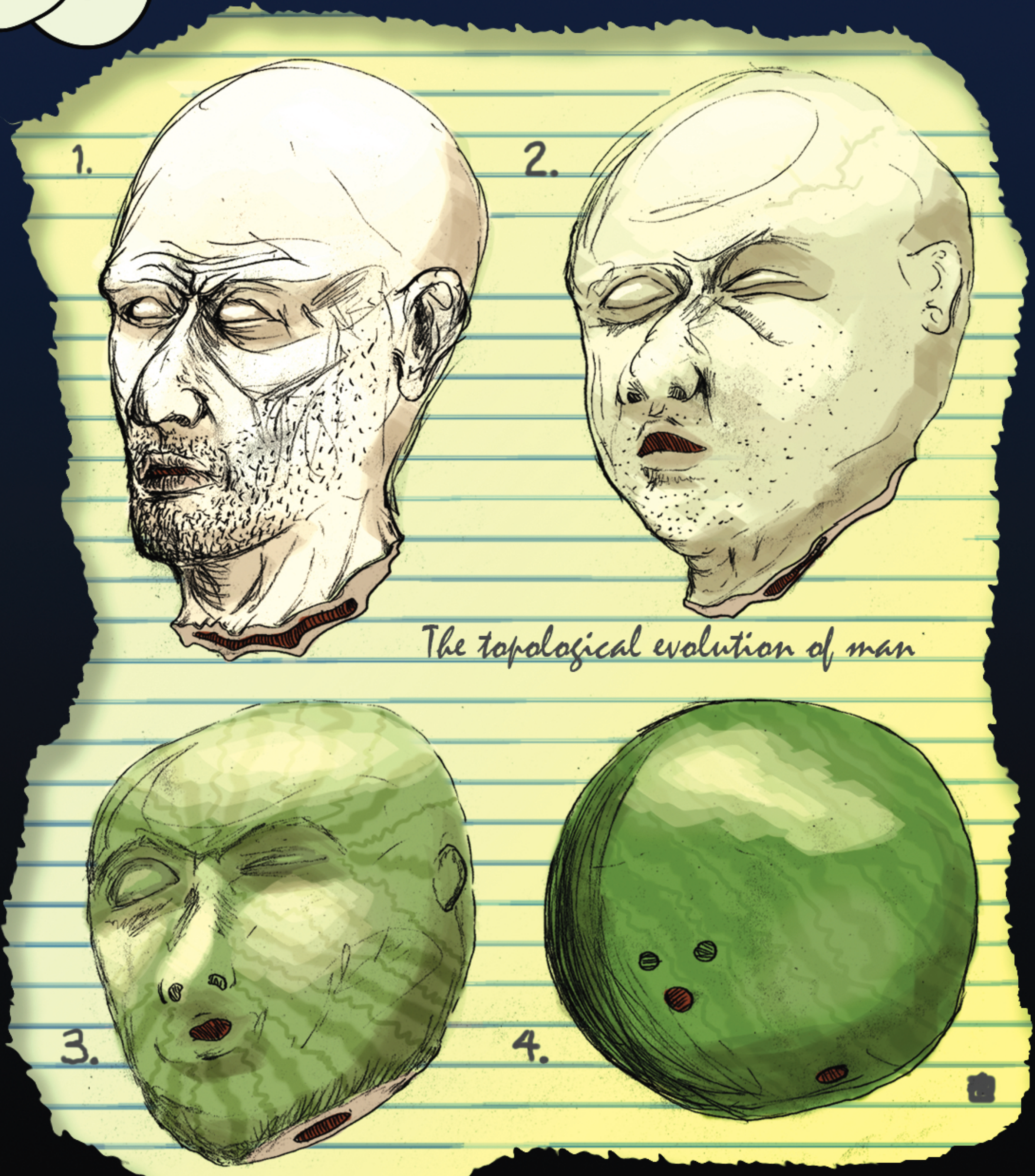
McGILL UNDERGRADUATE MATHEMATICS MAGAZINE

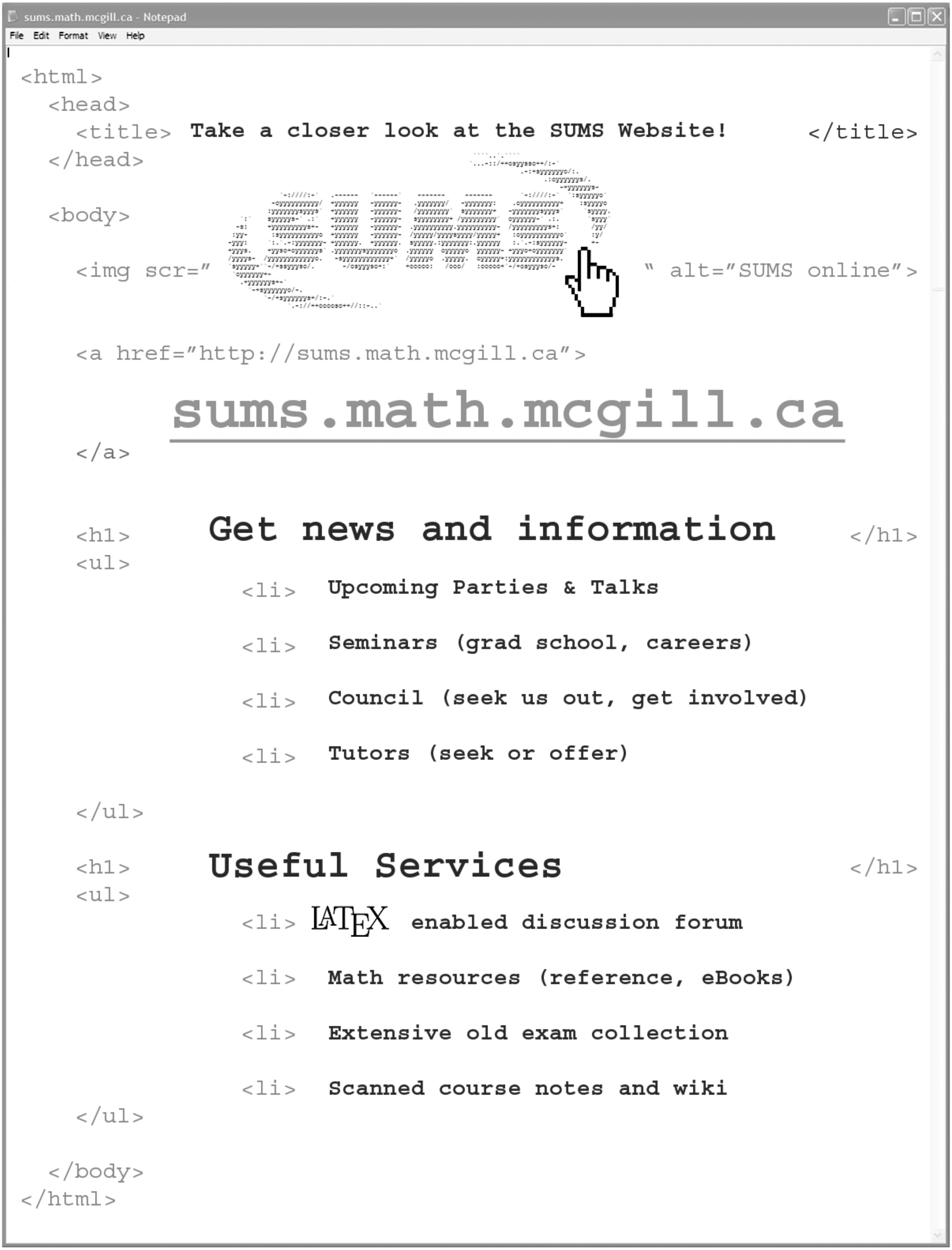
the δelta εpsilon

FIRST EDITION



The topological evolution of man

McGill SUMS



```
<html>
  <head>
    <title> Take a closer look at the SUMS Website!     </title>
  </head>

  <body>

    <img scr="                                           " alt="SUMS online">

    <a href="http://sums.math.mcgill.ca">

         sums.math.mcgill.ca

    </a>

    <h1>     Get news and information            </h1>
    <ul>
             <li>  Upcoming Parties & Talks

             <li>  Seminars (grad school, careers)

             <li>  Council (seek us out, get involved)

             <li>  Tutors (seek or offer)

    </ul>

    <h1>     Useful Services                     </h1>
    <ul>
             <li>  LaTeX enabled discussion forum

             <li>  Math resources (reference, eBooks)

             <li>  Extensive old exam collection

             <li>  Scanned course notes and wiki
    </ul>

  </body>
</html>
```

Contents

Letter From The Editors

Dear mathematics students of McGill,

One fine morning last April, as the snows were melting, the birds were singing and the ODE exam was nearing, a few math undergraduates turned envious eyes towards the Faculty of Arts and asked “Why do they have journals and not us? Aren’t we part of that faculty too?”. Half a year later, you’re holding the pilot edition of The Delta-Epsilon. Ain’t time remarkable.

This magazine is intended as a place to publish summer research by undergraduates, to learn what your professors or fellow students are doing, and to share ideas about course material. There are also articles on the history and culture of mathematics, the required jokes and book reviews, and some interesting puzzles that may keep you busy in (or out) of class. Some of the articles are more involved than others, and some are aimed at people who’ve taken a specific course; however, our main objective was to produce a magazine which you could keep on your bookshelf and consult from time to time as the content becomes relevant (or understandable) to you.

Also, we would like to encourage all undergraduates to think about writing an article for next year’s edition, scheduled for release in September, 2007.

So enjoy the articles and let us know what you think.

The Delta-Epsilon Editing Board

Letter from SUMS

On behalf of the Society of Undergraduate Math Students (SUMS) I would like to thank the staff of The Delta-Epsilon for their hard work and commitment to putting together this exciting undergraduate mathematics journal for the McGill mathematics community. It’s hard work from students such as these that enriches the McGill mathematics experience.

SUMS is an organization aimed at improving the undergraduate experience for mathematics students. We offer many services including the organization of lectures, social events, tutorials, a tutor list, a website and, starting this year, a notes scanning project.

I would like to encourage everyone with even a passing interest in mathematics to come join us at 1B20 in Burnside Hall in our lounge area anytime you want to connect. If you need help with your work, chances are you’ll find someone there who has been through the same trials and tribulations, and will be willing and able to help you out. On the other hand, if you’ve finished your work for the day and just want to unwind and play games or discuss current topics, we’re there for that too.

Please check out our website at http://sums.math.mcgill.ca

Marc-André Rousseau
SUMS President

The beauty of math: An Interview with Professor Niky Kamran

Alexandra Ortan

"What makes math interesting is what's unexpected and harmonious at the same time. It is the same in music. The interesting parts in a piece of music are the singular points. It's when things modulate, when something happens that you didn't expect. In math, if you just stay in familiar territory where you can guess the answer to every question, it's rather boring." And Professor Niky Kamran, whose passion next to math is playing the violin, is well-placed to know how interesting math can get when you step out of that familiar territory.

What got prof. Kamran interested in mathematics is precisely this æsthetic component. "I thought it was beautiful" he says about the whole new body of mathematics he encountered in high school. "The thing that really amazed me when I was being taught the rudiments of calculus, was the realization that you could compute exactly the velocity of an object that's moving and twisting in space, by knowing its position as a function of time. I thought that was amazing. The point is that there is the physical intuition that you have, and then there comes both a set of hypotheses that you formulate and a framework. Within that framework, you're able to discover and confirm facts. The process whereby you do this is fun, because it's like playing a game. You're given the rules, you have an objective (or maybe you don't have an objective, but you surmise an objective) and then you use the rules of the game to get there. Then there are the facts that you've been taught and that you've in some cases discovered, which are really beautiful, and that's the æsthetic component to it. It's the combination of the two which appealed to me."

Thus, it is not surprising to find prof. Kamran, many years later, still trying to figure out how objects move in space, though we are no longer talking about the rigid bodies and Newtonian space of our high school years, but rather about wave-particles and curved space-time – a notch or two higher in complexity. In fact, when asked what particular branch of mathematics he is working on, prof. Kamran replies that his interests lie at the confluence of geometry, analysis and a bit of algebra. "What I like to do, for example, is to look at differential equations (DEs) through the prism of geometry, to see what geometry can tell me about the properties and the attributes of the DE. I'm also interested in looking at DEs that originate in physics, particularly in relativity, and seeing how the concept of the curvature of space-time will affect the behaviour of the solutions to a

TRIVIA

Erdös number: 3

Favourite millennium problem: "The Riemann hypothesis. It really fascinates me because the statement is so easy to understand; every math student has to be fascinated by that. But the millennium problem which has greater relevance to what I do is the problem of mass gap for Yang-Mills theories, because I've been thinking about the mass gap in a much simpler context."

DE governing, for example, the propagation of waves. So there are problems that have their roots in physics, but that I look at from the perspective of geometry, analysis and algebra."

"Here's a simple question on which I worked and for which my collaborators and I got a precise answer. Suppose you're in ordinary flat Minkowski space (a 4 dimensional metric space with complex time-dimension) and you look at the propagation of a wave. What you can show, and you learn this in a PDEs course, is that the amplitude of the wave at any location of space will decay proportionally to $t^{-3/2}$. Now if you're in Kerr geometry (see framed text), you wonder 'is the space-time curvature going to affect the rate at which your wave is going to decay?' We expected that the rate would be affected by the curvature, but we didn't know how. So the bets were open. For a long time, we thought that the amplitude should decay faster than in Minkowski space, but the answer is that it decays slower. In fact the rate of decay is $t^{-5/6}$. You can justify it by some non-rigorous arguments, but to give a real solid mathematical proof takes some effort, and this is what we did in some of our papers: we quantified precisely the effects of space-time curvature on the propagation of certain types of waves in the Kerr-geometry."

"Now this result works for the simplest kind of waves, scalar waves. A really interesting problem is to know what happens if you look at a wave that corresponds to a gravitational perturbation. So you've got the Kerr black hole and then you have a gravitational field which is felt by the space-time geometry of the Kerr black hole through incoming gravitational waves. Things are going to interact, the event-horizon is going to go hay-wire, all sorts of things are going to go wrong. The question is to know whether eventually this thing will settle down or not. This is called the black hole stability problem, and it's one of the problems I'm working on. It's a very fascinating problem, and for mathematicians it's a real challenge."

So now that you've found a fascinating problem, you might wonder how to go about solving something which, at least to the neophyte, seems out of reach. "Well, typically when you do research in mathematics, you don't take a huge leap in one shot" says prof. Kamran. "You build on a very large body of known results. I don't wake up one morning and while shaving decide that I'm going to solve the stability problem for the Kerr black hole. You spend a lot of time studying the literature and then you let things sit for a while. Then typically what you need first is a strategy. That's in some sense the most important part. So you try to map out what would be the main steps that you'd have to go through. Once that is done and you're fairly confident that your strategy is the right one, you pick up your shovel and you start digging. You try to go from step one to step two in a process which will take you through one hundred steps. And sometimes you take step 34 for granted to see what happens in step 71 and

KERR GEOMETRY

When a sufficiently massive star exhausts its nuclear fuel and collapses gravitationally, you sometimes end up with a black hole. The Kerr geometry is the exact solution of the Einstein equations of gravitation, which describes the space-time geometry outside the rotating black hole in equilibrium. So as a space-time geometry, it's axi-symmetric, because there is rotation. It's stationary because it's an equilibrium configuration, and it describes a black hole because it has an event horizon, which hides a singularity. But what's amazing is that it's an exact solution, i.e. a solution that you can write in closed form. What this solution means is that when the star collapses and you end up with a black hole, all the complicated degrees of freedom of the star have been radiated away and you're left with mass, angular momentum and charge - if it is charged. That's an absolutely amazing result. It's the uniqueness theorem for the Einstein equations. The history of its discovery is very interesting. The Kerr geometry was discovered by a mathematician who was interested in Lorentzian geometry in four dimensions, Roy Kerr. He discovered it in 1963, from purely mathematical premises. What happened afterwards is that in 1968 and 1972, Werner Israel and Brandon Carter proved that the parameterized family of Kerr solutions is the unique solution of the boundary value problem for the Einstein equations that corresponds to black hole equilibrium state. So when you're observing a black hole in equilibrium, it's characterized uniquely by these parameters (mass, angular momentum and charge if applicable).

if it looks like everything else is going to work then you move forward and you come back to try to fix things." As to the part played by intuition in the whole process, prof. Kamran declares it to be crucial. "You work with intuition, and intuition is based on the wealth of experience, it doesn't come directly from heaven. It comes from a collection of facts which are part of your memory, things that you've seen, that's what intuition is built upon. You can't do any math without intuition. Otherwise, you could just feed the statement of a theorem into a computer and the computer would go on and prove it for you. There is a branch of theoretical computer science called automated theorem proving where this type of process is looked at, but for results of the type that I'm interested in, it's useless."

As most of our readers are likely to be young aspiring mathematicians, undertaking yet another (or first) year on the long road of mathematical training, and are perhaps still not sure what it is they've gotten themselves into, we asked prof. Kamran what it's really like to be a mathematician, and whether the urban legends speaking of a reclusive, solitary being have any foundation. "Not necessarily" he answers. "More and more people do mathematics in collaboration, but I think it's fair to say that when you think hard about a problem, you're on your own. The fun part however comes when you're able to collaborate with someone who's expertise is complementary to your own, then you realize that what you've been thinking about resonates into a broader context. You can then bring two points of view together and you can prove good theorems and solve interesting problems. So there is a reclusive, solitary component to it, but it's not the whole story by a long shot. Although mathematicians can be a bit strange sometimes, there is a whole social component to the activity."

