign one outcome to each person. Then continue tossing to double, quadruple, etc. the number of remaining outcomes, etc., until only one outcome remains. The number of outcomes at various stages are given by (the number following the arrow is the remainder when the preceding value is divided by n; then after the asterisk the remainder is doubled, quadrupled, etc.)

$$2^4 \to 7 * 2 \to 5 * 2 \to 1.$$

Invoking the renewal property, the expected number of tosses until a winner is determined is

$$\mu_9 = 4 + \frac{7}{2^4} + \frac{5}{2^5} + \frac{\mu_9}{2^6} = \frac{294}{63} = 4.67.$$

Each person's probability of winning is

$$\left(\frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{2^6}\right)\left[1 + \frac{1}{2^6} + \left(\frac{1}{2^6}\right)^2 + \ldots\right] = \left(\frac{7}{2^6}\right)\frac{2^6}{2^6 - 1} = \frac{7}{63} = \frac{1}{9}.$$

When $n = 11$, toss the coin four times so that the number of outcomes exceed 11; then assign one outcome to each person. Then continue tossing to double, quadruple, etc. the number of remaining outcomes, etc., until only one outcome remains. The number of outcomes at various stages are given by (recall the arrow notation)

$$2^4 \to 5 * 2^2 \to 9 * 2 \to 7 * 2 \to 3 * 2^2 \to 1.$$

Invoking the renewal property, the expected number of tosses until a winner is determined is

$$\mu_{11} = 4 + \frac{5}{2^4} * 2 + \frac{9}{2^6} + \frac{7}{2^7} + \frac{3}{2^8} * 2 + \frac{\mu_{11}}{2^{10}} = \frac{4960}{1023} = 4.85,$$

where the powers of 2 in the denominators of the summands above keep track of the number of tosses made so far. Each person's probability of winning is

$$\left(\frac{1}{2^4} + \frac{1}{2^6} + \frac{1}{2^7} + \frac{1}{2^8} + \frac{1}{2^{10}}\right) \left[1 + \frac{1}{2^{10}} + \left(\frac{1}{2^{10}}\right)^2 + \ldots\right] = \left(\frac{93}{2^{10}}\right) \frac{2^{10}}{2^{10} - 1} = \frac{1}{11}.$$

When $n = 13$, toss the coin four times so that the number of outcomes exceed 13; then assign one outcome to each person. Then continue tossing to double, quadruple, etc. the number of remaining outcomes, etc., until only one outcome remains. The number of outcomes at various stages are given by

$$2^4 \to 3 * 2^3 \to 11 * 2 \to 9 * 2 \to 5 * 2^2 \to 7 * 2 \to 1.$$

Invoking the renewal property, the expected number of tosses until a winner is determined is

$$\mu_{13} = 4 + \frac{3}{2^4} * 3 + \frac{11}{2^7} + \frac{9}{2^8} + \frac{5}{2^9} * 2 + \frac{7}{2^{11}} + \frac{\mu_{13}}{2^{12}} = \frac{19278}{2^{12} - 1} = 4.707692.$$

Each person's probability of winning is

$$\left(\frac{1}{2^4} + \frac{1}{2^7} + \frac{1}{2^8} + \frac{1}{2^9} + \frac{1}{2^{11}} + \frac{1}{2^{12}}\right) \left[1 + \frac{1}{2^{12}} + \left(\frac{1}{2^{12}}\right)^2 + \ldots\right] = \frac{315}{2^{12} - 1} = \frac{1}{13}.$$

We suggest the readers manually compute μ_n for $n = 15, 17, 19$, etc. and check that the probability of winning for each person is indeed $1/n$. R codes to compute μ_n and each contestant's winning probability are given in Appendix 1. Figure 1 shows the expected number of tosses needed to determine the winner by tossing an unbiased coin. In particular, if $n = 2^m$, then $\mu_n = m$; and if $2^{m-1} < n < 2^m$, then $m < \mu_n < m + 1$.

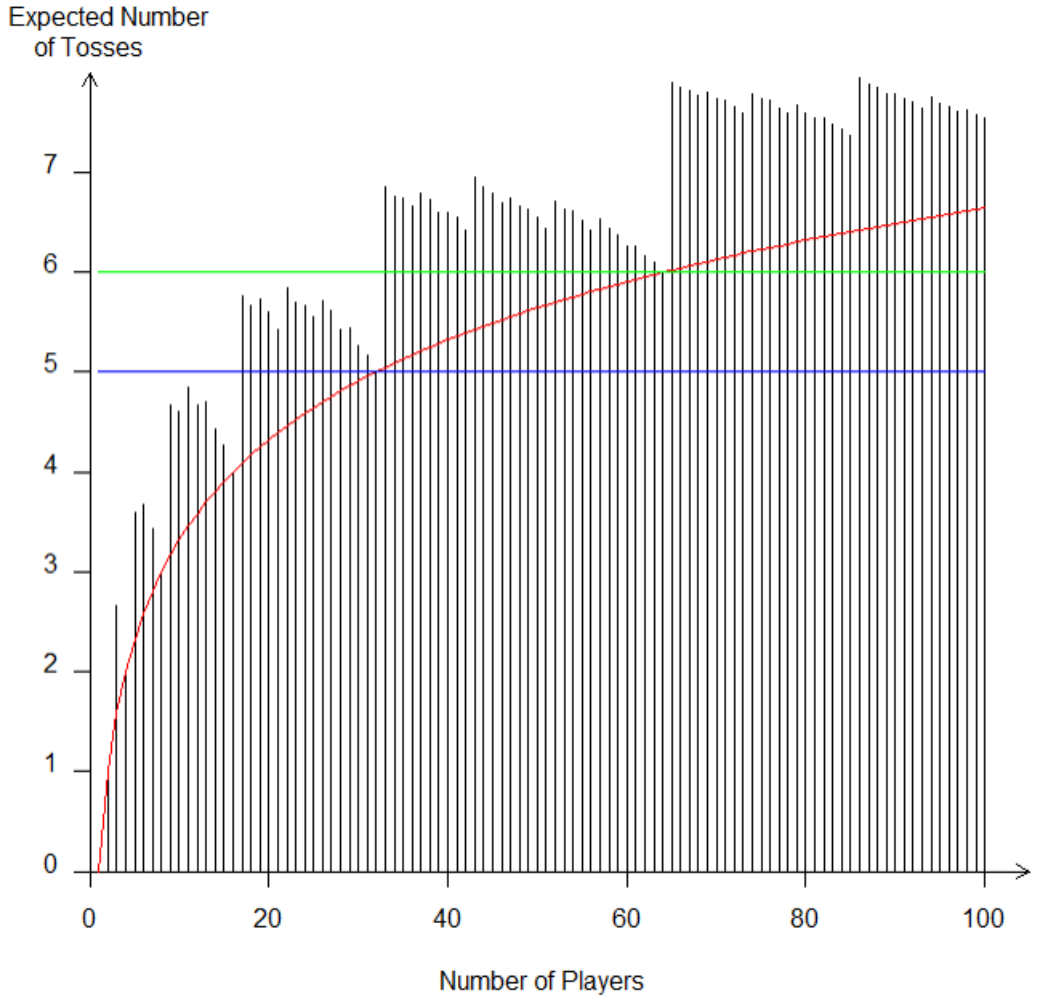


Figure 1: The average number of tosses of a fair coin until a winner is determined among $n \leq 100$ friends, giving each friend an equal chance of winning.

3 To Efficiently Determine the Winner Equitably Using an Unbiased Die

As a randomizer, an unbiased die is more informative than an unbiased coin. Any efficient algorithm that uses an unbiased coin can be implemented using an unbiased die: Classify any three outcomes as 1 and the remaining three as 0. For example, define an odd outcome as 1, and an even outcome as 0. Hence, an unbiased die is an unbiased coin (and much more). In particular, the expected number of rolls of a fair die to determine a winner among two contestants is $\theta_2 = 1$. Moreover, an unbiased die often offers more efficient algorithms requiring fewer rolls than an unbiased coin. For example, to determine the prize winner from among three hikers, one roll of a fair die suffices,

since each person can be allocated exactly two outcomes of the fair die: For example, 1 and 4 to A; 2 and 5 to B; 3 and 6 to C. Thus, the expected number of rolls of a fair die is to choose a winner equitably from among three hikers is $\theta_3 = 1$, which is smaller than the expected number of tosses of a fair coin $\mu_3 = 4/3$.

If $n = 4$, roll the fair die once; allocate outcomes 1 through 4 one to each of the four people; if the outcome is 5 or 6 then roll the die again to form 12 outcomes and allocate three outcomes to each person: For example, allocate (5, odd) to A; (5, even) to B; (6, odd) to C and (6, even) to D. Thus, one roll suffices with probability 2/3 and two rolls with probability 1/3, for an expected number of rolls $\theta_4 = 4/3$, which is smaller than $\mu_4 = 2$.

If $n = 5$, roll the fair die once; allocate outcomes 1 through 5, one to each of the five people; if the outcome is 6, then start over. Hence, the number of rolls of a fair die until a winner is determined is a geometric random variable with success probability 5/6. In particular, $\theta_5 = 1 + (1/6) * \theta_5 = 6/5 = 1.2$, which is smaller than $\mu_5 = 3.6$. The case for $n = 6$ is trivial with $\theta_6 = 1$. For $n = 7$, after one roll of the fair die, we cannot assign any outcome to anyone since then the assignee will have 1/6 (which is more than 1/7) chance of being the winner. Instead, we must roll the die twice; allocate to each person any set of five outcomes; and only if the single unallocated outcome occurs, restart the process. Hence, $\theta_7 = 2 + \theta_7/36 = 72/35 = 2.057143$.

For $n = 8$, we must roll the die twice; allocate to each person a disjoint set of four pairs of outcomes; and only if one of the four unallocated pairs of outcomes occurs, roll the die a third time to create $4 * 6 = 24$ possible triplets of outcomes, of which we can assign three triplets to each person. Hence, $\theta_8 = 2 + (4/36) * 1 = 19/9$. The case for $n = 9$ is trivial, since 9 divides 6^2, with $\theta_9 = 2$. For $n = 10$, we must roll the die twice; allocate to each person a disjoint set of three pairs of outcomes; and only if one of the six unallocated pairs of outcomes occurs, roll the die a third time to create $6 * 6 = 36$ possible triplets of outcomes, of which we can assign three triplets to each person; etc. Hence,

$$\theta_{10} = 2 + \frac{6}{36} * (\theta_{10} - 1) = \frac{11}{6} * \frac{6}{5} = 2.2.$$

The case for $n = 11$ is faced by the scouts and the scout master. When $n = 11$, roll the die twice so that three outcomes can be assigned to each person. In case one of the three unassigned outcomes happens, roll the die again; allocate one outcome to each person. And so on, until the process stops or renews. The number of outcomes at various stages are as follow (recall the arrow notation from the previous section):

$$6^2 \to 3 * 6 \to 7 * 6 \to 9 * 6 \to 10 * 6 \to 5 * 6 \to 8 * 6 \to 4 * 6 \to 2 * 6 \to 1.$$

The expected number of tosses until a winner is determined is

$$\theta_{11} = 2 + \frac{3}{6^2} + \frac{7}{6^3} + \frac{9}{6^4} + \frac{10}{6^5} + \frac{5}{6^6} + \frac{8}{6^7} + \frac{4}{6^8} + \frac{2}{6^9} + \frac{\theta_{11}}{6^{10}} = \frac{128436780}{6^{10} - 1} = 2.12411.$$

which is much less than $\mu_{11} = 4.85$. Each person's probability of winning is

$$\left(\frac{3}{6^2}+\frac{1}{6^3}+\frac{3}{6^4}+\frac{4}{6^5}+\frac{5}{6^6}+\frac{2}{6^7}+\frac{4}{6^8}+\frac{2}{6^9}+\frac{1}{6^{10}}\right)\left[1+\frac{1}{6^{10}}+\left(\frac{1}{6^{10}}\right)^2+\ldots\right]$$
$$=\left(\frac{5496925}{6^{10}}\right)\frac{6^{10}}{6^{10}-1}=\frac{1}{11}.$$

The case for n=12 is also trivial, since 12 divides 6^2, with $\theta_{12} = 2$. More generally, when there are a composite number of friends $n = p * 6^m$, where 6 does not divide p, then split them into 6^m teams of p friends each. Determine the winning team in m rolls, and then determine the winning member of the winning team, needing on average a total of $\theta_{p*6^m} = m + \theta_p$ tosses and giving each person a probability $(1/6^m)(1/p) = 1/(p * 6^m) = 1/n$ of winning. For example, $\theta_{12} = 1 + \theta_2 = 2; \theta_{18} = 1 + \theta_3 = 2;\ \theta_{24} = 1 + \theta_4 = 7/3;$ and $\theta_{30} = 1 + \theta_5 = 2.2$. Therefore, it suffices to solve the problem for non-multiples of 6.

