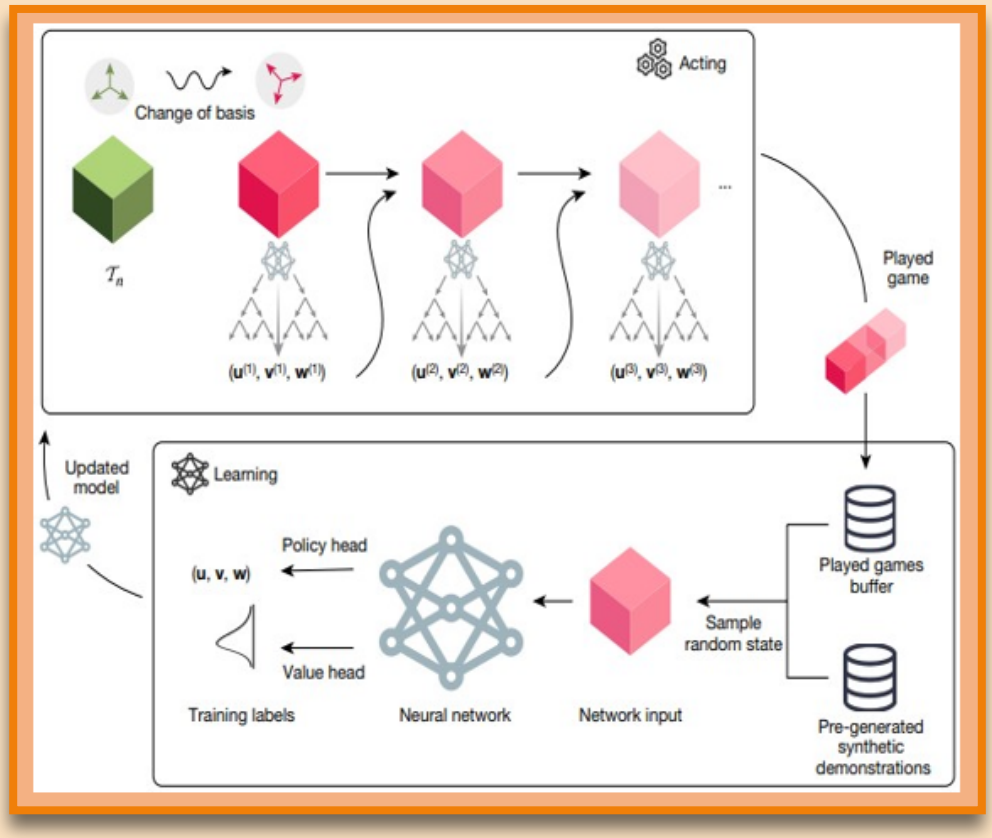
The Mathematics Consortium

BULLETIN

October 2023 | A TMC Publication | Vol. 5, Issue 2

Overview of Alphatensor



An AI System Autogenerating Matrix Multiplication Algorithms

The Mathematics Consortium

Bulletin

October 2023 Vol. 5, Issue 2

Contents

About the Cover Page: The figure on the cover page is Figure 2 from Reference [1] in Article-1 of this issue. The bottom box in the figure is a Deep neural network that takes as input a tensor S_t, and outputs triplets of vectors (**u, v, w**) a decomposition of S_t. It also outputs estimate of the future returns (for example, estimates of Rank(S_t)). The network is trained on two data sources: previously played games and synthetic demonstrations. The top box represents an actor (system) which plays the game. It receives the updated network which is used by the Monte Carlo Tree Search planner to generate new games.

Typeset and designed by: Praanjala Typesettings, Contact: +91 99603 03874, Email: mr.texdoc@gmail.com.

From the Editors' Desk

Recent developments in Artificial Intelligence (AI) are hitting the headlines of most popular and influential newspapers across the world. This has created an awareness among people at large and generated an intense discussion on whether AI is a boon or a curse for humanity.

All said and done, modern AI developments have the potential to bring about many positive changes in our lives. To name a few, AI can automate many tasks that are currently performed by humans, freeing up our time and resources for more creative and strategic work. AI can help to make better decisions in a variety of areas, such as business, finance, healthcare, environment preservation, accurate weather predictions etc. AI can be used to develop renewable energy sources, create new vaccines, and improve agricultural yields and thereby address some of the world's most pressing challenges.

However, it is also important to be aware of risks from these developments in order to ensure that they are used responsibly and ethically.

The major threats of AI include, widespread job displacement, particularly in industries that are heavily reliant on manual labour; AI systems could be exploited for malicious purposes, such as developing autonomous weapons or creating digitally altered sensitive videos that could be used to spread misinformation; AI could be used to create mass surveillance systems that track and monitor people's every move, which could lead to a loss of privacy and freedom. Some experts worry that AI could eventually become so powerful that it may surpass human intelligence and control, which could lead to catastrophic consequences.

The following are some measures recommended by AI researchers to mitigate the risks of AI:

- Invest in education and training to help people develop the skills they need to succeed in the AI economy.
- Evolve ethical guidelines for the development and use of AI.
- Regulate AI to prevent it from being used for harmful purposes.
- Research ways to make AI systems more transparent and accountable.

In Article 1, Prof. Ramamohan and Prof. Pathak have discussed the historical perspective of AI. They highlight the significant role played by Mathematics in evolution of modern AI techniques and discuss how these techniques are reconfiguring the intellectual and physical work spaces throughout the world. Specifically, they have given illustrations of some complex mathematical problems, resolution of which was significantly impacted by advanced computational techniques and modern AI systems.

The second Article contains excerpts from a conversation of Professor Ramachandran Balasubramanian, a renowned number theorist and a former Director of The Institute of Mathematical Sciences, Chennai, with Prof. Ambat Vijayakumar and his students, at CUSAT, Cochin.

In Article 3, Dr. D. V. Shah gives an account of significant developments in the Mathematical world during recent past, including brief write-ups on the winners of the 2024 Breakthrough Prize, the 2024 Maryam Mirzakhani New Frontiers Prize and 2024 New Horizons in Mathematics Prize. An obituary note on renowned mathematicians Prof. J. Tinsley Oden, Prof. Melvin Gordon Rothenberg and Prof. T. Parthasarathy, who passed away recently, is also included in this article.

We also pay tributes to India's legendary Statistician PROF. C. R. Rao, who passed away on 22 August, 2023.

In the Problem Corner, Dr. Udayan Prajapati presents two solutions to the problem posed in the July 2023 issue. These solutions are given by Prof. J. N. Salunke and by a student Pranjal Jha. A problem on Geometry is also posed, for our readers. Dr. Ramesh Kasilingam gives a calendar of academic events, planned during November, 2023 to January, 2024 in Article 5.

We are happy to bring out this second issue of Volume 5 in October, 2023. We thank all the authors, all the editors, our designers Mrs. Prajkta Holkar and Dr. R. D. Holkar, and all those who have directly or indirectly helped us in bringing out this issue on time.

Chief Editor, TMC Bulletin

1. Evolution of AI and its Impact on Computational Mathematics: Part - 1

S. Ramamohan and Vijay D. Pathak
Retired faculty, M. S. University of Baroda, Vadodara
Email: srmmsu@gmail.com and vdpmsu@gmail.com

Abstract: The recent advancements in the field of Artificial Intelligence (AI) based on Deep Artificial Neural Networks has made it possible to develop AI systems like AlphaGo, AlphaFold, AlphaTensor, which are reconfiguring the intellectual and physical work spaces throughout the world. While Mathematics plays a significant role in these developments, many of the long pending unsolved problems in Mathematics have been solved as a result of the progress in Computing Technologies and AI. Specifically, the AlphaTensor system introduced in [1], has been successful in auto-generation of the matrix multiplication algorithms which are faster than the best known algorithms in the literature.

In this two-part article, we give an account of Evolution of AI systems and their impact on Computational Mathematics. In Part-1, we discuss important milestones in these developments starting from origins of Artificial Neural Networks (ANN) up to the recent Deep Neural Networks (DNN) and briefly discuss AI systems developed based on DNN. At the end of the first part we mention some complex mathematical problems, resolution of which was significantly impacted by these advanced computational techniques and AI systems. In Part 2, we will discuss how increasingly faster Matrix multiplication algorithms have evolved in the last few decades. We will explain how any matrix multiplication algorithm can be represented by a 3-dimentional tensor and how recently developed AI system "AlphaTensor" has been used for auto-generation of efficient matrix multiplication algorithms.

1.1 Introduction

Human efforts to design machines to carry out arithmetical operations on numbers dates back to early 1640s when Blaise Pascal designed his first ever mechanical adding machine -"Pascaline". Persistent efforts in this direction led to the development of efficient electronic calculators. This process culminated in the development of digital computers with the concept of stored programming principle introduced by Von Neumann in 1940s. Since then, there has been a rapid progress in the computing devices resulting in exponential reduction in their size and exponential growth in their computing power. These increasingly powerful computers were mainly programmed using formal deterministic algorithms based on available techniques, to resolve the problem under consideration. The development of computational techniques using this approach, have helped in tackling difficult problems in several fields including mathematics.

Artificial Neural Networks (ANN) which emerged in 1940s, adopted a different approach to problem solving. ANN had capabilities to learn Input-Output relationships from the available data regarding a phenomenon, and then generalize it to infer reasonable assertions on data that it had not been exposed to earlier.

The origins of Artificial Intelligence (AI) can be traced back to the early 20th century, when philosophers and scientists began to explore the possibility of creating machines that could think like humans. In 1950, Alan Turing published a landmark paper in which he proposed the Turing test, a test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human. In the 1950s and 1960s, AI researchers developed a number of early AI systems, including initial versions of expert systems, natural language processing systems, and machine learning systems. However, many of these early systems were limited in their capabilities.

The true challenge to AI proved to be solving the tasks that are easy for people to perform but hard for people to describe formally - problems that we solve intuitively, that feel automatic, like recognizing spoken words or faces in images, and generating innovative ideas based on current knowledge and available data, for tackling problems under consideration. Among several approaches the scientists employed to meet these goals of AI, one prominent approach was to

build systems mimicking the functioning of the human brain. The ANN architectures fitted naturally into this approach. The advancement in ANN methodologies led to Multilayer networks and to Deep Neural Networks, by the first decade of 21^{st} century. Currently several goals of AI have started to be realized by systems that have DNN in their core. Some such systems like ChatGPT, Bard, AlphaGO, AlphaTensor are making deep impact in all walks of human life.

In part 1 of this article, we will be discussing various milestones in the development of one set of ANN architectures called feedforward networks (in which signals propagate in the direction from inputs to outputs) leading to DNNs and evolution of new AI systems. We will also briefly discuss the mathematical theories backing these developments and various complex mathematical problems resolved using advanced computational techniques and the power of AI. Whereas in Part 2, we will give an account of various Matrix multiplication algorithms available in the literature and try to explain how the AI techniques can be used for auto generation of efficient and verifiable Matrix multiplication Algorithms.

1.2 Artificial Neural Networks(ANN)

The information processing cells of the brain are the neurons. Each neuron is a type of cell with nucleus and communication links called dendrites and axon. A neuron receives electrochemical signals through several dendrites and the axon carries an output signal to other neurons. Inspired by this microscopic behavior of real neurons, in early 1940's, Warren McCulloch and Walter Pitts [21] introduced the concept of an artificial neuron. McCulloch and Pitts aimed to create a simplified mathematical model of a neuron that could capture essential aspects of neural behavior and perform computations similar to those carried out by real neurons. They were influenced by the concept of a neuron as a basic building block of the brain, capable of receiving inputs, processing them, and producing an output signal based on a threshold or an activation function.

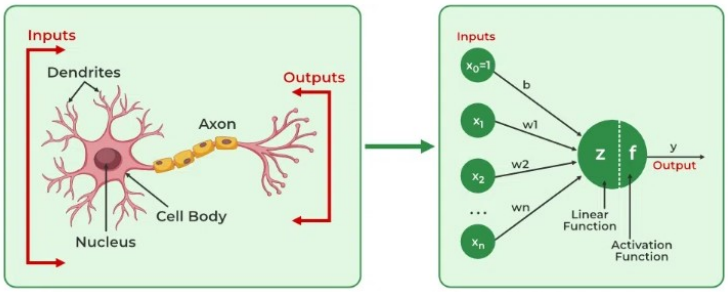


Figure 1: Biological Neuron to Artificial neuron (See [30])

One of the distinguished abilities of human intelligence is the ability to receive sensory inputs regarding a process or phenomenon and make useful inference regarding the real world, recognize patterns effortlessly and based on this recognition classify objects into categories. Human brain learns to classify data of a real-world system from the sample inputs it receives and then put the future unseen patterns in appropriate categories. Formally, the data classification problem for 2-class case [27] with linear separating hyperplanes in $\mathbb{R}^n$ can be stated as:

Given a set of Q data samples or patterns $X = \{x_1, x_2, \ldots, x_Q\}$, $x_i \epsilon \mathbb{R}^n$ drawn from 2 classes X_0, X_1, find an appropriate hyperplane that separates the two classes so that the resulting classification decisions on the unseen samples from these classes are on an average in close agreement with the actual outcome.

Mathematically, it can be proved that a hyperplane separating two sets X_0, X_1, exists if the sets are linearly separable, that is, if the convex hulls $Cov(X_0), Cov(X_1)$ of these sets are disjoint.

McCulloch and Pitts' artificial neuron model (see Figure 1) has been created with a modest objective to solve a 2-class linear classification problem in Boolean space $B^n = \{0,1\}^n$ which can be associated with a Boolean function g: $B^n \rightarrow \{0,1\}$ so that the two sets to be separated are $X_0 = g^{-1}(0)$ and $X_1 = g^{-1}(1)$. The model consisted of binary inputs $\{x_1, x_2, \ldots, x_n\}$, a bias x_0, output y, weights $\{w_0, w_1, \ldots, w_n\}$ and an activation function f. Here, x_0 is taken as a constant 1 or -1, and the weight w_0 (b in Figure 1) corresponds to the resting membrane potential

which is unknown to be determined. The neuron produced an output signal $y = f\left(\sum_{i=0}^{n} w_i x_i\right)$. Mathematically it is possible to decide whether such weights exist or not and if they exist, to compute appropriate weights corresponding to the classification problem (or associated Boolean function) under consideration. However, the question here is whether we can train the network to generate such appropriate weights from the known input-output patterns starting from some initial weights (normally all zero). The procedure that makes it possible for the ANN to learn this weight modification process is called a learning algorithm, which is an integral part of ANN and makes it a starting point of AI.

McCulloch and Pitts' work, published in 1943 [21], laid the foundation for the development of artificial neural networks and became a significant contribution to the field of computational neuroscience and artificial intelligence. It provided a starting point for subsequent researchers to explore more complex neural network architectures and learning algorithms, leading to the development of modern neural network models and deep learning techniques.

1.2.1 Learning in neural networks

Learning in neural networks refers to the ability of a neural network model to improve its performance on a task or problem through experience or exposure to training data. It involves adjusting the weights and biases of the network's connections to optimize its behavior and enhance its ability to more accurate predictions or classifications. The significance of learning in neural networks lies in its capacity to enable machines to acquire knowledge and skills from data without explicit programming. Neural networks can learn patterns, relationships, and representations directly from the data they are trained on.

In this perspective, ANNs can be regarded as model-free estimators with connectionist network of simple computing units (neurons) along with learning rules that train the network to discover inherent dynamics in the data exposed to the network.

Some key aspects of the learning in neural networks are as follows:

Adaptability and Generalization: Neural networks can adapt to changing environments or new input data by adjusting their internal parameters. Through learning, neural networks can generalize from a limited set of training examples to make accurate predictions or classifications on unseen data. This generalization ability is crucial for the network to be able to handle real-world scenarios beyond the specific examples it has been trained on.

Pattern Recognition: Learning allows neural networks to recognize complex patterns and extract meaningful features from raw data. This ability has broad applications in areas such as image processing and speech recognition, natural language processing, and recommendation systems.

Nonlinear Modelling: Neural networks can capture and model nonlinear relationships between input and output variables. This flexibility makes them suitable for solving complex problems that may involve intricate nonlinear interactions.

Automation and Efficiency: Neural network learning automates the process of knowledge acquisition and allows machines to learn and improve performance without human intervention. This can lead to more efficient and scalable solutions in various domains.