However prof. Kamran warns the aspiring mathematician that "First and foremost you have to love math. If you don't love it, you won't survive it. It's serious work, it's competitive. You have to sit down and spend a lot of time in apprenticeship. Secondly, you have to be able to be realistic and see if this is really for you or not. That can only come if you're a good apprentice and you learn to do mathematics. If it comes to you with some effort, but the pleasure exceeds the pain and you're getting something valuable out of it, then it's well worth it. It's a noble activity. It's the same thing for becoming a musician in some sense. I've seen many people who were good enough to become professional musicians but who've lost the sparkle and love for the activity, and that's one path to depression and unhappiness. It's the same thing for math. You need to keep loving doing mathematics."

Les Confessions, ou de la théorie des groupes en chromodynamique quantique

Michael McBreen

This article divides into two parts: a popularized introduction to gauge field theories for the mathematics student, followed by a technical exposition of an aspect of the authors NSERC-sponsored (USRA) work on gauge theory. None of the results presented here are original.

An Introduction to Gauge Theory

For pedagogical reasons, I will avoid mentioning quark fields until after the introduction of gauge symmetry. Any resulting inaccuracies should hopefully be rectified at that point.

The atom is composed of electrons and a nucleus, and this nucleus is made of protons and neutrons, collectively known as nucleons. According to the Standard Model[1], the nucleons themselves are made of particles called quarks. There are many species of quarks, each with its own mass, charge and so forth, but from now on you can assume I'm always talking about the same species.

Any given quark has a property called "colour", which corresponds to a direction in an abstract three-dimensional "colour-space" isomorphic to $\mathbb{C}^3$. I will sometimes say that a quark "points" in some direction, meaning its colour is given by that direction, or write $\hat{q}$ to indicate the corresponding unit vector. All colours or orientations are equivalent in that only the relative orientation (i.e. the inner product ($\hat{q}_2$, $\hat{q}_2$)) of two quarks will affect their interaction. In other words, there's no *absolute* orientation, just as there's no absolute position, orientation or velocity in ordinary space-time. This means that if we magically "rotate"[2] every quark by the same angle, no *observables* will change: nothing will get hotter or colder, go faster or slower, get bigger or smaller. We call the freedom to vary a parameter of our model without changing the physical situation a *symmetry.* It essentially means that the parameter is superfluous: it's an artifact of our formalism. We call the corresponding group of transformations the *symmetry group* ($SU(3)$ in this case). For instance, consider the interaction between two isolated stars: only the relative distance between the stars matters, not their "absolute" position. The latter is therefore a superfluous parameter, and the symmetry group is the group of translations.

In fact, quarks possess a stronger form of symmetry called *local gauge invariance.* Not only can we rotate all quarks by the same angle, we can rotate quarks at different positions by different angles (i.e. apply a *gauge transformation*). But wait - didn't I claim that the relative orientation of two quarks is important? In fact, there's an object called the gauge field U, which permeates all space, and which "remembers" the relative orientation of the quarks when we rotate them. More precisely, when we rotate the quarks we must also transform this field - I'll explain how in a second - and this transformation somehow cancels out the rotation of the quarks.

You might ask why we bother postulating gauge invariance, if we also need to postulate a field that cancels out any gauge transformations we might make. The justification is that this gauge field can be observed in nature. In fact, if we require U to evolve over time according to a certain equation of motion, we find that it reproduces the behaviour of the *strong force* that binds quarks together. Better still, it turns out that all fundamental forces seem to originate from gauge fields.

But back to U. What sort of a field do we need to cancel out gauge transformations? Say we have two quarks pointing in different directions, and we rotate them so the directions coincide. We want to recover the initial relative orientation, so we'd like the field to somehow encode a rotation that brings the vectors back to their initial (relative) positions. The field U does this as follows. $U = U(p)$ is a function

[1] The Standard model is the currently accepted description of particles and forces, though rumour has it that cracks are showing in the theory, and a replacement will urgently be needed.

[2] The analogue of the group of rotations relevant here is the group $SU(3)$ of unitary operators on $\mathbb{C}^3$ with determinant 1.

that associates a gauge group element g (a rotation) to each path p in spacetime. If p_2 begins where p_1 ends, then $U(p_1)U(p_2) = U(p_1 \cdot p_2)$, where $p_1 \cdot p_2$ is the path obtained by going along p_1 first and p_2 second. $U(p_{trivial}) = 1$ where $p_{trivial}$ is the trivial path from x to x, and finally $U(p)$ varies continuously as p is lengthened. Basically, U lets us compare two quarks at different points x and y by taking the quark at x, rotating it by $U(p_{x\to y})$ where $p_{x\to y}$ is some path[3] from x to y, and then taking the relative orientations. When we rotate the quarks at x and y by $R(x)$ and $R(y)$, $U(p_{x\to y})$ becomes $R(y)U(p_{x\to y})R(x)^{-1}$, so that we'll get the same answer if we compare our quarks after the rotation:

$$(U(p_{x\to y})v_x, v_y) = (R(y)U(p_{x\to y})R(x)^{-1} \\ R(x)v_x, R(y)v_y). \tag{0.1}$$

Now, I have a confession to make: the picture I've been painting isn't quite right. In fact, rather than having many point-like quarks flying about and interacting, we really have one big object called the quark field $\vec{\phi}$, which (roughly) associates a quark density with each point in space.[4] What we would call a quark is a bump in the field (in the density) that travels along, collides with other bumps, bounces off them, etc., like ripples on a lake. The time evolution of $\vec{\phi}$ is governed by an equation linking its spacetime "slope" and its absolute value at each point. For instance, an isolated bump will flatten out into expanding ripples, while a long straight wave will travel along at a constant speed. Why is $\vec{\phi}$ a vector field? Remember that each quark points in some direction. We therefore have three separate fields, ϕ_1, ϕ_2, ϕ_3, giving the quark density along each of the colour axes, and we choose to group them into the vector field $\vec{\phi} = (\phi_1, \phi_2, \phi_3)$. Local gauge invariance really means that you can rotate this vector by different angles at different points without affecting the physical state of the system.[5]

As before, the gauge field $U(p)$ allows local gauge invariance. The description of $U(p)$ above is still valid, but we need only consider infinitesimal paths $p_{r\to r+dr}$ since the motion of the field is governed by local laws.[6] We avoid worrying about the path independence of $U(p_{r\to r+dr})$ by always choosing the infinitesimal "straight" path. Of course, the gauge field itself must be gauge invariant (i.e. all measurable properties of the field should be invariant under local gauge transformations).

One more confession: what I've described up to now is a classical field, but we're really after the quantum field. The quantum field, much like the quantum particle, has a wavefunctional $\Phi(U(p))$ that associates a probability to each classical parameter (we'll simply call it a wavefunction). In this case, the function $U(p)$ is the classical parameter. The classical parameters of a system are grouped into subsets of *compatible* parameters, meaning they can simultaneously have definite measurable values, while non-compatible parameters satisfy a generalized version of the Heisenberg uncertainty principle.

That's gauge field theory for you – now comes my own summer research. We studied various gauge fields in a "vacuum", i.e. with no quarks present. The goal was to take a field with state U_0 at time $t = 0$, and determine U_T at $t = T$. We have an equation for this, but we can't solve it, so we make three major simplifying approximations and modifications:

1. **Space is discrete.** We replace the continuum of space with a lattice of discrete points (x, y, z) with spacing $\mathbf{a}$. The field $U(p)$ is defined on the set of edges (a edge is a straight path p joining two adjacent points). This way, each edge is associated with a group element g of $SU(3)$. The wavefunction $\Phi(U(p))$ associates a probability to each such configuration of the lattice. If we were including quarks in our model, we would define the field $\vec{\phi}$ on the vertices.

2. **We neglect the interaction of the field with itself.** In other words, any excitations of the field (waves or bumps) will travel straight through each other rather than bounce or deviate. To make this simplification, we remove all references to $U(p)$ as such (the potential terms) from the equation of motion, leaving only the derivatives of $U(p)$ (the kinetic terms).

3. **We use Euclidian time.** The equations of motion for the wavefunction involve a time variable t. We analytically continue the functions of this variable to the complex plane, so that $t \to z$, and evaluate the functions at $z = it$

[3] Don't worry about path independence - I'll get to that later.

[4] I'll explain the vector notation in a minute.

[5] We do require that the rotations change continuously as we go from point to point.

[6] In other words, $\frac{d\phi}{dt}|_x$ depends only on an infinitesimal spacetime region around x.

rather than t. We lose some information this way, but we can still obtain many interesting properties such as the energy of the field.

Within this simplified model, the equations of motion can be solved completely for important gauge fields such as the $SU(3)$ field discussed above, and the electromagnetic and weak fields, whose gauge groups are respectively $U(1)$ and $SU(2)$.

Lexicon

Symmetry group: Consider a model of a physical system in which every physical state is given by a (possibly infinite) number of parameters, such as position x, speed v, temperature T, etc. A symmetry of the system is a transformation of the parameters that leaves the physical state invariant. The set of all symmetries of the system form a group under composition, called the symmetry group.

Gauge invariance and gauge group: Consider a model which associates a set of parameters to each point in space - or spacetime, if you prefer. If we can transform the parameters at each point independently (with certain restrictions) while leaving the physical state invariant, we say the system is gauge invariant. The associated group of symmetries is the gauge group.

U(1): The multiplicative group of all complex numbers with norm 1. It is isomorphic to the group of rotations of the real plane about the origin.

SU(2): The group of unitary linear transformations with determinant 1 acting on $\mathbb{C}^2$. This can be thought of as a complex version of the group of rotations in 3D.

SU(3): The group of unitary linear transformations with determinant 1 acting on $\mathbb{C}^3$.

Group representation: A representation is an action of the group on a vector space V, i.e. a homomorphism $\rho : G \to \mathcal{L}(V)$. The homomorphism property notably implies $\rho(g_1 g_2) = \rho(g_1)\rho(g_2)$. If V is an n-dimensional complex space, $\rho : G \to GL_n(\mathbb{C})$ associates a matrix $\rho(g)$ with each group element g. Given a vector in $\mathbb{C}^n$, we can let the group act on it through some n-dimensional representation ρ, in which case we say the vector belongs to ρ.

There are a number of distinguished representations. For instance, the **trivial representation** is the homomorphism to $GL_1(\mathbb{C})$ that sends G to the identity. The **left regular**

Lexicon - continued

representation is the action of G on the space of functions on G defined by $\rho_{reg}(h) : f(g) \to f(h^{-1}g)$. It's an infinite dimensional reducible representation.

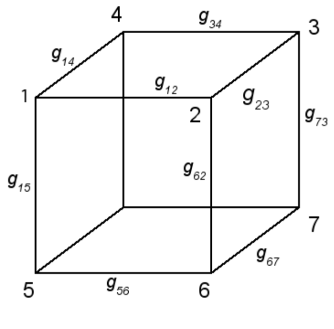
Irreducible representation: A representation that does not preserve any non-trivial proper subspaces of V. The name *irreducible* comes from the fact that non-irreducible (reducible) representations can often be expressed as "sums" of irreducible representations.

Representative functions of a representation: The functions $\rho(g)_{ij} : G \to \mathbb{C}$ that give the matrix elements $\rho(g)_{ij}$ of the matrix $\rho(g) \in GL_n(\mathbb{C})$. The Peter-Weyl Theorem tells us the representative functions of the irreducible representations are an orthogonal set of functions spanning the full space of "nice" functions on G. It applies to all compact Lie groups, including the gauge groups we're working with.

Harmonic Analysis on a Group Manifold

We take space to be a cubic lattice with side length N+1 vertices v_i (or N edges l_{ij}). Our field $U(p)$ is defined on the edges.

The wave function $\Phi : G \times G \cdots \times G \to \mathbb{C}$ associates a complex number with each classical configuration of the lattice (there's a copy of G in the function domain for each edge in the lattice). We denote a generic classical configuration, which associates some group element g_{ij} to each edge l_{ij} on the lattice, by $\overrightarrow{U(p)}$. We write $\Phi(\overrightarrow{U(p)})$ for the wavefunction, and the probability density of a configuration is $|\Phi(\overrightarrow{U(p)})|^2$. The set V all such wavefunctions is a vector space



over $\mathbb{C}$, with inner product defined as

$$(\Phi_1, \Phi_2) = \int_G dg_{12} \int_G dg_{23} \int_G dg_{34} \cdots \int_G dg_{nm}$$
$$\Phi_1^*(g_{12} \times g_{23} \times \cdots g_{nm}) \Phi_2(g_{12} \times g_{23} \times \cdots g_{nm})$$

where the star denotes complex conjugation. The integration measure dg is called the Haar measure, and is invariant under left and right group composition: $d(hg) = d(gh) = dg$ where h is a fixed group element.

A local gauge transformation Λ_3 at vertex v_3 sends g_{23} to $g_{23}\Lambda_3$, g_{43} to $g_{43}\Lambda_3$, etc. – the gauge group element at each edge connected to vertex i is transformed, as we can see from (0.1). Λ_3 accordingly sends $\Phi(\cdots \times g_{23} \times g_{43} \cdots)$ to $\Phi(\cdots \times g_{23}\Lambda_3 \times g_{43}\Lambda_3 \cdots)$. If Φ is gauge invariant, then it should be sent into itself. V contains both invariant and non-invariant wavefunctions, but the non-invariant ones are artifacts of our formalism and are not found in nature. What we are really interested in is the subspace V_{invar} of gauge invariant functions. I wont explain how to find this subspace here, but the hardy reader is referred to [4].

The time evolution of a wavefunction (its equation of motion) is given by

$$\Phi(t = T) = e^{-kT \sum_{ij} \Delta_{ij}} \Phi(t = 0) \qquad (0.2)$$

where k is a constant and Δ is a generalized version of the Laplacian called the Laplace-Beltrami operator on the group manifold. It can be thought of as measuring the convexity of a function of the group. Each Δ_{ij} acts on a single edge l_{ij}.

We want to find a basis of “harmonic” wavefunction with simple time evolution properties. This would allow us to find the evolution of an arbitrary wavefunction by decomposing it into harmonics. Luckily, the time evolution operator is self-adjoint, so we can find a basis of eigenfunctions for it spanning V.

To construct this basis, we need a central result of group representation theory: the Peter-Weyl Theorem. The PWT generalizes the theory of Fourier transformations to arbitrary group manifolds. Fourier analysis tells us that any sufficiently nice function of the real interval $[-2\pi, 2\pi]$ can be expressed as an (infinite) sum of orthogonal harmonics or sinusoidal functions. That is, the sinusoidal functions are in some sense an orthogonal basis for the space of functions with that domain. The PWT tells us that any nice function of the group manifold can be expressed as an (infinite) sum of representative functions of irreducible representations of the group, and that these representative functions are orthogonal.[7] For the gauge groups $U(1)$, $SU(2)$ and $SU(3)$, we have a simple classification of all non-isomorphic irreducible representations.

Now, the Laplace-Beltrami happens to be a Casimir operator for the group G, meaning that it commutes with all $\rho(g)$, no matter the representation. By a simple theorem of group representation theory called Schur’s Lemma, any operator that commutes with an irreducible representation of a group is a multiple of the identity on that representation. Using the fact that representations preserve the group action, we can easily show that the representative functions of any irreducible representation ρ transform under the action of G as the representation ρ itself. According to Schur’s Lemma, these are eigenvectors of all Casimir operators, including the Laplace-Beltrami operator.

We therefore have a basis of eigenvectors of Δ_{ij} spanning the functions on G. Φ is a function of the direct product $G \times G \times \cdots \times G$, which for physical reasons factors into a product of functions of G. We can hence decompose any Φ into a sum of products of representative functions using the PWT, and each product is an eigenvector of the sum of Laplace-Beltrami operators appearing in (0.2). In other words, we have a full basis of eigenvectors labeled by the representations and corresponding matrix indices appearing in the product. This was the objective.

References

[1] Montvay, I, Munster, G *Quantum Fields on a Lattice* , Cambridge University Press, Cambridge (1997)

[2] Creutz, M *Quarks, Gluons, and Lattices*, Cambridge University Press, (1986)

[3] Hall, B *Lie groups, Lie algebras, and representations. An elementary introduction.*, Graduate Texts in Mathematics, 222. Springer, (2003)

[4] Baez, J Spin Network States in Gauge Theory, arXiv:gr-qc/9411007 v1 2 Nov 94

[7] See the lexicon if you’re not familiar with group representation theory.

Predicting the Lifespan of AIDS Patients with Survival Analysis

Mireille Schnitzer

To infer expected survival time after the onset of a disease such as acquired immune deficiency syndrome (AIDS), a researcher will recruit a number of patients and eventually analyze their sample failure times. This paper will briefly describe the difficulties that arise when attempting a statistical analysis of such a study, and will explain one method of dealing with right-censorship. The product limit estimate (developed in 1958 by Kaplan and Meier) will be used to demonstrate a way of approximating expected failure time.