When $n = 13$, roll the die twice so that two outcomes can be assigned to each person. In case one of the ten unassigned outcomes happen, roll the die again; allocate four outcomes to each person. And so on, until the process stops or renews. The number of outcomes at various stages are as follow:

$$6^2 \to 10 * 6 \to 8 * 6 \to 9 * 6 \to 2 * 6^2 \to 7 * 6 \to 3 * 6 \to 5 * 6 \to 4 * 6 \to 11 * 6 \to 1.$$

The expected number of tosses until a winner is determined is

$$\theta_{13} = 2+\frac{10}{6^2}+\frac{8}{6^3}+\frac{9}{6^4}+\frac{2*2}{6^5}+\frac{7}{6^7}+\frac{3}{6^8}+\frac{5}{6^9}+\frac{4}{6^{10}}+\frac{11}{6^{11}}+\frac{\theta_{11}}{6^{12}}=\frac{842523983}{6^{12}-1}=2.3223.$$

which is much less than $\mu_{13} = 4.707692$. Again, note that the powers of 6 in the denominators of the summands above keep track of the number of rolls made so far. Each person's probability of winning is

$$\left(\frac{2}{6^2}+\frac{4}{6^3}+\frac{3}{6^4}+\frac{4}{6^5}+\frac{5}{6^7}+\frac{3}{6^8}+\frac{1}{6^9}+\frac{2}{6^{10}}+\frac{1}{6^{11}}+\frac{5}{6^{12}}\right)\times$$
$$\left[1+\frac{1}{6^{12}}+\left(\frac{1}{6^{12}}\right)^2+\ldots\right]=\left(\frac{167444790}{6^{12}}\right)\frac{6^{12}}{6^{12}-1}=\frac{1}{13}.$$

Again, we suggest the readers manually compute θ_{14}, θ_{15}, θ_{16}, θ_{17}, etc. and check the probability of winning for each person is $1/n$. R codes to compute μ_n and each contestant's winning probability are given in Appendix 2. Figure 2 shows the expected number of tosses needed to choose the winner by rolling an unbiased die.

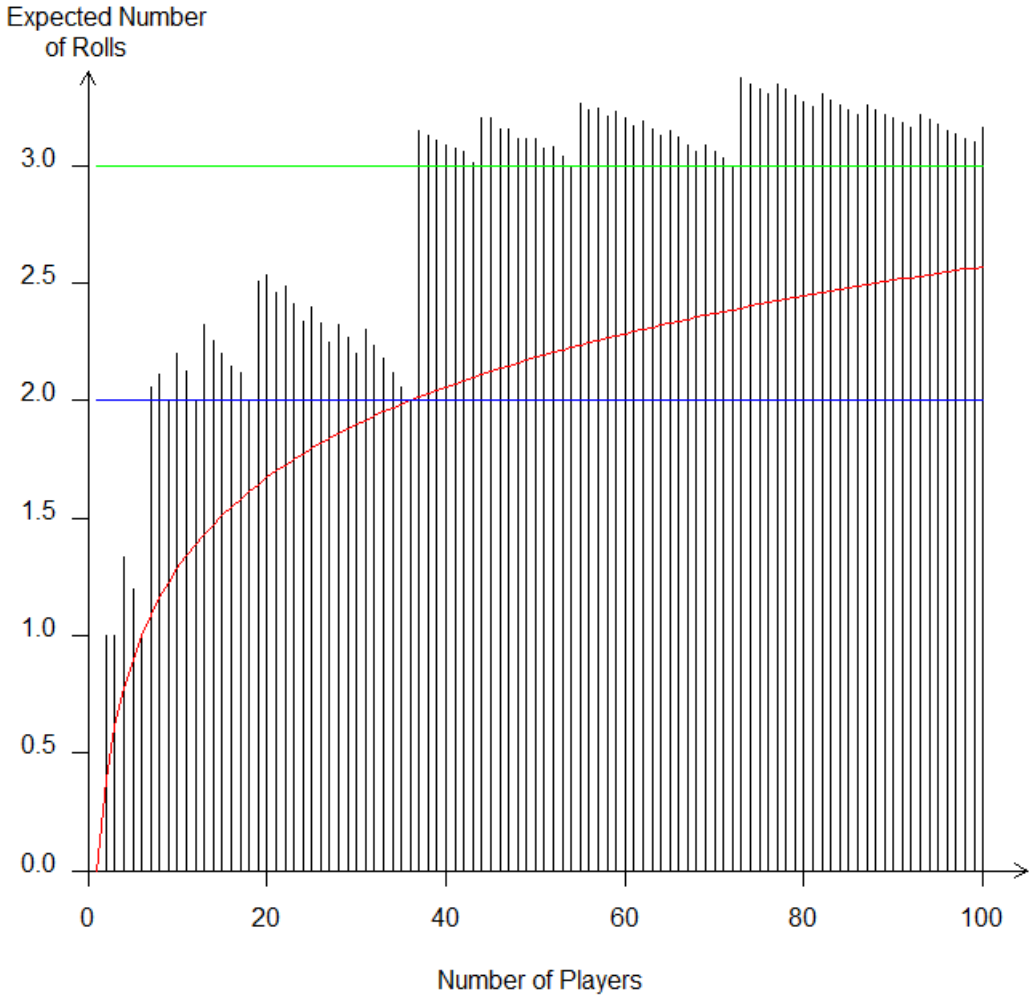


Figure 2: The average number of rolls of a fair die until a winner is determined among $n \leq 100$ participants, giving each participant an equal chance of winning.

4 Extracting Unbiased Bits from a Biased Coin

A fair coin is a utopia. It is much more realistic to assume that the coin is potentially biased. How can a biased coin be used to allocate a prize to n people equitably? Based on the strategy discussed in section 2, such an allocation is possible, if a sequence of independent and unbiased bits can be extracted from the outcomes of a biased coin. That such a thing is possible is quite startling "nay, mind boggling" to the common person. Nonetheless, the solution as stated by von Neumann (1951) is not complicated at all!

(Regular) von Neumann Procedure (to extract fair bits out of a biased coin):

> Suppose that the coin has probability of heads $P(1) = p$ and tails $P(0) = q = 1 - p$, where $0 < p < 1$ is unknown. Toss the coin twice: If the biased coin yields 01, which happens with probability qp, then define a new bit 1; and if the biased coin yields 10, which happens with probability pq, then define a new bit 0. On the other hand, if the biased coin yields 00 or 11, skip the attempt and define no new bit. Then repeat the process with new pairs of tosses. The new bits so defined form a sequence of independent and unbiased bits.

As simple as von Neumann's procedure is, a more efficient algorithm (that utilizes additional randomness inherent in the independent outputs of the biased coin) was found by Hoeffding and Simons (1970), and Perez (1992) proved the procedure optimal in the sense that the long run proportion of fair bits extracted per toss of the biased coin is the highest possible and equals the entropy function

$$H(p) = -\,p \log_2 p - q \log_2 q. \tag{3}$$

In particular, $H(1/2) = 1$, so that outcomes of a fair coin (though it is not known to be fair) yield about as many fair bits. The optimal procedure is appropriately named the iterated von Neumann procedure. Below we describe it.

Iterated von Neumann Procedure (IvNP):

1. Read the outputs of the biased coin in pairs.

2. Apply the regular von Neumann Procedure to the pairs (discarding the last toss, if unpaired) to extract (with probability $2pq$) a W-sequence of fair bits.

3. Create a new U-sequence of outcomes by writing down a 1 when the given pair is 11 and a 0 when the pair is 00. A U-outcome indicates which value occurred in a matching pair and contributed nothing to the W-sequence, and it equals 1 with probability $p^2/(p^2 + q^2)$ and 0 with probability $q^2/(p^2 + q^2)$.

4. Create another new V-sequence by writing down a 1 if the given pair is 01 or 10, and a 0 if the pair is 00 or 11. A V-outcome indicates whether or not a fair bit is extracted from the pair, and it equals 1 with probability $2pq$ and 0 with probability $p^2 + q^2$.

5. Apply steps 1–4 to the U-sequence, and the V-sequence and augment the new fair bits to the W-sequence obtained in step 2.

Perez (1992) proves the optimality of IvNP as follows: As reasoned in items (2)–(4), the rate of return function, reporting the long run proportion $r(p) = r(q)$ of fair bits

extracted per biased toss, satisfies $r(p) \leq H(p)$ given in (3), and

$$r(p) = pq + \frac{1}{2}\left(p^2+q^2\right)\ r\left(\frac{p^2}{p^2+q^2}\right) + \frac{1}{2}r\left(p^2+q^2\right). \tag{4}$$

However, since the entropy function $H(p)$ satisfies both the boundary conditions $H(0) = 0 = H(1)$ and the iterative condition (4), we must have $r(p) = H(p)$.

Note that the discarded unpaired outcomes cannot be augmented together to produce more fair bits as these unpaired outcomes have different probabilities of being 1. Note also that while $r(p)$ is the long run proportion of fair bits extracted per biased toss, for any finite sequence of biased tosses, the actual proportion of unbiased bits extracted maybe far less than $r(p)$. For instance, in Table 1, starting from 20 tosses of a biased coin with $p = .6$, we obtain the W(U)[V] sequences; then from the U (and V) sequences we obtain additional bits iteratively. All discarded outcomes are marked with an x underneath. Ultimately, we obtain only 12 fair bits (which are dropped down to the bottom row), even though $H(.6) = .9710$.

Table 1: Illustrating IvNP for a small sequence of unbiased bits

```
  biased: 10 11 11  01   11    10 10 00 01  10
W(U)[V]: 01 00 10 (11   10)  [10 01 01 10  11]
                   0(1)[01]   01 10(1)[11  11 0]
                     x  1[1]        x (11)[00]x
                          x            xx  xx
unbiased: 01 00 10  0    1     01 10
```

The R codes to implement IvNP are given in Appendix 3. The codes include a simple statistical test to check whether the algorithm indeed produces fair bits starting from unfair bits: In one application of the IvNP algorithm, with seed 99, starting from 1000 tosses of a biased coin with p=.6, we obtained 837 fair bits. For the 1000 generated biased outcomes (with 606 successes) a 95% confidence interval for the proportion of 1 was found to be $(0.5749, 0.6363)$; but for the 837 extracted bits (with 421 successes) the confidence interval was found to be $(0.4686, 0.5374)$, and the null hypothesis that the extracted bits are fair (against the two-sided alternative hypothesis) achieved a P-value of 0.890.

Starting from B tosses of a biased coin with p=.6, the number of fair bits extracted is a random variable, whose summary features we document by repeating the IvNP 100 times. The results shown in Table 2 demonstrate that as B increases, $r(p)$, the mean proportion of fair bits per biased toss, increases to $H(.6) = 0.9710$.

Table 2: The mean proportion of fair bits per biased toss increases to $H(p)$

```
p=.6, iter=100
  Min. 1st Qu.  Median    Mean    3rd Qu.    Max.   S.D.       B
    61      68      70    70.4       73      78   3.60     100
   799     830     837   835.5      842     856  10.45    1000
  8961    9018    9036    9037     9056    9119  30.79   10000
 93514   93685   93737   93735    93785   93979  87.99  100000
953759  954245  954436  954414   954611  954994 267.70 1000000
```

5 Extracting Equiprobable Outcomes from a Biased Die

If a fair coin is a utopia, then a fair die exists only in a fool's paradise. Every die is potentially biased with unknown probabilities of its six faces. How can we extract a sequence of fair rolls starting from a sequence of biased rolls? One simple way to do so is to extend von Neumann's procedure.

Extended (Regular) von Neumann Procedure (for a six-sided die):

> Roll the biased die three times. If the three outcomes are not all distinct, then skip and restart the process. When the three outcomes (x_1, x_2, x_3) are distinct, then define a new roll to be 1 through 6 according as the relative ranks of (x_1, x_2, x_3) is positioned among its six sorted permutations (in increasing order when thought of as a three-digit number): 123=1, 132=2, 213=3, 231=4, 312=5, 321=6. Since the six permutations are equiprobable, the new roll is equally likely to be 1 through 6; hence, fair.

For example, if $(x_1, x_2,\ x_3) = (3,\ 2,\ 5)$, then their relative ranks are (2, 1, 3); hence, define the new roll to be 3 (the third in the sorted list). Likewise, if $(x_1, x_2,\ x_3) = (4,\ 6,\ 1)$, then their relative ranks are (2, 3, 1); hence, define the new roll as 4 (the fourth in the sorted list).

Improvements are possible: (1) If the first two outcomes are identical, then why roll the biased die a third time only to discard the triplet? We might as well discard the two outcomes and start over. (2) Alternatively, we might choose to roll the biased die a third time (whether or not the first two outcomes are tied). If the three outcomes of the biased die consist of two distinct numbers, say (u, v, v) with $u \neq v$, we do not have to discard these outcomes altogether. Instead, we roll the die twice more. If we get distinct values (s, t) with $s \neq t$, (here s and/or t may tie with u or v) then we define a new outcome of a roll as follows:

$$(uvv, s < t) = 1,\ (vuv,\ s < t) = 2,\ (vvu,\ s < t) = 3,$$
$$(uvv, s > t) = 4,\ (vuv,\ s > t) = 5,\ (vvu,\ s > t) = 6.$$

If $s = t$, then we discard all five outcomes and start over. In either case, we save one roll of the biased die. Likewise, if the first three outcomes of the biased die are identical, we may wait until two other sets of three outcomes are also identical. Then combine these three triplets, if they are distinct, to define a new fair roll. Thus, the extended von Neumann procedure is inadmissible (there is a better procedure with a smaller expected length of the input sequence to determine one output value).

We won't pursue more efficient extraction of fair rolls from a biased die. Our purpose of assigning prizes equitably to n people is quite well served as soon we can extract, even if inefficiently, a sequence of fair rolls starting from the outcomes of a biased die. Thereafter, we simply augment the ideas of section 3 to determine the winner among n people based on a sequence of fair rolls of a die. Once a winner is determined, we repeat the process to determine another winner (with replacement)*ad infinitum*.

It is worth mentioning that if we replace the biased six-faced die by a biased 12- or 20-faced die (that is, if we use a dodecahedron or an icosahedron or a spinner with 12 or 20 values), then it is much more likely that the three rolls will be distinct. Hence, with a higher chance a triplet of rolls will yield an equiprobable value between 1 and 6, both inclusive.