Uncovering Hidden Information: Neural networks can learn to uncover hidden or latent representations in the data. They can discover underlying structures, detect anomalies, or reveal insights that may not be easily apparent through traditional analytical methods.

1.2.2 Types of ANN Learning

Learning in neural networks is very significant because it empowers machines to acquire knowledge, make informed decisions, and perform complex tasks by autonomously learning from data. Depending on the nature of the problem of learning we can classify learning algorithms into three categories: Supervised, Unsupervised and Reinforcement.

Supervised Learning

Many real world processes connect input parameters to the resulting output parameters. The exact relationship among the input and output parameters may not be known and may even be difficult to predict. However, a data set in the form of pairs of input-output vectors, $\{(x_i, y_i)|x_i \epsilon \mathbb{R}^n,\ y_i \epsilon\ \mathbb{R}^p, i = 1, 2, \ldots N\}$, is available and our task is to predict output vector y corresponding to a given (unseen) vector x. If ANN is designed to carry out this task, the available data is used for training the network. When an input x_i is presented to the system, it generates an output z_i, depending on the current values of the weights. Supervised learning computes the error $||y_i - z_i||$ (some norm quantifying the difference between predicted and desired outputs) and iteratively modify the weights so as to reduce this error. The implementation of the supervised learning algorithm is usually in the form of difference equations that are designed to work with such global information.

The set of data implicitly describes the behavior of the unknown function $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$. Supervised learning encodes this behavioristic pattern into the network by approximating the function f. Traditional techniques such as linear or non-linear (polynomial) regression assume some mathematical form and then attempt to estimate the function through determination of coefficients in the polynomial. ANN methods make no such assumptions.

Unsupervised Learning

Unsupervised Learning Algorithms learn patterns and structures from unlabeled data without any explicit target or output labels. An unsupervised learning system attempts to represent the entire data set which consists of only input vectors $\{x_i \epsilon \mathbb{R}^n, 1, 2, \ldots N\}$, by employing a small number of prototypical vectors - enough to allow the system to retain desired level of discrimination between samples. Note that there is no teaching input. In other words, the system attempts to answer a question: Given a set of data samples as above, is it possible to identify well-defined clusters, where each cluster defines a class of vectors which are similar in some broad sense. Clusters help to establish a classification structure within a data set that has no categories defined in advance. Learning in an unsupervised system is often driven by a complex competitive-cooperative process where individual neurons compete and cooperate with each other to iteratively update their weights based on the present input. Only winning neurons or clusters of neurons learn in each iteration.

The Reinforcement learning

The reinforcement learning is similar to supervised learning because it receives some feedback from its environment, not necessarily in terms of desired output values (which may or may not be known) corresponding to an input vector. The feedback obtained here is evaluative in terms of rewards from the environment as the consequences of its actions. This external feedback, called the reinforcement signal, is used to adjust the weights so as to get a better reward in next iteration. The agent does this by trial and error.

1.3 Evolution of ANN and Learning mechanisms

The models of ANN are specified by: (i) model's synaptic interconnections (ii) the training or learning rules used for updating the connection weights (iii) an activation function for each neuron.

An ANN consists of highly interconnected neurons such that each neuron's output is connected through weights to other neurons or to itself. The arrangements of neurons to form layers and connection pattern formed within and between layers is called network architecture. The ANN architectures can be classified as : (i) single layer or multilayer networks (ii) feed-forward, feedback or recurrent networks.

In single layer networks the neurons receiving inputs (input layer) are directly connected to neurons producing outputs (output layer). In multilayer networks between input and output layers

there are one or more layers which are not directly connected to the environment and hence called hidden layers.

A network is said to be a feed-forward network if output of a neuron in any layer is not an input to a neuron in the same layer or in the preceding layers. On the other hand, when output of a neuron is directed back as an input to a neuron in the preceding layers or the same layer or to the same neuron then the network is called a feedback/recurrent network. An activation function or a transfer function maps a net input to a neuron (weighted sum of all inputs to the node) into an output. Most commonly used activation functions are: (i) binary step function based on some threshold value (ii) binary sigmoid function based on some steepness parameter (iii) a Ramp function (iv) hyperbolic tangent function (v) Gaussian function based on two parameters (vi) Stochastic bipolar function associated with some probability distribution over input space and (vii) the more recent ReLU (Rectified Linear Unit).

We give below some milestones in the evolution of ANN leading to Deep neural networks.

1.3.1 Neuron and Hebbian Learning (1949)

The McCulloch-Pitts neuron takes binary inputs and produces a binary output based on a predefined activation function. The Hebbian learning rule, proposed by Donald Hebb in the late 1940s, is a fundamental concept in neural network learning. It is stated in the following passage from Donald Hebb's 1949 book [15] on page 62: "When an afferent (sensory) pathway repeatedly discharges into a cell, the synapse between them becomes more effective". The Hebbian learning rule can be expressed as "cells that fire together wire together". It suggests that when two connected neurons are activated simultaneously, the strength of the synaptic connection between them is increased. More precisely, in Hebb's rule, the weights are updated using the following formula: If w_j is a weight associated with connection between input x_j (j^{th} component of the input vector $\bar{x}$) and the targeted output y, then $w_j(\text{new}) = w_j + x_j y$, for $j = 0, 1, \ldots, n$, where $x_0 = 1$ or -1 and w_0 is the bias.

It is important to note that the Hebbian learning rule has limitations when applied over binary data and hence, representing data in bipolar form is advantageous.

1.3.2 Perceptron and Perceptron Learning Rule (1957)

In the 1950s and 1960s, the exploration and development of learning in neural networks gained significant attention, and one of the key contributors during this period was Frank Rosenblatt. Rosenblatt's pioneering work on perceptron culminated in two major publications:

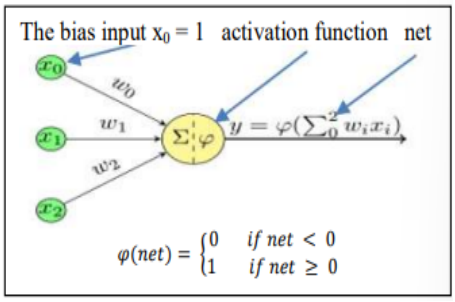


Figure 2: Perceptron

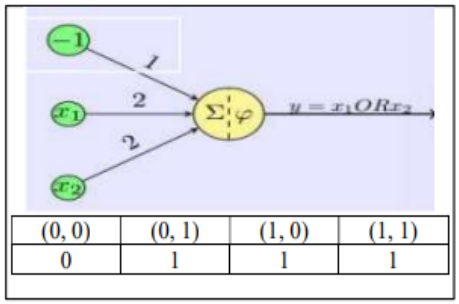


Figure 3: Perceptron for OR function

In 1958, Rosenblatt published a paper [24] which presented the perceptron as a mathematical model (see Figure 2) for simulating the information processing capabilities of biological neurons. It described the architecture of the perceptron, the learning algorithm based on the concept of error correction, and its potential applications in pattern recognition and information storage.

In 1962, Rosenblatt published a book [25] which expanded upon the ideas presented in his earlier paper and provided a comprehensive account of perceptrons, their structure, learning algorithms, and theoretical analysis. It also included practical examples and applications of perceptrons (such

as given in Figure 3), along with discussions on the limitations and potential future developments of the model.

Rosenblatt's work on perceptrons introduced the concept of supervised learning, proposed the Perceptron Learning Rule (1957) which was designed for single-layer perceptrons and proved its convergence.

The precise statement of the Perceptron Learning Rule and the Perceptron Convergence Theorem are given below:

- Perceptron Learning Rule: Given a binary classification problem where each input instance is represented by an associated feature vector $\bar{x} = (x_0, x_1, \ldots, x_n)$, with $x_0 = 1$ or -1 and $x_1, \ldots, x_n$ are components of input vector and a corresponding target output y (either 0 or 1), the Perceptron Learning Rule updates the associated weight vector $\bar{w} = (w_0, w_1, w_2, \ldots, w_n)$, with w_0 as bias, as follows: Initialize the weight vector w to small random values or zeros. For each training example $(\bar{x}, y)$, do the following:

  a. Compute the predicted output $z = \varphi(\bar{w} \cdot \bar{x})$, where activation function φ is the binary step function.

  b. Update the vector $\bar{w}$ using the following update rules: $\bar{w}(\text{new}) = \bar{w}$ - $\text{sgn}(\bar{w} \cdot \bar{x})\eta\bar{x}$ if $z \neq y$, otherwise $\bar{w}(\text{new}) = w$. Here, η is the learning rate, which controls the size of weight updates.

- Perceptron Convergence Theorem: The Perceptron Convergence Theorem, proved by Frank Rosenblatt in 1962, states that: Given a linearly separable dataset, there exists a choice of initial weights and learning rate that will make the Perceptron Learning Rule converge in a finite number of steps.

The theorem ensures that if the data is linearly separable, the Perceptron learning rule will eventually find a solution by adjusting the weights based on the error in the predictions. However, if the data is not linearly separable, the Perceptron learning rule may not converge and will not find a separating hyperplane. In 1969, Marvin Minsky and Seymour Papert published the book [25] "Perceptrons", which demonstrated some of these limitations and highlighted the challenges of training perceptrons to solve complex problems.

The weights for simulating the "OR" function, are depicted in Figure 3. These weights can be obtained using the Perceptron learning rule, starting with initial weight vector $\bar{0}$, and training the network iteratively by repeatedly giving 4 pairs of input vector $\bar{x}$ and corresponding outputs y in each iteration. This iterative process is said to converge if the weights do not change during an iteration. Observe that the sets to be separated (corresponding to OR function) in space B^2 are $X_0 = \{(0,0)\}$ and $X_1 = \{(0,1),(1,0),(1,1)\}$ and the separating hyperplane (line) is represented by equation: $2x_1 + 2x_2 - 1 = 0$.

The limitations of Perceptron can be realized if we try to simulate XOR function (see Figure 4). One can easily see that to simulate XOR function, the weights should satisfy the conditions: $-w_0 < 0$; $-w_0 + w_2 \geq 0$; $-w_0 + w_1 \geq 0$; $-w_0 + w_1 + w_2 < 0$; which is impossible to satisfy by any real values of weights. Here the sets to be separated are $X_0 = \{(0,0),(1,1)\}$ and $X_1 = \{(0,1),(1,0)\}$. The convex hulls of these two sets are line segment L_1 joining the two points $(0,0)$ and $(1,1)$ and the line segment L_2 joining the points $(0,1)$ and $(1,0)$ which are not disjoint and hence the sets are not linearly separable.

This led to a decline in interest in neural network research until the resurgence of the field in 1980s with the development of more advanced learning algorithms and architectures, such as backpropagation and multi-layer neural networks. However, Rosenblatt's work on perceptrons remains a landmark in the history of neural networks, laying the foundation for later developments in learning algorithms and inspiring further research in the field.

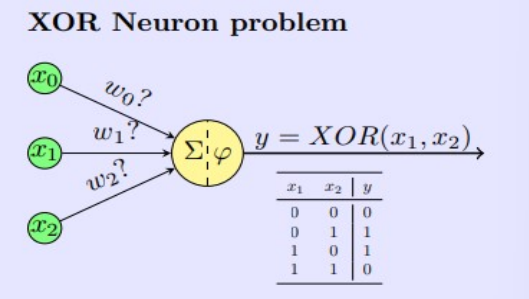


Figure 4

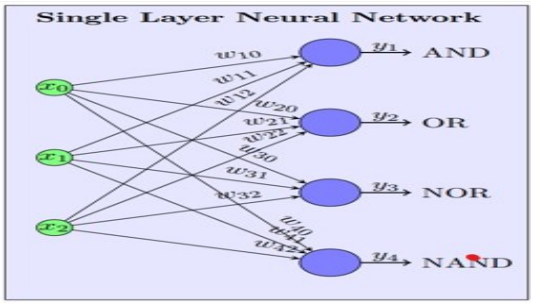


Figure 5

1.3.3 Adaptive Linear Systems and Delta Rule (1960)

Delta rule (also known as Widrow- Hoff rule or the Least Mean Square (LMS) rule) was developed by Bernard Widrow and Marcian Hoff [29]. It is widely used in the development of adaptive linear systems and commonly used for training linear regression models and single-layer neural networks (see Figure 5). The Delta Rule updates the weights and bias based on the difference between predicted output and the target output, scaled by the learning rate, so as to minimize the meansquared error between the predicted output and the target output. Specifically, consider a single layer network connecting n input neurons to p output neurons and consider a pair of input output vectors $(\bar{x}, \bar{y})$, $\bar{x} \in \mathbb{R}^{n+1}$, (where $x_0 = 1$ for all vectors x as mentioned earlier) $\bar{y} \in \mathbb{R}^p$. Let w_{ij} be a weight associated with the connection between i^{th} input neuron to j^{th} output neuron and weight vector $\bar{w}_j = (w_{0j}, w_{1j}, \ldots, w_{nj})$, $j = 0, 1, \ldots, p$, and w_{0j} is the bias associated with the j^{th} output neuron. Then $y_{in\,j} = \bar{x}.\bar{w}_j$ and the mean-square error E between actual output vector $\bar{y}$ and $\bar{y}_{in}$, is given by: $E = \frac{1}{2}\sum_{j=1}^{p}(y_j - y_{in\,j})^2$ which is a function of weight vector $\bar{w}_{j.}$ we have to find $\bar{w}_j$ so as to minimize E. This minimization problem can be resolved using gradient decent method which is based on the result in multivariate calculus which states that: The function E decreases most rapidly in the direction of negative gradient of E with respect to weights w_{ij} for $i = 1, 2, \ldots, n$. Weight adjustment procedure based on this fact is the delta rule. Thus, with every input-output pair given as data to the ANN the weights w_{ij} are modified as follows:
$w_{ij}(\text{new}) = w_{ij} + \eta(y_j - y_{in\,j})$, for $i = 0, 1, \ldots, n$. Note that $(y_j - y_{in\,j})x_i$ is the i^{th} component of gradient of E and η is the learning rate.

1.3.4 Multilayer Perceptron (MLP) and Backpropagation Rule

During the 1980s and 1990s, the development of Multi-Layer Perceptron (see Figure 6) witnessed significant advancements, starting from the discovery of the Backpropagation algorithm to the success of LeNet-5, designed by Yann LeCun. A multi-layer perceptron could represent the XOR function. A general 2-layer perceptron is illustrated in Figure 7.

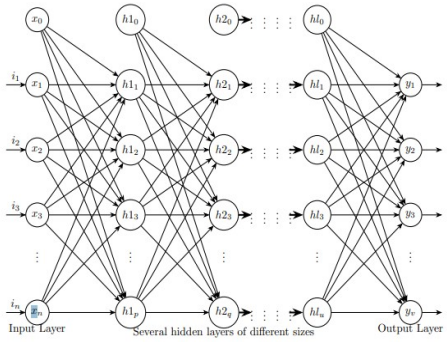


Figure 6: Multilayer Perceptron

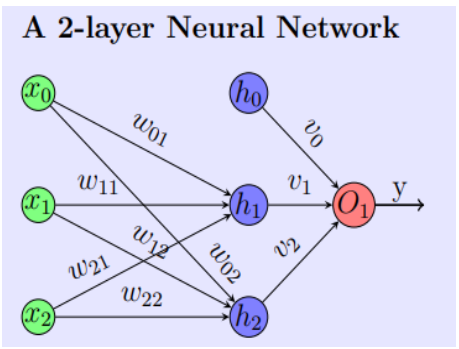


Figure 7: An XOR Neural Network

Here, $h_{in\,j} = \Sigma_{i=0}^{2} w_{ij}x_i$; $h_j = \varphi(h_{in\,j})$; $y_{in} = \Sigma_{j=0}^{2} v_j h_j$ and estimated y, $y_e = \varphi(y_{in})$, where

activation function φ is the binary step function.