Survival is a common concern in many scientific, social and industrial disciplines. For instance, a biologist might consider the average lifetime of wild deer, or an economist could study how long the average person remains on welfare. The field of *survival analysis* is a branch of statistics that evaluates the results of such studies and the data loss that might occur. Formally, survival analysis is the study of measuring a *time-to-event*, or the time lapse between two events.

One of the most important applications of survival analysis is in the medical field, where it plays a fundamental role in determining the estimated survival time after the onset of a disease. Such studies are designed to work in conjunction with hospitals or clinics where patients are asked to participate.

Let's consider the example of measuring the expected survival time after a patient develops AIDS (acquired immune deficiency syndrome). We will work with a hospital and acquire a sample of patients, recording their times of death as the study progresses. The notation generally used is as follows: we will observe n individuals whose failure times (random survival times after developing the disease) are the random variables $T_1, ..., T_n$. These will be assumed to be independent and identically distributed with common density $f(t)$ for time t. Note that we relabel the time of onset as $t_0 = 0$ so that the time of death corresponds to the failure time.

To simplify our analysis, we will assume that there is no bias in the way the patients are selected. This is usually a bad assumption to make in medical studies (as it is generally more likely to recruit patients with longer life spans), but it is necessary to simplify our analysis at this level. We will also assume that every patient has a known time of onset, such that we know the time at which they developed the disease.

The form of data loss that we will be concerned with is called *right-censoring* and it occurs when we never learn the failure time of an individual. The only information we have is the time at which we *lose* the individual, who is now referred to as being *censored*. This might arise if a patient chooses to end his visits to an AIDS clinic so that their previous doctor can no longer follow them. We will never know their failure time, but we do know a time at which they were still alive. Censoring could also occur if the study ends before all of our patients have died. In any case, a censoring at time a means that the only information we have is that $T_i > a$ for the ith individual.

A New Function

In statistics, we've been taught to find the cumulative density function, $P(X \leq x)$. But in survival analysis, it is more meaningful (and, eventually, neater) to study the *survivor function*:

$$F(t) = P(T > t)$$

which is the probability of an individual surviving past some time t. In the discrete case, where we have failure times $a_1, a_2, ...a_r$, this function becomes

$$F(t) = P(T > t) = \sum_{i|a_i > t} P(T = a_i).$$

The characteristics of the survivor function include:

1. The function is non-increasing (it cannot be more likely to survive past a later time)

2. $F(0) = 1$ (everyone is surviving initially)

3. The limit approaches zero (at time infinity, there are no survivors)

Clearly,

$$F(t) = 1 - P(T \leq t)$$

so that

$$\frac{dF(t)}{dt} = -f(t)$$

where $f(t)$ is the distribution function for T.

Since we are concerned with estimating this function using the data from a study, we must use empirical functions. Before we develop more sophisticated machinery, consider the situation where there is absolutely no data loss. To estimate the survivor function, we calculate the proportion of patients surviving after time t. To do so, consider a discrete model where the failure times $t_1, t_2, ...t_r$ are specifically the observed times where we have recorded a patient's death. Then, at any given time t_i, we can compute the number of deaths, d_i, and the number of survivors immediately after that time, n_i.

So, the *empirical survivor function* is

$$F_n(t) = \frac{no.\ t_i > t}{n}$$

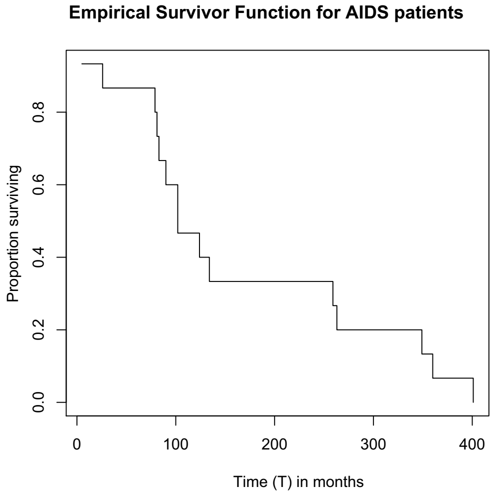
(which, you might notice, is also $1 - \overline{F_n(t)}$, the empirical distribution function).

The empirical survivor function is a step function that begins at 1 and decreases by $\frac{d_i}{n}$ at every t_i. When the population has expired, the curve is at zero.

To illustrate, suppose we have observed 15 individuals (Table 1) and witnessed each of their failure times in months (there is no data loss). Please note that these numbers have been arbitrarily selected and do not represent real measures of survival.

We can measure the survivorship at any given time simply by counting. Hence, we get the empirical survivor function in the next figure.

Notice that all members of the population survive at time zero, and that by 401 months, all have died. The average (mean) failure time is 163.87 months.



From the empirical survivor function, we can derive the *empirical failure distribution* by simply noting that $P(T = t_i) = F(t_i^-) - F(t_i) = d_i/n$ where $F(a_i^-) = \lim_{t \to a_i^-}$. That is, the probability of failure at any given time is literally the height of the step in the survivor function at that time (it's how many patients died after that many months).

Incorporating Right-Censorship

Assuming that we can collect every failure time is clearly incorrect, and we will now incorporate the possibility of right-censoring into our scenario. We have a pair of times for each individual; a failure time (T_i) and a censored time (C_i), and we assume that these are independent of each other. The independence here means that, for instance, patients do not change hospitals if their condition worsens (and their failure draws near). We only actually observe one of these times.

The notation is then as follows: we watch individuals $X_1, ...X_n$ assumed to be independent and identically distributed with common probability distribution function $f(t)$. For each random variable, we observe the data (δ_i, X_i). In studies with censoring, δ_i refers to the indicator that equals one if the individual's failure is observed, and zero if the individual is censored. X_i is then either the failure time T_i or the censored time C_i (the last time the individual was known to be alive), depending on which information we have, so that $X_i = \min(T_i, C_i)$.

Using this notation, we can use the likelihood to develop a maximum likelihood estimator for the survivor function for a parametric model. In order to build this estimator, we make the assumption that the censoring mechanism carries no information about the parameter. In other words, any function of the distribution of censoring is a constant with respect to θ, the parameter.

We observe the pair (X_i, δ_i) for every individual, and we consider the likelihood of such an observation. If a failure is not censored, its contribution to the likelihood is

$$P_\theta[X_i = t, \delta_i = 1]$$

where θ is the parameter. An uncensored observation means that C_i is larger than T_i (which is actually observed), so the corresponding contribution to the likelihood is,

$$P_\theta[T_i = t, C_i > t].$$

Now, using the assumption that the censoring scheme is independent, this is equal to,

$$f(t,\theta)G_i(t)$$

where f is the p.d.f. for the failure time T and G_i is the survivor function of the censored times. Recall that we've assumed that G_i holds no information about θ.

Now, suppose that we are considering a failure that is censored. The resulting probability of such an occurrence is

$$P_\theta[X_i = t, \delta_i = 0]$$

which, through the same reasoning, corresponds to,

$$P_\theta[C_i = t, T_i > t].$$

As before, this gives us

$$g_i(t)F(t,\theta)$$

where g_i is the distribution function of the censoring random variable for the ith individual.

So now we can construct the likelihood function:

$$\begin{aligned} L &= \prod_{i=1}^{n} P[X_i = t_i, \delta = \delta_i] \\ &= \prod_{i=1}^{n} (P[C_i = t_i, \delta_i = 0])^{1-\delta_i} \times \\ &\quad \times (P[T_i = t_i, \delta_i = 1])^{\delta_i} \\ &\propto \prod_{i=1}^{n} f(t_i,\theta)^{\delta_i} F(t_i,\theta)^{1-\delta_i}. \end{aligned}$$

The last line is proportional due to the assumption that the censoring is uninformative with respect to θ. It then does not influence the maximization of the function (since g_i and G_i are constant with respect to θ).

This demonstrates that the contribution of the censoring to the likelihood function is of the form $P(T > t)$. The contribution of a normal observed (uncensored) failure is $P(T = t)$.

Now, consider the results of a real study. We have $t_1 < t_2 < ... < t_k$ which are observed (uncensored) failure times. Then, at each of these observed failure times, we might have more than one failure, so we refer to the number of deaths at time t_j as d_j. In between these observed failures, there are a number of censorings that may occur. We will refer to the number of censorings in the interval $[t_j, t_{j+1}]$ as m_j. These censored times are $t_{j1}, ..., t_{jm_j}$ for $j = 0, ..., k$, $t_0 = 0$ and $t_{k+1} = \infty$. The number of individuals at risk just prior to time t_j is n_j (n_j is equal to all future deaths and censorings).

Recall that

$$P(T = t_i) = F(t_i^-) - F(t_i)$$

so that the proportional likelihood can be written as

$$L \propto \prod_{j=0}^{k} \{(F(t_j^-) - F(t_j))^{d_j} \prod_{l=1}^{m_j} F(t_{jl})\}.$$

For each observed failure, we must run through and multiply each censored contribution in the interval between failures.

To maximize this function for the parameter, we first find the non-parametric MLE for the survivor function that maximizes L. This function is then substituted into the likelihood. This simplifies the expression so that it can then be maximized in terms of its parameter. In the end, our MLE is the empirical survivor function written in terms of n_i and d_i. So, the *product limit estimate* or *Kaplan-Meier estimate* is

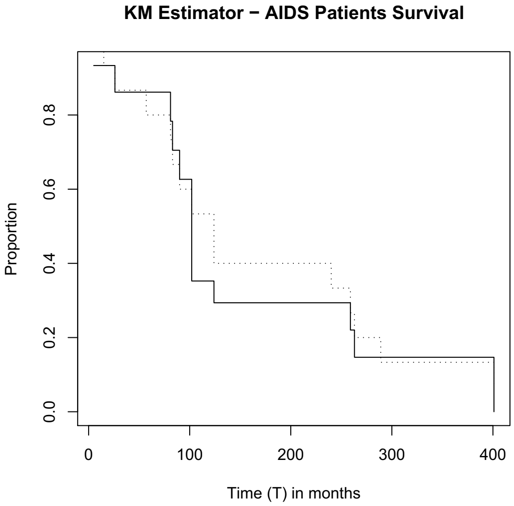
$$\hat{F}(t) = \prod_{j|t_j \le t} 1 - \frac{d_j}{n_j}.$$

(Note that the fraction $\frac{d_j}{n_j}$ is specifically the proportion of failures at every observed failure time.)

This function looks very much like the empirical survivor function and one can use it in the same way. One difference is that the Kaplan-Meier estimator may not go to zero if the last observation is a censoring. There are several suggested ways of dealing with this, but the most common is to take the function as undefined for time after the last censoring (Kalbfleisch and Prentice, 2002).

Now, as an illustration with right-censoring, consider the data set in Table 2.

Four of our 15 data points have been censored. Since censoring occurs more often as time passes, ignoring the censored individuals would be a bad idea as it would result in an underestimation. Furthermore, if we treated the censored times just like failure times, we would again be underestimating. The following graph compares the approximated survivor function found by using the KM estimator (solid line) to the empirical survivor function (dotted line) that treats the censored data as simple failure times.



If we do treat the censored times as failure times, we would expect the average survival time of patients ($E(T)$) to be just 142.47 months. But, using the KM estimator, we find that the expected failure time is 177.41 months (I used the fact that the jumps in the graph correspond to the p.d.f. of T, as mentioned earlier). By mistreating the data, we arrive at a value that is much smaller than the method that takes right-censoring into account.

The KM estimator is asymptotically unbiased, and its bias decreases exponentially as n increases (Zhou, 1988). Therefore, we can get very good results for high sample sizes. In a real medical study, the number of participants is far greater so that this method can be used to accurately predict an expected survival time after the onset of disease.

References

[1] A. G. Babiker, J. H. Darbyshire, T. E. A. Peto and A. Sarah, 1998, Issues in the Design and Analysis of Therapeutic Trials in Human Immunodeficiency Virus Infection: Walker Journal of the Royal Statistical Society. Series A (Statistics in Society), v. 161(2), pp. 239-249.

[2] P. Hougaard, 1999, Fundamentals of Survival Data: Biometrics, v. 55, pp. 13-22.

[3] J. D. Kalbfleisch and R. L. Prentice, 2002, *The Statistical Analysis of Failure Time Data*, second ed., John Wiley and Sons, Inc., New Jersey.

[4] E. L. Kaplan and P. Meier, 1958, Nonparametric Estimation from Incomplete Observations: Journal of the American Statistical Association, v. 53(282), pp. 457-481.

[5] M. Zhou, 1988, Two-Sided Bias Bound of the Kaplan-Meier Estimator: Probability Theory and Related Fields, v. 79, pp. 165-173.

Table 1:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
failure time	5	26	79	81	83	90	102	102	124	134	259	263	349	360	401

Table 2:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
failure time	5	26	15	81	83	90	102	102	124	57	259	263	289	240	401
δ_i	1	1	0	1	1	1	1	1	1	0	1	1	0	0	1

Benford's Law

Joël Perras

Benford's law states that for large sets of data, the distribution of the First Significant Digits (FSD) within this data follows a logarithmic relationship. The FSD frequency is determined by $P(FSD = d) = \log_{10}\left(1 + \frac{1}{d}\right)$, where $d = 1, 2, 3, \ldots, 8, 9$. Moreover, Benford's Law may be generalized to find the probability for the n^{th} significant digit or combinations of significant digits.

Discovery

Simon Newcomb, the original discoverer of Benford's Law, was an astronomer and a physicist who lived from 1835-1909. At that time, most calculations of the movements and precessions of celestial bodies involved long and complex derivations, which frequently required the use of Napier tables, now more commonly referred to as log tables. After extensive use of these tables, Newcomb came to the realization that, for some reason or another, the actual pages in his copy of logarithmic tables were not being used equally; instead, some of the pages were more stained and folded, showing an increase in usage. Upon further analysis, Newcomb postulated that the distribution of significant digits of his collected data was non-uniform; rather, the resulting *logarithms* of those digits was. In fact, Newcomb was able to publish an empirically-derived equation which described for this behaviour. However, without a proper proof or even possibility of application, no one paid attention to this very interesting result.

Frank Benford, who independently discovered the phenomenon in 1938, in contrast to Newcomb, did not simply publish his empirically derived formula; rather, he "compiled a list of nearly 20 000 different observations, covering everything from river drainage areas and the addresses of notable scientists to numbers appearing in an issue of Reader's Digest" [1]. Using this data, he was able to empirically verify this empirically determined phenomena; while this was not a ground-breaking proof, it was a step in the right direction. The actual proof of Benford's Law was found nearly sixty years after Benford's rediscovery of the phenomena, and was published by T.P. Hill [3], who had already written several other papers on the remarkable properties of this phenomenological law.

What is Benford's Law?

Surprisingly, the actual formulation of Benford's Law is quite simple. Moreover, Benford's Law itself is merely a generalization of a distribution pattern that arise from mixtures of uniform distributions [2], and is most often the null hypothesis in testing for human influence on data.

Benford's Equation

Originally, Benford's empirical formula was put forth to determine the distribution pattern of first significant digits (FSDs) from a collection of data sources, whose probability distributions were non-uniform and varied in nature:

$$P(FSD = d) = \log_{10}\left(1 + \frac{1}{d}\right) \quad \text{where } d = 1, 2, 3, \ldots, 8, 9. \tag{0.3}$$

While this in and of itself is quite an achievement, the formula lacked any capability for determining the frequency of second, third or n^{th} significant digits, as well as combinations of two or more digits (such as the frequency of the two consecutive digits of 99 appearing). As can be observed from eq. (0.3), for any probabilistic prediction of an n^{th} significant digit other than the first, a more generalized equation must be put forth. This was precisely what Hill derived

[8] A σ-algebra U is defined as follows: Let S be a set, and U be a non-empty collection of subsets of S such that the following are true:

1. Closed under complements
2. Closed under countable unions
3. Should contain the total set

in 1996. Moreover, he also formulated the correct σ-algebra [8] set that described the probability domain of Benford's Law, thus establishing a proper mathematical proof for the empirically derived law.

Hill was able to show that the set $\mathcal{A}$ is the smallest σ-algebra generated by the countable unions of the i^{th} significant digits [3]. From this, he derived his *General Significant Digit Law*:

$$P\left[\bigcap_{i=1}^{k}\{D_b = d_i\}\right] =$$
$$= \log_b\left[1 + \left(\sum_{i=1}^{k} b^{k-i}\cdot d_i\right)^{-1}\right]$$
$$\text{where } k \in \mathbb{N} \textit{ and } d_i \in \{1,2,3,\ldots,9\}. \tag{0.4}$$

This equation, when compared with (0.3), has the remarkable property of being able to predict the frequency of appearance of *any* significant digit, or combination of digits. Furthermore, it also illustrates the fact that the n^{th} significant digit is *dependent* on the $n-1$ previous significant digits.

To calculate *unconditional* probabilities for the n^{th} significant digit, the sum of the probabilities for the digits before the n^{th} significant digit must be calculated:

$$P(n^{th}SD = d) = \sum_{k=1}^{9}\log_{10}\left(1 + \frac{1}{10k+d}\right)$$
$$\text{Where } d = 0,1,\ldots,9. \tag{0.5}$$

Using this formula, we find that the (unconditional) maximum probability for the second significant digit occurs when $d = 0$, with a probability of 0.1197. As can be discerned from (0.5), the significant digit probabilities move exponentially towards a uniform distribution as $i \to \infty$ [3].