6 Extracting Equiprobable Outcomes from a Biased Spinner

We must emphasize that when we split into two parts the problem of determining the winner from among $b \geq 2$ participants using a randomizer that yields $a \geq 2$ outcomes — first extract unbiased sequence of a digits as in section 4 and 5, and then determine the winner from among b participants using this extracted sequence as in section 2 and 3 — we obtain less than an efficient procedure. It is better to determine the winner directly by extracting an outcome with b equiprobable values out of the original biased sequence of a outcomes.

To appreciate the above remark, let us return to the hikers' problem. In the two-part algorithm, $\bar{\beta}_3$, the overall expected number of biased tosses needed until a winner among 3 hikers is determined approximately equals the product of the expected number

of biased tosses needed until an unbiased bit is extracted and the expected number of unbiased tosses needed until a winner among 3 hikers is determined. Therefore,

$$\bar{\beta}_3 \approx 2\left[1+\left(p^2+q^2\right)\right]\cdot\left[1+\left(p^{2^2}+q^{2^2}\right)+\left(p^{2^3}+q^{2^3}\right)+\ldots\right]*(8/3)\,. \tag{5}$$

We can reduce the expected number of biased tosses needed until a winner among 3 hikers is determined by using a single-part algorithm, suggested by Hoeffding and Simons (1969), as follows: Toss the biased coin three times; allocate $\{HTT, THH\}$ to A; $\{THT, HTH\}$ to B and $\{TTH, HHT\}$ to C. Thus, with three tosses, each hiker has a $pq^2+qp^2=pq\leq 1/4$ chance of winning. Only if $HHH=H^3$ or $TTT=T^3$ occurs, with probability (p^3+q^3), toss the coin three more times, and follow the above recipe of allocating outcomes to the three hikers. This adds an extra $(p^3+q^3)\,pq$ chance of winning to each hiker. Only if H^3 or T^3 occurs during both the first and the second triplets, with probability $(p^3+q^3)^2$, toss the coin three more times, and follow the above recipe of allocation of outcomes to the three hikers, adding an extra $(p^3+q^3)^2\,pq$ chance of winning to each hiker. If H^3 or T^3 occurs also during the third triplet, without having to toss the coin anymore, allocate additional outcomes to the three hikers: $\{H^3T^3T^3, T^3H^3H^3\}$ to A; $\{T^3H^3T^3, H^3T^3H^3\}$ to B and $\{T^3T^3H^3, H^3H^3T^3\}$ to C. Only if H^9 or T^9 occurs, continue. Etc. For this single-part algorithm, the expected number of biased tosses needed until a winner is determined among three hikers is

$$\widetilde{\beta}_3 = 3\left[1+\left(p^3+q^3\right)+\left(p^3+q^3\right)^2\right]\cdot\left[1+\left(p^{3^2}+q^{3^2}\right)+\left(p^{3^3}+q^{3^3}\right)+\ldots\right]$$

which is smaller than $\bar{\beta}_3$ given in (5) for all $0<p<1$, as shown in Figure 3 (with R codes given in Appendix 4).

The above example illustrates that we must extract $b\geq 2$ equiprobable outcomes starting directly from a biased spinner having $a\geq 3$ possible outcomes with unknown probabilities. Extending the Hoeffding-Simons (1970) procedure, Dwass (1972) extracts $b\geq 2$ equiprobable outcomes using an independent sequence of discrete random variables on a countable set. To implement the rule, one must ahead of time set up an elaborate listing of all possible sequences of outcome that will lead to extraction of an equiprobable outcome and designate which outcome that ought to be. Here we give a much simpler (*albeit* less efficient) procedure that requires no initial set up. In fact, it may be called an extended (regular) von Neumann algorithm. To avoid cumbersome notation, we will illustrate the method for some choices of $a,\ b\ \geq 3$.

Example 1: A spinner can stop at $a=11$ different positions with unknown probabilities. Starting from a sequence of outcomes of this biased spinner, how can we extract a sequence of fair values of an unbiased spinner that takes on $b=19$ possible values?

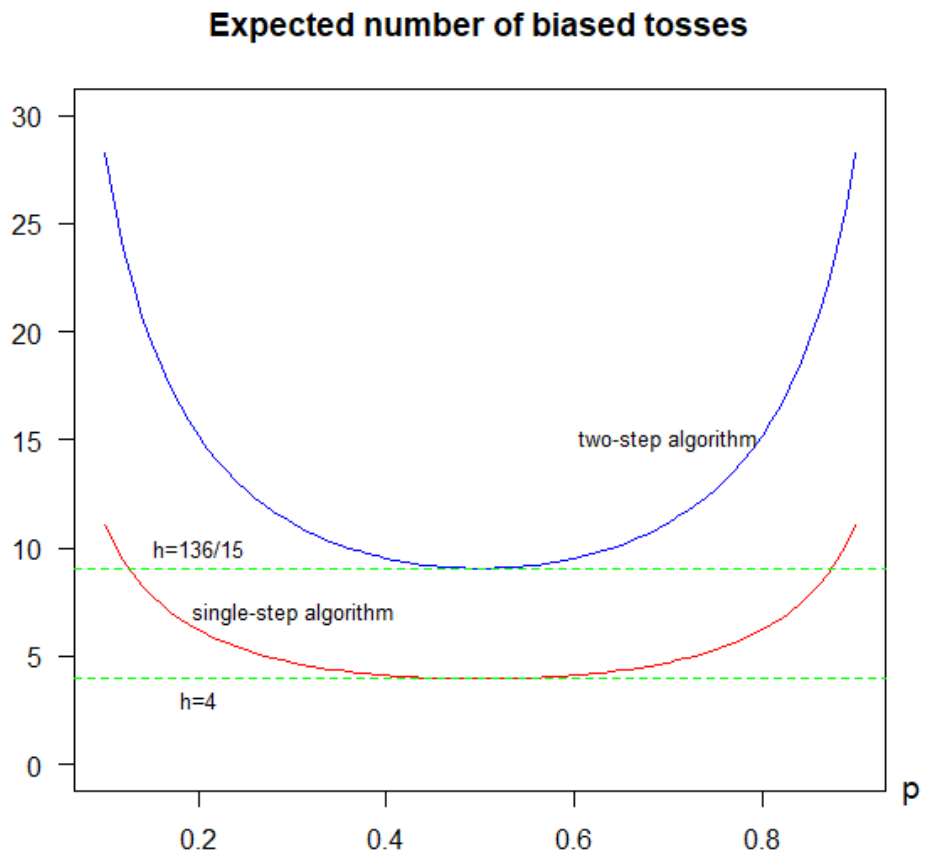


Figure 3: The expected number of biased tosses needed to equitably choose a winner from among three hikers is uniformly smaller when a single-step algorithm is used than when a two-step algorithm is used

Extended (Regular) von Neumann Procedure (for a spinner) (To extract one of 19 equiprobable values from a biased 11-outcome spinner):

> Since $4! = 24 \geq 19$, start with four independent outcomes of the spinner. If the four outcomes (x_1, x_2, x_3, x_4) are distinct, then define a new roll to be 1 through 24 according as the relative ranks of (x_1, x_2, x_3, x_4) is positioned among its 24 naturally sorted permutations (with positional values in increasing order). Since these 24 permutations are equiprobable, the new roll is equally likely to be 1 through 24: hence, fair. But if the new roll exceeds 19 or if the four outcomes are not all distinct, then skip and restart the process.

For example, if the outcomes are $(7,\ 3,\ 9,\ 2)$, then their relative ranks are $(3,\ 2,\ 4,\ 1)$

and the roll extracted is $(3-1)*3!+(2-1)*2!+(2-1)*1+1=16$. We beg your pardon for apparently pulling this conversion rule out of thin air. Please give us a moment to explain: The leading digit 3 exceeds two other digits 1 and 2, and the remaining three digits can be rearranged in 3! ways; hence, we get the first summand $(3-1)*3!$. Next, we rewrite the remaining three digits (2, 4, 1) in terms of their relative ranks as (2, 3, 1), and repeat the previous step: The leading digit 2 exceeds the digit 1, while the other two digits can be rearranged in 2! ways; hence, we get the second summand $(2-1)*2!$. We rewrite the remaining two digits (4, 1) in terms of their relative ranks as (2, 1), whence we get the third summand $(2-1)*1!$. Finally, the single remaining digit contributes the fourth summand 1.

Example 2: Next, suppose that the roles of 11 and 19 in the above example are interchanged. That is, consider a spinner that can stop at $a=19$ different positions with unknown probabilities. Starting from a sequence of outcomes of this biased spinner, how can we extract a sequence of fair values of an unbiased spinner that takes on $b=11$ possible values?

Start with four independent outcomes of the spinner. If they are all distinct, replace them by their relative rank and define a new roll value as its relative position (modulo 12) among all 24 sorted permutations of (1, 2, 3, 4), provided it is between 1 and 11, both inclusive; otherwise (that is, if it is 12 or 24), then discard and start over. For example, if the outcomes are (7, 17, 13, 2), then their relative ranks are (2, 4, 3, 1) and the proposed extracted value $1*3!+2*2!+1*1!+1=12$ must be discarded. But if the outcomes are (7, 5, 2, 13), then their relative ranks are (3, 2, 1, 4) and the roll extracted is $2*3!+1*2!+0*1!+1=15=3 \pmod{12}$.

7 Discussion and Open Questions

A friend (a medical technician by profession) suggested a simple idea to ensure that the input sequence generated by the tosses of a biased coin will be converted to an output sequence containing 0 and 1 in about equal proportion: "Switch the outcomes (0 to 1 and 1 to 0) occurring at even positions; keep unchanged the outcomes at odd positions." We agree that the proportions of 0 and 1 will be about equal. However, we must declare that the output sequence will not be unbiased, for unbiasedness means much more than balancing the proportions of 0 and 1, it means that all k-tuples (for $k \geq 1$) are equally likely. For this "switch-even-positions" transformation, if we read two successive outputs in odd and even positions, then $P(00)=qp=P(11), P(01)=q^2, P(10)=p^2$, which are not 1/4 each; and if we read two successive outputs in even and odd positions, then $P(00)=pq=P(11), P(01)=p^2, P(10)=q^2$, which are not 1/4 each either. Furthermore, the output sequence must be independent, which means irrespective of

what the previous outputs have been, the probability distribution for the next output must remain unchanged. Here, on the other hand, we have

$$P\left(x_{2i}=0|\ x_{2i-1}=0\right)=P\left(x_{2i}=0|\ x_{2i-1}=1\right)=p\neq 1/2,$$
$$P\left(x_{2i+1}=0|\ x_{2i}=0\right)=P\left(x_{2i+1}=0|\ x_{2i}=1\right)=q\neq 1/2.$$

Just as any coin is suspected to be possibly biased, so can the input sequence generated by tossing the coin can be suspected to be possibly dependent. Samuelson (1968) shows that if the input sequence exhibits a simple Markov dependence of each toss upon the outcome of the previous toss, then by concentrating only on tosses that follow a common outcome, one can extract independent, equally probable binary bits. Elias (1972) extended the von Neumann (1951) and Hoeffding and Simons (1970) results from independent binary input sequence to finite-states generated by a stationary process to extract independent b-valued output sequences. Whereas we described only an extended von Neumann procedure, which we also demonstrated to be inadmissible, we leave to the reader to discover more efficient procedures (and, if possible, the optimal procedure) to extract a sequence of independent, unbiased rolls starting from the outcomes of a biased die, or more generally, to extract b equiprobable outcomes using a biased spinner generating a distinct values with unknown probabilities.

Acknowledgements

We thank our colleagues with whom we had some productive lunchtime conversations on rolling dice. We also thank an anonymous referee and the editor.

Bibliography

[1] Dwass, M. (1972). Unbiased coin tossing with discrete random variables, *The Annals of Mathematical Statistics*, 43(3), 860–864.

[2] Elias, P. (1972). The efficient construction of an unbiased random sequence, *The Annals of Mathematical Statistics*, 43(3), 865–870.

[3] Hoeffding, W. and Simons, G. (1970). Unbiased coin tossing with a biased coin, *The Annals of Mathematical Statistics*, 41(2), 341–352.

[4] Perez, Y. (1992). Iterating von Neumann's procedure for extracting random bits, *The Annals of Statistics*, 20(1), 590–597.

[5] Samuelson, P.A. (1968). Constructing an unbiased random sequence, *Journal of American Statistical Association*, 63, 1526–1527.

[6] Von Neumann, J. (1951). Various techniques used in connection with random digits, *National Bureau of Standards Applied Mathematics Series*, 12, 36–38.

[7] Wikipedia: The Free Encyclopedia. (a) Fermat's little theorem; (b) Euler's theorem.