Taking $x_0 = h_0 = -1$, $w_{01} = 1.5$, $w_{02} = 0.5$; $w_{ij} = 1$ for $i = 1,2$ and $j = 1,2$; $v_0 = 0.5$, $v_1 = -2$ and $v_2 = 1$, this 2-layer perceptron can simulate XOR function. The computations are shown in Table 2. Observe that the estimated output by the MLP exactly matches with the output of the XOR function.

x_1	x_2	$h_{in\,1}$	h_1	$h_{in\,2}$	h_2	y_{in}	y_e
0	0	-1.5	0	-1.5	0	-0.5	0
0	1	-0.5	0	0.5	1	0.5	1
1	0	-0.5	0	0.5	1	0.5	1
1	1	0.5	1	1.5	1	-1.5	0

Note that the XOR function divides input data set $\{(0,0),(0,1),(1,0),(1,1)\}$ into two non-separable sets. In the corresponding MLP, the weights between the input layer and the hidden layer generates a mapping $\mathbf{B}^2$ to $\mathbf{B}^2$ given by $(x_1,x_2) \to (\varphi(x_1+x_2-1.5),\ \varphi(x_1+x_2-0.5))$. This mapping maps input data set to $\{(0,0),(0,1),(1,1)\}$ which can now be divided into separable sets as per the XOR function. Then the weights in hidden and output layers can give the desired classification. The weights given here are not generated through any iterative process. These specific values are used to illustrate the advantage of MLP.

Throughout the 1980s and 1990s, researchers focused on developing and refining the architecture, learning algorithms, and training techniques for MLPs. Various modifications and improvements were proposed to enhance their performance. However, the discovery of the Backpropagation algorithm by Paul Werbos in 1976 [28] and its independent rediscovery in 1986 by Rumelhart, Hinton and Williams [26] was a major breakthrough for training MLPs. Backpropagation algorithms are a set of methods used to efficiently train artificial neural networks following a gradient descent approach which exploits the chain rule. This discovery sparked renewed interest in neural networks.

To get the basic understanding of the Backpropagation algorithm, consider an MLP as in Figure 7 with 2 neurons in input layer, two neurons in the hidden layer, 1 neuron in output layer and biases for hidden and output layers. Let the activation function for the neurons in the hidden and the output layers be the Sigmoid function $f(t) = 1/(1+e^{-t})$ and let η in $[0,1]$ denote a learning rate. The neurons in input layer transfer the inputs to the hidden layer neurons without any modifications. Since the sigmoid function is smooth, it is possible to apply backpropagation algorithm to train the network. This process requires availability of training data consisting of input patterns (x_1,x_2) with corresponding desired output y. Let vector $\bar{x} = (x_0,x_1,x_2)$, where $x_0 = -1$ and a vector $\bar{h} = (h_0,h_1,h_2)$, where h_1,h_2 are outputs at the hidden layer and $h_0 = -1$. As shown in the figure 7, weights connecting i^{th} input neuron to j^{th} hidden layer neuron are denoted by w_{ij}, for $i,j = 1,2$ and weights connecting hidden layer neurons to output neuron by v_1,v_2. The biases for hidden nodes are denoted by w_{0j}, $j = 1,2$ and the bias to the output node is denoted by v_0. Let vectors $\bar{w}_j = (w_{0j},w_{1j},w_{2j})$, for $j = 1,2$ and $\bar{v} = (v_0,v_1,v_2)$. In the forward pass we compute $h_j = f(h_{in\,j})$, where $h_{in\,j} = \bar{w}_j.\bar{x}$, for $j = 1,2$ and estimated value of y as $y_e = f(y_{in})$, where $y_{in} = \bar{v}.\bar{h}$. Then the square error $E = \frac{1}{2}(y-y_e)^2$. For each I/O pair (x_1,x_2), and y, the weights are modified in the backward pass so as to minimize the error E using the steepest decent criteria as follows:

$$\begin{aligned} v_j(\text{new}) &= v_j + \eta\frac{\partial E}{\partial v_j} = v_j + \eta\frac{\partial E}{\partial y_e}\frac{\partial y_e}{\partial y_{in}}\frac{\partial y_{in}}{\partial v_j} = v_j - \eta(y-y_e)\frac{e^{-y_{in}}}{(1+e^{-y_{in}})^2}h_j \\ &= v_j - \eta(y-y_e)y_e(1-y_e)h_j,\ j = 0,1,2; \end{aligned}$$

$$\begin{aligned} w_{ij}(\text{new}) &= w_{ij} + \eta\frac{\partial E}{\partial w_{ij}} = w_{ij} + \eta\frac{\partial E}{\partial y_e}\frac{\partial y_e}{\partial y_{in}}\frac{\partial y_{in}}{\partial w_{ij}},\ i = 0,1,2 \text{ and } j = 1,2; \\ &= w_{ij} - \eta(y-y_e)y_e(1-y_e)v_j f'(h_{in\ j})x_i, i = 0,1,2 \text{ and } j = 1,2; \\ &= w_{ij} - \eta(y-y_e)y_e(1-y_e)v_j h_j(1-h_j)x_i, i = 0,1,2 \text{ and } j = 1,2; \end{aligned}$$

Here, $(y - y_e)y_e(1 - y_e)v_j$ can be regarded as error $(y - y_e)$ computed at the output node scaled by slope $y_e(1 - y_e)$ of the signal passed from the j^{th} hidden layer neuron to the output node, which is backpropagated to a j^{th} hidden layer neuron through the corresponding weight v_j.

Thus, the MLP is trained using training data set until the error is below the tolerance limit. After the training process is over, the MLP is assumed to have captured the inherent dynamics in the training data. Then the trained network can be used to predict a reasonable output for unseen inputs.

Note that the Perceptron learning rule, Delta rule and Backpropagation learning rule are examples of Supervised Learning Algorithms since those procedures depend upon training patterns with output class label. And the Hebbian rule illustrates an Unsupervised Learning Algorithm.

Other notable unsupervised learning algorithms are (i) Autoencoders (ii) Restricted Boltzmann Machines (RBMs) (iii) Kohonen Self-Organizing Maps (SOMs) (iv) Generative Adversarial Networks (GANs) (v) Deep Boltzmann Machines (DBMs) etc. These unsupervised learning methods for ANNs enable the discovery of patterns, structures, and representations in the input data and have wide applications in various domains. Discussion of these learning rules is beyond the scope of this article.

1.4 Deep Neural Networks and Deep Learning

While for ANNs discussed above, it is possible to solve easy mathematical questions, and computer problems, including basic gate structures with their respective truth tables, it is tough for these networks to solve complicated image processing, computer vision, and natural language processing tasks. For these problems, we utilize **deep neural networks**, which often have a complex hidden layer structure with a wide variety of different layers, such as a convolutional layer, max-pooling layer, dense layer, and other unique layers. These additional layers help the model to understand problems better and provide optimal solutions to complex projects.

1.4.1 Beginnings of Deep Networks

The 1990s saw advancements in MLPs with the development of more efficient training algorithms and the exploration of different architectures. One notable architecture is the LeNet-5, developed by Yann LeCun [19] and his colleagues in the early 1990s. LeNet-5 was designed specifically for handwritten digit recognition and played a crucial role in advancing the field of convolutional neural networks (CNNs).

LeNet-5 consisted of seven layers (see Figure 8), including convolutional layers, pooling layers, and fully connected layers. It demonstrated the power of deep learning in achieving high accuracy on challenging tasks. LeNet-5 pioneered the use of convolutional operations, weight sharing, and pooling, which are now fundamental components of modern CNN architectures.

Though LeNet-5 had many features of DNN, it is still referred to as an MLP. The first recognized deep neural network (DNN) is considered to be the AlexNet model, developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton [18], which won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2012. AlexNet is a deep convolutional neural network architecture that achieved a significant breakthrough in the field of computer vision and demonstrated the power of deep learning. The AlexNet consisted of 8 layers (see Figure 9).

These developments led to the exploration of more complex network architectures, theintroduction of additional techniques such as regularization and dropout, and the application of neural networks to a wide range of fields, including computer vision, natural language processing, and speech recognition.

Most of the deep neural network architectures that have been successful are based on either Convolutional Neural Networks (CNN) or Recurrent Neural Networks (RNN) methodologies. CNNs are well-suited for tasks that involve processing spatial data, such as images and videos. RNNs are well-suited for tasks that involve processing sequential data, such as text and speech.

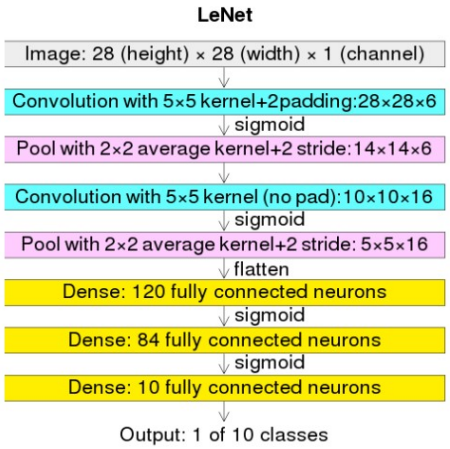


Figure 8: The 7 layers of LeNet-5 (See [31])

Figure 9: The 8 layers of AlexNet (See [31])

1.4.2 Convolutional Neural Networks

CNNs are a type of deep learning architecture designed for processing structured grid-like data, primarily used for tasks involving visual data such as images or videos. CNNs are composed of multiple layers of convolutional and pooling operations.

Convolutional Layers: Convolutional Layers are the fundamental building blocks of CNNs. They perform the convolution operation, which involves sliding a small window called a kernel or filter over the input data and computing element-wise multiplications and summations. Each convolutional layer consists of multiple filters, and each filter learns to detect a specific feature or pattern. The output of a convolutional layer is often referred to as feature maps or activation maps.

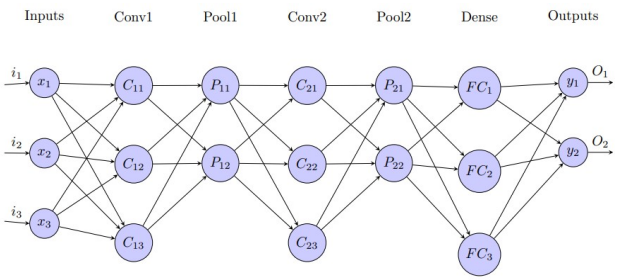


Figure 10: A Convolutional Neural Network

Pooling Layers: Pooling Layers are used in CNNs to reduce the spatial dimensionality of the feature maps generated by the convolutional layers. The most commonly used pooling operation is max pooling, where the input feature map is divided into non-overlapping regions, and the maximum value within each region is retained. This down sampling operation reduces the computational burden and makes the learned features more robust to small spatial translations or distortions. The pooling layers are typically inserted between consecutive convolutional layers to progressively reduce the spatial dimensions of the feature maps while retaining the most important features.

Fully connected (Dense) layers: The cascade of alternating convolutional and pooling layers in a CNN is followed by several fully connected layers in which every neuron in one layer connects to every neuron in the next layer, similar to a traditional neural network. They take the high-level features extracted by the earlier layers and use them to make predictions or classifications.

Some examples of successful deep neural network architectures that are based on CNNs are: (i) AlexNet (ii) VGGNet and (iii) ResNet.

1.4.3 Recurrent Neural Networks

RNNs are a type of neural network that excel in processing sequential data due to their ability to capture temporal dependencies. The structure of an RNN involves recurrent connections, creating

a loop that allows information to persist over time. This enables the network to analyze input sequences step by step, incorporating contextual information from previous steps. Variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) further enhance RNNs by addressing the vanishing gradient problem and introducing gating mechanisms to regulate information flow.

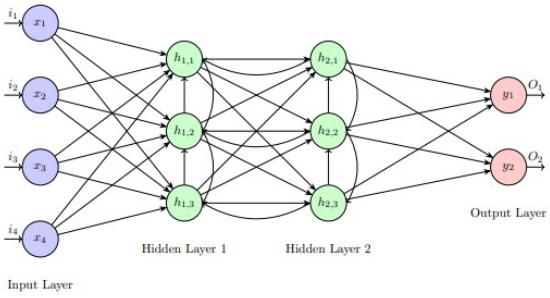


Figure 11: A simple Recurrent Neural Network

RNNs are particularly well-suited for tasks involving language modeling and generation. Their sequential nature allows them to capture the intricate patterns and dependencies present in natural language data. In large language models, RNNs are often used in a stacked or deep architecture, enabling the learning of hierarchical representations and capturing both low-level and high-level features. Training RNNs involves techniques like Backpropagation Through Time (BPTT), which extends the standard backpropagation algorithm to handle recurrent connections and learn from temporal dependencies.

RNNs have revolutionized the field of language modelling by providing a powerful framework for understanding and generating sequences of variable length. Their recurrent connections and variants like LSTM and GRU enable the modelling of long-term dependencies, making them instrumental in the development of large language models and applications such as text generation, language translation, and dialogue systems.

Some examples of successful deep neural network architectures that are based on RNNs are: (i) LSTM: Long Short Term Memory, (ii) GRU: Gated Recurrent Unit and (iii) Transformer.

The T in the famous Generative Pre-trained Transformer (GPT) series GPT-1, GPT-2, GPT-3 and GPT- 4 stands for Transformer and such systems are called Large Language Models (LLMs).

1.4.4 Factors contributing to the success of DNN

The development of DNNs really took off in the 2010s, with the introduction of new deep learning frameworks, and advancement in computing technology. Here are some of the key factors which are contributing to the success of DNN in accomplishing wide variety of AI tasks:

(i) Development of Learning Algorithms, tools and Libraries: The development of the backpropagation algorithm in the 1980s and Deep reinforcement learning algorithm (DRL) in early 2010s have made it possible to train DNN with large number of hidden layers. DeepMind's AlphaGo, AlphaFold and AlphaTensor systems which use DRL, have recently demonstrated capabilities of attacking difficult problems like playing games, medical diagnosis, financial trading, the Protein Folding Problem and discovering novel Matrix Multiplication Algorithms.

Building DNNs from scratch is time-consuming and requires enormous effort. To make deep learning simpler, several tools and libraries have been developed to yield an effective deep neural network model capable of solving complex problems with a few lines of code.

The most popular deep learning libraries and tools utilized for constructing deep neural networks are TensorFlow, Keras, and PyTorch.

(ii) The availability of large datasets: Data is the most critical component in constructing a highly accurate DNN model. During training, the model might also encounter issues such as underfitting or overfitting. Underfitting usually occurs due to a lack of data, while overfitting is a more prominent issue that occurs due to training data consistently improving while the test data remains constant. Hence, the training accuracy is high, but the validation accuracy is low, leading to a highly unstable model that does not yield the best results. These

data requirements are now fulfilled due to availability of large datasets in open domain and development of data augmentation techniques.

(iii) The availability of more powerful computational resources: Apart from a large amount of data, one must also consider the high computational cost/time of computing using the deep neural network. For example, Models like the Generative Pre-trained Transformer 3 (GPT-3) have 175 billion parameters, the original AlphaGo used a 120-layer deep convolutional neural network with 30 million parameters and the latest AlphaTensor system can use deep neural networks with up to 100,000 layers and billions of parameters. The compilation and training of models for complex tasks have become possible because of availability of resourceful Graphics processing units (GPU). Models can often be trained more efficiently on GPUs or Tensor processing units (TPU) rather than CPUs. Even with these resources, many AI tasks require number of days/weeks to yield satisfactory results.

1.4.5 The Mathematics of ANN and DNN

In general, most of the advanced technologies have good amount of mathematical theory backing them. The mathematical theories discussing the capabilities of ANN may be traced back to the Cover's Theorem in 1965 [8].

Cover's theorem states that any continuous function on a compact subset of Euclidean space can be uniformly approximated by a two-layer feedforward neural network with a single hidden layer having a finite number of hidden units. Cover's theorem is a fundamental result in the theory of neural networks.

Cover's theorem was extended in 1989 by Cybenko [9] for networks with three hidden layers and in 1991 by Hornik [16] for multilayer networks provided that the activation functions of the hidden units are non-linear.

These theorems provide a theoretical foundation for the use of deep neural networks (DNNs) for a wide variety of tasks. DNNs have been shown to be very effective at learning complex relationships between the inputs and outputs, and they have been used to achieve state-of-the-art results on tasks, such as image recognition, natural language processing, and machine translation.

It is important to note that the theorems on the universal approximation capability of ANNs do not provide any guarantees about the number of hidden units or the depth of the network that is required to approximate a given function. In practice, it is often necessary to experiment with different network architectures and hyperparameters to find a network that is able to accurately approximate the desired function.