References

[1] R. Matthews, "Benford Bend", May/June 2000: EBSCO Research Database.

[2] R. J. Rodriguez, Ricardo, "First Significant Digit Patterns From Mixtures of Uniform Distributions", The American Statistician, 58.1 February 2004. 64-71.

[3] T.P. Hill, "The Significant Digit Phenomenon", The American Mathematical Monthly, 102.4 April 2004, 322-327.

Jokes

Q: How can you tell a sailor used to be a mathematician?
A: Instead of saying "aye, aye, captain" he says "negative one, captain!" □

A math professor organized the seminar in hydrodynamics in his University. Among the regular attendees there were two men in uniform, obviously military engineers. They never discussed the problems they were working on. But one day they asked the professor to help them with a math problem. They explained that the solution of a certain equation oscillated and asked how they should change the coefficients to make it monotonic. The professor looked at the equation and said, "Make the wings longer!" □

A chemist, a physicist and a mathematician are stranded on an island when a can of food rolls ashore. The chemist and the physicist comes up with many ingenious ways to open the can. Then suddenly the mathematician gets a bright idea: "Assume we have a can opener..." □

A tragedy of mathematics is a beautiful conjecture ruined by an ugly fact. □

The dean, to the physics department: "Why do I always have to give you guys so much money for laboratories and expensive equipment and stuff? Why couldn't you be like the math department – all they need is money for pencils, paper and waste-paper baskets; or even better, like the philosophy department. All they need is pencils and paper!" □

Generators of $SL(2,\mathbb{Z})$

Agnès F. Beaudry

We first prove that $SL(2,\mathbb{Z})$ is generated by two elements. Then we motivate the study of this group by describing its action on the upper-half plane.

Why $SL(2,\mathbb{Z})$?

If you studied algebra, you have probably encountered the group of matrices called the general linear group of degree 2, $GL(2,F)$, the set of invertible 2×2 matrices with entries in a field F. One important subgroup of $GL(2,F)$ is the special linear group, $SL(2,F)$, the set of matrices with determinant equal to one. Now, if $F=\mathbb{R}$, the set of matrices in $SL(2,\mathbb{R})$ with integer entries, denoted $SL(2,\mathbb{Z})$, is also a subgroup and plays an important role in the geometry of the upper-half plane

$$\mathcal{H}=\{z=x+yi\in\mathbf{C}\mid y>0\}.$$

I will start by discussing the structure of the group $SL(2,\mathbb{Z})$, showing that it is generated by the matrices $T=\begin{pmatrix}1&1\\0&1\end{pmatrix}$ and $S=\begin{pmatrix}0&-1\\1&0\end{pmatrix}$. Then I will quickly try to convince you that this group is interesting.

The Generators

$SL(2,\mathbb{Z})$ is closed under matrix multiplication, since the determinant of a product is the product of the determinants. With identity $I=\begin{pmatrix}1&0\\0&1\end{pmatrix}$, all elements $\begin{pmatrix}a&b\\c&d\end{pmatrix}$ have an inverse, $\begin{pmatrix}d&-b\\-c&a\end{pmatrix}$ in $SL(2,\mathbb{Z})$. This gives $SL(2,\mathbb{Z})$ the desired group structure.

Theorem. *Let $T=\begin{pmatrix}1&1\\0&1\end{pmatrix}$ and $S=\begin{pmatrix}0&-1\\1&0\end{pmatrix}$. Any matrix in $SL(2,\mathbb{Z})$ can be written as a product of positive powers of T and S.*

This is stronger than asking that T and S generate $SL(2,\mathbb{Z})$, so it implies it. The proof will use the fact that all elements of $SL(2,\mathbb{Z})$ are invertible, so they can be brought to the identity by a series of row operations. Then we will use the fact that $\mathbb{Z}$ is Euclidean, i.e. that there is a Euclidean division algorithm on $\mathbb{Z}$. As a reminder, this means that for any pair of integers α and β, there exists a list $[q_1,...,q_n,q_{n+1}]$ and $[r_1,...r_n]$ such that

$$\begin{aligned}\alpha &= q_1\beta+r_1\\ \beta &= q_2r_1+r_2\\ &\vdots\\ r_{n-2} &= q_nr_{n-1}+r_n\\ r_{n-1} &= q_{n+1}r_n,\end{aligned}$$

with $0\le r_1<\beta, 0\le r_i<r_{i-1}$ and $r_n=\gcd(\alpha,\beta)$.

Before we start, it is useful notice that multiplication on the right by S interchanges the columns of the matrix (changing signs to preserve the determinant), while multiplication on the left interchanges the rows.

Proof. If any $B^{-1}\in SL(2,\mathbb{Z})$ can be written as a product of powers of S and T, then since each $A\in SL(2,\mathbb{Z})$ is the inverse of $A^{-1}\in SL(2,\mathbb{Z})$, all A can be expressed as such a product. In the following, we denote the desired matrix by B and let its inverse be A. Writing A^{-1} as a product of S and T is just writing B as such. What we will do is equivalent to transforming A into the identity using a product of elementary matrices. The whole set of elementary operations, i.e. interchanging rows, multiplying by a non-zero scalar or multiplying a row and adding it to another, cannot be used because in the two first cases, the relevant elementary matrices are not in $SL(2,\mathbb{Z})$ (unless the scalar is one). So we must only allow the third kind of operation, together with multiplication by powers of S. For this we need the matrices $T^z=\begin{pmatrix}1&z\\0&1\end{pmatrix}$ and $T^{zt}=\begin{pmatrix}1&0\\z&1\end{pmatrix}$ for $z\in\mathbb{Z}$, M^t denoting the transpose of the matrix M. The following identities together with the fact that $T^n=\begin{pmatrix}1&n\\0&1\end{pmatrix}$ (and similarly for the other matrices below), show that these elementary matrices are products of powers of S and T.

- $S^{-1} = S^3$
- $T^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = S^2 \cdot STSTS$
- $T^{-1^t} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = S^2 \cdot STS$
- $T^t = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = TST.$

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Because $\det(A) = ad - bc = 1$, we know that $\gcd(a,b) = \gcd(a,c) = \gcd(b,c) = \gcd(c,d) = 1$ (if $x|a$ and $x|b$, then $x|(ad-bc) = 1$, thus $x = \pm 1$. The same argument holds for the other pairs.) First suppose that $c = 0$, then $A = \begin{pmatrix} \pm 1 & b \\ 0 & \pm 1 \end{pmatrix}$. Then $\begin{pmatrix} 1 & \mp b \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} \pm 1 & b \\ 0 & \pm 1 \end{pmatrix} = \pm I$. Since $S^2 = -I$ we are done.

Now suppose $c \neq 0$. Because $\mathbb{Z}$ is euclidean and $\gcd(a,c) = 1$, we have a list of integers $[q_1, \ldots, q_n, q_{n+1}]$ which gives us the above equations with last remainder $r_n = 1$. The following procedure is just the division algorithm carried out on a column of A.

$$T^{-q_1} A = \begin{pmatrix} r_1 & b_1 \\ c & d_1 \end{pmatrix}$$

where $r_1 = a - q_1 c$ is the first remainder and b_1 and d_1 are the appropriate new entries (note that $d_1 = d$, but we re-label for simplicity of notation). The rows remain relatively prime because the determinant of the two matrices is 1. Then

$$T^{-q_2 t} A = \begin{pmatrix} r_1 & b_2 \\ r_2 & d_2 \end{pmatrix}.$$

As above, $r_2 = c - q_2 r_1$. Because $\gcd(a,c) = 1$, this process must terminate with $(1,0)^t$ or $(0,1)^t$ in the first column of A. In the latter case, multiplication on the left by S^3 finally brings A to $A' = \begin{pmatrix} \pm 1 & b'_n \\ 0 & \pm 1 \end{pmatrix}$ (because $det(A') = 1$). From here it is obvious that multiplication by possibly S^2, and then T^{-b_n} gives the identity. □

The Upper-half plane

One of the interesting things about $SL(2,\mathbb{Z})$ is its group action on the upper-half plane $\mathcal{H}$. A *group action* of a group G on a set X is a function $G \times X \Rightarrow X$, denoted $g \cdot x = y$ for $g \in G, /x, y \in X$, satisfying both $g \cdot (h \cdot x) = (g \circ h) \cdot x$ and $i \cdot x = x$, where g, h are in G, i is the identity element of G and $\circ$ is the composition of the group. The action of $SL(2,\mathbb{Z})$ on $\mathcal{H}$ is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot (z) = \frac{az+b}{cz+d},\ z \in \mathcal{H}.$$

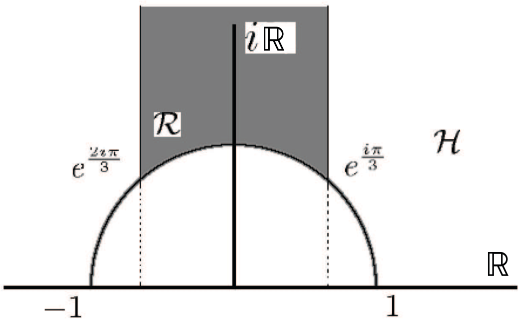
We need to mod out by $\pm I$ since $-IS \cdot z = S \cdot z$. We denote the "quotient" by $PSL(2,\mathbb{Z})$, often called the *modular group*.

Now, a *fundamental region* $\mathcal{R}$ for the action of $PSL(2,\mathbb{Z})$ in $\mathcal{H}$ is a region of $\mathcal{H}$ such that, for every $z \in \mathcal{H}$, there exists a unique $A \in PSL(2,\mathbb{Z})$ and $z' \in \mathcal{R}$ such that $A \cdot z' = z$. If we want to fix $\mathcal{R}$, the region

$$\mathcal{R} = \{z = x + yi | x^2 + y^2 \geq 1 \ \text{ and } \ |x| \leq \frac{1}{2}\},$$

with some restrictions on the boundaries, is the usual choice.

The interesting point, which you should look up, is that by applying elements of $PSL(2,\mathbb{Z})$ to both the vertices of $\mathcal{R}$ in $\mathcal{H}$ ($e^{\frac{\pi i}{3}}$ and $e^{\frac{2\pi i}{3}}$), and to ∞, which we view as the third vertex of the "triangle" $\mathcal{R}$, (defining $A \cdot \infty = \frac{a}{c}$), we get a tessellation of the upper-half plane, i.e. the upper-half plane is divided into hyperbolic triangles. Of course, since S and T generate the group, it is sufficient to do the exercise with powers of these two elements (try it with T to see how $\mathcal{R}$ is just translated throughout $\mathcal{H}$). You will notice that this extends $\mathcal{H}$ to include $\mathbb{Q}$. Accordingly, we call $\mathcal{H} \cup \infty \cup \mathbb{Q}$ the *extended upper-half plane*. It needs to be shown that this is indeed a tessellation, i.e. that it is *space-filling* and that there are no overlaps. I hope this is enough to spark your interest.



ISM

Institut des sciences mathématiques

The Institut des Sciences Mathématiques (ISM) is a consortium of seven Quebec universities (Concordia University, Laval University, McGill University, Université de Montréal, UQAM, UQTR and Université de Sherbrooke) whose three main aims are to:

1) Enhance research performance by integrating member researchers into ten inter-university scientific groups, hiring exceptional postdoctoral fellows, and organizing seminars and conferences.

2) Contribute to a top level graduate education by coordinating advanced Master's and Ph.D courses, and by encouraging excellence among graduate students and initiating gifted undergraduates to mathematical research through a variety of scholarships.

3) Promote and spread mathematical knowledge among teachers, young students and the general public by publishing and distributing the magazine *Accromath*, and by organizing conferences in cegeps.

Student Programmes:

- Inter-university ISM courses
- ISM Student Seminar
- ISM Quebec Student Conference
- CRM-ISM Mathematics Colloquium
- CRM-ISM-GERAD Statistics Colloquium
- ISM Scholarships
- ISM Travel Bursaries
- Undergraduate Research Scholarships

CONTACT INFO

Postal address
Institut des sciences mathématiques
Case postale 8888, succursale Centre-Ville
Montréal (Québec) H3C 3P8 Canada

Civic address
Pavillon Président-Kennedy
201, Avenue du Président Kennedy
Office PK-5211

Telephone: (514) 987-3000, x 1811
Fax: (514) 987-8935
Email : ism@math.uqam.ca
www.math.uqam.ca/ism

Mathematical digest

Michael McBreen

The mathematical digest is a collection of articles on undergraduate course material. The articles aim to give an intuitive understanding of the material, not to prove anything rigorously. Enjoy.

The Fundamental Theorem of Calculus

The FTC states that

$$\int_a^b f(x)dx = F(b) - F(a)$$

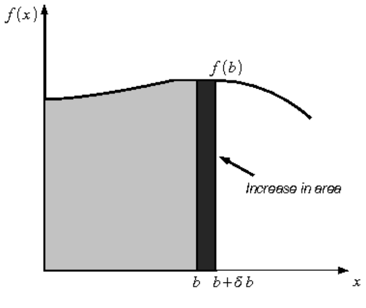
where $\frac{dF(x)}{dx} = f(x)$. We can also write this as

$$\frac{d}{db}\int_a^b f(x)dx = f(b)$$
$$\frac{d}{da}\int_a^b f(x)dx = -f(a)$$

We're not going to prove the FTC here. We're going to show why it makes sense. We only consider the positive $f(x)$ case for simplicity. Consider the area under a curve $f(x)$ bounded by lines at $x = a$ and $x = b$, or $\int_a^b f(x)dx$. If we keep a fixed, the area is obviously a function of b, which we'll call **Integral(b)**. As we increase b, the right line moves along the axis and the area increases.

The FTC tells us what the derivative of **Integral(b)** is. In other words, how fast does **Integral(b)** increase when we increase b? More precisely, how many times faster (or slower) than b does it increase?

Well, the bigger $f(b)$ is, the faster it increases. In fact, if you increase b by an infinitesimal amount δb, you're increasing the area by adding a rectangle[9] with height $f(b)$ and width δb, as shown in the following figure:



Symbolically,

$$\text{Integral}(b + \delta b) = \text{Integral}(b) + f(b)\delta b$$

Of course, this means that **Integral(b)** increases $f(b)$ times faster than b, or in symbolic language,

$$\frac{d}{db}\int_a^b f(x)dx = f(b)$$

And that's the FTC. Well, that's half of it. Hopefully, you can explain the other half, concerning $F(a)$, to yourself - exactly the same reasoning applies.

Linear Differential Equations and the Minimal Polynomial

We're taught in ODE how to solve equations of the form

$$P(D)y = 0$$

where $P(D)$ is a non-trivial polynomial in the linear differential operator $D = \frac{d}{dx}$. Our aim is to find the solution space $V = \ker(P(D))$ of all functions annihilated by $P(D)$. V happens to be a vector space, because if y and z are solutions, then so are $y - z$ and λy for $\lambda \in \mathbb{C}$. To find V, we simply factor $P(D)$ into mutually prime components $p_i(D)$ such as $(a + D)^m$, find the solution space $\ker(p_i(D))$ of each individual factor and then take V to be the direct sum of all the individual solution spaces: $V = ker(p_1(A)) \oplus ker(p_2(A)) \oplus \ldots \oplus ker(p_n(A))$.

Why can we do this? Recall the Primary Decomposition Theorem from Linear Algebra: *Let $P(A)$ be the minimal polynomial of a linear operator A acting on a finite dimensional vector space W. Let $P(A)$ be the product of mutually prime factors $p_1(A)p_2(A)\cdots p_n(A)$. Then $W = ker(p_1(A)) \oplus ker(p_2(A)) \oplus \ldots \oplus ker(p_n(A))$*

Clearly, if we can show that V is finite dimensional and that $P(D) = \lambda M(D)$ where $M(D)$ is the minimal polynomial of D over V,

[9] It's really a trapezium, but from the picture you can see that the difference in area is negligible.

then the two situations will be completely equivalent.

To show $dim(V) < \infty$, recall that if $P(D)$ is of degree n, then we can reduce the equation

$$P(D)y = 0$$

to a system of n first order equations which we write concisely as

$$D\vec{y} = \vec{F}(\vec{y}, x)$$

where $\vec{y}$ is the n-component vector of unknown functions of x. Given an initial vector $\vec{y}_0$, this equation obviously has a single solution. In other words, n initial conditions (the n components of $\vec{y}_0$) suffice to determine a solution completely. If $dim(V) = \infty$, then n initial conditions would be insufficient to pick out a single function, so $dim(V)$ must be finite.

To show $P(D) = \lambda M(D)$, note that $P(D)$ annihilates everything in V, so $M(D)$ must divide $P(D)$. We just have to show that if $P(D) = Q(D)M(D)$ with deg$Q(D) > 0$, then $P(D)$ annihilates some functions that $M(D)$ leaves standing, which contradicts $M(D)$ being the minimal polynomial of D for $V = \ker(P(D))$. This can easily be proven using the fundamental existence and uniqueness theorem; I won't do it here.

And that's why we can do what we do in ODE.