8 APPENDIX

8.1 R codes for Figure 1

```
### Allocating a prize equitably to n contestants using a fair coin
### for 1<=n <=1000

# the smallest power of 2 that reaches or exceeds n>=3
sp2=function(r, n){
   for(i in 1:10){if (n<=r*2^i){m=i; break}}
   c(m, (r*2^m)%/%n, (r*2^m)%%n) }

# compute power of 2=#tosses, quotient, and remainder
pqr=function(r, n){
   if(n%%2 == 0){print("work with n/2")}
   else{
      for(i in 1:n){
         mat=cbind(mat, sp2(r, n))
         r=mat[3, ncol(mat)]
         #if(r==0){print("n is a power of 2"); break}
         if(r==1){break}
      } # end for
      as.matrix(mat[,-1], nrow=3)
   } # end else
}

## expected number of tosses until winner
mat=matrix(c(0,0,1), nrow=3)
ent=function(n){
if(n==1){out=0}
if(n>=2){
   if(n%%2 == 0){out=ent(n/2)+1}
   else{
      mat=matrix(c(0,0,1), nrow=3);
```

```
      (out=pqr(1,n))
      p=out[1,]
      q=out[2,]
      r=out[3,]
      P=cumsum(p)
      (one=n*sum(q/2^P)) # 1=n*probability of win
      Pr=rev(cumsum(rev(p)))
      l=length(p)
      out=sum(2^Pr*p*c(1,r[-l]))/(2^Pr[1]-1) # expected #tosses
   } # end else
} # end if
out
}

## For n=1, 2, 3, ..., 100 print expected number of tosses
ev=c(0); for(i in 2:100){ ev=c(ev, ent(i)) }; ev
n=seq(1, 100)

plot(n, ev, type='h', las=1, xlab="Number of Players",
    frame.plot=FALSE, ylab="",xaxt="n", yaxt="n", xaxs="i", yaxs="i")
text(0,8.5, expression("Expected Number \nof Tosses"), xpd=TRUE)

arrows(0,0,105,0, code = 2, xpd = TRUE, length=.10)
arrows(0,0,0,8, code = 2, xpd = TRUE, length=.10)

axis(1, pos=0, at=c(seq(0,100,20)),padj=0,cex=1,xpd=T)
axis(2, pos=0, at=c(seq(0,8,1)),padj=0,cex=1, las=1)
points(n, log2(n), type='l', col="red")
abline(h=5, col="blue")
abline(h=6, col="green")
```

8.2 R codes for Figure 2

```
### Allocating a prize equitably to n contestants using a fair die
### for n >= 2

# the smallest power of 6 that reaches or exceeds n
sp6=function(r, n){
   if(r==0){c(0, 0, 0)}
   else{
   for(i in 1:n){if (n<=r*6^i){m=i; break}}
   c(m, (r*6^m)%/%n, (r*6^m)%%n)
```

```
   } # end else
}

# compute power of 6=#tosses, quotient, and remainder
mat=matrix(c(0,0,1), nrow=3);
pqr=function(r, n){ rv=c()
   if(n%%6 == 0){print("work with n/6")}
   else{
     for(i in 1:n){
         mat=cbind(mat, sp6(r, n))
         r=mat[3, ncol(mat)];
   if(r %in% rv){rv=c(rv,r); break};
   rv=c(rv,r)
    } # end for
   mat[,-1]
   } # end else
}

## expected number of tosses until winner
enr=function(n){
if(n==1){ex=0}; if(n==2){ex=1}; if(n==3){ex=1}
if(n>=4){
   if(n%%6 == 0){ex=enr(n/6)+1}
   else{ mat=matrix(c(0,0,1), nrow=3);
      out=pqr(1,n); print(out)
      p=out[1,]; q=out[2,]; r=out[3,]; P=cumsum(p)
     m=ncol(out); k=match(r[m], r)
      p1=p[1:k]; p2=p[(k+1):m];
     P1=rev(cumsum(rev(p1))); P2=rev(cumsum(rev(p2)))
     r1=c(1); if(k>1){r1=c(1, r[1:(k-1)])}; r2=r[k:(m-1)]
     if(r2[1]==0){ex=sum(r1*p1*6^P1)/6^P1[1]}
     else{
      rec=sum(r2*p2*6^P2)/(6^P2[1]-1); print(rec/r1[k]) # mu_k
      ex=(rec+sum(r1*p1*6^P1))/6^P1[1] # expected #tosses
     q1=q[1:k]; q2=q[(k+1):m];
      one=n*(sum(q1/6^cumsum(p1))+sum(q2/6^(P[k]
            +cumsum(p2)))/(1-6^P[k]/6^P[m]));
     print(one) # 1=n*probability of win
     } # end else
 } # end else
} # end if
ex
}
```

```
## For n=1, 2, 3, ..., 100 print expected number of tosses
ev=c(0); for(i in 2:100){ ev=c(ev, enr(i)) }; ev
n=seq(1, 100)

plot(n, ev, type='h', las=1, xlab="Number of Players",
    frame.plot=FALSE, ylab="",xaxt="n", yaxt="n", xaxs="i", yaxs="i")
text(-.5,3.5, expression("Expected Number \n of Tosses"),xpd=TRUE)

arrows(0,0,105,0, code = 2, xpd = TRUE, length=.10)
arrows(0,0,0,3.4, code = 2, xpd = TRUE, length=.10)
axis(1, pos=0, at=c(seq(0,100,20)),padj=0,cex=1,xpd=T)
axis(2, pos=0, at=c(seq(0,3.6,.5)),padj=0,cex=1, las=1)
points(n, log(n)/log(6), type='l', col="red")
abline(h=2, col="blue")
abline(h=3, col="green")
```

8.3 R Codes for Iterated von Neumann Procedure

```
# initialize
p=.6 # to generate biased tosses (thereafter, forget the value)
b=100 # block size, nb of tosses of the biased coin
biased=rbinom(b, 1, p) # generate tosses of a biased coin

# generate unbiased bits from biased tosses (von Neumann)
bit=function(x){  w=c()
  if(length(x)>1){
    l=floor(length(x)/2)
    for(i in 1:l){
      if(x[2*i-1]==0 & x[2*i]==1){w=c(w,1)}
      if(x[2*i-1]==1 & x[2*i]==0){w=c(w,0)}
    } # end for
  } # end if
  w
} # end function

# save randomness via matched value w.p. p^2/(p^2+q^2)
mat=function(x){  u=c()
  if(length(x)>1){
    l=floor(length(x)/2)
    for(i in 1:l){
      if(x[2*i-1] +  x[2*i]==0){u=c(u,0)}
      if(x[2*i-1] +  x[2*i]==2){u=c(u,1)}
```

```
    } # end for
  } # end if
  u
} # end function

# save randomness via matching indicator w.p. p^2+q^2
dec=function(x){  v=c()
  if(length(x)>1){
    l=floor(length(x)/2)
    for(i in 1:l){
      if(x[2*i-1] + x[2*i]==1){v=c(v,1)}
      else{v=c(v,0)}
    } # end for
  } # end if
  v
} # end function

# augment all unbiased bits from iterations
tuv=function(x){
  out=c()
  if(length(x)>1){
    out=c(bit(x), tuv(mat(x)), tuv(dec(x)))
  } # end if
  out # augmented unbiased bits
} # end function

p=.6; b=10^3
biased=rbinom(b, 1, p) # biased tosses generated
unb=tuv(biased) # fair bits extracted
length(unb) # nb of fair bits extracted

# test proportion/unbiasedness
prop.test(sum(biased), length(biased)) # P{CI includes p}=.95
prop.test(sum(unb), length(unb))  # P{CI includes 1/2}=.95

### Measure efficiency: mean(bitsize)/b
p=.75; b=10^4; N=10^2
bitsize=rep(0,N)
for (i in 1:N){
  biased=rbinom(b, 1, p)
  bitsize[i]=length(tuv(biased))
}
hist(bitsize)
summary(bitsize)
```

```
c(sd(bitsize), b)
```

8.4 R codes for Figure 3

```
# compare two-step algorithm to single-step
p=seq(0.1, 0.9, .01)
q=1-p # *(1+p^9+q^9+p^27+q^27+p^81+q^81)
plot(p, 3*(1+p^3+q^3+(p^3+q^3)^2)/(1-p^9-q^9), type='l',
col="red", ylab="", xlab="", ylim=c(0,30), las=1,
main="Expected number of biased tosses", cex=.8)
#sub="to equitably choose a winner among 3 hikers")
#points(p, 8/(3*p*q), type='l', col="blue")
points(p, (16/3)*(1+p^2+q^2)*(1+p^4/(1-p^4)+q^4/(1-q^4)),
    type='l', col="blue")
text(.7,15,"two-step algorithm", cex=.8)
text(.3,7,"single-step algorithm", cex=.8)
abline(h=4, lty=2, col="green")
abline(h=8*17/15, lty=2, col="green")
text(.2, 3,"h=4", cex=.8)
text(.2,10,"h=136/15", cex=.8)
```

5 Going Beyond Turing: A Fractal Approach Towards Decision Making

Mayukh Mukhopadhyay

Assistant Consultant, Tata Consultancy Services

Doctoral Candidate, IIM Indore
Indore – 453556

Email: mayukh.mukhopadhyay@tcs.com, ef22mayukhm@iimidr.ac.in

Abstract: Scholars in industry and academia have been working to replicate the human brain's capacity for making decisions for a very long time. We have observed in this competition that the logic-driven approach based on the Turing philosophy has been encouraged to mature into a field in terms of machine learning and artificial intelligence. The initial goal of this movement has been relegated into the spiritual and philosophical worlds, despite the fact that it has produced numerous significant advancements that have benefited humanity as a whole. This article traces the recent advances of a non-turing fractal approach capable of brain inspired code-less decision making, pioneered by an interdisciplinary network of Japanese funded Indian scholars in the field of pure mathematics, quantum physics, material science, and neuroscience.

1 Introduction

First conceived in the 1940s, the Turing philosophy posits that everything that happens to humans may be represented by a tape or a set of logically defined processes. All key computing endeavors and algorithmic models adhere to the Turing computing methodology, either by hinting that their model might eventually be reduced to a logic gate or as the outcome of a series of consecutive occurrences. The drawback of Turing tape-based computing is that the entire processing scheme must be painstakingly built as the

result of a step-by-step logical reduction approach, and all potential arguments must be understood in advance. We get "it from bit" because every piece of knowledge that science develops is the result of an experiment that poses a question whose answer is "yes" or "no" (Wheeler, 1990). "Every 'thing' gets its function, its meaning, and its very existence solely from the apparatus-elicited responses to yes or no queries, binary options, and bits" (Tukey, 1984). Even quantum experiments assume that nature's truth could be disclosed as a "yes" or "no" response to a question (Wheeler, 1984). The universe is not merely participative; both the observer and the environment contribute to the universe's information content. Hence the information is geometric (Ghosh, 2014).

Several mathematicians advocate for a fractal universe (Margolus, 2003). No natural event could be recreated by stringing together a series of elementary events. An event could be a closed three-dimensional topological structure (Figure 1). If nature integrates subevents as topology, the path to integrate information would be within and above, never next to one another (Bandyopadhyay, 2020). If we could translate data into a topology, then we could apply pure physics: how topological structures change, violate symmetry, and experience a phase transition. We may therefore employ physicists to predict threats and other decision-making without needing to locate a programmer to decipher concepts in "bits". Then, it is possible to make decisions without coding a single line of an algorithm, just like our brain always does. Geometric entities may be employed in computations (Forrest, 1971; Preparata and Shamos, 1985). However, the usage of geometric shapes as building blocks for mathematical structures did not exist at the time (Bandyopadhyay, 2020).

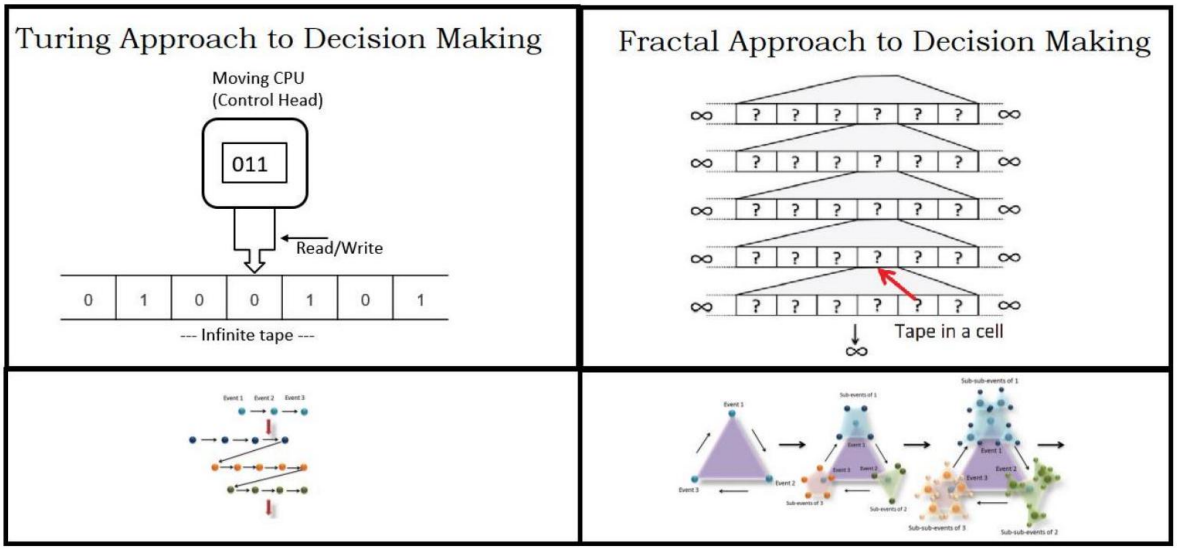


Figure 1. Decision making of events using Turing approach vs fractal approach (Bandyopadhyay, 2020)

Three sections make to the paper's structure. The literature review section provides

an overview of how the structure and properties of microtubules inspired a group of Indianorigin scientists to take up a non-Turing approach to computation, that led to the development of Fractal Information Theory using a universal geometric musical language and a 12-dimensional geometric algebra that generates a space-time-topology-prime metrics which can trigger a self-operating mathematical universe. The application section explores how these experimental outcomes and theoretical models can be used in the fields of business analytics, security for risk mitigation, and healthcare. We conclude with the limitations of this field along with possible avenues of future work.

2 Literature Review

Hameroff and Penrose (2014) have argued against the Hodgkin-Huxley type neuron bursts as the basis for the brain's information processing (1952). According to the Orchestrated objective reduction model proposed by Hameroff and Penrose (2012), cognitive abilities are caused by quantum level activities occurring within neurons, specifically the dynamic instability of microtubules. This contrasts with the Hodgkin-Huxley model, which proposed that cognition results from an increase in neuronal connections. As the brain is an organ that exists at room temperature, quantum phenomena are impractical because they require cryogenic temperatures to generate superconductivity, a claim that has been vehemently questioned by several academics.