There are still many open questions about the capabilities of ANNs and DNNs. For example, it is not fully understood why DNNs are able to learn complex relationships between the inputs and outputs so effectively. Additionally, it is not fully understood how the depth of the network affects its ability to learn complex relationships and the role of data in learning.

The mathematical theory of deep learning and deep neural networks is still in its infancy. Some of the reasons why developing a complete mathematical theory for deep learning has been challenging are :

1. Complexity of Networks: Deep neural networks can have millions or even billions of parameters, and the interactions between these parameters can be highly non-linear. This complexity makes it difficult to derive closed-form mathematical solutions.

2. Non-Convex Optimization: Training deep neural networks involves solving nonconvex optimization problems, which are notoriously challenging. This means that there can be many local minima, and finding the global minimum (optimal solution) is often not guaranteed.

3. Data Dependence: Deep learning's performance is highly dependent on the quantity and quality of training data. The theoretical underpinnings of how data affects learning and generalization are still an active area of research.

4. Architecture Variability: There are numerous neural network architectures, each with its own set of hyperparameters and characteristics. Developing a unified theory that encompasses all possible architectures is a formidable task.

5. Empirical Nature: Many advancements in deep learning have been driven by empirical experimentation and engineering rather than theoretical breakthroughs. This has made it challenging to develop a purely theoretical framework.

While there may not be a complete, all-encompassing mathematical theory for deep learning yet, significant progress has been made in understanding various aspects of deep neural networks. Researchers have made strides in areas such as optimization algorithms, generalization, adversarial robustness, and transfer learning. Additionally, there are mathematical frameworks that describe specific aspects of deep learning, such as the theory of overparameterization and the study of neural network expressiveness.

It's important to note that the absence of a complete mathematical theory does not diminish the practical value and impact of deep learning. Deep learning has proven to be highly effective in a wide range of applications, and researchers continue to make important advancements in both theory and practice. The development of a comprehensive mathematical theory for deep learning is an ongoing and complex research endeavour, and it may take more time to achieve a complete understanding of this transformative technology.

1.5 Impact of Computers on Mathematics/Scientific Research

Human beings have been using various devices to aid computation since ancient times. These devices may be physical like the Abacus, Pascaline, Napier's Bones, Slide Rule, Electrical calculators and Computers or these devices may be algorithms like the Sulvasutra procedure for computing square root of 2 or Euclid's algorithm or the Decimal Calculation methods or the modern computer programs. It seems that mainstream mathematics was not affected intimately by these devices until recent times. But in the past few decades, we are seeing instances of use of digital computers in mathematical discovery. The following are a few mathematical results that have been partially or completely proved using mathematical insight and computational resources including AI systems. Apart from these specific results, there are many instances of extensive use of computers in resolution of mathematical problems. For example, proof of nonexistence of a projective plane of order 10, solution of the Navier-Stokes equations for a wide range of problems in fluid mechanics, etc.

1.5.1 The Four Color Map Theorem

Arguably, the first famous case of using the assistance of digital computers in mathematical proofs was the resolution of the Four Color Map problem. In 1976 Kenneth Appel and Wolfgang Haken [2] published a two-page proof of the Four Color Map Theorem. The list of all possible configurations of regions in a map in Appel and Haken's approach contained 1,936 cases. These cases were generated by a computer program and it is claimed that any configuration can be reduced to one of these cases. The proof showed that for each of these cases, it was possible to color the regions using four colors without any adjacent regions having the same color.

Appel and Haken's proof was controversial when it was first published. Some mathematicians argued that the use of a computer made the proof invalid. Appel and Haken presented the details of their work in 4 papers published in 1977, in the Illinois Journal of Mathematics [3], [4], [5] and [6] totaling 263 pages. However, the proof has since been accepted by the mathematical community, and it is now considered to be a major achievement in mathematics.

1.5.2 The Kepler conjecture

This conjecture was first proposed by Johannes Kepler in 1611, which states that no arrangement of equally sized spheres filling space has a greater average density than that of the cubic close packing (face-centred cubic) and hexagonal close packing arrangements. The density of these arrangements is around 74.05%.

Thomas Hales [13] proved the Kepler conjecture in 1998 using a proof by exhaustion involving the checking of 12,000 cases. Each case is a configuration of spheres that could potentially have a higher density than the cubic close packing. Hales used a computer to check each case and showed that none of them had a higher density.

Hales' proof was not without controversy. Some mathematicians argued that the use of computers in the proof made it invalid. However, the vast majority of mathematicians accepted the proof.

In 2014, the Flyspeck project team, headed by Hales, announced the completion of a formal proof of the Kepler conjecture using a combination of the Isabelle and HOL Light proof assistants [10]. This formal proof was accepted by the journal Forum of Mathematics, Pi in 2017.

The formal proof of the Kepler conjecture is a major achievement in the field of mathematics. It affirms that it is possible to use computers to verify complex mathematical proofs, and it opens up the possibility of formalizing other mathematical proofs.

1.5.3 The classification of finite simple groups

The classification of finite simple groups is one of the most important and significant results in modern mathematics. It states that there are only 18 infinite families of finite simple groups, and 26 sporadic groups. (See [7], [11], [12], [14]).

The classification was completed in the 1980s by a team of mathematicians led by Robert Griess, John Conway, and Martin Gardner. The proof is over 15,000 pages long and uses a wide range of mathematical techniques, including group theory, representation theory, and algebraic geometry.

Computers were used extensively in the proof of the classification of finite simple groups. For example, computers were used to check the hundreds of cases that were involved in the proof. Computers were also used to develop new mathematical tools that were needed for the proof.

One of the most important computer-assisted proofs in the classification of finite simple groups is the proof of the Feit-Thompson theorem which states that every finite group of odd order is solvable.

This theorem was a major breakthrough in the classification, and it allowed the mathematicians to focus on groups of even order.

Another important computer-assisted proof is the proof of the classification of 2-transitive groups. 2-transitive groups are groups that act transitively on ordered pairs of elements of some set in a faithful manner. The classification of 2-transitive groups was a difficult problem, and it required the development of new mathematical tools.

Computers have played a vital role in the proof of the classification of finite simple groups.

1.5.4 The Protein Folding Problem

The protein folding problem [20] in Mathematical Biology refers to the challenge of understanding and predicting how a protein molecule acquires its unique three-dimensional structure, known as its folded conformation, from its linear sequence of amino acids. Proteins are fundamental biomolecules that perform a wide range of functions in living organisms, and their function is intricately tied to their structure.

AlphaFold is able to predict the structure of proteins [17], [22] with a high degree of accuracy in a wide variety of environments. This is a major breakthrough, as it will allow researchers to study the structure and function of proteins in much greater detail. AlphaFold was able to predict

the structure of the SARS-CoV-2 virus which was critical to the development of vaccines and treatments for COVID-19.

AlphaFold has the potential to revolutionize many areas of biology and medicine. For example, it could be used to design new drugs that target specific proteins, or to develop new treatments for diseases caused by protein misfolding.

Apart from these problems, one of the major developments in 2022, which has inspired authors to write this article, was the use of AI system AlphaTensor for auto generation of Matrix multiplication algorithms which are faster than best known algorithms in the literature. We will discuss this auto generation process along with the account of various matrix multiplication algorithms available in the literature, in the second part of this article.

Acknowledgement: We are extremely thankful to reviewers Mr. Venkat Kandru, Chief Technology Advisor, Pagewood Real Estate Services, USA, Prof. R. K. George, IIST, Trivandrum, Prof. M. C. Joshi, IIT Gandhinagar and formerly at IIT Bombay and Prof. Amit Nanavati, Ahmedabad University, and Prof. S. A. Katre, SPPU, Pune, for their valuable critical review and important suggestions in shaping this article.

Bibliography

[1] Alhussein Fawzi, Matej Balog, Aja Huang1, Thomas Hubert1, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J. R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, David Silver, Demis Hassabis & Pushmeet Kohli; Discovering faster matrix multiplication algorithms with reinforcement learning, Nature Volume 610, pages 47– 53 (2022).

[2] Kenneth Appel and Wolfgang Haken; Every planar map is four colorable, Bulletin of the American Mathematical Society, 82(5): 711–712, 1976.

[3] Kenneth Appel and Wolfgang Haken; Every planar map is four colorable. Part i: Discharging, Illinois Journal of Mathematics, 21(3): 429–490, 1977a.

[4] Kenneth Appel and Wolfgang Haken; Every planar map is four colorable. Part ii: Reducibility, Illinois Journal of Mathematics, 21(3): 491–567, 1977b.

[5] Kenneth Appel and Wolfgang Haken; Every planar map is four colorable. Part iii: The Birkhoff Grothendieck theorem, Illinois Journal of Mathematics, 21(3): 582–615, 1977c. 19 Kenneth

[6] Kenneth Appel and Wolfgang Haken; Every planar map is four colorable. Part iv: The proof, Illinois Journal of Mathematics, 21(3): 645–734, 1977d.

[7] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, R. A. Wilson; ATLAS of finite groups. Maximal subgroups and ordinary characters for simple groups. With computational assistance from J. G. Thackray, Oxford University Press ISBN:0-19-853199-0

[8] Thomas M. Cover; Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, IEEE Transactions on Electronic Computers, EC-14(5):326–334, 1965.

[9] George Cybenko; Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems, 2(4):303–314, 1989.

[10] Flyspeck project; "The Kepler conjecture", Flyspeck project website. 2014. https://flyspeck.org/wiki/Main_Page

[11] Robert L. Griess, Jr.; My life and times with the sporadic simple groups, Notices of the International Consortium of Chinese Mathematicians, Volume 9 (2021), Number 1, pp 11- 46.

[12] Daniel Gorenstein, Richard Lyons, and Ronald Solomon; The Classification of the Finite Simple Groups, AMS- Mathematical Surveys and Monographs, vol. 40. (40.1 (1994), 40.2 (1995), 40.3 (1997), 40.4 (1999), 40.5 (2002), 40.6 (2002), 40.7 (2018), 40.8 (2018), 40.9 (2021)).

[13] Hales, Thomas C; A proof of the Kepler conjecture, Annals of Mathematics (2005), pp. 1063–1183.

[14] Daniel Hartwig and Jenny Johnson; Guide to the Martin Gardner Papers SC0647, URL: http://library.stanford.edu/spc

[15] Donald O. Hebb; The Organization of Behavior: A Neuropsychological Theory, Wiley, 1949.

[16] Kurt Hornik; Approximation capabilities of multilayer feedforward networks, Neural networks, 4(2), 251–257, 1991.

[17] Jumper John, Evans Richard, et al.; Highly accurate protein structure prediction with AlphaFold, Nature,596(7873), 589–596, 2021.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton; ImageNet Classification Deep Convolutional Neural Networks, Part of Advances in Neural Information Processing Systems 25 (NIPS 2012).

[19] Yann LeCun, Leon Bottou, Yoshua Bengio and Patrick Haffner; Gradient-based Learning applied to Document Recognition, Proc. of the IEEE, Vol. 86, No. 11, 1998, 2278-2324.

[20] Cyrus Levinthal; How to fold a protein, Journal of Molecular Biology, 35(3): 629–637, 1969.Bull. Am. Math. Soc. 82, 126–128 (1976).

[21] Warren S McCulloch and Walter Pitts; A logical calculus of the ideas immanent in nervous activity, TheBulletin of Mathematical Biophysics, 5(4):115–133, 1943.

[22] Mikko Vihinen. Alphafold 2; A major breakthrough in protein structure prediction, Trends in Biochemical Sciences, 46(7): 529–531, 2021.

[23] Marvin Minsky and Seymour Papert; “Perceptrons”, MIT Press, Cambridge, MA, 1969.

[24] Rosenblatt; The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Psychological Review [5], 1958.

[25] Rosenblatt; Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, 1962.

[26] David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Willims; Learning representations by back-propagating errors, Nature Vol. 323, No. 6088, 1986: 533-536.

[27] Satish Kumar; Neural Networks – A classical approach, Second Ed., Tata McGraw Hill Education Pvt. Ltd., 2013.

[28] Paul J. Werbos, Beyond Regression: New tools for prediction and Analysis in Behavioural Sciences, Ph.D.Thesis, Harvard University,1974.

[29] Widrow, B. and Hoff, M.: Adaptive switching circuits, Western Electronic Show and Convention, Institute of Radio Engineers, 4, 1960, 96-104.

[30] https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/

[31] https://en.m.wikipedia.org/wiki/File:Comparison_image_neural_networks.svg

□□□

2. In Conversation with the Number Theorist- R. Balasubramanian

Amritha Varrier, Athira K., Adwaitha M. S.,(Final year M. Sc. Mathematics)
Ambat Vijayakumar, Emeritus Professor (vambat@gmail.com). Department of Mathematics,
Cochin University of Science and Technology, Cochin.

Professor Ramachandran Balasubramanian (aka Balu in the closer circles), a renowned number theorist and a former Director of The Institute of Mathematical Sciences, Chennai visited CUSAT recently.
He is a recipient of the Shanti Swarup Bhatnagar Prize for Science and Technology in 1990.
The French government's Ordre National du M´erite for "furthering Indo-French cooperation in the field of Mathematics" in 2003.
The Padma Shri in 2006 and The Lifetime Achievement Award of The Department of Atomic Energy, awarded by Dr. Manmohan Singh, the former Prime Minister of India in 2013.
He is also an inaugural class of Fellow of the American Mathematical Society in 2012.

Prof. Balasubramanian receiving Padma Shri from Dr. APJ Abdul Kalam

Prof. Balasubramanian with the students

Prof. Balasubramanian unassumingly spared a few minutes to answer the following questions on his mentors, research etc.
1. What were your memories from school or college related to mathematics? How did you realize Mathematics is your way?
⋆ I never decided mathematics was my way. I grew up in a village and nobody around me knew much about research or anything. But I liked mathematics. The teachers I had in high school were passionate about imparting whatever knowledge they had and creating enthusiasm among students for mathematics. Then I continued mathematics through college, but I had no clear idea about research. I studied at A. V. V. M. Sri Pushpam College, Poondi, Thanjavur district Tamil Nadu and completed my Master's there. The aim of every one around me was to clear some competitive examination and I was one of them. At that time, the HOD was V. Krishnamurthy who had just moved from Loyola College, Chennai to Sri Pushpam college. Now let me say a few things about Loyola college. That is the time when the Tata Institute of Fundamental Research (TIFR) had just started. Many young students from Loyola college had joined TIFR and later turned out to be top mathematicians (including Professors C. S. Seshadri and M. S. Narasimhan). Since Professor Krishnamurthy was in Loyla he was aware of TIFR. I remember it was around 1972 May, when I didn't know much about research, and was busy looking for clearing government examinations, and one day Prof. Krishnamurthy visited my house. He saw an advertisement for the Tata Institute and was very sure that I wouldn't have seen it at all, so with the help of his friend, he found my address and came to my house. He told me to apply for it and added a good recommendation letter also. Then I attended the interview and got selected.

2. Could you please comment on the contrast in academic and personal relationships between the research scholars and their guides as of now compared to the 80s?

⋆ If you are looking for the interaction between students and the faculty, my knowledge is practically based upon what I have experienced in my institute. My friends had told me that they learned more mathematics from their fellow research scholars compared to classrooms or textbooks. The student seminars and one to one interaction between the graduate students is one best way of learning. Unfortunately, I now see it is declining.

3. You were a visiting scholar at the Institute of Advanced Studies, Princeton, USA. What was your experience there like?

⋆ There I could talk to great mathematicians, like Professor A. Selberg (who was awarded the Fields Medal in 1950 and an honorary Abel Prize in 2002) and E. Bombieri (currently Professor Emeritus in the School of Mathematics at the Institute for Advanced Study in Princeton, who won the Fields Medal in 1974). I could befriend many number theorists. In particular I should mention Prof M. Ram Murty (currently at Queen's University, Canada) and James Haffner. The friendship with Ram Murty and his brother Kumar Murty (Currently the Director of the Fields Institute, Canada) has grown stronger in the subsequent years. Similarly next year, I went to the University of Illinois where I was fortunate to meet Professors Bruce Berndt (well known for the 'Ramanujan Notebooks') and H. Halberstam (renowned number theorist specialised in Sieve Theory, who passed away in 2014).