The first isomorphism theorems for rings and groups

If you look at the FIT the right way, it seems almost trivial. First, the claim for rings: *The image of a surjective homomorphism $\rho : R \to K$ is isomorphic to $R/\ker(\rho)$.*

Now, the explanation. Consider a finite ring. Any finite ring is completely described by its addition and multiplication tables. Since an isomorphism preserves these tables, the *only* thing it can do to a ring is change the names of the elements (i.e. $\mathbf{1} \to \mathbf{a}\,, \mathbf{2} \to \mathbf{b}\,, \mathbf{3} \to \mathbf{c}\,, \mathbf{4} \to \mathbf{d}$,etc). Since an isomorphism is reversible, it must give distinct names to distinct elements. It's simply a name switch. Now, consider a surjective homomorphism. What can this homomorphism do that an isomorphism can't? It still can't actually modify the + and × tables, but it can *forget* that certain elements are distinct: it can give the same name to distinct elements. There are certain obvious consistency requirements: if you forget that **a** and **b** are distinct, then **a - b** must be sent to **0** (i.e. you must also forget that **a - b** and **0** are distinct). Likewise, if you send **a - b** to **0**, then **a** and **b** must be merged together. Finally, if you send **a** to **0**, you have to send **ab** to **0** as well. These consistency requirements imply that what your homomorphism "forgets" is *entirely determined* by what it sends to **0**, i.e. by its kernel, i.e. your whole homomorphism is determined by its kernel. Now, where else have we taken a ring, sent some of its elements to **0** and then studied the result? You will kindly recall that we do the exact same thing when we take the quotient of a ring by an ideal. Note that the conditions that define an ideal correspond precisely to the consistency requirements on our homomorphism. For instance,

$$\mathbf{a} \in \mathbf{I} \to \mathbf{ba} \in \mathbf{I}$$

simply means that if you send **a** to **0**, then you must also send **ba** to **0**. Once we've specified a kernel for our homomorphism, i.e. which ideal we'll be modding out with, we have only one freedom left: we can name the resulting elements as we wish (provided we give distinct names to distinct elements in the target ring-we've already chosen which elements we want to merge). That's why we say that the target ring is identical to the quotient ring "up to an isomorphism" (a name switch). And that, essentially, is the first isomorphism theorem for rings. In the group case, there's only one major difference: the "kernel" is the set mapped to **1** and not **0**. This is why the consistency requirements on a normal subgroup are different from those on an ideal. For instance,

$$\mathbf{g} \in \mathbf{N} \to \mathbf{hgh^{-1}} \in \mathbf{N}$$

means (roughly) that if you send **g** to **1**, you must also send $\mathbf{hgh^{-1}}$ to **1** , which makes perfect sense.

Vanilla Taylor Series

This article is about the *meaning* of Taylor series, strange as it may seem. For a more conventional proof of the formula, see the box at the end of the article. We start with some context.

Picture yourself gazing wistfully through your window at a narrow stretch of road below, with your chronometer on (Why? That's not for me to say). A car passes by at time $t = 100$ seconds sharp, and then disappears from view.

Now, the question is: how far from the house will the car be at $t = 105$ sec? How far was it at 95 sec? In other words, what is $f(t) = d$, the function that gives car distance d with respect to time t? Using the magic of Taylor series, we can approximate $f(t)$ step by step by replacing it with simpler functions $P_n(t)$ – *polynomials* in t, more precisely.

1. 0^{th} **order**[10] **approximation in** t **:** We can say that 5 sec later, the car will still be roughly in front of the house (at $d = 0$ m). In other words,

$$P_0(t) = f(100) = 0.$$

Not so great.

2. 1^{st} **order approximation in** t **:** Knowing the car's speed $f'(t)$ as it passed (at $t = 100$ sec), we can assume it keeps that speed and let

$$P_1(t) = f(100) + [f'(100)](t - 100)$$

This is better. Notice that we must subtract 100 sec from t to get the actual time the car spent driving away from us.

3. 2^{nd} **order approximation in** t **:** What if the car is accelerating? In other words, what if $f''(100) \neq 0$? In that case, we can let

$$\begin{aligned} P_2(t) = f(100) &+ [f'(100)](t - 100) \\ &+ \frac{1}{2}[f''(100)](t - 100)^2 \end{aligned}$$

What do the factor of $\frac{1}{2}$ and the square mean?[11] The square means that at large times (eg. 1000 sec), your rate of acceleration will have a lot more impact than your initial speed. A car may zoom past your window while a bus crawls slowly by, but if the car is constantly slowing down while the bus is constantly accelerating, the bus will eventually leave the car far behind. As for the $\frac{1}{2}$ factor, it (very, very roughly) means that in the short term, speed is more important than acceleration.

4. 3^{rd} **order approximation in** t **:** Here we depart from high-school physics. If the car is accelerating faster and faster, we can take it into account by setting

$$\begin{aligned} P_3(t) = f(100) &+ [f'(100)](t - 100) \\ &+ \frac{1}{2}[f''(100)](t - 100)^2 \\ &+ \frac{1}{6}[f'''(100)](t - 100)^3 \end{aligned}$$

The meaning of the cube and $\frac{1}{6}$ factor are much the same as above.

5. n^{th} **order approximation in** t **:** We can continue this process forever, taking into account ever higher order rates of change. Each successive $P_n(t)$ is a polynomial in t, with coefficients determined by successive derivatives of $f(t)$ at 100 sec. Higher derivatives correspond to more removed forms of "acceleration" and hence are multiplied by smaller and smaller fractions. If you don't see why, think of it like this: imagine three cars at the starting line of a race. One starts at speed $200km/h$, the second at acceleration $200km/h^2$ and speed 0, and the third starts with 0 speed, acceleration 0 and rate of increase of acceleration $200km/h^3$. In the short term, the first car will lead. After a while the second car will have gained enough speed to overtake the first, and will eventually leave it far behind. But later still, the third car will lead, because its acceleration is always increasing. The higher the order of the rate of change, the later the effect.

The formula for the n^{th} order approximation, should you really want it, is:

$$\begin{aligned} n^{th}\text{order } P_n(t) = f(100) &+ [f'(100)](t \\ -100) &+ \frac{1}{2}[f''(100)](t - 100)^2 \\ &+ \ldots + \frac{1}{N!}[f^{(N)}(100)](t - 100)^N \end{aligned} \tag{0.6}$$

If we're lucky[12], as we let n go to infinity, $P_n(t)$ should accurately give us the car's position at any time. We call this $P_\infty(t)$ the *Taylor series* of $f(t)$ around $t = 100$ sec. Something could go wrong, though. Let's see what.

Possible Failure and Analytic Functions

First off, what if the driver does something unexpected? We only have information about

[10] The order, in this case, is simply the degree n of the polynomial $P_n(t)$.

[11] You can explain the square by noting that acceleration has dimensions of m/s^2 (if you believe the physicists), but it's not a particularly intuitive reason.

[12] I'll explain what this means in a moment.

what he's doing as the car passes our window. If he chooses to stop at Roarin' Willy's Roadside Pub 15 seconds later, our predictions fail: no amount of Taylor series can save him. In fact, if our Taylor series really does predict what the driver will do in the far future, we have a very strange driver on our hands. The $f(t)$ of this predictable driver is called an *analytic* function: its value at any point is fully determined by its behaviour in some small region.

Textbook Taylor Series

Here's the usual proof that *if* $f(x)$ has a power series expansion around $x = c$ (i.e. can be expressed as an "infinite polynomial" in $(x - c)$), then that power series is given by the Taylor series

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''}{2}(c)(x-c)^2 + \ldots + \frac{f^{(n)}}{n!}(c)(x-c)^n + \ldots$$

Let $f(x)$ be expressed by the following power series:

$$f(x) = a_0 + a_1(x-c) + a_2(x-c)^2 + \ldots + a_3(x-c)^n + \ldots \tag{0.7}$$

Set $x = c$ to get $f(c) = a_0$.

Then, differentiate (0.7) once and set $x = c$ to get $f'(c) = a_1$ In general, differentiate (0.7) n times and set $x = c$ to get $f^{(n)}(c) = n!a_n$ or

$$a_n = \frac{f^{(n)}(c)}{n!}$$

which completes the proof. To prove that some function $f(x)$ does have a power series expansion is more involved, and we won't go into it here.

However, even if our driver does act predictably, our predictions could yield infinite quantities (gibberish) past a given time (eg. 200 sec). The causes of that failure, non-convergence, are beyond the reach of this article.

The Change of Variables Formula

Before we begin: infinitesimals. I'm going to throw a lot of infinitesimals around: $dx, dy, d\theta$ and so on. If you like, you can think of them as extremely small numbers. They're meant to help your intuition along, not to be elements of a rigorous proof, so don't worry too much about the fine print. If the change of variables formula makes sense to you by the end of this article, then all is well.

We use the change of variables formula (CoVF) when we're integrating some function of 3-space, and wish to go from Cartesian coordinates (x, y, z) to spherical coordinates (ρ, θ, ϕ), for instance. We're going to stick to that example for the rest of the article, but it should be obvious that nothing I say will be specific to these two coordinate systems or to 3D space.

Now, the CoVF tells us that when we change coordinates, we have to multiply our integrand by $|\det J_{C\to S}(\rho,\theta,\phi)|$ where $J_{C\to S}(\rho,\theta,\phi)$ is the Jacobian matrix of our coordinate change (I write $C \to S$ for "Cartesian to Spherical"):

$$\int_V f(x,y,z)\,dx\,dy\,dz = \\ = \int_{V'} f(r,\theta,\phi)\ |\det J_{C\to S}(\rho,\theta,\phi)|\ dr d\theta d\phi$$

where

$$J_{C\to S}(r,\theta,\phi) = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{bmatrix}$$

Why does this Jacobian factor arise? We're integrating over a volume, so the integrand $f(x, y, z)$ can be thought of as the "density" of some substance (call it *gob*). When we integrate, we divide our integration region into tiny boxes with sides dx, dy and dz. You build such a box by choosing a point (x, y, z), varying x by dx, y by dy and z by dz and taking the volume you've just "swept out" to be your box. This box contains an amount $f(x,y,z)\,dx\,dy\,dz = f(x,y,z)\,dV$ of gob, and the integral sums the contributions from all boxes to give the total gob in the region.

Now, let's change variables to ρ, θ and ϕ. We get our new boxes by varying ρ by $d\rho$, θ by $d\theta$ and ϕ by $d\phi$. The problem is that these boxes will no longer all have the same volume. If you make the same small variations at a point with small ρ and at a point with large ρ, your second box will be bigger, and chances are that neither will be the size of the Cartesian boxes they're replacing. Clearly, we need a function that gives the change in volume of the new boxes with respect to the old boxes. We're going to show that this function is the Jacobian $J_{C\to S}(\rho,\theta,\phi)$.

The first thing to note is that $J_{C\to S}(\rho,\theta,\phi)$ is a coordinate change matrix: it takes an infinitesimal vector with spherical coordinates and re-expresses it in Cartesian coordinates. To see this, stare hard at the following equation:

$$\begin{aligned} J_{C\to S}(\rho,\theta,\phi)\begin{pmatrix} d\rho \\ d\theta \\ d\phi \end{pmatrix} &= \\ &= \begin{bmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{bmatrix}\begin{pmatrix} d\rho \\ d\theta \\ d\phi \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial x}{\partial \rho}d\rho + \frac{\partial x}{\partial \theta}d\theta + \frac{\partial x}{\partial \phi}d\phi \\ \frac{\partial y}{\partial \rho}d\rho + \frac{\partial y}{\partial \theta}d\theta + \frac{\partial y}{\partial \phi}d\phi \\ \frac{\partial z}{\partial \rho}d\rho + \frac{\partial z}{\partial \theta}d\theta + \frac{\partial z}{\partial \phi}d\phi \end{pmatrix} \\ &= \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} \end{aligned}$$

The last equality follows from the chain rule (or simply from the fact that a smooth function - such as $x(\rho,\theta,\phi)$ - varies linearly when we make infinitesimal changes in its arguments). We need to show that when this coordinate change matrix J acts on a box with sides dx, dy, dz, the box's volume increases by a factor $|\det J|$. We can use the following theorem

Theorem. *Polar Decomposition Theorem Any real or complex matrix **S** can be expressed as a product of the self-adjoint operator $\sqrt{S^*S}$ and some isometry*[13] *U :*

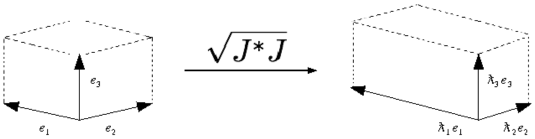
$$S = U\sqrt{S^*S}$$

I will only sketch the proof and leave the details to you. First note that

$$\begin{aligned} \|\sqrt{S^*S}v\|^2 &= \langle\sqrt{S^*S}v, \sqrt{S^*S}v\rangle = \langle S^*Sv, v\rangle \\ &= \langle Sv, Sv\rangle = \|Sv\|^2 \end{aligned} \tag{0.8}$$

Define $U' : \operatorname{Im}(\sqrt{S^*S}) \to \operatorname{Im}(S)$ by $U'(\sqrt{S^*S}v) = Sv$. Using (0.8), you can check that U' is in fact an isometry. We can easily extend U' to an isometry U of the full vector space. This proves the theorem.

We can therefore write $J = U\sqrt{J^*J}$ for some U. Since U is an isometry (analogous to a rotation), it obviously preserves volumes. Hence, $\text{Volume}(J\,dV) = \text{Volume}(U\sqrt{J^*J}\,dV) = \text{Volume}(\sqrt{J^*J}\,dV)$. But since $\sqrt{J^*J}$ is self-adjoint (as you should see), the real spectral theorem tells us that it has a basis of orthonormal eigenvectors! Such an operator expands or contracts volumes by a factor equal to the product of its eigenvalues $\lambda_1\lambda_2\ldots\lambda_n$, or more precisely $|\lambda_1\lambda_2\ldots\lambda_n|$ (since a negative factor simply means certain directions have been reversed). Look at the following figure to see why.



Now, the determinant of a diagonal matrix is simply the product of its eigenvalues, so we have

$$\begin{aligned} |\lambda_1\lambda_2\ldots\lambda_n| &= |\det\sqrt{J^*J}| = \sqrt{\det(J^*J)} \\ &= \sqrt{\det(J^*)\det(J)} \\ &= \sqrt{(\det J)^*\det(J)} \\ &= |\det J| \end{aligned} \tag{0.9}$$

Equation (0.9) gives precisely the change of variables formula.

[13] An isometry is a linear operator that preserves the length of vectors or more generally the inner product of two vectors. Rotations and reflections are isometries acting on $\mathbb{R}^3$.

The Birch and Swinnerton-Dyer Conjecture: An Interview with Professor Henri Darmon

Agnès F. Beaudry

> *If you made a poll of number theorists and asked them, "What's your favorite problem in number theory?", you would probably have it equally divided between the Riemann Hypothesis and the Birch and Swinnerton-Dyer conjecture, which are two of the Millennium Prize problems in number theory. My favorite problem is the Birch and Swinnerton-Dyer conjecture.*
>
> -Prof. Henri Darmon

The Birch and Swinnerton-Dyer conjecture (BSD) was stated in the sixties by Peter Swinnerton-Dyer and Bryan Birch who gathered considerable evidence, based on numerical data from the EDSAC computer at Cambridge, suggesting a relation between the rational solutions of special Diophantine equations called *elliptic curves* and their solutions in $\mathbb{Z}_p$ for different prime values. In the early eighties, the work of Victor Kolyvagin, Benedict H. Gross and Don Zagier, combined with that of Andrew Wiles finally created tools to approach the previously obscure problem, eventually bringing it to the forefront, the Millennium prize stamping it as one of the most important problems of the century. prof. Henri Darmon has been working on this problem. We met with him to try to understand what the BSD really means and perhaps get a glimpse of his work on the problem.

Elliptic curves and projective models

In the official description by Andrew Wiles, the BSD is described as a relation between the *L-function* of an *elliptic curve*, terms I will clarify below, and its *rank*, a number that, to some extent, measures the size of the set of rational solutions (solutions in $\mathbb{Q}$) of that elliptic curve. Pr. Darmon explains:

"It started without involving L-functions at all, these are just part of the baggage that you need to make this conjecture very precise. But the idea is simple. You start with an an elliptic curve

$$E = y^2 = x^3 + ax + b,\ a, b \in \mathbb{Q}.$$

The set of rational solutions of these equations, $E(\mathbb{Q})$, has a very nice structure, an abelian group law. To obtain this group, one needs to look at the projective model of the equation."

The projective model of an equation is obtained by adding an extra variable, z, in order to transform it into a homogenous equation of degree three:

$$y^2z = x^3 + axz^2 + bz^3.$$

This equation has a trivial solution, $(x, y, z) = (0, 0, 0)$, which we ignore. Also, if (x, y, z) is a solution to this equation, then so is $(\lambda x, \lambda y, \lambda z)$. We let two solutions be equivalent if they differ by a non-zero scalar.