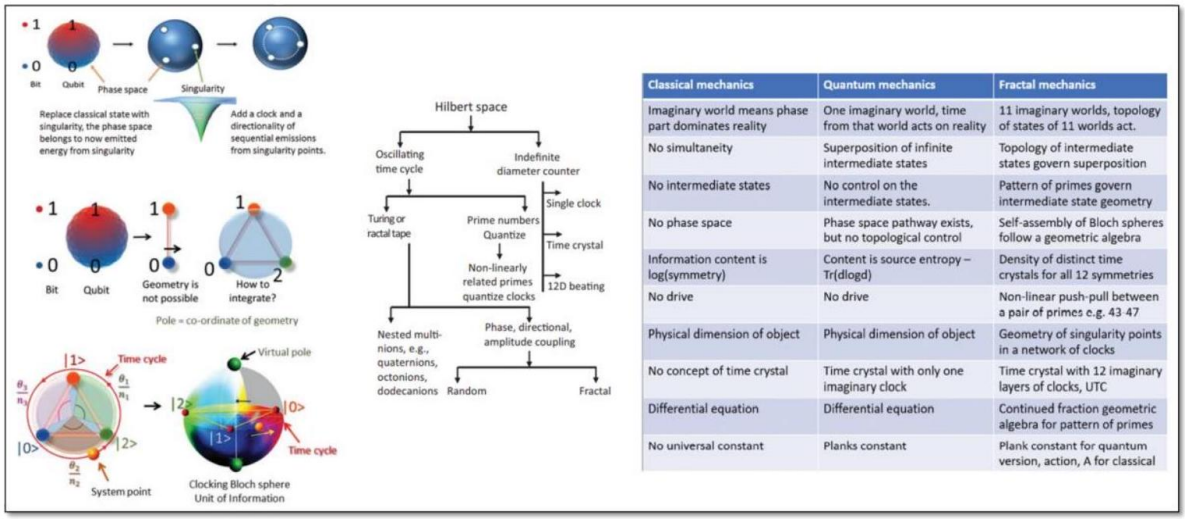
2.1 Microtubules

A single brain microtubule was found to have multi-level memory-switching properties by Bandyopadhyay and his team in 2013 while they were studying supramolecular electronics in the material science lab of Tsukuba, Japan. They also discovered a monomolecular water channel inside the microtubule that allows microtubules with 40,000 tubulins to exhibit conductivity thousand times more than the single tubulin protein (Sahu et. al, 2013). This study experimentally verified the orchestration postulate of Hameroff and Penrose, which states that resonant vibrations can exist amongst all neurons in the entire brain. By traversing the fatty myelin sheath, an axon within a neuron is not required to convey extraordinarily powerful signals wirelessly across the brain. It would just take a small amount of conical radiation or absorption near a neuron's dual polar ends to start a communication chain that would spread across the entire brain (Ghosh et. al., 2014). As a result of these experimental findings about the microtubular properties, a new paradigm for examining information theory evolved, which is presented in the next section.

2.2 Fractal Information Theory

Every natural phenomenon and event have been decoded since the introduction of Existing Information Theory, which consists of Claude Shannon's Classical Information Theory and later Quantum Information theory, as a sequential sum of elementary "true" and "false" decisions, only to find that the length of the sequential chain has increased to astronomical proportions. (Agrawal et.al., 2018). Although a quantum computer could be able to read this set of instructions right away (Shor, 1997), both quantum and classical computers have the drawback of needing precise details about the path and an internal circuit to work.

Equations and signal bursts are linked in modern science. The intervals between signals are not now a major concern in models. Phase or "quiet" is a topic of quantum information theory (QIT), but it is utterly compromised in real values. The "silent" network wasn't impacted at all. By utilizing the network of "silence," Agarwal et al. (2018) are developing a revolutionary technique for recording, processing, and remembering information so that we will never have to discretize nature and lose everything just to recreate it afterwards. There is no classical parallel for the fundamental geometric patterns in a Bloch sphere, such as quantum mechanics, according to the fractal information theory (Bandyopadhyay, 2020). Everything that happens is a change in the 3D network of rhythms that includes both the user and the environment. Even space and mass are converted into a web of time cycles that store geometric shapes. Bandyopadhyay and his team developed a new approach to information processing and decision-making within the framework of this fractal information theory (FIT), in which there are no logic gates, switches, or any other elements that made up the probability-based existing information theory (EIT) foundation (Figure 2).



Classical mechanics	Quantum mechanics	Fractal mechanics
Imaginary world means phase part dominates reality	One imaginary world, time from that world acts on reality	11 imaginary worlds, topology of states of 11 worlds act.
No simultaneity	Superposition of infinite intermediate states	Topology of intermediate states govern superposition
No intermediate states	No control on the intermediate states.	Pattern of primes govern intermediate state geometry
No phase space	Phase space pathway exists, but no topological control	Self-assembly of Bloch spheres follow a geometric algebra
Information content is log(symmetry)	Content is source entropy – Tr(dlogd)	Density of distinct time crystals for all 12 symmetries
No drive	No drive	Non-linear push-pull between a pair of primes e.g. 43·47
Physical dimension of object	Physical dimension of object	Geometry of singularity points in a network of clocks
No concept of time crystal	Time crystal with only one imaginary clock	Time crystal with 12 imaginary layers of clocks, UTC
Differential equation	Differential equation	Continued fraction geometric algebra for pattern of primes
No universal constant	Planks constant	Plank constant for quantum version, action, A for classical

Figure 2. Visualizing Fractal Information along with comparison of Fractal Mechanics

(Agarwal et. al, 2018)

However, to process such an entity of information, a new kind of geometrically rhythmic language needs to be developed in lieu of the logic-based building blocks which has been explored in the next section.

2.3 Geometric Musical Language (GML)

No algorithm is directing the reader's brain as they pursue these lines. As we read, proteins in the neurons are measuring time in picoseconds, nanoseconds, and microseconds so that membrane firing on a millisecond scale is accomplished properly. Instantaneous perception is a circular arrangement of occurrences that contradicts the universal progression of time (Zeh, 1989). Without a single line of software code or algorithm, the brain's materials are performing calculations, modeling the future, and making judgments; everything a computer can do and much more.

Due to a unique twist in the spiral symmetry, microtubules and protein research suggests that devices can gather noise and turn it into a clean signal (Sahu et al.,2014). Thus, the necessity for engineering singularity, a language that communicates through blinking singularities, namely GML, arose. A single letter, a time cycle, a rhythm, a clock, and a unitary operator make up this language. At the moment, information is represented by a clocking-geometry Bloch sphere (Figure 2). Multiple spheres of this type self-assemble and expand like a balloon to store and process complex data. The "bings" are singularity glue used to incorporate clocked Bloch spheres. Adjusting the "silence" or pause between "bings" to preserve the square and triangle's geometrical properties. The term "geometric musical language" (GML) describes the process of turning five sensory inputs into musically inspired geometric forms and rhythms. New information is incorporated as a guest into a single Bloch sphere that is constantly increasing. "Situation" is written as geometric forms and always combined in a temporal cycle, either side by side or one inside the other, to obliterate the distinction between questions and answers. FIT-GML hypercomputing does not require an algorithm or programming, and it uses fractal beats, or geometric nesting within a Hilbert space, like the human brain. If the clocking geometry in the Bloch sphere and virtual poles are removed, converting FIT to a classical state, QIT is the result. Figure 3 depicts the GML dimensions. The first column describes "who jumps?" The response describes how to actualize a dimension that is one level higher. The second row illustrates how the data appears in the operational brain. The third row represents the decisions as quaternions.

Geometric musical language (GML)	Turning machine based Software/Algorithm
Close loop of arguments, clocks, time crystal	Linear flow of arguments, switch, bits/qubits
Arranged geometry of arguments sensed and 1D-11D shapes arrange to build an argument	How arguments are arranged has no role in the intelligence building
Every argument has structure of arguments inside, all statement undefined, all linked.	Every argument is static, complete, nothing inside, guest-host concept does not exist
Data structure: triplet of linguistics, Subject, clause, verb-adverb: who-why-what-how	Data structure: If-then statement links events as if one happens the other would follow
A catalogue of linking symmetries (Phase prime metric, PPM) look into shape and transform-link discrete events into one.	Human free will uses its senses, intelligence to link events, deep learning or fitting tools designed by human bias also links events.
Shape similarity links uncorrelated events	Waits for human/statistical validity i.e., fitting
Temporal editing may take to different paths, logical tree is purely temporal, instant.	Flow of logic defines paths, condition sets it, not a change in geometric shape or an instant shape
Superposition of geometries, composition of shape making a new shape, make new.	Superposition of arguments cannot happen as linearity does not support simultaneity
Drive to symmetry create/transform logic	There is no natural drive, phase transition big no
Topological constraints sets boundary	Instructions set boundary, limit, halting
Ten dimensional data needs fractal hardware	1D data needs linear hardware
Converts to resonance bands for machine	Converts to 0 and 1 for running in a machine

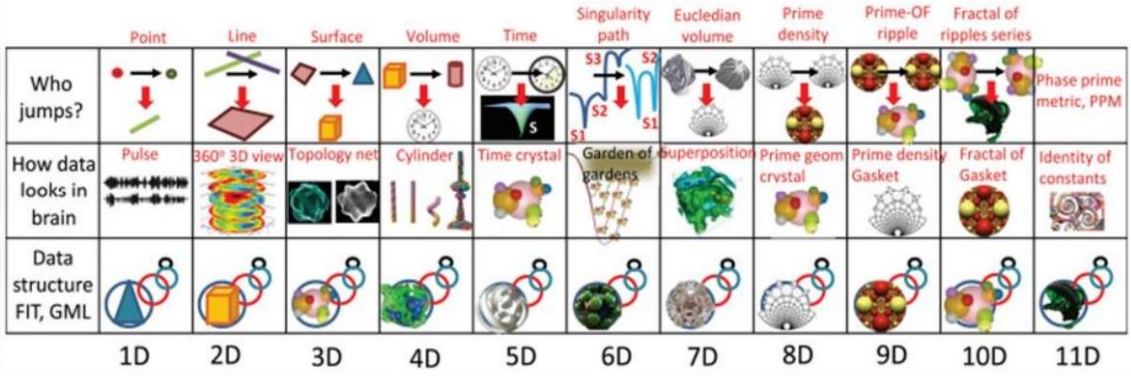


Figure 3. GML vs Turing Machine alongside Dimensional Description for GML (Bandyopadhyay, 2020) However, to deconstruct the silence and singularity of GML, we require a higher order conformal geometric algebra. In the next section, we discuss how a twelve-dimensional geometric algebra was developed for this purpose.

2.4 Dodecanion Algebra

While the other dimensions were obtained from the 4D and 8D tensors, the four and eight dimensions (D) were historically explored as quaternions and octonions. The different stereographic projections for each dimension, from one to twelve, were carefully examined by Singh et al. in 2021. The difference between eight and twelve hypothetical worlds is that instead of the Fano plane, which governs the products of hypothetical vectors, a trio of manifolds that could coexist in three different configurations replaces it (Figure 4).

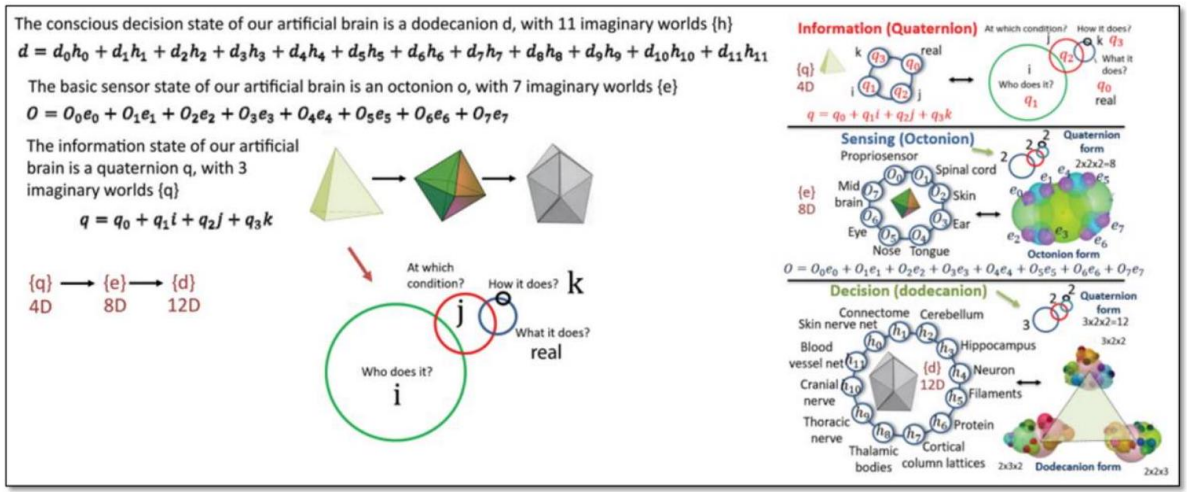


Figure 4. Information-Sensing-Decision Layers using Dodecanion Geometric Algebra (Singh et. al., 2021)

In addition, they developed a new metric, space-time-topology-prime (stTS), for a mathematical universe with n layered imaginary worlds by combining two ideas from the current Space-time metric: 15 geometric shapes as topology (T) and 15 primes as symmetry (S). A decision is presented as a time-varying geometry using the stTS metric. In the past, space and time were the most important ideas in astrophysics and general relativity. A quaternion could describe these ideas well by using a 3-by-3 space tensor and a 1-by-3 time tensor. It means that a 4-by-4 tensor has been split into a space-time tensor. But when they made a mathematical world where tensors could have between 2 and 20 dimensions so that they could support topologies up to anicosanion, they had to come up with two more things. So, space-time has been replaced with space-time-topology-prime, which can be used to build a new 11-dimensional metric. Dodecanion has a 12D tensor, but its dynamics are only 11D. (Singh et. al., 2021).