4. You are a great number theorist, how would you explain the beginning of your research? And the famous Waring's problem also.

⋆ You cannot always determine your research areas in the beginning. Of course, there are exceptions, like I have a friend K. Soundararajan who joined Princeton University for Ph. D. and he already had some research problems in his mind. But I didn't have any. My guide Prof. Ramachandra was interested in the Riemann zeta function at the time of my joining him, having established himself as a master in algebraic Number theory and Transcendental number theory by then. He didn't give me any specific problems, but just an overview. At that time I started thinking about problems on my own. I started to explore the zeros of the Riemann zeta function on the critical line and the mean square estimate of the Riemann zeta function on the critical line. After I was through with my thesis, he asked me to look at Waring's problem for biquadrates. I was able to prove that g(4) is at most 20. I should add here that I was relying on some calculations which was done by earlier researchers on this problem and it turned out that the calculations needed some modifications (as pointed out later by my coauthor). Around that time, Professor Ramachandra went abroad and met Professor Deshouillers who was also working in the problem along with Professor Dress (both are currently Professors at the University of Bordeaux, France). But we were working on different lemmas with different lines of thought. Professor Deshouillers realized that we can get the final result by combining both of our approaches.

5. You are a recipient of so many awards like the Padma Shri. Has it made a difference in your life?

⋆ The day I received the Padma Shri and the previous day when I did not have it, I do not feel much difference. I am happy to have gotten the Padma Shri. I am even happier because I got it from Dr. A. P. J. Abdul Kalam. In fact, I knew him already because both of us were interested in cryptology. His acquaintance is the one I still cherish.

Prof. K Ramachandra

6. Do you consider anybody as your mentor or guru?

⋆ I had so many good teachers in my school and college. If I have to rank them, then V. Krishnamurthy who was my teacher in Sri Pushpam college will get a high rank. I should add that I owe all my success to Professor K. Ramachandra. He was not only my teacher, he treated me as his own son.

7. What are your hobbies other than mathematics?

⋆ Like everyone else, I used to listen to music and read fiction and non-fiction. I used to play Bridge, a card game. When I became the Director of IMSC, I had to stop it and now after retiring, I am planning to start again.

8. How was your experience in IMSC?

⋆ My experience at IMSC has been wonderful. First of all, there are many advantages of being in a small Institute. Being an institute funded by DAE, we never had any serious issues with funding. I should say I was lucky in many respects during my tenure as the Director. First is the unstinted support I got from the chairmen of DAE at various times (I should thankfully remember Dr. R. Chidambaram, Dr. Anil Kakodkar and Dr. Srikumar Banerji in this connection) and the Joint Secretaries (like Mrs Sudha Bhave). The second is the support and the confidence I enjoyed from the Governing Council, in particular the chairman Dr. S. K. Joshi. I was fortunate to inherit a cohesive group of faculty (with absolutely no internal dissensions) from my predecessor Professor Ramachandran. The institute also had been fortunate to have good Registrars (Mr. Manja, Mr. Vishnu Prasad) who took care of most of the administration and the rest by the efficient P.A.'s, Mrs. Indra and Mrs. Vidyalakshmi.

9. What is your perception of the role of mathematicians in our society? What advice would you like to give to young mathematicians?

⋆ What I say will have so many exceptions, for a person who just started research, it is hard to know how difficult a problem is. At that time you have to consult your guide and he will tell you which one is easy or which one is difficult or which one is not so easy and not so difficult. Scholarship is a must to attack good problems. This is true of even Ramanujan, contrary to popular myth. And none of us is a Ramanujan. So you should go to your guide and ask for the materials needed to attack the problem and only after mastering it, you will have an idea. It is also possible when your teacher gives a problem and while reading the relevant materials you come up with another problem that is more appealing to you than the original one. For research, you should have adequate knowledge. You may not start your research on day 1 or day 100 but maybe you can start on day 181.

10. We know that your Erdös number is 2. How was your experience of communicating with Erdös?

⋆ As you know, the Erdös number describes the 'collaborative distance' between the Hungarian mathematician Paul Erdös and another person, as measured by authorship of mathematical papers. An author who has directly collaborated with Erdös has an Erdös number 1. The people who have collaborated with them have Erdös number 2. I have met professor Erdös, once in TIFR and many times during my visits abroad and have discussed with him. I have also corresponded with him, but it is rare.

11. Can you please enlighten us on the famous Balu-Koblitz theorem?

⋆ Professor Neal Koblitz of University of Washington visited IMSc a few years back and gave a course of lectures. We became close after the visit (even we went for a movie). Later, I met him in Bangalore in a conference on cryptology. In this lecture, he explained his result on the field of definition of Weil pairing of an elliptic curve. He suggested that the result should have an impact on Menezes Okamoto Vanstone (MOV) algorithm for attacking elliptic curve cryptosystems. After the lecture, we were walking towards the guest house for lunch. During the walk and during the lunch, we discussed about MOV and we realised that his result on the field of definition of 'Weil pairing' (a bilinear form, though with multiplicative notation, on the points of order dividing n of an elliptic curve E, taking values in nth roots of unity) in fact, proves that the MOV is not sub exponential in general. Rather surprisingly it turned to be one of my well cited papers.

12. In your long academic career, you would have come across many young math aspirants. Any one you still remember?

⋆ I am happy that I have played some role in the future of many young aspirants. The most

prominent is Kannan Soundarajan (recipient of several international recognitions including the Infosys Prize (2011) is currently at the Stanford University). Also many students from Chennai Mathematical Institute attended my course on Number theory and a lot of them did Ph.D. in Number theory.

13. About your family?
⋆ I am married to Lakshmi. I have a son and a daughter. My son Ravi is working in Qualcomm in Sandiego. Daughter Aruna and our son in law Niranjan teach computer science in Stonybrook university. So far we have one granddaughter through Aruna.

Waring's Problem

Waring's problem: Whether each natural number k has an associated positive integer s such that every natural number is the sum of at most s natural numbers raised to the power k. For example, every natural number is the sum of at most 4 squares, 9 cubes, or 19 fourth powers. Waring's problem was proposed in 1770 by Edward Waring FRS (1736 - 1798) who was a British mathematician after whom it is named. Its affirmative answer, known as the Hilbert-Waring theorem, was provided by the celebrated mathematician David Hilbert in 1909.
For every k, let $g(k)$ denote the minimum number s of k-th powers of naturals needed to represent all positive integers. Every positive integer is the sum of one first power, itself, so $g(1) = 1$. Some simple computations show that 7 requires 4 squares, 23 requires 9 cubes, and 79 requires 19 fourth powers; these examples show that $g(2) \geq 4$, $g(3) \geq 9$ and $g(4) \geq 19$. Waring conjectured that these lower bounds were in fact exact values.
Lagrange's four-square theorem of 1770 states that every natural number is the sum of at most four squares. Since three squares are not enough, this theorem establishes that $g(2) = 4$. That $g(3) = 9$ was established from 1909 to 1912 by Wieferich and A. J. Kempner, $g(4) = 19$ in 1986 by *R.* Balasubramanian, F. Dress, and J. M. Deshouillers, $g(5) = 37$ in 1964 by Chen Jingrun, and $g(6) = 73$ in 1940 by S. S. Pillai.

□□□

3. What is Happening in the Mathematical World?

Devbhadra V. Shah
Department of Mathematics, VNSGU, Surat; Email: drdvshah@yahoo.com

3.1 Celebrating 200^{th} Birth Anniversary of Gotthold Eisenstein

This year marks the 200^{th} birth anniversary of one of the renowned mathematicians *Ferdinand Gotthold Max Eisenstein.*

Eisenstein was born on April 16, 1823 in Berlin, Germany and made important contributions to Number Theory and Analysis. He worked on a variety of topics including quadratic and cubic forms, the reciprocity theorem for cubic residues, quadratic partition of prime numbers and reciprocity laws. He proved several results that eluded even Gauss. His name is associated with various concepts of mathematics, like:

- Eisenstein numbers / primes: The Eisenstein integers, sometimes also called the Eisenstein-Jacobi integers, are numbers of the form, $a + b\omega$ where a and b are integers, and $\omega = \frac{1}{2}(-1 + i\sqrt{3})$ is a primitive cube-root of 1. The Eisenstein primes are Eisenstein integers that cannot be written as products of two Eisenstein integers of absolute vale (or equivalent norm) > 1.

- Eisenstein criterion: Consider a polynomial $Q(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, of degree n, with integer coefficients. If there exists a prime number p satisfying the following three conditions (i) p divides each ai for $0 \leq i < n$, (ii) p does not divide a_n, and (iii) p^2 does not divide a_0, then Q is irreducible over the rational numbers.

- Eisenstein series: Eisenstein series are particular modular forms with infinite series expansions that may be written down directly. Originally defined for the modular group as follows: Let τ be a complex number with strictly positive imaginary part, then the holomorphic Eisenstein series $G_{2k}(\tau)$ of weight $2k$, where $k \geq 2$ is an integer, is $\sum_{(m,n)\, \epsilon\, \mathbb{Z}^2 - \{(0,0)\}} \frac{1}{(m+n\tau)^{2k}}$. This series absolutely converges to a holomorphic function of τ in the upper half-plane. Eisenstein series can be generalized in the theory of automorphic forms.

His academic advisors were renowned mathematicians like Carl Friedrich Gauss, Peter Gustav Lejeune Dirichlet, Ernst Kummer and Nikolaus Wolfgang Fischer. Bernhard Riemann was one of his notable students.

Eisenstein suffered all his life from bad health but at least he survived childhood which none of his five brothers and sisters succeeded in doing; they all died of meningitis. He showed considerable talent for music from a young age and played the piano and composed music throughout his life.

At the age of seventeen, when he was still at school, he began to attend lectures by Dirichlet and other mathematicians at the University of Berlin. He wrote in his autobiography about the reasons that he was so attracted to mathematics:

"What attracted me so strongly and exclusively to mathematics, apart from the actual content, was particularly the specific nature of the mental processes by which mathematical concepts are handled. This way of deducing and discovering new truths from old ones, and the extraordinary clarity and self-evidence of the theorems, the ingeniousness of the ideas …had an irresistible fascination for me. Beginning from the individual theorems, I grew accustomed to delve more deeply into their relationships and to grasp whole theories as a single entity. That is how I conceived the idea of mathematical beauty …"

In 1843, Eisenstein entered the Friedrich Wilhelm University and in the following year, he published several papers and 2 problems in the reputable Crelle's journal.

In February 1845, Eisenstein received an honorary Doctorate from the University of Breslau. He became a professor of mathematics at Berlin in 1847 and was elected to the Royal Prussian Academy of Sciences and Humanities.

In 1848, Eisenstein was imprisoned by the Prussian army for his revolutionary activities. After his arrest, he was ill-treated. The harsh treatment he received had a bad effect on his delicate health.

Despite his health, Eisenstein continued writing papers on quadratic partitions of prime numbers and the reciprocity laws. Gauss proposed him for election to the Göttingen Academy and he was elected in 1851. Early in 1852, at Dirichlet's request, he was elected to the Berlin Academy.

Eisenstein's work continued till his death, totaling 40 mathematical articles and more than 700 printed pages. He died of tuberculosis, on Oct. 11, 1852 in Berlin, when he was only 29 years old. Like Galois and Abel before him, Eisenstein died before the age of 30.

Sources:

1. *https://mathworld.wolfram.com/EisensteinInteger.html*
2. *https://en.wikipedia.org/wiki/Eisenstein%27s_criterion*
3. *https://en.wikipedia.org/wiki/Eisenstein_series*

3.2 Major Progress towards The Kakeya Conjecture

A new proof marks major progress toward solving the Kakeya conjecture, a simple question that formed the base of a hierarchy (a tower) of three conjectures namely, "restriction" conjecture, Bochner-Riesz conjecture and at the very top sits the local smoothing conjecture. Since each statement in the hierarchy implies the one below it, if the Kakeya conjecture is false, none of the other conjectures are true. The entire tower will come crashing down.

In 1917, the Japanese mathematician *Sōichi Kakeya* posed a following problem: What is the smallest amount of area required to continuously rotate a unit line segment (an infinitely thin needle) in the plane by a full rotation?

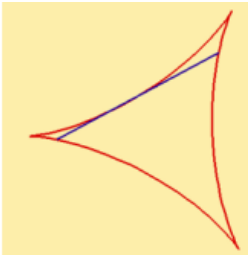


Figure 1

If you simply spin the needle around its center, you will get a circle of radius $1/2$, which has an area $\pi/4$. But it is possible to move the needle in innovative ways, so that you sweep a much smaller amount of area. It is possible to rotate the needle using a "three-point U-turn" inside a deltoid which has an area $\pi/8$. Observe that at every stage of its rotation (except when an endpoint is at a cusp of the deltoid), the needle is in contact with the deltoid at three points: two endpoints and one tangent point.

In 1928, Besicovitch proved that one could in fact rotate a needle in arbitrary small amounts of area. The proof relied on two observations. (i) First, one could translate a needle by any distance using as little area as one pleased (ii) one could find sets of arbitrarily small area that contained line segments (or thin triangles or rectangles) in every direction.

This led to the definition of a Kakeya set in $\mathbb{R}^n$ to be a set which contained a unit line segment in every direction. Besicovitch's construction showed that Kakeya sets in $\mathbb{R}^2$ could have arbitrarily small measure; in fact, one can construct Kakeya sets which have Lebesgue measure zero. While they know that such sets can be small in terms of area (or volume when needles are arranged in three or more dimensions), they believe the sets must always be large if their size is measured by the Hausdorff dimension or the Minkowski dimension.

Kakeya set conjecture: A Kakeya set in $\mathbb{R}^n$ has Hausdorff and Minkowski dimension n.

In fractal geometry, Minkowski dimension is a way of determining the fractal dimension of a set S in a Eucliden space R^n, or more generally in a metric space (X, d). It is named after the Polish mathematician Hermann Minkowski. Suppose that $N(\varepsilon)$ is the number of boxes of side length ε required to cover the set. Then the Minkowski dimension is defined as $\lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{log(1/\varepsilon)}$.

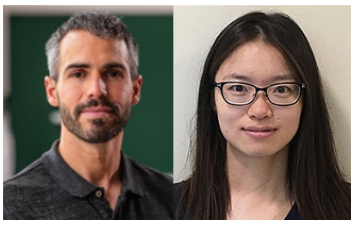
Hausdorff dimension is a measure of roughness, or more specifically, fractal dimension, that was introduced in 1918 by mathematician Felix Hausdorff. To formally define the Hausdorff dimension we first define the d-dimensional Hausdorff outer measure as follows:

Let X be a metric space. If $S \subset X$ and $d \in [0\ \infty)$, $H^d_\delta(S) = inf\{\sum_{i=1}^{\infty}(diamU_i)^d |\ U_{i=1}^{\infty}U_i \supseteq S,\ diam\ U_i < \delta\}$, where infimum is taken overall countable covers U of S. The Hausdorff d-dimensional outer measure is then defined as $H^d(S) = \lim_{\delta \to 0} H^d_\delta(S)$. And the restriction of the mapping to measurable sets justifies it as a measure, called the d-dimensional Hausdorff Measure. The Hausdorff dimension $diamH(X)$ of X is defined as $inf\ \{d \geq 0 | H^d(X) = 0\}$.

The Kakeya conjecture was solved for $n = 2$ by Davies in 1971, but remained open for $n \geq 3$. While it is apparently a simple question about needles, the geometry of these Kakeya sets uncovered surprising connections to partial differential equations, harmonic analysis, number theory and even physics.

Figure 2

In 1995, Thomas Wolff proved that the Minkowski dimension of a Kakeya set in 3D space has to be at least 2.5. That lower bound turned out to be difficult to increase. Then, in 1999, the mathematicians Nets Katz, Izabella ŁLaba and Terence Tao managed to beat it. Their new bound is 2.500000001. Despite how small the improvement was, it overcame a massive theoretical barrier. Their paper was published in the *Annals of Mathematics*. Katz and Tao later hoped to apply some of the ideas from that work to attack the 3D Kakeya conjecture in a different way. They hypothesized that any counterexample must have three particular properties, and that the coexistence of those properties must lead to a contradiction. They (along with other mathematicians) could show that any counterexample must have two of the three properties. It must be "plany", which means that whenever line segments intersect at a point, those segments also lie nearly in the same plane. It must also be "grainy", which requires that the planes of nearby points of intersection be similarly oriented. However, they couldn't prove that all counterexamples must be sticky. In a "sticky" set, line segments that point in nearly the same direction also have to be located close to each other in space which in turn force a lot of overlap among the line segments, thereby making the set as small as possible – precisely what you need to create a counterexample.