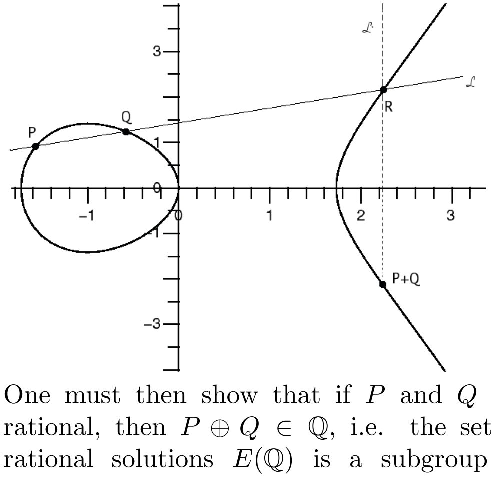
There are two possibilities, either $z \neq 0$ or $z = 0$. If $z \neq 0$ for a solution $P = (x, y, z)$, then P is equivalent to a solution $(x', y', 1)$, because we can multiply P by z^{-1}. This solves the original equation, thus we get a bijection between the solutions of the projective model with $z \neq 0$, and the solutions original elliptic curve called the *affine model*. On the other hand, if $z = 0$, things get interesting: "Here we get new solutions which are of the form $Q = (x, y, 0)$. If we substitute $z = 0$ into the equation, it becomes $x^3 = 0$, and x also has to be zero. Therefore $Q = (0, y, 0)$, where $y \neq 0$ (since we are not allowing the zero solution), which is equivalent to $(0, 1, 0)$. We therefore have this new solution called the *point at infinity* of the projective model, which did not appear in the affine model.

THE GROUP LAW

To find the group law on the set of solutions, we first look at the given elliptic curve

$$E = y^2 = x^3 + ax + b,\ a, b \in \mathbb{Q}$$

over $\mathbb{C}$. Given two points P and Q on E and the line determined by these two points, $L = y = mx + n$, L must intersect E at another point R since $\mathbb{C}$ is algebraically closed (and the intersection $(mx+n)^2 = x^3 + ax + b$ is of degree three.) Note that these three points need not be distinct, i.e. we count multiplicities. Now we draw the line L' connecting R and ∞ (which lies on E since we are looking at the projective model.) This line intersects E at a third point, $P \oplus Q$, and the operation $\oplus$ thus defined is the composition law of the group with the identity element $\infty = (0, 1, 0)$. It satisfies all the axioms of an abelian group. Strangely enough, the hard thing to verify is the associativity property.



One must then show that if P and Q are rational, then $P \oplus Q \in \mathbb{Q}$, i.e. the set of rational solutions $E(\mathbb{Q})$ is a subgroup of $E(\mathbb{C})$. To verify this, you can write the equations of the E, L and L' and verify that the intersections must be in $\mathbb{Q}$. For more details, see [1].

That extra point is a distinguished point, which plays the role of the identity element for the addition law of the group of rational solutions. That's why, for the elliptic curve, we always consider the projective model. The hard thing to show is that this group is finitely generated, it's that finiteness result."

The rank and its role in the BSD

That $E(\mathbb{Q})$ is finitely generated was proved by Louis Mordell in 1922. It implies that

$$E(\mathbb{Q}) \simeq \mathbb{Z}^r \oplus T,$$

where T is the torsion part of the group, i.e. the set of elements of $E(\mathbb{Q})$ with finite order, and r is the smallest number of elements needed to generate the non-torsion part of $E(\mathbb{Q})$. We call r the *rank*. It's similar to the dimension of a vector space: it measures the size of the space of solutions. "What you want to be able to compute is this rank, how many solutions you need to generate all the other ones by repeated applications of these group laws." The BSD proposes an answer to this question.

Now, Diophantine equations in $\mathbb{Q}$ are not as malleable as solutions in certain other fields, so one hopes to find tools in these nicer fields to get to the rational solutions:
"You have an equation, you try to understand it by studying its complex solutions, its real solutions maybe, and its solutions over a finite field. These are very easy solutions, much less subtle than those over $\mathbb{Q}$. Over $\mathbb{C}$ you get this nice surface with topological invariants, and somehow you just care about the shape. Over the finite fields you have the finite sets, the cardinality. Then you'd like to understand the solutions over $\mathbb{Q}$, and the principles that allow you to derive this information are very deep. This is the main philosophical question in the study of Diophantine equation, this passage from understanding the behaviour of the solutions in finite fields to their behaviour over $\mathbb{Q}$.

Birch and Swinnerton-Dyer's first insight was thinking that maybe you could get at the rank by counting the number of solutions over $\mathbb{Z}_p$. So they defined N_p to be the number of (x, y) where $p | f(x, y)$, in other words, where $f(x, y) \equiv 0 \pmod{p}$. A priori, that has nothing to do with it, because here you are looking at the solutions over $\mathbb{Q}$ and there you are looking at the corresponding congruence classes. Their insight was that if r is big, then you get a systematic contribution to $E(\mathbb{Z}_p)$, or the solutions over $\mathbb{Q}$ reduced modulo p.

Take the solutions over $\mathbb{Q}$. If $r = 0$, then you have essentially no solutions, i.e. finitely many, but if $r = 1, 2$, then you have many, many rational solutions. You take these rational solutions (x, y), (supposing that p does not divide the denominator of x and y) and reduce them mod p. If r is large, then the number of solutions over the different $\mathbb{Z}_p$ should have a tendency to be large. What Birch and Swinnerton-Dyer did is they tried to make that precise and quantitative: they looked at the product of N_p over p.

Roughly, how big is N_p? You can let x range from 0 to $p - 1$. For every value of x, you get a number modulo p. You're asking whether this equation, $y^2 = x^3 + ax + b$ has a solution in $\mathbb{Z}_p$. You're asking whether this number is a square or not (a quadratic residue or non-residue). Half of the elements in $\mathbb{Z}_p$ are squares, and the other half are not. Thus when you run over all the values of x, roughly one out of two times you are going to win, you'll get two solutions, the one and its negative, i.e. if (x, y) is a solution to $Y^2 = X^3 + aX + b$ then so is $(x, -y)$. The other half of the time you'll loose, you'll get no solutions. On average you expect to get roughly p solutions, but there's an "error term" defined to be

$$a_p = 1 + p - N_p.$$

It turns out that $a_p < p$ since in the worst of cases, for every trial you'll get two points. It can actually be shown that $|a_p|$ is at most $2\sqrt{p}$.

Birch and Swinnerton-Dyer looked at this product over the primes less than a given χ. They observed that this quantity grows roughly as a constant that depends on E, times a power of $\log \chi$, and that the exponent seems to be the rank.

$$\prod_{p<\chi} \frac{N_p}{p} \sim C_E (\log \chi)^r$$

In particular, they found that the product seems to be bounded when $r = 0$, in other words when $E(\mathbb{Q})$ is finite. When $r = 1$, it appears to grow as $\log \chi$, when $r = 2$ it appears to grow like $\log \chi^2$ and so on. This means that if you know everything there is to know about this equation over the congruence equations, then you will know something about the equation over $\mathbb{Q}$. This was found experimentally, and that's the BSD, that's really what it's about. It's about relating the solutions over $\mathbb{Q}$ to the solutions of the corresponding congruence equations. If you look at finitely many of these N_p, you're not going to be able to tell what the rank is, but here you're looking at the asymptotic of these finite products, and in the asymptotic the N_p do know about the rank."

L-functions

To attack the problem, mathematicians introduced analytical tools to study the behaviour of the solutions over the $\mathbb{Z}_p$. That's where the L-functions come in. There's no real definition of a general L-funtion. Roughly, L-functions are infinite products indexed by the primes, but they can't be any such product, they have to be somehow "natural". The best way to describe them is to give examples. Historically, the first L-function was the Riemann zeta-function.

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{0.1}$$

"The main result about the Riemann zeta-function is that it factors into a product over all the primes."

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1} \tag{0.2}$$

This equality, together with the fact that (0.1) converges for $s > 1$ is a restatement of the unique factorization theorem.

"This was proven by Euler, the first to have given a factorization formula for an L-function. These factors, which are indexed by the primes, are therefore called Euler factors. For Diophantine equations it's the same idea as for the Riemann zeta-function, you look at solutions over various finite fields and you package together this information by making a generating series.

There is a very concrete recipe which takes N_p and converts it into some polynomial. The interesting case is when we have an elliptic curve equation. Here the polynomial is defined to be $P_p(x) = 1 - a_p x + p x^2$, with the L-function defined as

$$L(E, s) = \prod_p P_p(p^{-s}),$$

(convergent for $s > 3/2$). This definition of the L-function is the object of interest in the BSD. The very deep fact, which was proved by Wiles, is that it actually extends to an analytic function of the complex variable s over the entire complex plane."

THE ζ-FUNCTION AND UPF

$$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_{p}(1-p^{-s})^{-1} \tag{0.3}$$

if and only if every $n \in \mathbb{N}$ has a unique prime factorization.

Proof. First assume unique prime factorization (UPF). We can expand each factor in the left of (0.3) into a geometric series

$$\frac{1}{1-p^{-s}}=1+p^{-s}+p^{-2s}+\ldots+p^{-ns}+\ldots$$

Let p_i be the primes. If we expand the product in (0.2), we get exactly one term $(p_1^{m_1}p_2^{m_2}p_3^{m_3}\cdots p_k^{m_k}\ldots)^{-s}$ for each list of positive integers m_j with a finite number of nonzero entries. Unique prime factorization tells us that every $n \in \mathbb{N}$ can be expressed as one of these combinations in one and only one way. Also, every such combination corresponds to an integer. Therefore there is exactly on term $\frac{1}{n^s}$ for each n in $\mathbb{N}$ on the left of (0.3), and the equality follows.

Now suppose that (0.3) holds. Then expanding the product as above, we get that $\prod_p(1-p^{-s})^{-1}=\sum_{n=1}^{\infty}\frac{f_n}{n^s}$ where f_n denotes the number of distinct prime factorizations for n. Therefore, when the sum converges, i.e. when $s>0$, we can write (0.3) as $T(s)=\sum_{n=1}^{\infty}\frac{a_n}{n^s}\equiv 0$ with $a_n=1-f_n$. Letting $s\to\infty$, $T(s)=a_1+\frac{a_2}{2^s}+\ldots=a_1$. Thus $a_1=0$ and $T(s)=\frac{a_2}{2^s}+\frac{a_3}{3^s}+\ldots=0$. A simple inductive argument shows that $a_n=0$ for all n, and therefore $f_n=1$. This proves the theorem. □

The Conjecture and the Corollary

With this fact, the BSD can be formulated in its current form:

Conjecture. *For an elliptic curve E, the Taylor expansion of the L-function $L(E,s)$, $s\in\mathbb{C}$, around $s=1$ is*

$$L(E,s)=c(s-1)^r+\ \textit{higher order terms}$$

where r is the rank of $E(\mathbb{Q})$ and $c\neq 0$.

From this we get the immediate corollary that

$$L(E,1)=0 \Leftrightarrow E(\mathbb{Q})\ \text{ is infinite,}$$

because if $E(\mathbb{Q})$ is finite (i.e. $r=0$), and the BSD implies that $L(E,1)=c$. Conversely, if $E(\mathbb{Q})$ is infinite, then $r>0$ and all terms of the Taylor expansion vanish.

"This is one of the most striking parts of the BSD, because it's actually easy to test numerically whether $L(E,1)$ is 0 or not. Also, there *is* a theorem (independent of the conjecture) saying that if $L(E,s)\neq 0$, the group of solutions $E(\mathbb{Q})$ is finite. The converse is really, really exciting, since with a computer, it's easy to compute $L(E,1)$ and see whether it's zero or not. If it's zero, this implies that there is a solution. So even if it's really hard to compute it, you know it's there, and that would already be an amazing result. Just that implication would already be amazing."

It is because of the proved implication that prof. Darmon began working on the BSD: "My advisor at Harvard, Benidict H. Gross, proved part of the first implication. Then there was a Russian mathematician called Victor Kolyvagin who brought another piece of the puzzle. The combination of these two results, Gross and Kolyvagin, proved that implication, called the "easy" implication. That was around my second year of graduate school and was very exciting. It brought the BSD at the forefront. Suddenly, there were tools and techniques to explore it. Then, when I was a post doc at Princeton, Wiles proved Fermat's Last Theorem by showing that the L-function had the analytic continuation. That was another big piece of the puzzle for the BSD. When you're a graduate you always try to gravitate towards the areas where there's a lot of activity, because there's a lot of action."

His Work on the BSD

When asked about his contribution to the BSD problem, prof. Darmon laughed and explained his work:

"It's a bit technical. There are all kinds of variants and generalizations for either cases of the BSD. The case I proved has to do with elliptic curves which are defined over number fields. Instead of looking at $\mathbb{Q}$ you can look at extensions of the rational numbers. The number fields I was considering were quadratic fields, i.e. the extension of the rationals by $\sqrt{d}$ for some $d\in\mathbb{Z}$. You can have an elliptic curve over that field and look at the N_p's. They're indexed by the primes of that field rather than the primes of $\mathbb{Q}$. You can do exactly the same thing and make the same conjecture. For quadratic number fields, it's even more mysterious. Even that "easy" implication is not completely understood, but there were certain cases where we were able to prove it."

I asked prof. Darmon what draws him to the BSD:

"I think it's always important to work on problems that are very central and have a lot of mystery. The important thing about the BSD is that even the major ideas have been found in the last twenty years and there's still a lot of uncertainty in the conjecture, it's by no means a done deal. When you're doing research, you always try to gravitate towards problems that have this mystery. Even if you can't prove the Riemann Hypothesis, even if you can't prove the BSD, you keep being led to rich and useful results. It's a very motivating idea, because you feel like you're really digging into something that we absolutely don't understand.

However, you want to make a compromise between that and not working on a problem which hasn't been understood at all and where absolutely no progress has been made. There are a lot of questions in number theory of that sort, we can come up with all sorts of questions that no one has any idea how to solve, where no structure has ever been approached that would say anything about the problem. So one tries to avoid those also. The BSD is a great problem: on the one hand, it's a fundamental mystery and at the same time there are all kinds of incredibly rich structures that have been developed to say something about it. All the fundamental concepts of number theory have been used here at some time or another.

Of course, how do you determine what's an important problem? It comes from experience. When a problem seems to be very difficult, and at the same time seems to create a lot of interesting mathematics, then you get this feeling that it's a fundamental problem. There's a feeling about the BSD that if we understood it, if we were able to prove it or even prove some special case, then that would lead to a lot of progress, we would understand a lot of other things as well. It's like the Riemann hypothesis, to which it is related by the L-functions. The BSD tries to link two objects that seem to be part of different worlds. One of them is an analytically defined object, which is the L-function, the other is an algebraically defined object which is the elliptic curve and the solutions over $\mathbb{Q}$. We don't really know how to make a bridge between these two and if we understood that mechanism, I think the insight that would emerge would say something about the Riemann Hypothesis. Working on the BSD or the Riemann Hypothesis, one is bound to find something very rich and intricate."

References

[1] Silverman JH. The Arithmetic of Elliptic Curves, Springer, 1985

A Field of Six Elements?

Agnès F. Beaudry

Have you ever tried to find a field of six elements? Well now is a good time to stop searching, because I will show here that there is no such field.

First, I will prove that any field contains a subfield which is isomorphic to $\mathbb{Q}$ or $\mathbb{Z}_p$ for a prime p. Then, recall from your algebra class that any field K containing another field F can be viewed as an F-vector space, i.e. we can see K as vector space over F. We will use this fact to prove that no field can have six elements.

Theorem. *Any field K contains a subfield F isomorphic to either $\mathbb{Q}$ or $\mathbb{Z}_p$.*

Proof. I'll sketch the proof and leave the details to you. You can map $\mathbb{N}^0 = \mathbb{N} \cup \{0\}$ into the field K through the function $\phi : \mathbb{N}^0 \to K$, $\phi(n) = n \cdot 1$ where $n \cdot 1$ means addition of the multiplicative identity n times with itself in K. Two things can happen, either $\phi(n) = 0$ if and only if $n = 0$, or there exists an n in $\mathbb{N}$ such that $\phi(n) = 0$. In the first case, we can see that there is a subring of K isomorphic to $\mathbb{Z}$. The field generated by this subring must also be in K and is isomorphic to $\mathbb{Q}$. I'll let you think about this and move on.

If the second case holds, then take the smallest such n. If n was not prime, i.e. $n = qs$ for neither $q, s \neq 1$, then the elements $q \cdot 1$ and $c \cdot 1$ would be zero divisors. You can figure this out easily. Since K is a field, this can not happen, so n must be prime. Thus there is a subfield of K isomorphic to $\mathbb{Z}_p$ for some prime p, namely $\{0, 1, 2 \cdot 1, ..., (p-1) \cdot 1\}$, and our claim is proven. □

This theorem tells us that our field K contains the field $\mathbb{Z}_p$ for some prime p (if you're not convinced, think really hard how a field of six elements could contain a copy of $\mathbb{Q}$. If you find an answer, please write to us at mcgillmathmagaz@gmail.com.) Thus we can view K as a vector space over $\mathbb{Z}_p$. Since K is finite, this vector space is finite dimensional, thus it has a basis, $\alpha_1, \ldots, \alpha_k$. Every element of K is determined uniquely by a linear combination, $p_1\alpha_1 + \ldots + p_k\alpha_k$ with coefficients $p_i \in \mathbb{Z}_p$. Also, every such linear combination determines a unique element of K. There are p^k choices for these, therefore the cardinality of K is p^k. Now I leave it to the reader to prove that there does not exist a prime p such that $p^k = 6$.