With this robust framework of mathematical foundations and empirical results we explore a few key possible applications of the fractal-based decision approach in three key areas of healthcare, security, and business analytics in the upcoming section.

3 Applications

On the healthcare front, many non-invasive devices and therapy can be planned where the dynamic instability of microtubules can be altered using specific overtones or rhythms. Mental wellbeing can also be a serious area of research for fractal-based clocking devices. In terms of security, prime numbers already provide the backbone of cryptography in the Turing machines. With the advent of topology, such algorithms can be made more robust against quantum attacks. On the business analytics front, pattern match and search problems are considered hard problems in Turing based computer algorithms. But with fractal devices, we can achieve speed with no codebase as the self-organizing structures will make heuristic decisions rather than complex linear transformations (Ghosh et. al., 2019a, 2019b).

4 Limitations and Future Work

During the review we have noted three key limitations of this approach that can culminate into future research agendas.

Analog Approach - As the devices are non-Turing, they are specifically designed to meet certain experimental requirements. Hence, devising a minimal viable product for retail consumers is a distant possibility as it involves proprietary technology.

Open sourcing challenges - As the devices do not have codebase, open sourcing such projects to the outer world becomes a challenge. Scalability then takes a hit which leads to delay in adoption of the technology.

Hindered innovation diffusion - Due to the above two factors, coupled with the unconventional nature of the knowledge base it requires, diffusion of this innovation becomes hindered and has a possibility to fall into a chasm or suffer stalling in long winters, like the AI revolution after the 1950s.

Hence, as a future agenda, we need to view fractal-based devices not as a problem only related to high-tech innovation, but also as a marketing problem. Only then such unique innovations grounded firmly on empirically tested outcomes can flourish to its true potential in the real markets. Else, they might quietly die in the chasm or get acquired and repackaged by multinationals, analogous to the intellectual properties of Xerox palo alto labs which got distributed among Apple and Microsoft (Sirk, 2020). That said, these approaches by Indian origin scholars are a commendable feat that strengthens the oriental knowledge system where the Vedic philosophy acts as the epistemological lynchpin.

References

Agrawal, L., Chhajed, R., Ghosh, S., Ghosh, B., Ray, K., Sahu, S., ... & Bandyopadhyay, A. (2018). Fractal Information Theory (FIT)-Derived Geometric Musical Language (GML) for Brain-Inspired Hypercomputing. In Soft Computing: Theories and Applications (pp. 343-372). Springer, Singapore.

Bandyopadhyay, A. (2020). Nanobrain: the making of an artificial brain from a time crystal. CRC Press. Forrest, J., Menser, M., & Burgess, J. A. (1971). High frequency of diabetes mellitus in young adults with congenital rubella. The Lancet, 298(7720), 332-334.

Ghosh, S. et al. (2014) Design and operation of a brain like computer: A new class of frequency-fractal computing using wireless communication in supramolecular organic, inorganic systems. Information 5, 28 − 99.

Ghosh, S., Fujita, D., Bandyopadhyay, A. (2019a) Universal Geometric-musical language for big data processing in an assembly of clocking resonators, JP-2017-150171, 8/2/2017: World patent received February 2019, WorldPatent: WO 2019/026983.

Ghosh, S., Fujita, D., Bandyopadhyay, A. (2019b) Human brain like intelligent decision-making machine JP-2017150173, 8/2/2017; World patent WO 2019/026984.

Ghosh, S., Sahu, S., & Bandyopadhyay, A. (2014). Evidence of massive global synchronization and the consciousness: Comment on "Consciousness in the universe: A review of the 'Orch OR' theory" by Hameroff and Penrose.Physics of Life Reviews, 11(1), 83-84.

Hameroff, S. (2012). How quantum brain biology can rescue conscious free will. Frontiers in Integrative Neuroscience, 6, 93.

Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the 'Orch OR' theory. Physics of Life Reviews, 11(1), 39-78.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of Physiology, 117(4), 500.

Margolus, N. (2003). Looking at Nature as a Computer. International Journal of Theoretical Physics, 42(2), 309327.

Preparata, F. P., & Shamos, M. I. (1985). Convex hulls: Basic algorithms. In Computational Geometry (pp. 95-149). Springer, New York, NY.

Sahu, S., Ghosh, S., Ghosh, B., Aswani, K., Hirata, K., Fujita, D., & Bandyopadhyay, A. (2013). Atomic water channel controlling remarkable properties of a single brain microtubule: correlating single protein to its supramolecular assembly. Biosensors and Bioelectronics, 47, 141-148.

Sahu, S., Ghosh, S., Hirata, K., Fujita, D., & Bandyopadhyay, A. (2013). Multi-level memory-switching properties of a single brain microtubule. Applied Physics Letters, 102(12), 123701.

Shor, P., & Laflamme, R. (1997). Quantum analog of the MacWilliams identities for classical coding theory. Physical Review Letters, 78(8), 1600.

Singh, P., Sahoo, P., Saxena, K., Ghosh, S., Sahu, S., Ray, K., ... & Bandyopadhyay, A. (2021). Quaternion, octonion to dodecanion manifold: stereographic projections from infinity lead to a self-operating mathematical universe. In Proceedings of International Conference on Trends in Computational and Cognitive Engineering (pp. 55-77). Springer, Singapore.

Sirk, C. (2020). Xerox PARC and the Origins of GUI.

Tukey, J. W. (1984) "Sequential conversion of continuous data to digital data," Bell Laboratories memorandum, Annals Hist Computing 6, 152-155.

Turing, A. M. (1940). Mathematical theory of enigma machine. Public Record Office, London, 3, 150.

Wheeler, J. A. (1984) Bits, quanta, meaning. In: Problems in Theoretical Physics.

Wheeler, J. A. (1990) Information, physics, quantum: The search for links. In: Complexity, Entropy, and the Physics of Information, W. H. Zurek, Ed. Addison-Wesley, Redwood City, CA.

Zeh, H. D. (1989). The Direction of Time. Springer-Verlag, Berlin.

6 Two Consecutive Positive Integers Cannot Both Be Perfect

Shailesh Shirali

Sahyadri School KFI
Rajgurunagar, Khed
Pune – 410513

Email: shailesh.shirali@gmail.com

Theorem. *Two consecutive positive integers cannot both be perfect.*

Proof

This write-up is based on the article [1], "Can Two Consecutive Numbers Both Be Perfect?" We do not know if the result has any significant implications. However, the reasoning used in proving the result is elegant and worth studying.

Let denote $s(n)$ the sum of the divisors of a number $n \in \mathbb{N}$, including n itself. Then, by definition, n is perfect if $s(n) = 2n$.

If two consecutive numbers are both perfect, then one of them must be even. It is well-known (Euler) that all even perfect numbers are of the form $2^{p-1}(2^p - 1)$ where both p and $2^p - 1$ are prime (see, e.g., [2], [3]; we do not reproduce the proof here). So it *suffices* to prove that the following two numbers M, N are both not perfect:

- $M = 2^{n-1}(2^n - 1) + 1$ where $n > 2$ is odd;
- $N = 2^{n-1}(2^n - 1) - 1$ where $n > 2$ is odd.

Let $n > 2$ be an odd number; then $n - 1$ is even.

First consider M. Since $2^k \equiv (-1)^k \pmod 3$, it follows that

$$2^{n-1}(2^n - 1) + 1 \equiv 1 \cdot (-1 - 1) + 1 \pmod 3 \equiv -1 \pmod 3,$$

so $M \equiv -1 \pmod 3$.

Next consider N. Since $4 \mid 2^{n-1}(2^n - 1)$, it follows that $N \equiv -1 \pmod 4$.

The main result now follows from the following two lemmas.

Lemma 1. *A number M of the form $-1 \pmod 3$ cannot be perfect.*

Lemma 2. *A number N of the form $-1 \pmod 4$ cannot be perfect.*

Proof of Lemma 1

Since $M \equiv -1 \pmod 3$, it follows that M is not a perfect square. Now consider the value of $s(M)$. We have:

$$s(M) = \sum_{d|M} d = \sum_{d|M,\ d<\sqrt{M}} \left(d + \frac{M}{d}\right) \qquad \text{(since } M \text{ is not a perfect square).}$$

Since $M \equiv -1 \pmod 3$, all divisors d of M are of the form $\pm 1 \pmod 3$.

Since $d \cdot \frac{M}{d} \equiv -1 \pmod 3$ for all $d \mid M$, it follows that:

- If $d \mid M$ and $d \equiv 1 \pmod 3$, then $\frac{M}{d} \equiv -1 \pmod 3$.
- If $d \mid M$ and $d \equiv -1 \pmod 3$, then $\frac{M}{d} \equiv 1 \pmod 3$.

This implies that $d + \frac{M}{d} \equiv 0 \pmod 3$ for all $d \mid M$. Therefore $s(M)$ is a multiple of 3.

Since $M \equiv -1 \pmod 3$, it follows that $s(M) \neq 2M$. So M is not perfect. ■

Proof of Lemma 2

Since $M \equiv -1 \pmod 4$, it follows that N is not a perfect square. Now consider the value of $s(N)$. We have:

$$s(N) = \sum_{d|N} d = \sum_{d|N,\ d<\sqrt{N}} \left(d + \frac{N}{d}\right) \qquad \text{(since } N \text{ is not a perfect square).}$$

Since $N \equiv -1 \pmod 4$, all divisors d of N are of the form $\pm 1 \pmod 4$.

Since $d \cdot \frac{N}{d} \equiv -1 \pmod 4$ for all $d \mid N$, it follows that:

- If $d \mid N$ and $d \equiv 1 \pmod 4$, then $\frac{N}{d} \equiv -1 \pmod 4$.
- If $d \mid N$ and $d \equiv -1 \pmod 4$, then $\frac{N}{d} \equiv 1 \pmod 4$.

This implies that $d + \frac{N}{d} \equiv 0 \pmod 4$ for each $d \mid N$ (as earlier). Hence $s(N)$ is a multiple of 4.

But N is odd, so $s(N) \neq 2N$. Hence N is not perfect. ■

Bibliography

[1] Luca, Florian, and Francis B. Coghlan. "Can Two Consecutive Numbers Both Be Perfect?: 10711." *The American Mathematical Monthly*, vol. 108, no. 1, 2001, pp. 80â€"81. JSTOR, https://doi.org/10.2307/2695692. Accessed 7 Jun. 2022.

[2] Caldwell, Chris K., "Characterizing all even perfect numbers" from https://primes.utm.edu/notes/proofs/EvenPerfect.html

[3] Wikipedia, "Even perfect numbers" from https://en.wikipedia.org/wiki/Perfect_number#Even_perfect_numbers

7 The Moore-Penrose Inverse and Its Applications

J. K. Verma

Department of Mathematics
IIT Bombay
Mumbai – 400076

Email: jkv@iitb.ac.in

1 Generalized inverse and pseudoinverse of a matrix

In this expository article, we discuss the Moore-Penrose inverse, also called the pseudoinverse of a complex matrix.[1] Eliakim H. Moore [10] in 1920 and Sir Roger Penrose [8] in 1955, introduced the pseudoinverse of a complex matrix. It is a good analogue of the usual inverse of an invertible matrix. If A is an $m \times n$ complex matrix then an $n \times m$ complex matrix B is called a **Moore-Penrose inverse** or a **pseudoinverse** of A if it satisfies the following four conditions:

$$ABA = A, BAB = B, (AB)^* = AB, (BA)^* = BA.$$

Here A^* denotes the conjugate transpose of A. It is clear that if A is an invertible matrix then A^{-1} is its pseudoinverse. We shall prove that any complex matrix has a unique pseudoinverse which is denoted by A^+. We shall prove its existence and describe methods of computation.

The **length of a vector** $v \in \mathbb{C}^n$ is the real number $\sqrt{v^*v}$ and it is denoted by $||v||$. We say that $u, v \in \mathbb{C}^n$ are orthogonal if $u^*v = 0$. If W is a subspace of $\mathbb{C}^n$ and $u \in \mathbb{C}^n$ then we can write $u = v + w$ where v and w are orthogonal vectors and $w \in W$. In this case

[1] These notes are based on lectures delivered in various teacher's training programmes of the National Centre for Mathematics. The material is freely borrowed from the textbooks and papers listed in the references, specially the lecture notes of the author and Murali K. Srinivasan [11]. No claim of originality is made.

w is unique and it is called the **orthogonal projection** of u in W. The column space of a matrix A is the subspace spanned by the column vectors of A. It will be denoted by $C(A)$. A square complex matrix A is called **Hermitian** if $A = A^*$.

The pseudoinverse of a matrix has remarkable properties.

If a system of linear equations $Ax = b$ has a solution then A^+b is a solution with smallest length. If $Ax = b$ has no solution then A^+b is a least-squares solution in the sense that $||AA^+b - b||$ is the smallest among $||Ax - b||$ for all $x \in \mathbb{C}^n$. We shall prove that AA^+u is the orthogonal projection of a vector u onto the column space of A and A^+Av is the orthogonal projection of a vector v onto the column space of A^*. Moore used these two properties of A^+ to define the pseudoinverse of a matrix.