Now, mathematicians *Joshua Zahl* (left) of the University of British Columbia and *Hong Wang* (right) of New York University, have moved the needle, so to speak. Their new proof strikes down a major obstacle that has stood for decades - renewing hope that a solution might be in sight.

They started by assuming the existence of a sticky counterexample with a Minkowski dimension of less than 3. They knew from previous work that such a counterexample had to be plany and grainy. Now they needed to show that the plany, grainy and sticky properties played off each other and led to a contradiction, which would mean that this counterexample couldn't actually exist.

To get that contradiction, however, Wang and Zahl turned their attention in a direction that had not been anticipated – toward an area known as projection theory. They used a "stickiness" to prove that such a paradoxical-sounding set cannot exist - meaning that there are no sticky counterexamples to the Kakeya conjecture. Wang and Zahl's work strongly suggests that the Kakeya conjecture is true. While it only applies to the three-dimensional case, some of its techniques might be useful in higher dimensions also.

Sources:

1. *https://www.quantamagazine.org/a-tower-of-conjectures-that-rests-upon-a-needle-20230912/*
2. *https://www.quantamagazine.org/new-proof-threads-the-needle-on-a-sticky-geometry-problem-20230711/*
3. *Terence Tao: Recent progress on the Kakeya conjecture, University of California, Los Angeles*
4. *https://en.wikipedia.org/wiki/Hausdorff_dimension*
5. *https://en.wikipedia.org/wiki/Minkowski%E2%80%93Bouligand_dimension*

3.3 A Key Möbius Strip Problem, Solved After Almost 50 Years of Search

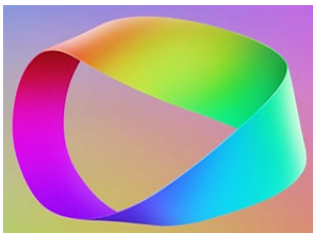


Figure - 1

A Möbius strip (or band) is both a physical and mathematical object. Möbius strip can be constructed by twisting a simple strip of paper one time and then taping the ends together. Since they were first discovered back in the mid-1800s, mathematicians have been scratching their heads trying to determine one simple constraint - how small can you make a Möbius strip without it intersecting itself? Back in the late 1970s, a pair of mathematicians, Charles Sidney Weaver and Benjamin Rigler Halpern, found that the problem could be made simpler by allowing self-intersections - that changed the problem to one that involved seeking the minimum amount of strip needed to avoid self-intersections. For nearly fifty years, mathematicians have puzzled over the misleadingly simple question.

Now, *Richard Schwartz*, a mathematician at Brown University, Providence, Rhode Island has proposed an elegant solution to this problem, which was originally posed by mathematicians *Charles Weaver* and *Benjamin Halpern* in 1977. In their paper, Halpern and Weaver pose a limit for Möbius strips based on the familiar geometry of folded bits of solid paper - that the ratio between the length and width of the paper must be greater than $\sqrt{3}$, or around 1.73.

Schwartz had several attempts at solving it over the years and published a paper in 2021 with a promising approach that ultimately fell short. When he resumed investigating the problem, he noticed a mistake in a "lemma"-an intermediate result-involving a "T- pattern" in his previous paper.

The lemma begins with one basic idea: Möbius bands, have straight lines on them, so as to form a ruled surface. You can imagine drawing these straight lines so that they cut across the Möbius band and hit the boundary at either end. In his earlier work, Schwartz identified two straight lines that are perpendicular to each other and also in the same plane, forming a T-pattern on every Möbius strip.

The next step was to set up and solve an optimization problem that entailed slicing open a Möbius band at an angle (rather than perpendicular to the boundary) along a line segment that stretched across the width of the band and considering the resulting shape. For this step, in Schwartz's 2021 paper, he incorrectly concluded that this shape was a parallelogram. It's actually a trapezoid.

Then with some help from a few colleagues - Schwartz corrected his error and found a really nice proof for the intermediate step that greatly simplified the paper.

Sources:

1. *https://www.sciencealert.com/mathematicians-solve-a-key-mbius-strip-problem-after-almost-50-years-of-searching*
2. *https://www.scientificamerican.com/article/mathematicians-solve-50-year-old-moebius-strip-puzzle1/*

3.4 Two Students Disprove A Widely Believed Local-Global Conjecture

Mathematicians thought they were on the point of proving a conjecture about the ancient structures known as Apollonian circles. But now a new work reveals it to be false.

About 2,200 years ago, the Greek geometer *Apollonius* inquired about how circles would fit together if they all touched each other.

Imagine arranging three coins so that each one touches the others. You can always draw a circle around them that touches all three from the outside. Then you can start to ask questions:

How does the size of that bigger circle relate to those of the three coins? What size circle will fit into the gap between the three coins? And if you start to draw circles that fill in progressively smaller and smaller gaps between circles - creating a fractal pattern known as a packing - how do the sizes of those circles relate to one another?

Figure 1

Rather than think about the diameter of these circles, mathematicians use a measure 'curvature' - the inverse of the radius.

Renaissance mathematicians proved that if the first four circles have a curvature that is an integer, the curvatures of all the subsequent circles in the packing are guaranteed to be whole numbers. That is remarkable on its own. But mathematicians have taken the problem a step further by asking questions about which integers show up as the circles get smaller and smaller and the curvatures get larger and larger.

In 2010, *Elena Fuchs*, a number theorist now at the University of California, observed that if you divided each curvature by 24, a rule emerged for the possible remainders. Some packings only have curvatures with remainders of 0, 1, 4, 9, 12 or 16, for example. Others only leave remainders of 3, 6, 7, 10, 15, 18, 19 or 22. There were six different possible groups. In other words, curvatures follow a particular relationship that forces them into certain subsets, modulo 24.

Soon mathematicians became convinced that not only must the curvatures fall into one of the 6 subsets, but also that every possible number in each subset must be represented. This idea came to be known as the 'local-global conjecture'.

In 2012, *Kontorovich* and *Jean Bourgain* proved that virtually every number predicted by the conjecture does occur. But "virtually all" does not mean "all". Mathematicians thought the rare counterexamples that remained possible after Kontorovich and Bourgain's paper did not actually exist, mostly because the two or three most well-studied circle packings seemed to follow the local-global conjecture so well.

Now, *Summer Haag* and *Clyde Kertzer* alongwith their number theorist supervisor *Katherine Stange* (from left to right) from the University of Colorado, Boulder, US proved the conjecture to be false.

Haag and Kertzer clustered around charts that demonstrated how a few buckets seemed to be missing certain numbers. Numbers they expected to appear never did. They proved that the pattern they observed continues indefinitely, disproving the conjecture.

The proof centers on a centuries-old principle called quadratic reciprocity which arises from certain indirect factorization patterns involving perfect square numbers. Stange's team discovered how reciprocity applies to circle packings. It explains why certain curvatures cannot be tangent to each other.

Source: *https://www.quantamagazine.org/two-students-unravel-a-widely-believed-math-conjecture-20230810/?mc_cid=d0ada5007c&mc_eid=df6d259cb7*

3.5 Unexpected Link Between Number Theory and Genetics Discovered

Number theory, the study of the properties of integers, is perhaps the purest form of mathematics. At first sight, it may seem far too abstract to apply to the natural world. And yet, again and again, number theory finds unexpected applications in science and engineering, from leaf angles that (almost) universally follow the Fibonacci sequence, to modern encryption techniques based on factoring prime numbers.

Now, an interdisciplinary team of mathematicians, engineers, physicists, and medical scientists have uncovered an unexpected link between Number Theory and Evolutionary Genetics, that exposes key insights into the structure of neutral mutations and the evolution of organisms.

Specifically, the team of researchers (from Oxford, Harvard, Cambridge, GUST, MIT, Imperial, and the Alan Turing Institute) lead by *Vaibhav Mohanty* (Harvard Medical School) have discovered a deep connection between the sums-of-digits function $S_k(n)$, (defined as the sum of the digits of a natural number n in base k) from number theory and a key quantity in genetics, the phenotype mutational robustness which is defined as the average probability that a point mutation does not change a phenotype (a characteristic of an organism).

The discovery may have important implications for evolutionary genetics. Many genetic mutations are neutral, meaning that they can slowly accumulate over time without affecting the viability of the phenotype. These neutral mutations cause genome sequences to change at a steady rate over time. Because this rate is known, scientists can compare the percentage difference in the sequence between two organisms and infer when their latest common ancestor lived.

But the existence of these neutral mutations posed an important question: What fraction of mutations to a sequence is neutral? This property, called the phenotype mutational robustness, defines the average amount of mutations that can occur across all sequences without affecting the phenotype.

It is precisely this question that the team has answered. They proved that the maximum robustness is proportional to the logarithm of the fraction of all possible sequences that map to a phenotype, with a correction which is given by the sums of digits function $S_k(n)$.

Another surprise was that the maximum robustness also turns out to be related to the famous Takagi function, a strange function that is continuous everywhere, but differentiable nowhere. This fractal function is also called the blancmange curve, because it looks like the French dessert.

What is most surprising is that the team found clear evidence in the mapping from sequences to RNA secondary structures that nature in some cases achieves the exact maximum robustness bound. It is as if biology knows about the fractal sums-of-digits function. This study will help to find many fascinating new links between number theory and genetics.

Source: *https://www.news-medical.net/news/20230808/Unexpected-link-between-pure-mathematics-and- genetics-discovered.aspx*

3.6 New Estimate of The Size of Triangles Created by Packing Points into A Square

A new proof breaks a decades-long drought of progress on the problem of estimating the size of triangles created by packing points into a square.

Consider a square with a bunch of points inside. Take three of those points, and you can make a triangle. Four points define four different triangles. Ten points define 120 triangles. The numbers grow quickly from there - 100 points define 1,61,700 different triangles. Each of those triangles, of course, has a particular area.

Hans Heilbronn, a German mathematician thought of these triangles in the late 1940s when he saw a group of soldiers outside his window. The soldiers did not appear to be in formation, which got him thinking: If there are n soldiers inside a square, how large can the smallest triangle be, for a chosen arrangement? Heilbronn wondered how one might go about arranging the soldiers (or, for mathematical simplicity, points) to maximize the size of the smallest triangle. By placing three points very close together, you can easily make the smallest triangle in an arrangement arbitrarily small. But trying to keep the smallest triangle big, is trickier. As you keep adding in more dots, the smallest triangle is forced to be pretty small - new dots can only be so far from existing ones. It is relatively easy to show that the smallest triangle can't have an area any bigger than $1/(n-2)$ by splitting the square into nonoverlapping triangles, n being the number of points.

But Heilbronn thought that the limit was even tinier than that. He guessed that no matter how the dots were arranged in the square, there could not be a smallest triangle with an area larger than around $1/n^2$, a number which shrinks much faster as n grows.

In 1980, the Hungarian mathematicians János Komlós, János Pintz and Endre Szemerédi found a pattern of dots whose smallest triangle had an area ever so slightly larger than $1/n^2$, proving

Heilbronn wrong. In a separate paper published around the same time, they also showed that it is impossible to arrange n dots to create a smallest triangle that is bigger than around $1/n^{8/7}$. When n is large, this is much smaller than $1/n$, but much bigger than $1/n^2$.

The problem is simple to state, but progress on the Heilbronn triangle problem, as it came to be called, has been halting, and results dried up entirely in the 1980s. Researchers kept working on the Heilbronn triangle problem over the years, despite the long wait for progress, motivated by its several links with other areas of mathematics. It is closely related to problems about intersecting shapes, which in turn connect to both number theory and Fourier analysis.

Now mathematicians *Alex Cohen*, *Cosmin Pohoata* and *Dmitrii Zakharov* (from left to right) of Massachusetts Institute of Technology, Cambridge announced a new cap on the size of the smallest triangle. This trio has shown that for sufficiently large n, in every configuration of n points chosen inside the unit square there exists a triangle of area less than $1/n^{8/7+1/2000}$. The new result has revived the long-languishing Heilbronn triangle problem.

Source: *https://www.quantamagazine.org/the-biggest-smallest-triangle-just-got-smaller-20230908/*

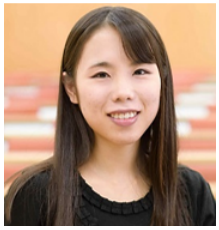
3.7 Awards

3.7.1 Maryam Mirzakhani New Frontiers Prize 2024 to be Awarded to Three Women Mathematicians

Maryam Mirzakhani New Frontiers Prize is awarded to three women mathematicians for early- career achievements each receiving $50,000 prize. The prize is presented to women mathematicians who have completed their Ph.D. within the past two years.

The first recipient is *Dr. Hannah Larson*, an assistant professor of mathematics at University of California, Berkeley, US, who will be awarded the 2024 Maryam Mirzakhani New Frontiers Prize for advances in Brill-Noether theory and the geometry of the moduli space of curves. Larson joined UC Berkeley in 2023 and is also a Clay Research Fellow for 2022-2027.

The second recipient is *Dr. Laura Monk*, research associate of University of Bristol, UK, who will be awarded the 2024 Maryam Mirzakhani New Frontiers Prize for advancing our understanding of random hyperbolic surfaces of large genus. Monk's research is in the field of spectral geometry, an area of mathematics studying the relationship between the vibrational modes of surfaces and their geometry. The novelty of her approach lies in bringing new probabilistic methods in a long-established field of pure mathematics.



Japanese mathematician and mathematical physicist *Dr. Mayuko Yamashita*, associate professor, Department of Mathematics, Kyoto University, Japan, is the third recipient of this prize. She will be awarded 2024 Maryam Mirzakhani New Frontiers Prize for contributions to mathematical physics and index theory. She is working on algebraic topology and differential cohomology and their relation with quantum field theories. She represented Japan in the 2013 International Mathematical Olympiad, earning a silver medal.

All the three laureates will be honored at gala award ceremony in Los Angeles on April 13, 2024.

Sources:

1. *https://breakthroughprize.org/News/83*
2. *https://finance.yahoo.com/news/breakthrough-prize-announces-2024-laureates-130600650.html*

3.7.2 2024 New Horizons in Mathematics Prize to be Awarded to Three Mathematicians

This year, $1,00,000 New Horizons in Mathematics Prize is awarded to three mathematicians.

The first recipient is *Dr. Roland Bauerschmidt* from New York University, US, for his work in probability theory and the renormalization group - a concept that emerged from the quantum field theories studied by this year's Breakthrough Prize in Fundamental Physics winners, and has become an important object of study in mathematics. Earlier, he was professor of probability at the University of Cambridge, and post-doc at Harvard University, Cambridge, and the Institute for Advanced Study, US.

The second recipient is German applied mathematician *Dr. Angkana Rüland*, a professor in mathematics of University of Bonn, Germany, honored for work also touching on ideas derived from physics, such as transitions between states of matter, which are now studied in mathematical fields including analysis. Her research has included work on the mathematical modeling of shape-memory alloys and on the inverse problems arising in animal echolocation. She is holder of a Hausdorff Chair in mathematics at the Hausdorff Center for Mathematics of the University of Bonn.

Dr. Michael Groechenig, associate professor of University of Toronto, Ontario, Canada receives the Prize for his insights into arithmetic geometry. He is awarded this prize for contributions to the theory of rigid local systems and applications of p-adic integration to mirror symmetry and the fundamental lemma. His research interest includes problems in arithmetic geometry related to Higgs bundles, p-adic or motivic integration, algebraic K-theory. Most of his research is devoted to finding arithmetic approaches to problems in geometry. Homotopy theory also plays an important role in his work. He is recipient of Alfred P. Sloan fellowship ($75,000) for 2022-2024 and also Marie Sklodowska-Curie individual fellowship (EU grant amounting to a total of 1,59,460 Euro) for 2016-2018.