Remark. Nothing special about the number 6 was used in this proof. This means that for any finite field K, the cardinality of K is a prime power. I think this is a very interesting result!

Jokes

Q: Why did the chicken cross the Möbius strip?
A: To get to the other... uh.... □

A logician sees a sign on his way to fish that reads, "All the worms you want for 1 dollar." He stops his car and orders 2 dollars' worth. □

A physicist has been conducting experiments and has worked out a set of equations which seem to explain his data. He asks a mathematician to check them. A week later, the mathematician calls: "I'm sorry, but your equations are complete nonsense." "But these equations accurately predict results of experiments. Are you sure they are completely wrong?" asks the physicist. The mathematician replies, "To be precise, they are not always a complete nonsense. But the only case in which they are true is the trivial one where the field is Archimedean..." □

A team of engineers is trying to measure the height of a flag pole, but they can't keep the measuring tape on the pole, since it kept falling off. A mathematician passes by, asks them what the problem was, then proceeds to remove the pole and lay it on the ground. After he leaves, an engineer says to another, "Just like a mathematician! We need the height, and he gives us the length!" □

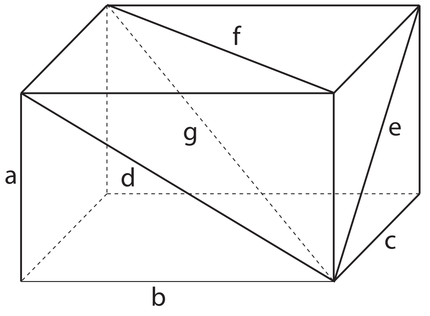
Euler's brick

Alexandra Ortan and Vincent Quenneville-Bélair

An Euler brick is a parallelepiped with integer sides whose face diagonals are also integers. Euler is the first to have thoroughly investigated the problem and has thus bequeathed it his name. If such a parallelepiped's body diagonal also happens to be an integer, it is called an Euler integer brick, or perfect cuboid. In spite of numerous efforts, no perfect cuboid has yet been discovered.

Finding instances of an Euler brick (EB) is equivalent to finding solutions to the first three of the following system of Diophantine equations. If the last one is also satisfied, one has a perfect cuboid (PC).

$$\begin{aligned} a^2+b^2 &= d^2 \\ a^2+c^2 &= e^2 \\ b^2+c^2 &= f^2 \\ a^2+b^2+c^2 &= g^2. \end{aligned}$$



If the edges a, b and c are not relatively prime, i.e. there exists a t such that $t|a,b,c$, then the equations above can all be divided by t^2 on both sides and thus the perfect cuboid is scaled down to a smaller one whose edges are relatively prime. Such cuboids are called primitive cuboids, and are the most interesting ones, since any other can be obtained by taking a primitive cuboid and scaling it up.

Pythagorean triples

Before attempting to solve the whole Diophantine system for the perfect cuboid, it is instructive to take a closer look at the individual equations. One observes that the first three equations describe Pythagorean triples (PT), i.e. triples of positive integers that form a right-angled triangle. Here's a way to generate these triples.

Consider the integers x, y, z such that $x^2 + y^2 = z^2$. Once again, one is only really interested in primitive triples, so the set of solutions can be restricted to triples such that $\gcd(x,y,z) = 1$. This implies the following:

Theorem 1. $\gcd(x,y) = \gcd(y,z) = \gcd(x,z) = 1$.

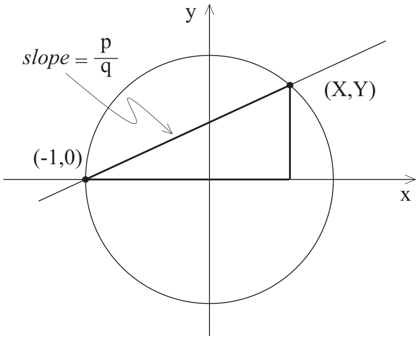
Proof. Suppose $t|x,y$; then $x^2 + y^2 = z^2 \Rightarrow t^2(\frac{x^2}{t^2} + \frac{y^2}{t^2}) = z^2 \Rightarrow t|z$. □

Theorem 2. *z is odd and x, y have opposite parity.*

Proof. Assume $x = y = z = 1(mod2)$; this implies that $x^2 = y^2 = z^2 = 1\ (mod2)$ which leads to $1+1 = 1\ (mod2)$, a contradiction. The same argument holds for one side odd and two even. Thus, only one side can be even (if they were all even, they would not be a primitive triple). Assume $z = 0\ (mod2)$ and $x = y = 1\ (mod2)$; then $z^2 = 0\ (mod4)$ and $x^2 = y^2 = 1\ (mod4)$, which is a contradiction because $1 + 1 \neq 0\ (mod4)$. Hence, the hypotenuse of a primitive PT is odd and one of the legs is even. □

Theorem 3. *All primitive PTs can be generated by two integers p, q of opposite parity whose* gcd *is 1.*

Proof. Finding PT is equivalent to finding integer points on the circle $x^2 + y^2 = z^2$, or rational points on the circle $X^2 + Y^2 = 1$ (where $X = \frac{x}{z}$ and $Y = \frac{y}{z}$). Any rational point $(\frac{a}{b}, \frac{c}{d})$ can be joined to the point $(0,1)$ by a line with rational slope - just let $p = cb$ and $q = ad + bd$. Then $Y = \frac{p}{q}(X+1)$ is the equation of that line, where $\gcd(p,q) = 1$. This line intersects the circle precisely where $X^2 + \frac{p^2}{q^2}(X+1)^2 = 1$ or $X = \frac{\pm q^2 - p^2}{p^2+q^2}$; since only positive solutions are of interest, we will keep the + sign. Together with $Y = \frac{2pq}{p^2+q^2}$, this forms a rational point on the circle $X^2 + Y^2 = 1$. To recover integer points (x,y), one need only choose $z = p^2 + q^2$.



Since $\gcd(p, q) = 1$, p and q cannot both be even. Assume $p = q = 1\ (mod 2)$; then $p^2 = q^2 = 1\ (mod 2)$, but $z = p^2 + q^2 = 0\ (mod 2)$ which implies that the PT is not primitve and contradicts statement 2. Hence, p and q have opposite parity. □

To sum up, here are a few useful properties of primitive Pythagorean triples.

Property 1. One leg is odd, the other is even and the hypotenuse is odd.

Property 2. One leg is divisible by 3.

Proof. Assume neither x nor y are divisible by 3. Then $x^2 = y^2 = 1\ (mod 3)$, and $z^2 = 2\ (mod 3)$. However, this equation has no solutions in $\mathbb{Z}_3$. □

Property 3. One leg is divisible by 4.

Proof. The PT's formula states that $y = 2pq$, where p and q have opposite parity. Thus, $4|y$. □

Property 4. One member of the triple is divisible by 5.

Proof. Assume any two are not, so $x \neq 0\ (mod 5)$ and $y \neq 0\ (mod 5)$ where x and y can be any two edges. Thus the third member is $z^2 = \pm x^2 \pm y^2$. We have $x^2, y^2 \in \{0, 1, 4\}\ (mod 5) \Rightarrow z^2 \in \{\pm 1 \pm 1, \pm 1 \pm 4, \pm 4 \pm 4\}\ (mod 5)$, so z^2 can only take on the values 0,2,3. However, among those, only 0 is a square in $\mathbb{Z}_5$, which implies that z must be divisible by 5. Therefore, at least one edge is divisible by 5. On the other hand, if 5 divides two or more edges, the triple is no longer primitive. □

Euler bricks

Having thus gleaned some information about the divisibility of PTs, one can now wonder how this brings one any closer to finding an Euler brick. In fact, it is now possible to derive a series of properties of the edges of Euler bricks. Here are a few of these properties, along with their proofs.

Theorem 4. *There is exactly one odd edge.*

Proof. Every pair of edges is also a pair of legs of a PT. Since at least one edge must be even in any PT, exactly two edges must be even in a primitive EB. □

Theorem 5. *One edge must be divisible by 4 and another by 16.*

Proof. Exactly two edges of the EB are divisible by 4, by the same argument as above. Consider the PT formed by these two edges and divide it by the gcd of its sides - one obtains a primitive PT, which must still have one leg divisible by 4. Therefore, one of the edges of the EB must be divisible by 16. □

Theorem 6. *One edge must be divisible by 3 and another by 9.*

Proof. One can apply the same method as above. □

Theorem 7. *One edge must be divisible by 5.*

Proof. Suppose none are. Property 10 implies that all face diagonals of this EB must be divisible by 5. An edge can be congruent to either ± 1 or $\pm 2\ (mod 5)$. Two edges cannot both be congruent to $\pm 1\ (mod 5)$ since their corresponding face diagonal would not be an integer, $d^2 = a^2 + b^2 = 2\ (mod 5)$ not having any solutions. The same argument stands for two edges both congruent to $\pm 2\ (mod 5)$. Some edge must therefore be congruent to $0\ (mod 5)$. □

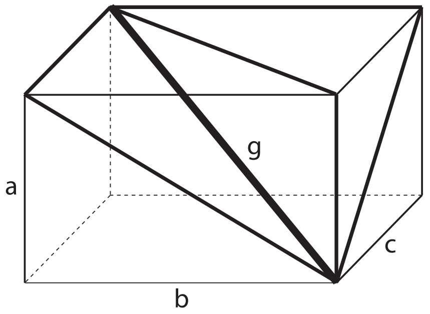
Theorem 8. *One edge must be divisible by 11.*

Proof. In $\mathbb{Z}_{11}$, the only perfect squares are $\{0, 1, 3, 4, 5, 9\}$, so the square of the hypotenuse of a PT can only take on those values. A bit of calculation reveals that the only pairs of legs whose squares add up to one of those numbers are $\{(0, \pm 1), (0, \pm 2), (0, \pm 3), (0, \pm 4), (0, \pm 5), (\pm 1, \pm 2), (\pm 1, \pm 5), (\pm 2, \pm 4), (\pm 3, \pm 4), (\pm 3, \pm 5)\}$. Supposing no edge is divisible by 11 reduces those possibilities to $\{(\pm 1, \pm 2),$

$(\pm1,\pm5)$, $(\pm2,\pm4)$, $(\pm3,\pm4)$, $(\pm3,\pm5)\}$. However, one can verify that if an edge takes on any of those values, the two other edges cannot themselves also form a pair of legs of a PT. Hence no such EB exists. □

Perfect cuboids

In addition to being an Euler brick, a perfect cuboid's body diagonal must also be an integer. This diagonal would have to be the hypotenuse of three Pythagorean triples (not necessarily primitive), each of which must have a leg which is in turn the hypotenuse of another PT.



The first condition is in principle easily complied with: if an integer has n distinct prime factors of the form $4k+1$, then it can be the hypotenuse of 2^{n-1} distinct primitive PTs [9]. The second condition, however, is more elusive: three of those PTs must have an edge of the cuboid as one of their legs, such that the sum of the squares of those legs is precisely $a^2+b^2+c^2=g^2$.

The problem can be viewed the other way around: find a number whose square is the sum of three squares which in turn, paired two by two, generate three PTs. In this case, one would have to first look at integers g such that $g^2=a^2+b^2+c^2$. A derivation similar to that for PTs, but involving rational points on a sphere this time, shows that for integers p,q,r with $\gcd(p,q,r)=1$, $a=r^2-p^2-q^2$, $b=2qr$, $c=2pr$ and $g=r^2+p^2+q^2$ satisfy $g^2=a^2+b^2+c^2$. It is still unclear precisely when the integers a,b,c thus obtained are the legs of three different right-angled triangles.

Results to date

The smallest Euler brick is (240,117,44) and was discovered by Paul Halcke in 1719. Later, Saunderson found a parametric solution that always generates Euler bricks, but not all of them. Starting with a PT (x,y,z) he showed that $(a,b,c)=(x(4y^2-z^2),y(4x^2-z^2),4xyz)$ is an Euler brick with face diagonals $d_{ab}=z^3$, $d_{ac}=x(4y^2+z^2)$, $d_{bc}=y(4x^2+z^2)$. Lagrange, however, proved that none of those, nor any derived cuboids are perfect cuboids [6]. Some two centuries later, Korec used a computer to show that the smallest edge of a perfect cuboid can be no smaller than 10^6, proving along the way a few numTheorems to speed up the algorithm. More recently, Rathbun has increased the lower bound to 2^{32}.

Theorem 9. *The following is equivalent to finding the PC:*

1. *Let $a=\frac{p^2+q^2}{2pq}$ and $b=\frac{p^2-q^2}{2pq}$. The question is to know whether it is possible to get ab and $\frac{a}{b}$ in the form of b.*

2. *Find non-trivial integer solutions to $(a^2c^2-b^2d^2)(a^2d^2-b^2c^2)=(a^2b^2-c^2d^2)^2$*

Theorem 10. *If there exists a PC, the following also exist:*

1. *A classical rational cuboid with edges x_1, x_2, x_3 and a square z^2 such that $z^2-x_i^2$ are all squares.*

2. *A set of four non-zero squares whose differences are all squares.*

3. *Two ratios of the form $\frac{p^2-q^2}{2pq}$, whose product and ratios are of this form.*

4. *A set of four squares whose sums in pairs are also square.*

5. *An arbitrarily long or infinite sequence of squares, where the sum of any two (or three) adjacent members is also a square.*

6. *A cycle of integer solutions to $\frac{\alpha_1^2-\beta_1^2}{2\alpha_1\beta_1}\frac{\alpha_3^2-\beta_3^2}{2\alpha_3\beta_3}=\frac{\alpha_2^2+\beta_2^2}{2\alpha_2\beta_2}$ with $\frac{\alpha_1}{\beta_1}=\frac{p^2-q^2}{2pq}$ and $\frac{\alpha_2}{\beta_2}=\frac{r^2+s^2}{2rs}$.*

7. *The validity of Conjecture C in [3] (which is far beyond the scope of this article)...*

Solutions to the PC with relaxation (one edge or one face diagonal irrational) have also been studied.

The problem of finding a PC with one edge irrational is equivalent to finding an integer solution to $x_1^2+x_2^2=y_3^2$ with $t+x_1^2$, $t+x_2^2$ and $t+y_3^2$ where t is an integer – the square of the irrational edge. One solution (x_1, x_2, y_3) is (124, 957, 13852800). It can be extended to a one-parameter family of solutions.

Finding a PC with one face diagonal irrational corresponds to solving the following system of equations: $x_1^2+x_2^2 = y_3^2$, $x_1^2+x_3^2 = y_2^2$ and $x_1^2 + x_2^2 + x_3^2 = z^2$. There are no conditions on $x_2^2 + x_3^2$. One can see that the equations imply that $2(z^2 + x_1^2)$, $2(z^2 - x_1^2)$ and $2\left|x_2^2 - x_3^2\right|$ have their sums and differences square. The sums in pairs are $2z$, $2x_1$, $2y_2$, $2y_3$, $2x_2$ and $2x_3$. Note that the differences of z, y_2, x_3 and z, y_3, x_2 are all squares.

Conclusion

If the problem of finding a perfect cuboid has remained the same for many centuries, the techniques employed to tackle it have evolved much since Euler first gave it some serious thought. More recently it has been viewed through the lens of algebraic geometry, where solutions to the perfect cuboid translate into rational points on curves, a question much beyond the scope of this article.

If such pursuits as this appear appealing to the reader, the following problems have a similar flavor. With money helping motivation, there is a million dollars price for the solution of the Birch and Swinnerton-Dyer Conjecture. [http://www.claymath.org/millennium/Birch_and_Swinnerton-Dyer_Conjecture]. For only $100,000, the Beal Conjecture is a good deal. It goes as follows: let A, B, C, x, y and z be positive integers where x, y and z are all greater than 2 such that $A^x + B^y = C^z$, then A, B and C have a common divisor. [http://www.bealconjecture.com/]. There is also a small hundred dollars from Martin Gardner for finding a 3x3 magic square with 9 distinct square entries. Moreover, Sierpinski asks if there are non-trivial solutions to $(x + y + z)^3 = xyz$. Finally you can try to find the solutions for $1^n + 2^n + ... + k^n = (k + 1)^n$ with $n > 1$ and a characterization of the integers d for which $x^2 - dy^2 = -1$ has an integer solution.

References

[1] N. Saunderson, The Elements of Algebra, Vol 2, Cambridge, 1740.

[2] J. Lagrange, Sur le dérivé du cuboïde Eulérien, Canadian Mathematical Bulletin, Vol 22, 1979.

[3] J. Leech, The Rational Cuboid Revisited, The American Mathematical Monthly, Vol 84, No 7, p. 518-533, 1977.

[4] P. Halcke, Deliciae Mathematicae; oder, Mathematisches sinnen-confect. Hamburg, Germany, N. Sauer, 1719.

[5] M. Kraitchik, On certain Rational Cuboids, Scripta Math. 11, 1945.

[6] I. Korec, Nonexistence of small Perfect Rational Cuboid I/II, Acta Mathematica Universitatis Comenianae, 1983/1984.

[7] Jay R. Goldman, The queen of mathematics : an historical motivated guide to number theory, Wellesley, 1998.

[8] R. Luijk, On Perfect Cuboids, doctoral thesis, University of Ultrecht, 2000.