Definition 1.1. *Let $A \in \mathbb{R}^{m\times n}$. A matrix $G \in \mathbb{R}^{n\times m}$ is called a* ***generalised inverse*** *of A if*

$$AGA = A.$$

Proposition 1.2. *If a matrix A is invertible then A has a generalised inverse and $G = A^{-1}$.*

Proof. If A^{-1} exists then $G = A^{-1}$ satisfies the equation $AGA = A$. Let A have a generalized inverse G and A be invertible, then $AGA = A$ implies that $G = A^{-1}$. ■

Example 1.3. A matrix may have infinitely many generalized inverses. Consider the matrix $A = [1, 2]$. Let $G = \begin{bmatrix} x \\ y \end{bmatrix}$ be a generalized inverse of A. Then

$$A = [1, 2] = AGA = [1, 2] \begin{bmatrix} x \\ y \end{bmatrix} [1, 2] = [x + 2y, 2x + 4y].$$

Hence G is a generalized inverse if and only if $x + 2y = 1$. Hence there are infinitely many generalized inverses of A.

We can use generalized inverse to find a particular solution to a consistent system of linear equations.

Proposition 1.4. *Let A be an $m \times n$ matrix and $Ax = b$ be a consistent system of linear equations. Let G be a generalized inverse of A. Then $x = Gb$ is a solution of $Ax = b$.*

Proof. Multiply the equation $Ax = b$ by AG to get

$$A(Gb) = A(GAx) = (AGA)(x) = Ax = b.$$

Therefore $x = Gb$ is a solution of $Ax = b$. ■

Example 1.5. We have seen that the row vector $A = [1, 2]$ has a generalized inverse $G = [x, y]^t$ if and only if it satisfies $x + 2y = 1$. For G to be a pseudoinverse we need to check the four conditions stated above:

$$[x, y]^t = G = GAG = [x, y]^t[1, 2][x, y]^t = (x + 2y)[x, y]^t$$

$$x + 2y = (AG)^t = AG = [1, 2][x, y]^t = x + 2y$$

$$(GA)^t = \begin{bmatrix} x & y \\ 2x & 2y \end{bmatrix} = GA = \begin{bmatrix} x \\ y \end{bmatrix}[1, 2] = \begin{bmatrix} x & 2x \\ y & 2y \end{bmatrix}$$

$$\implies 2x = y, x + 2y = 1 \implies x = 1/5, y = 2/5.$$

Hence A has a unique pseudoinverse $[\frac{1}{5}, \frac{2}{5}]^t$.

Theorem 1.6. *Let A be any complex matrix. Then its pseudoinverse is unique.*

Proof. Let B and C be two pseudoinverses of A. Consider the matrix $X = AC - AB$. Then X is Hermitian since both AC and AB are. Note that

$$X^2 = ACAC - ACAB - ABAC + ABAB = AC - AB - AC + AB = 0.$$

Thus $||Xu||^2 = (Xu)^*Xu = 0$. Hence $Xu = 0$ for all $u \in \mathbb{C}^n$. Hence $X = 0$. Thus $AC = AB$. Similar argument applied to $CA - BA$ shows that $CA = BA$. Thus

$$C = CAC = (BA)C = B(AC) = B(AB) = B.$$

■

Theorem 1.7. *Let A be an $m \times n$ complex matrix and* $\operatorname{rank} A = n$. *Then A has a pseudoinverse*

$$A^+ = (A^*A)^{-1}A^*.$$

Proof. We check the four conditions.

(1) $AA^+A = A((A^*A)^{-1}A^*)A = A$.
(2) $A^+AA^+ = (A^*A)^{-1}A^*A(A^*A)^{-1}A^* = A^+$.
(3) $AA^+ = A((A^*A)^{-1}A^*)$ is Hermitian.
(4) $A^+A = (A^*A)^{-1}A^*.A = I$

Therefore $A^+ = (A^*A)^{-1}A^*$ is the pseudo-inverse of A. Note that A^+ is a left inverse of A in this case. ■

Theorem 1.8. *Let A be an $m \times n$ complex matrix whose rows are linearly independent. Then*

$$A^+ = A^*(AA^*)^{-1}.$$

Proof. Since rank A = rank $AA^* = m$, AA^* is invertible. We check the four axioms.

(1) $AA^+A = AA^*(AA^*)^{-1}A = A$. (2) $A^+AA^+ = A^*(AA^*)^{-1}AA^*(AA^*)^{-1} = A^+$.
(3) $AA^+ = AA^*(AA^*)^{-1} = I$ (4) $A^+A = A^*(AA^*)^{-1}A$ is Hermitian.

Note that in this case A^+ is the right inverse of A. ■

Recall that a real square matrix Q is called **orthogonal** if $QQ^t = I$. A complex square matrix Q is called **unitary** if $QQ^* = I$. can be written as $A = QR$ where Q is an orthogonal matrix and R is an upper triangular matrix. This decomposition is called a QR -decomposition of A. If A is complex then we can write $A = QR$ where Q is unitary and R is upper-triangular. We refer the reader to [12] for basic results in linear algebra.

Theorem 1.9. *Let $A \in \mathbb{R}^{m\times n}$ and rank $A = n$. Let $A = QR$ be its QR-decomposition. Then*

$$A^+ = R^{-1}Q^t.$$

Proof. Let $A = QR$ be its QR-decomposition. Then

$$A^tA = (QR)^t(QR) = R^tQ^tQR = R^tR$$

Therefore A^+ is given by

$$A^+ = (A^tA)^{-1}A^t = (R^tR)^{-1}(QR)^t = R^{-1}(R^t)^{-1}R^tQ^t = R^{-1}Q^t.$$

■

Theorem 1.10. *Let A be a diagonal matrix. Then the pseudoinverse of A is also a diagonal matrix B so that if $B = (b_{ij})$ then $b_{ii} = 1/a_{ii}$ if $a_{ii} \neq 0$ and $b_{ii} = 0$ if $a_{ii} = 0$.*

Proof. We verify this in an example. The proof of the general fact is similar. Let

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } BA = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are both symmetric. The two other conditions are also satisfied:

$$ABA = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} = A \text{ and } \quad BAB = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{bmatrix} = B.$$

■

One of the most useful decomposition of a real matrix is its **singular value decomposition**. Any $m \times n$ real matrix can be written as $A = UDV^t$ where U and V are orthogonal matrices and $D = (d_{ij})$ is an $m \times n$ diagonal matrix whose diagonal entries d_{ii} are nonnegative. We refer the reader to [12] for more on singular value decomposition.

Theorem 1.11. *Let A be any matrix with a singular value decomposition $A = U\Sigma V^t$. Then a pseudoinverse of A is given by*

$$A^+ = V\Sigma^+ U^t.$$

Proof. Verify the four conditions.

$$AA^+A = U\Sigma V^t \cdot V\Sigma^+ U^t \cdot U\Sigma V^t = U\Sigma\Sigma^+\Sigma V^t = U\Sigma V^t = A$$

$$A^+AA^+ = V\Sigma^+ U^t \cdot U\Sigma V^t \cdot V\Sigma^+ U^t = V\Sigma^+\Sigma\Sigma^+ U^t = V\Sigma^+ U^t = A^+$$

$$AA^+ = U\Sigma V^t \cdot V\Sigma^+ U^t = U\Sigma\Sigma^+ U^t \quad \text{(Symmetric)}$$

$$A^+A = V\Sigma^+ U^t \cdot U\Sigma V^t = V\Sigma^+\Sigma V^t \quad \text{(symmetric)}\ .$$

■

Theorem 1.12. *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Let $Ax = b$ have a solution. Then $x^* = A^+b$ is a solution of $Ax = b$ and $||x^*|| \leq ||x||$ for all solutions x of $Ax = b$.*

Proof. Since A^+ is a generalized inverse, A^+b is a solution of $Ax = b$. We shall show that it has smallest possible norm among all solutions of $Ax = b$. First note that

$$x = (A^+A)x + (I - A^+A)x$$

is an orthogonal decomposition of x. Recall that A^+A and AA^+ are symmetric. Therefore

$$(A^+Ax)^t(x - A^+Ax) = x^t(A^+A)^tx - x^t(A^+A)^tA^+Ax = x^tA^+Ax - x^tA^+AA^+Ax = 0.$$

Therefore

$$||x||^2 = ||A^+Ax||^2 + ||x - A^+Ax||^2 \geq ||A^+b||^2.$$

Hence we conclude that $||x|| \geq ||A^+b||$ and equality holds if and only if $x = A^+b$. ■

Example 1.13. Consider the plane $Ax = b$ where $A = [-1,\ 2,\ 2]$, $x = [x_1,\ x_2,\ x_3], b = 18$. The solution closest to the origin is given by A^+b. An SVD of A is given by

$$\begin{bmatrix} -1 & 2 & 2 \end{bmatrix} = [1] \begin{bmatrix} 3 & 0 & 0 \end{bmatrix} \frac{1}{3} \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}.$$

Therefore the pseudoinverse A^+ is given by

$$A^+ = \frac{1}{3} \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ 0 \\ 0 \end{bmatrix} [1] = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{bmatrix}.$$

Hence $A^+b = [-2, 4, 4]$ is the point on the plane $Ax = b$ that is closest to the origin.

Proposition 1.14. *Let A be an $m \times n$ complex matrix. Then*
(1) *The vector AA^+u is the orthogonal projection of $u \in \mathbb{C}^m$ onto $C(A)$.*
(2) *The vector A^+Av is the orthogonal projection of $v \in \mathbb{C}^n$ onto $C(A^*)$.*

Proof. (1) Since $A^+u \in \mathbb{C}^n$, $AA^+u \in C(A)$. Let us prove that $u = AA^+u + (u - AA^+u)$ where $AA^+u \perp (u - AA^+u)$. Indeed,

$$(AA^+u)^*(u - AA^+u) = u^*AA^+u - u^*AA^+AA^+u = u^*AA^+u - u^*AA^+u = 0.$$

(2) Let $v \in \mathbb{C}^n$. In order to show that $A^+Av \in C(A^*)$, we prove that $A^+Av \in N(A)^\perp$. Let $Aw = 0$ for some $w \in \mathbb{C}^n$. Then $(A^+Av)^*w = v^*A^+Aw = 0$. We have the orthogonal decomposition of v given by

$$v = A^+Av + (v - AA^+v).$$

■

2 The least squares problem

Suppose we have a large number of data points $(x_i, y_i),\ i = 1, 2, \ldots, n$ collected from some experiment. Frequently there is reason to believe that these points should lie on a straight line. So we want a linear function $y(x) = s+tx$ such that $y(x_i) = y_i,\ i = 1, \ldots, n$. Due to uncertainity in data and experimental error, in practice the points will deviate somewhat from a straightline and so it is impossible to find a linear $y(x)$ that passes through all of them. So we seek a line that fits the data well, in the sense that the errors are made as small as possible. A natural question that arises now is: how do we define the error?

Consider the following system of linear equations, in the variables s and t, and known coefficients $x_i, y_i,\ i = 1, \ldots, n$:

$$\begin{aligned} s + x_1 t &= y_1 \\ s + x_2 t &= y_2 \\ &\vdots \\ s + x_n t &= y_n \end{aligned}$$

Note that typically n would be much greater than 2. If we can find s and t to satisfy all these equations, then we have solved our problem. However, for reasons mentioned above, this is not always possible. For given values of s and t the error in the ith equation is $|y_i - s - x_i t|$. There are several ways of combining the errors in the individual equations to get a measure of the total error. The following are three examples:

$$\sqrt{\sum_{i=1}^{n}(y_i - s - x_i t)^2}, \quad \sum_{i=1}^{n} |y_i - s - x_i t|, \quad \max_{1 \le i \le n} |y_i - s - x_i t|.$$

Both analytically and computationally, a nice theory exists for the first of these choices and this is what we shall study. The problem of finding s, t so as to minimize

$$\sqrt{\sum_{i=1}^{n}(y_i - s - x_i t)^2}$$

is called a **least squares problem**. Let

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{pmatrix},\ b = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix},\ \text{and } x = \begin{pmatrix} s \\ t \end{pmatrix},\ \text{so that } Ax = \begin{pmatrix} s + tx_1 \\ s + tx_2 \\ . \\ . \\ s + tx_n \end{pmatrix}.$$

The least squares problem is finding an x such that $||b-Ax||$ is minimized, i.e., find an x such that Ax is the best approximation to b in the column space of A. This is precisely the problem of finding x such that $b - Ax$ is orthogonal to the column space of A.

A straight line can be considered as a polynomial of degree 1. We can also try to fit an mth degree polynomial

$$y(x) = s_0 + s_1 x + s_2 x^2 + \cdots + s_m x^m$$

to the data points $(x_i, y_i),\ i = 1, \ldots, n$, so as to minimize the error (in the least squares sense). In this case $s_0, s_1, \ldots, s_m$ are the variables and we have

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & . & . & x_1^m \\ 1 & x_2 & x_2^2 & . & . & x_2^m \\ & . & . & . & . & . \\ & . & . & . & . & . \\ 1 & x_n & x_n^2 & . & . & x_n^m \end{pmatrix},\ b = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix},\ x = \begin{pmatrix} s_0 \\ s_1 \\ . \\ . \\ s_m \end{pmatrix}.$$

Example 2.1. Find s, t such that the straight line $y = s + tx$ best fits the following data in the least squares sense:

$$y = 1 \text{ at } x = -1, \quad y = 1 \text{ at } x = 1, \quad y = 3 \text{ at } x = 2.$$

We want to project $b = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$ onto the column space of $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$. Now $A^tA = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}$ and $A^tb = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$. The normal equations are

$$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}.$$

The solution is $s = 9/7,\ t = 4/7$ and the best line is $y = \frac{9}{7} + \frac{4}{7}x$.