Source: *https://finance.yahoo.com/news/breakthrough-prize-announces-2024-laureates-130600650.html*

3.7.3 Breakthrought Prize 2024 Will be Awarded to Prof. Simon Brendle

The Breakthrough Prize Foundation announced the winners of the 2024 Breakthrough Prizes, honoring an esteemed group of the world's most brilliant minds for impactful scientific discoveries. These prizes recognize "the world's top scientists" in the fields of life sciences, fundamental physics and mathematics. The annual Breakthrough Prize – popularly known as the "Oscars of Science" - was created in 2012 to celebrate the wonders of our scientific age, by founding sponsors *Sergey Brin*, *Priscilla Chan* and *Mark Zuckerberg*, *Julia* and *Yuri Milner*, and *Anne Wojcicki*. It comes with a $3 million award.

Berkeley researcher Simon Brendle, professor of mathematics, has been awarded the Breakthrough Prize 2024. He is recognized for "a series of remarkable leaps in differential geometry, a field that uses the tools of calculus to study curves, surfaces and spaces. Many of his results concern the shape of surfaces, as well as manifolds in higher dimensions than those we experience in everyday life".

He is awarded the prize for transformative contributions to differential geometry, including sharp geometric inequalities, many results on Ricci flow and mean curvature flow and the Lawson conjecture on minimal tori in the 3-sphere. His ongoing research on differential geometry and nonlinear partial differential equations is of vital importance for the field.

The Prize will be awarded on April 13, 2024 at the 10^{th} annual Breakthrough Prize ceremony, to be held in Los Angeles. The Breakthrough Prize ceremony is the only one of its kind that places scientists on center stage, and is attended by celebrities in film, sports, comedy, and music, to lend their spotlight to shine on scientists.
Source: *https://breakthroughprize.org/News/83*

3.8 Obituary

3.8.1 Pioneer of Computational Mechanics Prof. J. Tinsley Oden Passes away at the age of 86

Prof. J. Tinsley Oden, world-renowned pioneer in computational mechanics died on Aug. 27, 2023 at the age of 86.

His revolutionary treatise, "Finite Elements of Nonlinear Continua", first published in 1972, has not only demonstrated the great potential of computational methods but established computational mechanics as a new intellectually rich discipline built upon concepts in mathematics, computer sciences, physics and mechanics. Computational mechanics has since become a fundamentally important discipline, affecting engineering practice and education worldwide, and laying the foundations for the flourishing field of computational science and engineering.

Oden was born on Dec. 25, 1936 at Alexandria, Louisiana, US. He received a B.S. degree in civil engineering in 1959 and a Ph.D. in engineering mechanics from Oklahoma State University (OSU), US in 1962. He taught at OSU and the University of Alabama in Huntsville, US where he was the head of the Department of Engineering Mechanics prior to going to Texas in 1973. He has held visiting professor positions at other universities in the United States, England, and Brazil.

Oden was an Honorary Member of the American Society of Mechanical Engineers and was a Fellow of six international scientific/technical societies: IACM, AAM, ASME, ASCE, SES, and BMIA. He was a Fellow, founding member, and first President of the US Association for Computational Mechanics and the International Association for Computational Mechanics. He was a Fellow and past President of both the American Academy of Mechanics and the Society of Engineering Science.

Oden was awarded the A. Cemal Eringen Medal in 1989, the Worcester Reed Warner Medal, the Lohmann Medal, the Theodore von Karman Medal, the John von Neumann medal, the Newton/Gauss Congress Medal, and the Stephan P. Timoshenko Medal. He was also knighted as "Chevalier des Palmes Academiques" by the French government and he held four honorary doctorates, honoris causa, from universities in Portugal, Belgium, Poland and the United States. Oden was also elected a member of the US National Academy of Engineering in 1988.

A prolific writer and researcher, Oden was author or editor of more than 800 scientific works including 57 books. He educated and advised more than 45 doctoral students and dozens of postdoctoral researchers. He published extensively in this field and in related areas over the last three decades.

Known for his legendary work ethic, Oden often came to the office on Sundays. He continued to come to the institute daily even just a few weeks prior to his death.
Sources:

1. *https://en.wikipedia.org/wiki/J._Tinsley_Oden*
2. *https://news.utexas.edu/2023/08/30/ut-mourns-pioneer-of-computational-mechanics- and -founder-of-oden-institute/*

3.8.2 Renowned Topologist Prof. Melvin Rothenberg Passes away at the age of 89

Melvin Gordon Rothenberg, a professor emeritus of mathematics who spent more than four decades making ground breaking mathematical discoveries at the University of Chicago, US, died on Aug. 1, 2023 at the age of 89. He made multiple contributions in the mathematical fields of algebraic and geometric topology that form the foundation of the work still ongoing today.

Born in Boston in 1934, Rothenberg was raised in Cleveland. He studied philosophy and mathematics at the University of Michigan, US, and was recruited by the University of California, Berkeley, US for his master's and doctorate degrees. There he studied under *P. Emery Thomas*, wrote his dissertation "On the Milnor Construction of Universal Bundles", and received Ph.D. in 1962. He joined the University of Chicago, US, as a mathematics instructor the year prior.

Rothenberg's early work focused on algebraic topology, related to the unstable version of the J-homomorphism. With Norman Steenrod, he found a spectral sequence for the cohomology of the classifying space of an H-space.

He then moved to work in geometric topology, where his contributions were fundamental and groundbreaking. His two largest collaborations were with *Dick Lashof* and with *Ib Madsen*. With Lashof, he made important early contributions to smoothing and triangulation theory and equivariant triangulation theory. His work with Madsen showed that it was possible to understand odd order group actions by surgery theoretic means.

Rothenberg is fondly remembered as the model of an absent-minded professor. He was also committed to social justice and political activism.

Source: *https://news.uchicago.edu/story/prof-emeritus-melvin-rothenberg-uchicago-mathematician -and- activist-1934-2023*

3.8.3 Renowned Game Theorist Prof. T. Parthasarathy Passes away at the age of 84

Distinguished Indian mathematician and renowned game theorist *Thiruvenkatachari Parthasarathy* died on Sept. 22, 2023 at the age of 83.

T. Parthasarathy was a co-author of a book on game theory with *T. E. S. Raghavan*, and of two research monographs, one on optimization and one on univalence theory, published by Springer-Verlag. He was a former president of the Indian Mathematical Society.

He was born on Feb. 29, 1940, and received his B.Sc. and M.Sc. degrees from University of Madras. He worked on the topic "Minimax Theorems and Product solutions for simple games" under the guidance of *C. R. Rao* and received Ph.D. in 1967 from the Indian Statistical Institute, Kolkata. After several years in academic positions at various institutions in USA he joined the Indian Statistical Institute (ISI), Delhi, as Professor in 1979, where he served until retirement in the normal course. Subsequently he was also affiliated with the Chennai Mathematical Institute (CMI) for several years.

Prof. Parthasarathy received Shanti Swaroup Bhatnagar Award for Mathematical Sciences in 1986. He was elected as Fellow of the Indian Academy of Sciences in 1988 and Indian National Science Academy in 1995.

Source: *https://en.wikipedia.org/wiki/Thiruvenkatachari_Parthasarathy*

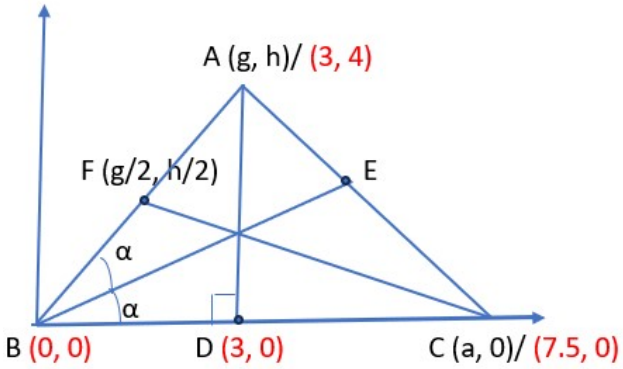
□□□

4. Problem Corner

Udayan Prajapati
Mathematics Department, St. Xavier's College, Ahmedabad
Email: udayan.prajapati@gmail.com

In the August 2023 issue of TMC Bulletin, we posed a problem from Geometry for our readers. We have received four solutions to that problem from Prof. J N Salunke, Latur, Maharashtra, a student Mr. Pranjal Jha, Kota, a student Ms. Saee Patil, Pune and Dr. M. R. Modak, Pune.
Also, in this issue we pose a problem from Geometry for our readers. Readers are invited to email their solutions to Dr. Udayan Prajapati (Udayan.prajapati@gmail.com), Coordinator, Problem Corner, before 15^{th} December, 2023. Most innovative solution will be published in the subsequent issue of the bulletin.

Problem posed in the previous issue: Does there exist a scalene acute angled triangle ABC with an altitude AD, an angle bisector BE and a median CF such that AD, BE and CF are concurrent?
(We present first two solutions in this Issue and other two solutions in the next Issue.)



Solution by Prof. J. N. Salunke: Consider an acute-angled scalene triangle ABC with an altitude AD, an angle bisector BE and a median CF. Take B as origin, point C on the positive X-axis and the vertex A in the first quadrant. Let the coordinates of C be $(a,0)$ and A as (g,h) with $a = BC$. So g and h are positive real numbers. Let angle $B/2 = \alpha$.
The equation of the line containing the angle bisector BE is $y = x\ tan\ \alpha$. —-(1)
The equation of the altitude AD is $x = g$. —-(2)
The equation of the line containing the median CF is $hx + (2a-g)y = ha$. —-(3)
The slope of BA is, $\frac{h}{g} = \tan 2\alpha = \frac{2\tan\alpha}{1-\tan^2\alpha}$. Hence $\left(\tan\alpha + \frac{g}{h}\right)^2 = \frac{g^2+h^2}{h^2}$. Taking non-negative square root we get $\tan\alpha + \frac{g}{h} = \frac{\sqrt{(g^2+h^2)}}{h}$. That is $\tan\alpha = \frac{\sqrt{(g^2+h^2)}-g}{h}$.
Now, the lines mentioned in (1), (2) and (3) are concurrent if and only if

$$\begin{vmatrix} \tan\alpha & -1 & 0 \\ 1 & 0 & -g \\ h & 2a-g & -ha \end{vmatrix} = 0,$$

i. e. $g(2a-g)\tan\alpha + h(g-a) = 0$

i. e. $g(2a-g)\left(\frac{\sqrt{(g^2+h^2)}-g}{h}\right) = h(a-g)$,

i. e. $g(2a-g)\sqrt{(g^2+h^2)} = g^2(2a-g) + h^2(a-g)$.

i. e. $g^2(2a-g)^2(g^2+h^2) = g^4(2a-g)^2 + h^4(a-g)^2 + 2h^2g^2(2a-g)(a-g)$.

i. e. $(2a-g)^2 = h^2(a-g)^2 + 2g^2(2a-g)(a-g)$.

i. e. $g^2(4a^2-4ag+g^2) = h^2(a^2+g^2-2ag) + 2g^2(2a^2-3ag+g^2)$.

i. e. $g^2[-g^2+2ag] = h^2(a^2+g^2-2ag)$.

i. e. $a^2h^2 - 2g(h^2+g^2)a + g^2(h^2+g^2) = 0$.

It is quadractic equation in a. Its larger root is,

$$\frac{2g(h^2+g^2) + \sqrt{4g^2(h^2+g^2)^2 - 4h^2g^2(h^2+g^2)}}{2h^2} = \frac{g(h^2+g^2) + g^2\sqrt{h^2+g^2}}{h^2}.$$

Taking this as a, the concurrency condition is satisfied.

For example, if $g = 1$, $h = 2$ then $a = \frac{5+\sqrt{5}}{4} = BC$ and $AB = \sqrt{5}$, $CA = \sqrt{4 + \frac{(1+\sqrt{5})^2}{16}}$. One can verify in this case that $AB < BC < CA$, triangle ABC is scalene and acute angled, the altitude AD, the angle bisector BE and the median CF are concurrent.

Editor's Note: It can be proved using the above calculations that there are non-similar infinitely many triangles satisfying the required criteria.

Since the larger root mentioned above satisfies $a - g = \frac{[g^3+g^2\sqrt{g^2+h^2}]}{h^2}$, we see that $a > g$. Now, for a fixed g, the expression on the right tends to 0 as h tends to infinity, so for large h, it will be less than 1. Take also $h > g + 1$.

Now the side BC is $a < g + 1$. Since the height of the triangle is $h > g + 1$,the sides AB and AC are $> g + 1 > BC$. Hence angles C and B are bigger than angle A. As angles B and C are already acute angles, so is angle A. Thus, when h is $(> g + 1)$ sufficiently large, all angles are acute angles. Already $BC < g + 1$ and both AB and $AC > g + 1$. To make sure $AB \neq AC$, take $g > 1$. As AD is the altitude, we have $BD = g > 1$, and as $BC < g + 1$, we must have $DC < 1$. Thus $BD \neq DC$, so $AB \neq AC$. Thus, all the sides of triangle ABC are distinct and all angles are acute angles. Also, by the choice of a, the concurrency condition is satisfied. This gives infinitely many non-similar triangles of the required type by fixing $g > 1$ and taking h sufficiently large.

Solution by Mr. Pranjal Jha: We show that such a triangle does exist.

Construct a triangle ABC (as in above figure with co-ordinates of points in red) with $\cos B = 3/5$ and $\tan C = 8/9$. Since $\cos B$ and $\tan C$ are greater than $0, B$ and C, are acute angles. Also, $\tan A = -\tan(B + C) = (\tan B + \tan C)/(\tan B \tan C - 1) > 0$, so A is also an acute angle. Hence, ABC is an acute angled triangle and angles A, B, C are different as values of their tangents are different.

Now by converse of Ceva's theorem, we are required to show that: $\frac{BD}{DC} \times \frac{CE}{EA} \times \frac{AF}{FB} = 1$. (1)

As CF is median, $\frac{AF}{FB} = 1$. By the angle bisector theorem, $\frac{CE}{EA} = \frac{BC}{AB}$.

Finally, $\frac{BD}{DC} = \frac{AB\cos B}{AC\cos C}$. Hence by (1), we are required to show that $\frac{AB\cos B}{AC\cos C} \times \frac{BC}{AB} = 1$, that is $BC\cos B = AC\cos C$. But by sine rule, $BC/\sin A = AC/\sin B$, so it is enough to show that $\sin A\cos B = \sin B\cos C$. Also, $\sin A = \sin(B + C)$ and so we have to prove that $(\sin B\cos C + \cos B\sin C)\cos B = \sin B\cos C$. i. e. to show that (dividing by $\cos C$)

$$\sin B\cos B + \cos^2 B\tan C = \sin B. \tag{2}$$

Now, $\cos B = 3/5$, so $\sin B = 4/5$. And $\tan C = 8/9$,

So, LHS of (2) $= (4/5)(3/5) + (9/25)(8/9) = 4/5 =$ RHS of (2). Hence the required triangle exists.

Solution by Dr. M. R. Modak: With usual notation, by Ceva's theorem and its converse, lines AD, BE, CF are concurrent if and only if

$$\frac{BD}{DC}\cdot\frac{CE}{EA}\cdot\frac{AF}{FB} \quad \text{or} \quad \frac{c\cos B}{b\cos C}\cdot\frac{a}{c}\cdot 1 = 1,$$

$$\text{or} \quad a\cos B = b\cos C \quad \text{or} \quad a\cdot\frac{c^2+a^2-b^2}{2ca} = b\cdot\frac{a^2+b^2-c^2}{2ab},$$

$$\text{or} \quad a^3 + a(c^2-b^2) + c^3 - c(a^2+b^2) = 0 \quad \text{or} \quad \text{with} \quad a/c = x, b/c = y,$$

$$x^3 + x(1-y^2) + 1 - (x^2+y^2) = 0,$$

$$\text{or} \quad y^2 = \frac{x^2(x-1)}{x+1} + 1. \tag{1}$$

Here $x, y > 0$. So by (1), $x = 1 \Leftrightarrow y = 1 \Leftrightarrow \triangle ABC$ is equilateral. Also, $y < x \Leftrightarrow y^2 < x^2 \Leftrightarrow x^3 - x^2 + x + 1 < x^3 + x^2 \Leftrightarrow 2x^2 - x - 1 > 0 \Leftrightarrow x > 1$ or $x < -1/2$.

So, (i) $x > 1 \Rightarrow x > y > 1 \Rightarrow a > b > c$. Also, in this case, $y+1 > x \Leftrightarrow y > x-1 \Leftrightarrow y^2 > (x-1)^2 \Leftrightarrow x^3 - x^2 + x + 1 > (x^2-1)(x-1) \Leftrightarrow x^3 - x^2 + x + 1 > x^3 - x^2 - x + 1 \Leftrightarrow 2x > 0$, which is true. Hence $x, y, 1$ are sides of a triangle.

Similarly, by the above, (ii) $x < 1 \Rightarrow x < y < 1 \Rightarrow a < b < c$. Also, $x + y > 1 \Leftrightarrow y > 1 - x \Leftrightarrow y^2 > (1-x)^2 \Leftrightarrow x^3 - x^2 + x + 1 > (1-x^2)(1-x) \Leftrightarrow 2x > 0$. Hence $x, y, 1$ are again sides of a triangle. Hence the solutions (x, y) of (1) with $x > 0$, $x \neq 1$, yield the family of all the scalene triangles with the given property.

For example, for $x = 3$, (1) gives $y^2 = 11/2$. So, taking $c = \sqrt{2}$, we get $a = xc = 3\sqrt{2}$, $b = \sqrt{11}$. Here $x > y > 1$ and $\cos A < 0$ and so the triangle is obtuse-angled.

Next, for $a = 4$, $c = 5$, we get $x = 0.8$ and (1) gives $y^2 = 209/225$ so that $b = \sqrt{209}/3$. Here $x < y < 1$ and the triangle is acute-angled.

For variations of this problem refer to Problem 790, Mathematics Magazine, Jan. 1972, pages 40-51.

Editorial Note: A criterion for a required *acute angled* scalene triangle can be obtained as follows:

(i) If $x > 1$ the by above, $a > b > c$. So, A is the largest angle. So ABC is an acute angles triangle if and only if A is an acute angle, if and only if $\cos a > 0$, i. e. $b^2 + c^2 > a^2$, i. e. $y^2 + 1 > x^2$. If $y^2 + 1 = x^2$, ABC is a right-angled triangle. If $y^2 + 1 < x^2$, ABC is an obtuse angled triangle. Now, $y^2 = x^2(x-1)/(x+1) + 1$. For an acute angled triangle we require $y^2 + 1 > x^2$, i. e. $x^2(x-1)/(x+1) + 1 + 1 > x^2$, i. e. $x^2(x-1)/(x+1) > x^2 - 2$, i. e. $x^2(x-1) > (x^2-2)(x+1)$ i. e. $x^3 - x^2 > x^3 + x^2 - 2x - 2$, i. e. $2x^2 - 2x - 2 < 0$, i. e. $x^2 - x - 1 < 0$, i. e. $x \in \left(\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2}\right)$. As already $x > 1$, we get the condition that $x \in \left(1, \frac{1+\sqrt{5}}{2}\right)$.

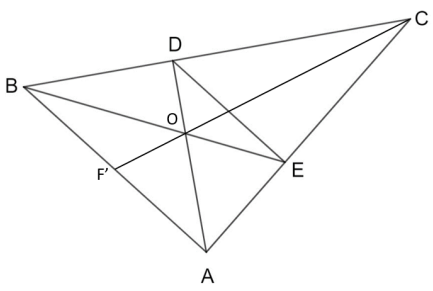
(ii) If $x < 1$, then by above, $a < b < c$. So, C is the largest angle. So ABC is an acute-angled triangle if and only if C is an acute angle if and only if $\cos C > 0$, i. e. $a^2 + b^2 > c^2$, i. e. $x^2 + y^2 > 1$. Again, if $x^2 + y^2 = 1$, then ABC is a right-angled triangle. If $x^2 + y^2 < 1$, ABC is an obtuse angled triangle.

Now, using, $y^2 = x^2(x-1)/(x+1) + 1$, $x^2 + y^2 > 1$ becomes, $x^2 + x^2(x-1)/(x+1) + 1 > 1$, i. e. $x^2 + x^2(x-1)/(x+1) > 0$. Cancelling $x^2, 1 + (x-1)/(x+1) > 0$, i. e. $x + 1 + x - 1 > 0$, i. e. $2x > 0$, i. e. $x > 0$, which is true. Thus if $x < 1$, the condition is always satisfied and the triangle is acute-angled.

Thus, ABC is also acute angled if and only if either $0 < x < 1$ or $1 < x < (1+\sqrt{5})/2$, and y satisfies $y^2 = x^2(x-1)/(x+1) + 1$. If $x = 1$, the triangle is equilateral.

Solution by Saee Patil: We show that the answer is yes and give a way to construct such triangles.

- Start with an arbitrary isosceles triangle $\triangle BDE$ with $BD = DE$ and such that $45^\circ > \angle DBE > \frac{1}{2}\cos^{-1}\left((\sqrt{5}-1)/2\right)$ and $\angle DBE \neq 30^\circ$.
- Denote by l the line parallel to DE passing through B and let m be the line perpendicular to BD at D.
- A is the intersection of l and m, and C is the intersection of lines BD and AE.
- The desired triangle is $\triangle ABC$ as shown in the figure below.



To prove that gives a required triangle, note that AD is the altitude from A and BE bisects $\angle ABC$. Let AD, BE intersect at O and let CO meet AB in F'.

Then by Ceva's theorem, $1 = BD/DC \cdot CE/EA \cdot AF'/F'B = c\cos B/b\cos C \cdot a/c \cdot AF'/F'B = a\cos B/b\cos C \cdot AF'/F'B$. So that to prove that $AF' = F'B$, it is enough to prove that $a\cos B = b\cos C$. Now, $EC = ab/(a+c)$ and since triangles ABC and EDC are similar, $a/b\cos C = b/EC$, so that $b\cos C = a^2/(a+c)$. Hence, $a\cos B = (a/c)c\cos B = (a/c)(a - b\cos C) = (a/c)(a - a^2/(a+c)) = a^2/(a+c) = b\cos C$. Thus CF' is the median from C.

Also, by the above, $\cos B = a/(a+c) = x/(x+1)$, taking $x = a/c$, as in earlier solution.

Now by data, $90^\circ > \angle B = 2\angle DBE > cos^{-1}\left((\sqrt{5}-1)/2\right)$. Hence, angle B is acute and $x/(x+1) = \cos B < (\sqrt{5}-1)/2$, so that $x < (\sqrt{5}-1)/(3-\sqrt{5}) = (\sqrt{5}+1)/2$. Since, $\angle B \neq 60^\circ, x \neq 1$ and hence by editorial note above, $\triangle ABC$ is the required triangle.

Problems for this issue

Proposed by Dr. Vinaykumar Acharya

1. Determine all the triangles with integer sides having semi-perimeter same as its area.
2. 5, 5, 6 and 5, 5, 8 are isosceles triangles with integer sides and with equal integer area. Determine all such pairs of isosceles triangles with integer sides having equal integer area.

□□□

5. International Calendar of Mathematics Events

Ramesh Kasilingam
Department of Mathematics, IITM, Chennai
Email: rameshk@iitm.ac.in

November 2023

- November 6-8, 2023, The First Sharjah International Conference on Mathematical Sciences, Department of Mathematics, College of Science. Sharjah. United Arab Emirates. *www.sharjah.ac.ae/en/Media/Conferences/SICMS23/Pages/default.aspx*
- November 3-5, 2023, 2023 Field of Dreams Conference, Atlanta, GA. *www.mathalliance.org/field-of-dreams-conference/index.html*
- November 6-9,2023, Algorithms, Approximation, and Learning in Market And Mechanism Design, SLMath (Formerly MSRI), 17, Gauss Way, Berkeley, CA 94720. *www.slmath.org/workshops/1082*
- November 10-12, 2023, The Fall 2023 edition of the Texas Geometry and Topology conference, Rice University, Texas, USA. *https://sites.google.com/view/tgtc-fall-2023/home*
- November 11-12, 2023, BUGCAT Conference 2023, SUNY Binghamton. *seminars.math.binghamton.edu/BUGCAT/index.html*

- November 17-19, 2023, Workshop on Geometric Representation Theory and Moduli Spaces, University of North Carolina, Chapel Hill, NC. *tarheels.live/grtm/*
- November 27-29, 2023, Computational Algebra, Magma University of Sydney, Australia. *www.maths.usyd.edu.au/u/comalg*
- November 27 - December 1, 2023, Workshop IV: Topology, Quantum Error Correction and Quantum Gravity, Institute for Pure and Applied Mathematics (IPAM), Los Angeles, CA. *www.ipam.ucla.edu/programs/workshops/workshop-iv-topology-quantum-error- correction - and-quantum-gravity/*

December 2023

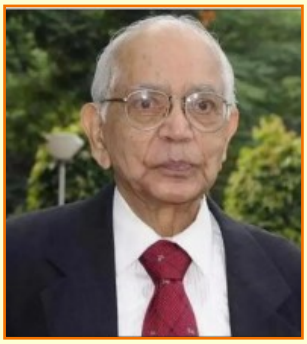
- December 4-7, 2023, Mathematical Relativity: Past, Present, Future, Erwin Schroedinger Institute, Vienna. *www.esi.ac.at/events/e526/*
- December 4-8, 2023, Hot Topics Workshop: Recent Progress in Deterministic and Stochastic Fluid-Structure Interaction, SLMath (Formerly MSRI), 17, Gauss Way, Berkeley, CA 94720. *www.slmath.org/workshops/1048*
- December 8-10, 2023, Tech Topology Conference, Georgia Institute of Technology, Atlanta, Georgia. *etnyre.math.gatech.edu/TechTopology/2023*
- December 22-23, 2023, National Conference of Mathematics and its Application in Science (NCMAS-2022), Department of Mathematics, School of Science, Uttarakhand Open University, Haldwani, Uttarakhand, India, 263139. *uou.ac.in/ncmas-2022*
- December 22-24, 2023, 38th Annual Conference of Ramanujan Mathematical Society, IIT Guwahati. *https://event.iitg.ac.in/rms2023/registration.php*

January 2024

- January 3-6, 2024, 2024 Joint Mathematics Meetings, Moscone North/South and The San Francisco Marriott Marquis, San Francisco, CA. *www.jointmathematicsmeetings.org//jmm*
- January 4-7, 2024, International Conference on Mathematics and Computing - Icmc 2024 Kalasalingam Academy of Research and Education, Krishnankoil-626126, Tamilnadu, India. *icmc2024.kalasalingam.ac.in/*
- January 7-10, 2024, ACM-SIAM Symposium on Discrete Algorithms (SODA24), Westin Alexandria Old Town Alexandria, Virginia, USA. *www.siam.org/conferences/cm/conference/soda24*
- January 8-10, 2024, International Symposium on Artificial Intelligence and Mathematics (ISAIM2024), Fort Lauderdale, FL. *isaim2022.cs.ou.edu/*
- January 8-10, 2024, International Workshop on Combinatorial Image Analysis (IWCIA'24), Fort Lauderdale, FL – Collocated With ISAIM. *isaim2024.cs.ou.edu/iwcia.html*
- January 18-19, 2024, Connections Workshop: Commutative Algebra, SLMath 17 Gauss Way, Berkeley, CA 94720. *www.msri.org/workshops/1052*
- January19-21, 2024, International Conference on History of Mathematics (ICHM2023-24), Under the patronage of Indian Society for History of Mathematics, IIT Guwahati. *https://indianshm.org/index.php/conferences/401-ishm-conf-jan-2024*
- January 29 - Febroary 2, 2024, AIM Workshop: Analytic, Arithmetic, and Geometric Aspects of Automorphic Forms, American Institute Of Mathematics, Pasadena, California. *aimath.org/workshops/upcoming/aagaautomorphic/*

□□□

TRIBUTES TO LEGENDARY STATISTICIAN PROF. C. R. RAO



On a day India marked its presence on the Moon, it lost one of its brightest mathematical stars. Prof. Calyampudi Radhakrishna Rao, one of India's greatest mathematicians and statisticians, died in US on Aug. 22, 2023, about two weeks before his 103rd birthday.

He was a former director of Indian Statistical Institute (ISI) and had hit the headlines earlier this year after he was awarded the International Prize in Statistics, which many consider equivalent to the Nobel Prize.

Having taught and researched at the Indian Statistical Institute (ISI), Kolkata, Dr. Rao pioneered several fundamental statistical concepts such as the Cramer-Rao inequality and Rao-Blackwellization, concepts that now appear in undergraduate textbooks on Statistics and Econometrics.

C. R. Rao was born on Sept. 10, 1920, in Hadagali, Bellary district, in a Telugu family. He joined the ISI as a student when it was set up by Dr. P. C. Mahanobilis, when statistics as a subject was still in its early years and yet to be taught as a distinct subject at the post-graduate level.

He was sent by Mahalanobis to Cambridge University to study use of certain statistical techniques for anthropological analysis, and from where he earned a doctorate under the supervision of Ronald Fischer, who is among the pioneers of the field. On returning to the ISI in India, where Rao spent the next 40 years of his career, he established several UG and graduate programmes in statistics and was instrumental in establishing several bureaus of the Indian Statistical Institute in various states.

Dr. Rao was also a member of several government committees for the development of national statistical systems, statistical education and research in India. He served as chairman of the committee on Statistics (1962-69), chairman of the Demographic and Communication for Population Control (1968-69), chairman of the committee on mathematics, Atomic Energy Commission (1969-78), member of the committee on Science and Technology (1969-71).

He was a pillar of the ISI, inspiring generations of students and researchers who themselves became iconic figures like S. R. Srinivasa Varadhan and K. R. Parthasarathy. He had won every conceivable award and honour.

After his retirement, Rao moved to the United States and worked at several universities. Former US President George Bush conferred on him the National Medal of Science. He was awarded India's civilian honours the Padma Bhushan and Padma Vibhushan in 1969 and 2001, respectively.

On behalf of TMC, we pay our tributes to this great son of our soil.

Source: https://www.buffalo.edu/ubnow/stories/2023/08/obit-cr-rao.html

Shiing-Shen Chern (28 Oct. 1911 - 03 Dec. 2004)

A Chinese-American mathematician and poet. Made fundamental contributions to differential geometry & topology. Called the "father of modern differential geometry". Won many awards including the Wolf Prize and the Shaw Prize. The IMU established the Chern Medal in 2010. At UC Berkeley, he Co-founded the Mathematical Sciences Research Institute in 1982.

Akshay Venkatesh, FRS (born 21 Nov. 1981)

Australian-Indian mathematician. Worked in the fields of automorphic forms, representation theory, ergodic theory, and algebraic topology. Won medals at both the International Physics and Mathematical Olympiad at the age of 12. Was awarded the Fields Medal for his synthesis of analytic number theory, homogeneous dynamics, topology, and representation theory.

Jacob Alexander (born 07 Dec. 1977)

An American mathematician, a 2014 MacArthur Fellow. Won a gold medal with a perfect score in IMO-1994. Known for his work on infinity categories and derived algebraic geometry. Also worked on elliptic cohomology and topological field theories. The winners of the Breakthrough Prize in Mathematics and a MacArthur "Genius Grant" Fellowship both in 2014.

Publisher

The Mathematics Consortium (India),
(Reg. no. MAHA/562/2016 /Pune dated 01/04/2016),
43/16 Gunadhar Bungalow, Erandawane, Pune 411004, India.
Email: tmathconsort@gmail.com Website: themathconsortium.in

Contact Persons

Prof. Vijay Pathak Prof. S. A. Katre
vdpmsu@gmail.com (9426324267) sakatre@gmail.com (9890001215)

Printers

AUM Copy Point, G-26, Saffron Complex, Nr. Maharana Pratap Chowk,
Fatehgunj, Vadodara-390001; Phone: 0265 2786005;
Email: aumcopypoint.ap@gmail.com

Annual Subscription for 4 issues (Hard copies)

Individual	: TMC members: Rs. 700;	Others: Rs. 1000.
Societies/Institutions	: TMC members: Rs. 1400;	Others: Rs. 2000.
Outside India	: USD ($) 50.	

The amount should be deposited to the account of
"The Mathematics Consortium", Kotak Mahindra Bank, East Street Branch,
Pune, Maharashtra 411001, INDIA.

Account Number: 9412331450, IFSC Code: KKBK0000721