Suppose that $A \in \mathbb{R}^{m \times n}$ is a tall matrix. This means the system of linear equations $Ax = b$ has more equations than variables. We know that the system $Ax = b$ has a solution if and only if $b \in C(A)$. Due to experimental errors in the data matrix A and the data vector b, the system $Ax = b$ may be inconsistent. In this case we want to find an x so that the error $||Ax - b||$ is minimised. Any vector $\hat{x}$ that satisfies

$$||A\hat{x} - b|| \leq ||Ax - b|| \text{ for all } x$$

is called a solution of the least squares problem. This problem was independently solved by C. F. Gauss and A. M. Legendre in the beginning of the nineteenth century. In fact

Gauss used this method to predict the appearance of the largest asteroid Ceres in 1802. Ceres was found on January 1, 1801 by an Italian astronomer G. Piazzi. Gauss predicted its orbit using the method of least squares. Based on his prediction, it was again sighted on January 1, 1802.

Let the columns of A be the vectors $a_1, a_2, \ldots, a_n \in \mathbb{R}^m$. The least squares problem is the problem of computing a vector in the column space of A that is closest to the vector b.

Example 2.2. Let us solve the least squares problem for the data

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

The system of linear equations $Ax = b$ is given by

$$2x = 1, -x + y = 0 \text{ and } 2y = -1.$$

Check that the above system has no solution. The least squares problem is

$$\text{minimize } f(x, y) = (2x - 1)^2 + (-x + y)^2 + (2y + 1)^2.$$

The critical points of this function are found by solving the equations:

$$f_x = 4(2x - 1) + 2(-x + y)(-1) = 0 \text{ and } f_y = 2(-x + y) + 4(2y + 1) = 0.$$

These equations have a unique solution, namely $(x, y) = (1/3, -1/3)$.

Solution via geometry of inner product spaces. Let U be a subspace of an inner product space V. The orthogonal complement of U is the subspace

$$U^\perp = \{v \in V \mid \langle v, u\rangle = 0 \text{ for all } u \in U\}.$$

Theorem 2.3. *Let V be an inner product space and U be a subspace of V. Then*
(1) $U \cap U^\perp = \{0\}$.
(2) *If V is finite dimensional then* $V = U \oplus U^\perp$ *and* $\dim U + \dim U^\perp = \dim V$
(3) *If V is finite dimensional then* $(U^\perp)^\perp = U$.

Proof. (1) If $w \in U \cap U^\perp$ then $\langle w, w\rangle = 0$. Hence $w = 0$.
(2) Let $\dim V = n$ and $\dim U = k$. Let $u_1, u_2, \ldots, u_k$ be an orthonormal basis of U. Extend this to an orthonormal basis $u_1, \ldots, u_n$ of V. Therefore $u_{k+1}, \ldots, u_n \in U^\perp$. Any $v \in V$ can be expressed uniquely as

$$v = x_1u_1 + x_2u_2 + \cdots + x_ku_k + x_{k+1}u_{k+1} + \cdots + x_nu_n.$$

Hence $V = U + U^{\perp}$. Since $U \cap U^{\perp} = \{0\}$, it follows that $V = U \oplus U^{\perp}$. This also proves that $\dim V = \dim U + \dim U^{\perp}$.
(3) Let $v \in (U^{\perp})^{\perp}$. Then $\langle v,\, u_i\rangle = 0$ for all $i = k+1, k+2, \ldots, n$. Since $v = \sum_{i=1}^{n} x_i u_i$ for some scalars $x_1, \ldots, x_n$, it follows that $x_i = 0$ for all $i = k+1, \ldots, n$. Hence $v \in U$. Conversely let $u \in U$. Then $\langle u,\, u_i\rangle = 0$ for all $i = k+1, \ldots, n$. Hence $u \in (U^{\perp})^{\perp}$. ■

Definition 2.4. *Let U be a subspace of an inner product space V. Let $v \in V$. If we can write $v = p + (v - p)$ for a vector $p \in U$ and $v - p \in U^{\perp}$ then p is called an orthogonal projection of v in U and we write it as $p = proj_U v$.*

By Theorem 2.3, If V is finite dimensional then $p = \mathrm{proj}_U v$ is uniquely determined. As $v = p + (v - p)$, it follows that $||v||^2 = ||p||^2 + ||(v-p)||^2$. Hence $||p|| \leq ||v||$ and equality holds if and only if $v = p$.

If $\dim U = 1$ and U is the linear span of a unit vector $u \in U$ then any $v \in V$ can be written as $v = \langle v,\, u\rangle u + (v - \langle v,\, u\rangle u)$. Note that $u \perp v - \langle v,\, u\rangle u$. This construction works for any finite-dimensional subspace of an inner product space.

Theorem 2.5. *Let U be a nonzero finite dimensional subspace of an inner product space V. Let $\{u_1, u_2, \ldots, u_n\}$ be an orthonormal basis of U. Let $v \in V$. then*

$$proj_U v = \sum_{j=1}^{n} proj_{u_j} v = \langle v,\, u_1\rangle u_1 + \langle v,\, u_2\rangle u_2 + \cdots + \langle v,\, u_n\rangle u_n.$$

Let $p = proj_U v$. Then $||v - p|| < ||v - u||$ for all $u \in U$ and $u \neq p$.

Proof. Set $p = \sum_{j=1}^{n} \mathrm{proj}_{u_j} v$. We show that $p = \mathrm{proj}_U v$. This follows from $v - p \in U^{\perp}$. Indeed,

$$\langle p,\, u_j\rangle = \langle \sum_{k=1}^{n} \mathrm{proj}_{u_k} v,\, u_j\rangle = \langle v,\, u_j\rangle.$$

Let $u \in U$ and $u = \sum_{k=1}^{n} x_k u_k$. Thus

$$\langle v,\, u\rangle = \sum_{k=1}^{n} x_k \langle v,\, u_k\rangle = \sum_{k=1}^{n} x_k \langle p,\, u_k\rangle = \langle p,\, \sum_{k=1}^{n} x_k u_k\rangle = \langle p,\, u\rangle.$$

Therefore $\langle v - p,\, u\rangle = 0$ for all $u \in U$. Hence $v - p \perp U$. Thus $p = \mathrm{proj}_U v$.

Now let $u \in U$. Then $p - u \in U$ and $v - p \in U^{\perp}$. Thus

$$||v - p||^2 + ||p - u||^2 = ||v - u||^2.$$

Therefore $||v - p|| \leq ||v - u||$ and equality holds if and only if $u = p$. ■

Solution via calculus. The function $f(x) = ||Ax-b||^2$ is a linear function of n variables $x_1, x_2, \ldots, x_n$. Let us find the critical points of f. These are the solutions to the system of linear equations

$$\frac{\partial f}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n.$$

These n equations as one vector equation $\nabla f(x) = 0$. The function $f(x)$ is given by

$$f(x) = ||Ax - b||^2 = \sum_{i=1}^{m} \left[\sum_{j=1}^{n} A_{ij} x_j - b_i \right]^2.$$

Calculate the partial derivatives $\frac{\partial f}{\partial x_k}$:

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= \sum_{i=1}^{m} 2 \left[\sum_{j=1}^{n} A_{ij} x_j - b_i \right] A_{ik} \\ &= \sum_{i=1}^{m} 2(A^T)_{ki}(Ax - b)_i \\ &= 2(A^t(Ax - b))_k. \end{aligned}$$

Hence the critical points are solutions of the linear equations,

$$\nabla f(x) = 2A^t(Ax - b) = 0.$$

These equations are called the **normal equations**:

$$A^t A x = A^t b.$$

If the column vectors of A are linearly independent then $A^t A$ is invertible. Hence there is a unique solution to the least squares problem

$$\hat{x} = (A^t A)^{-1} A^t b = A^+ b.$$

Theorem 2.6. *Let Let $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and $x \in \mathbb{R}^m$. Then $AA^+b \in C(A)$ and it is the vector in $C(A)$ that is closest to b.*

Proof. It is clear that $AA^+b \in C(A)$. For any $x \in \mathbb{R}^m$ we have the orthogonal decomposition

$$x = AA^+x + (x - AA^+x).$$

Note that $AA^+ \cdot AA^+ = A(A^+AA^+) = AA^+$. Hence

$$(AA^+x)^t(x - AA^+x) = x^tAA^+x - x^tAA^+ \cdot AA^+x = 0.$$

Hence AA^+b is the projection of b in $C(A)$. Thus AA^+b is the best approximation of b in $C(A)$. ■

Acknowledgement. Thanks are due to Geetha Venkataraman for a careful reading of the manuscript.

3 Exercises

1. Show that the pseudoinverse of $0 \neq u \in \mathbb{C}^n$ is u^*/u^*u. If $u = 0$ then $u^+ = u^t$.

2. Prove that if A has orthogonal columns then $A^+ = A^*$.

3. Let $A \in \mathbb{C}^{m\times n}$ and $A = BC$ be its rank decomposition. This means if $\operatorname{rank} A = r$ then $B \in \mathbb{C}^{m\times r}$ and $C \in \mathbb{C}^{r\times n}$ have rank r. Show that
$$A^+ = C^+B^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*.$$

4. Let $A \in \mathbb{C}^{m\times n}$. Prove the following properties of A^+ : (a) $(A^+)^+ = A$, (b) $(A^+)^t = (A^t)^+$, (c) $\overline{A^+} = (\overline{A})^+$, (d) $(A^+)^* = (A^*)^+$, (e) $(zA)^+ = z^{-1}A^+$ for $z \neq 0$.

5. Show that $(AB)^+$ need not be B^+A^+.

6. Show that $(AA^*)^+ = (A^*)^+A^+$.

7. Let $A \in \mathbb{C}^{m\times n}$. Let I_n denote the $n \times n$ identity matrix. Let $P_1 = I_n - A^+A$ and $P_2 = I_m - AA^+$. Prove the following:
(a) Show that P_1 and P_2 are orthogonal projectors, i. e. $P_k^2 = P_k$ and P_k is Hermitian for $k = 1, 2$.
(b) $\ker(A) = C(P_1), C(A) = \ker(P_2), \ker(A^+) = C(P_2), C(A^+) = \ker(P_1)$.
(c) $C(A) = \ker(A^+)^\perp$ and $C(A^+) = \ker(A)^\perp$.
(d) $\ker(A) \oplus C(A^+) = \mathbb{C}^n$ and $\ker(A^+) \oplus C(A) = \mathbb{C}^m$ both being direct sums of orthogonal subspaces.

8. Let A be an $m \times n$ real matrix with orthogonal column vectors. Let $b \in \mathbb{R}^m$. Show that $\hat{x} = A^tb$ minimises $||Ax - b||^2$.

9. Let A be an $m \times n$ real matrix and $b \in \mathbb{R}^m$. Let $\operatorname{rank} A = n$. Let $\hat{x} = A^+b$ be the least squares approximate solution of $Ax = b$. (a) Show that for any $x \in \mathbb{R}^n$, $(Ax)^tb = (Ax)^t(A\hat{x})$. (b) Show that $x = \hat{x}$ minimises the angle between Ax and b.

10. Suppose that a tall $m \times n$ matrix A has linearly independent columns.Then it does not have a right inverse, i. e. There is no matrix X So that $AX = I$. Show that $||AA^+ - I| \leq ||AX - I||$ for any $n \times m$ matrix X. Here $||B||$ denotes any matrix norm of a matrix B.

11. Let A be an $m \times n$ matrix with linearly independent columns. Suppose that $A = QR$ is a QR factorization of A. Let $b \in \mathbb{R}^m$. Let $\hat{x}$ be the least squares solution of $Ax = b$. (a) Show that $A\hat{x} = QQ^tb$. (b) Show that $||A\hat{x} - b||^2 = ||b||^2 - ||Q^tb||^2$.

Bibliography

[1] O. M. Baksalary and G. Trenkler, *The Moore-Penrose inverse: a hundred years on a frontline of physics research*, Eur. Phys. J. H. **46**, (2021), 1-10.

[2] J. C. A. Barata and M. S. Hussein, *The Moore-Penrose Pseudoinverse. A Tutorial Review of the Theory,* arxiv: 1110.6882 (2011), 23 pages.

[3] S. L. Campbell and C.D. Meyer, Jr. (1991). Generalized Inverses of Linear Transformations, Dover (1991).

[4] Adi Ben-Israel and N. E. Thomas Greville, Thomas N.E. , Generalized inverses: Theory and applications (2nd ed.). New York, NY: Springer (2003).

[5] Stephen Boyd and Lieven Vandenberghe, Introduction to Applied Linear Algebra, vectors, matrices and least squares, Cambridge Univ. Press, 2018.

[6] G. Chen, Mathematical methods for data visualisation, Lecture Notes of MATH 253, San José State University, 2021.

[7] E. H. Moore, *On the reciprocal of the general algebraic matrix*, Bull. Amer. Math. Soc. **26** (1920), 394-395.

[8] Roger Penrose, *On best approximate solutions of linear matrix equations,* Math. Proc. Cambridge Philos. Soc. **51** (1955), 406-413.

[9] C. R. Rao, and S. K. Mitra, Generalized Inverse of Matrices and its Applications. New York: John Wiley and Sons. (1971).

[10] Helene Shapiro, Linear algebra and matrices, topics for a second course, Pure and Applied Undergraduate Texts 24, Amer. Math. Soc. 2015.

[11] Murali K. Srinivasan and Jugal K. Verma, Lecture notes on basic linear algebra (unpublished), (2021).

[12] G. Strang, Linear Algebra and its Applications, Fifth Edition, Wellesley-Cambridge Press, 2016.