[9] A. H. Beiler, The Eternal Triangle, Ch. 14 in Recreations in the Theory of Numbers: The Queen of Mathematics Entertains, New York: Dover, 1966.

Jokes

"Have solved the Riemann Hypothesis" – G. H. Hardy □

A physicist and an engineer are in a hot-air balloon. Soon, they find themselves lost in a canyon somewhere. They yell out for help: "Hellllooooooo! Where are we?" 15 minutes later, they hear an echoing voice: "Hellllooooooo! You're in a hot-air balloon!" The physicist says, "That must have been a mathematician." The engineer asks, "Why do you say that?" The physicist replies, "The answer was absolutely correct, and it was utterly useless." □

Ten Proofs of the Infinitude of Primes

Nan Yang

A **prime number** is a counting number, greater than 1, that cannot be divided except by 1 and itself. Prime numbers have fascinated amateur and professional mathematicians for thousands of years. While I am far from being qualified even as an amateur mathematician, I would like to share my fascination with these figures by presenting a collection of proofs of the following theorem, which has been known since ancient times:

Theorem. *There exists an infinite number of primes.*

We will begin with a modernized version of Euclid's original proof which appeared as proposition 20 in book 9 of *The Elements* over 2000 years ago. Although Euclid used slightly different notations (he proved the case where $n = 3$), the idea of the proof has not changed since that time:

Proof. 1 Suppose there are only n primes $p_1, ..., p_n$. Then $p_1...p_n + 1$ is either prime or not. If it is prime, then we have found another prime, hence the hypothesis must be false. If it is not prime, then it must be divisible by some prime p_i where $1 \leq i \leq n$. But since p_i also divides $p_1...p_n$, it must divide the difference, 1, which is impossible, and hence the hypothesis must be false. Therefore the number of primes cannot be finite. □

By using ideas similar to those used in the first proof, it is possible to produce a much shorter proof:

Proof. 2 Any prime divisor of $n! + 1$ is greater than n. □

Hidden in the statement of the previous proof the fact that a divisor of any number is obviously less than or equal to the number itself. Hence there is at least one prime between n and $n! + 1$, and another between $n! + 1$ and $(n! + 1)! + 1$, etc.

It is possible to produce a proof based on Euclid's method, but without adding a unit. Stieltjes first published such a proof, of which a variant is shown here:

Proof. 3 If there are only n primes $p_1, ..., p_n$, then $p_1...p_i + p_{i+1}...p_n$ is not divisible by any of the n primes, since each prime divides exactly one of the summands. □

Métrod gave a proof similar to that of Stieltjes:

Proof. 4 Suppose there are only n primes $p_1, ..., p_n$. Let $N = \prod_{i \leq n} p_i$ and let $Q_i = N/p_i$. Thus each of the n primes does not divide exactly one of the summands of $S = \sum_{i \leq n} Q_i$; therefore, S is not divisible by any of the primes $p_1, ..., p_n$. □

Hardy and Wright's *Introduction to the Theory of Numbers* contains a very intricate proof which uses ideas that are different from those of Euclid:

Proof. 5 Suppose that $2, 3, ..., p_j$ are the first j primes and let $N(x)$ be the number of $n \leq x$ such that n is not divisible by any prime $p \geq p_j$. Any such n can be expressed in the form $n = a^2 b$ where b is squarefree, i.e. $b = 2^{e_1} 3^{e_2} ... p_j^{e_j}$ where $e_i = 0$ or 1. Hence, there are 2^j possible values of b. Since $a \leq \sqrt{n} \leq \sqrt{x}$, there are not more than $\sqrt{x}$ different values of a. Therefore $N(x) \leq 2^j \sqrt{x}$.

Now suppose there are only j primes, then $N(x) = x$ for all x. This implies that $x \leq 2^{2j}$, which is false for $x \geq 2^{2j} + 1$. □

One way of proving that the primes are infinite is by constructing an infinite sequence of numbers that are pairwise coprime, that is, the prime factors of an element of that sequence is unique to that element. A very well known such sequence is the **Fermat numbers** F_n, which are defined as $F_n = 2^{2^n} + 1$.

Proof. 6 Suppose F_n and F_{n+k} have a common factor m. Let $x = 2^{2^n}$. We then have

$$\frac{F_{n+k} - 2}{F_n} = \frac{x^{2^k} - 1}{x + 1} = x^{2^k - 1} - x^{2^k - 2} + ... - 1.$$

Hence m divides $F_{n+k} - 2$, which means $m = 2$, but this is impossible since all Fermat numbers are odd. □

A number is a **Fermat prime** if it is a Fermat number and is prime. An open question about Fermat primes is that whether they are infinite. Although Fermat himself conjectured that *all* numbers of the form $2^{2^n}+1$ are prime, this is now known to be false. In fact, so far only the first five are known to be prime, the largest being $F_4 = 65537$. Many mathematicians believe that all Fermat numbers greater than F_4 are composite; the largest known such composite is F_{23471}.

It is also possible to prove the infinitude of primes by constructing *arbitrarily long* sequences whose elements are pairwise coprime. Schorn produced one such sequence as follows:

Proof. 7 If $1 \leq i < j \leq n$, then any divisor of $n!i+1$ and $n!j+1$ must divide the difference, which is $n!(j-i)$. However, by proof 2, any divisor of $n!+1$ can not divide $n!$ and must be greater than n. Therefore

$$\gcd(n!i+1, n!j+1) = 1,$$

and the n integers $n!i+1 (i = 1, 2, ..., n)$ are pairwise coprime. □

Recall the following lemma: $gcd(a,b) = 1 \Leftrightarrow$ there exists s, t such that $sa + tb = 1$. With this in mind:

Proof. 8 Let $q_1 = 3, q_{n+1} = q_1 ... q_n - 1$. Without loss of generality let $i < j$. Then

$$\prod_{k=1}^{j} q_k - q_j = 1.$$

Let $A = \prod_{k=1}^{j} q_k / q_i$. We thus have $Aq_i - q_j = 1$. Therefore, any two elements q_i, q_j are coprime by the converse of the previous lemma. Since (q_n) is increasing for $n > 2$, there exists infinitely many primes. □

A more exotic proof came from Fürstenberg, based on topological ideas, in 1955. Here is a variant:

Proof. 9 Define a topology on the integers $\mathbb{Z}$ by taking the set of arithmetic progressions from $-\infty$ to ∞ as a basis. One can check that this gives an actual topological space. Note that the complement of an arithmetic progression is the union of the other arithmetic progressions with the same step size, so that arithmetic progressions are both closed and open. Now, consider $A = \bigcup_p A_p$ where p runs through all the primes ≥ 2, and A_p is the (closed) set of all multiples of p. The complement of A is $\{-1, 1\}$, because all other numbers are primes or multiples of primes. Since $\{-1, 1\}$ is obviously not open, its complement A cannot be closed. However, a finite union of closed sets is closed, so there must be infinitely many A_p, i.e. there are infinitely many primes. □

Recall the geometric series identity

$$\sum_{k=0}^{\infty} \frac{1}{p^k} = \frac{1}{1-(1/p)}.$$

If p, q are two primes, then

$$1 + \frac{1}{p} + \frac{1}{q} + \frac{1}{p^2} + \frac{1}{pq} + \frac{1}{q^2} + ... = \frac{1}{1-\frac{1}{p}} \times \frac{1}{1-\frac{1}{q}}.$$

By the unique factorization theorem and by multiplying the sums of the reciprocals of every possible prime to every possible power, we obtain[14]

$$\sum_{n\in\mathbb{N}} \frac{1}{n} = \prod_{p\in\mathbb{P}} \frac{1}{1-(1/p)}.$$

Euler was the first person to discover this identity. The left-hand side of Euler's identity is a sum taken over the counting numbers while the right-hand side is a product taken over the primes. We will thus conclude with an analytic proof of the infinitude of primes:

Proof. 10 By comparing the harmonic series to an integral, we obtain:

$$\int_1^x \frac{dn}{n+1} < \sum_{n=1}^{x} \frac{1}{n}$$

for all x. Since the integral diverges as $x \to \infty$, so does the sum. By Euler's identity, the number of primes cannot be finite. □

References

[1] Hardy, G. H. and Wright, E. M. "An Introduction to the Theory of Numbers." Oxford at the Clarendon Press, 1960.

[2] Ribenboim, Paulo. "The New Book of Prime Number Records." Springer, 1996.

[3] Ore, Oystein. "Number Theory and Its History." Dover Publications, 1988.

[14] See A. Beaudry, *Interview with prof. Henri Darmon* for a proof.

Living without Math

Vincent Quenneville-Bélair

Living without math *is* possible – as surprising as it seems. The 200 members of the Pirahã tribe, gathered in groups of ten to twenty near the Amazon, live without the concepts of numbers and counting.

Daniel Everett, an American linguistic anthropologist, and Peter Gordon, a psycholinguist from New York's Columbia University, studied the Pirahã. The former lived with this tribe for 27 years; the latter did some experiments with them over a three-year period.

No Math

Gordon's study was conducted on a group of men only, because cultural taboos excluded women and children. Everett tried to teach them how to count for eight months, but "in the end, not a single person could count to ten." One of the experiments consisted in duplicating a line of batteries. Beyond two or three, the men started making mistakes. The problem seemed to come from their lack of words for counting. The word they use for "one" is closer to "a small amount" and the word for "two" is like "a relatively bigger amount". It is impossible to know if a Pirahã is designating one fish, a small fish or two fishes. They also have trouble drawing: "Producing simple straight lines was accomplished only with great effort and concentration, accompanied by heavy sighs and groans." (Gordon)

Other characteristics

The Pirahãs do not have a written language; they communicate by singing, whistling and humming. Their pronouns seem to originate from another language, and some of them combine singular and plural: "he" and "they" are the same; "more", "several" and "all" are inexistent. Furthermore, their collective memory does not extend more than two generations back. No equivalent to our art and fiction seems to exist. Also, they simply say, when urged to explain their history, that "everything is the same". Everett explains that the Pirahã culture limits "communication to non-abstract subjects in immediate experience for the interlocutors".

Notable Facts

The particularities of the Pirahãs are not explained by social or genetic isolation, because they have been trading goods and women with Brazilians for 200 years. Their motivation to learn came from such exchanges with outsiders: they wanted to know if they were being cheated because they did not understand non-barter economic relations. Their enthusiasm made them attend daily classes given by Everett and his family with great interest. However, adding 3 to 1 remained impossible for them, and the Pirahã concluded that they could not learn the material.

References

[1] D. Everett. "Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language", Current Anthropology 46 (4): 621-46, 2005.

[2] J. Crow, 2006. [www.jcrows.com/ withoutnumbers.html]

Interview with professor Nilima Nigam

Nilima Nigam is a professor of mathematics at McGill, where she teaches differential equations and numerical analysis to terrified (but increasingly enthusiastic) undergraduates. She was kind enough to grant the Delta-Epsilon an interview about her work and – yes – life as a mathematician.

The δelta-ϵpsilon: What area of research are you working on now? Can you tell us about some of your projects? What led you to that specific area and project?

Nilima Nigam: This year, I'm focusing on four projects, three of which are in the area of numerical analysis (the study of algorithms and their convergence), and one of which is a collaboration with Prof. Komarova in the Faculty of Dentistry. The latter is an investigation into the growth dynamics of bone cells. It's a fun project- my collaborator has actual cultures of these cells, and lots of experimental data. My job is to build mathematical models to describe what we see, and help predict specific things about the system. The models lead to new biological questions, which in turn suggest new experiments. The results of these experiments may confirm our model, or require us to refine it. This summer, the SUMS treasurer, Tayeb, was deeply involved with this project.

In the long term, I'm deeply interested in the mathematics of wave interactions with bounded objects. The waves could be sound, electromagnetic, elastic, gravitational or pressure waves. This field is quite old, and has motivated the development of many branches of mathematics. I've been thinking about problems in this area for well over a decade, and every time I think something is settled, another interesting question arises.

While performing numerical simulations of wave-obstacle interactions, one is constrained to describe the physical problem on a bounded region. Think of sound scattering off a hedgehog. In theory, the scattered wave can propagate forever. Computationally, though, we need to put a box around the hedgehog, and try to capture the behaviour of the scattered wave inside the box. This process introduces errors; it's not at all obvious what boundary conditions to prescribe on the box, and finally creating algorithms which are provably convergent and accurate is hard. Two of my projects concern techniques for this problem; I'm collaborating with a graduate student, Simon Gemmrich, on one of these. There are many interesting analytical questions at the PDE level, and then a whole host of other questions concerning the numerical analysis of the methods. This work is strictly of a theorem-proof nature, though of course I design experiments to test the methods.

Associated with scattering problems, but also more generally applicable, are a class of numerical methods for PDE known as finite element methods. These work by approximating the solution in terms of compactly supported (polynomial) basis functions. Recently, much work has been done to develop a finite element exterior calculus using the tools of differential geometry and homological algebra. In this framework, discretizations of PDE are designed to be compatible with the geometric, algebraic and topological properties of the actual solutions of the PDE. Existing work requires the polynomial basis functions to be defined on simplicial objects–tetrahedra or boxes in 3D. In work with Joel Phillips, another graduate student, we're trying to extend this framework to non-simplicial objects such as pyramids.

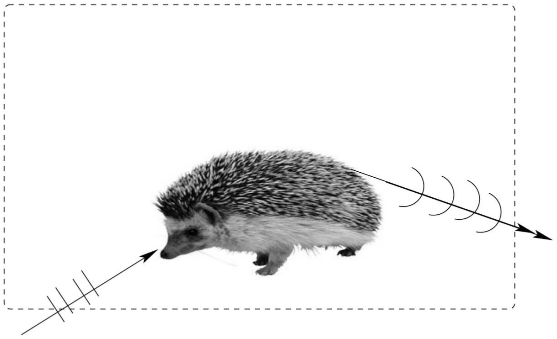


Fig.1 A plane wave incident on a hedgehog and the associated scattered wave. The dashed line indicates the artificial boundary we'd put around Hedgie while computing the scattered wave.

$\delta - \epsilon$: How does your research relate to what is taught in undergraduate courses here?

NN: I admire the undergraduate program broadly for instilling a sense of mathematical fearlessness. On the many occasions when I'm stuck on a research problem, I think to myself - 'if this were a homework problem, and I were a McGill undergrad, when would I quit?' A lot of my work requires functional analysis and PDE (not just separation of variables!). You see some of the analytical tools involved in courses like Math 564/565. At an advanced level, the study of PDE merges with analysis (Math 575/580/581). In addition, courses in numerical analysis or matrix computation (Math 317/387, 327/397, 578/579) contain many of the key concepts–stability, accuracy and convergence of numerical algorithms- which I use. Increasingly, algorithms for PDE incorporate tools from differential geometry, which is another field you may see during your studies here.

I'm not sure if students in Mathematics take courses in the Physics department. In an ideal world, budding mathematicians interested in the mathematics of scattering theory would see classical mechanics, electrodynamics and quantum mechanics.

$\delta - \epsilon$: Why did you choose to become a mathematician? What kind of a life is it?

NN: Becoming a mathematician didn't really involve a choice. I'm very fortunate to be able to do what I love, and would be miserable doing anything else. I started off as a student of Physics, realized I loved the mathematical aspects of my training most, and ended up pursuing a career in mathematics. Physics has big problems, which filter into our collective consciousness - this decade, quantitatively inclined dreamers want to work in String Theory. A millennium ago, High Energy Physics and Cosmology captured my imagination. By the time I got the necessary educational background to begin to understand the science behind these areas, I realized mathematics was my deeper passion.

It's a great life. I'm fortunate enough to enjoy the various aspects of my chosen career- doing research, teaching, and interacting with other mathematicians and scientists. Some may favour one part of this life, and regard the other bits as distractions. I'm energized by mathematics, and by the belief that fun mathematical questions can be found everywhere. This makes teaching and interacting with people part of the larger search for interesting mathematical questions and their resolution - any given lecture, a student could ask me something thought-provoking.

$\delta - \epsilon$: Are there any particular open problems you'd like to see the solutions to?

NN: There are several open and rich questions in mathematics, and several technical questions in my field of interest I'd like to see resolved. However, a particularly challenging mathematical question concerns the existence and regularity properties of solutions of the Navier-Stokes equations in $\mathbb{R}^3$. This system of partial differential equations governs the motion of fluids; the study of their solutions will require huge advances in the analysis of PDE. The question evades standard methods of attack, and a successful approach will likely be both surprising, and intimately connected to many other branches of mathematics. The problem was recently classified as one of the Millennium problems by the Clay Institute; the precise problem statement due to Charles Fefferman is available at http://www.claymath.org/millennium/Navier-Stokes-Equations/navierstokes.pdf.

$\delta - \epsilon$: What major mathematical event (beyond your own work) do you remember most vividly? Alternatively, what such event had the biggest impact on your life as a mathematician?

NN: This one's tricky... I cannot think of a single formative experience. Rather, many chance encounters with mathematicians I admire have impacted my career. Reading about the lives of famous mathematicians and their work habits has always inspired me.

The Birth of Quaternions

Michael McBreen

> Where do vector spaces come from? Did Gauss wake from troubled slumber one night, spring from his bed and shout, "Let there be a set V and a field F, and let $+ : V \times V \to V$ be an associative, commutative operation such that ..."? No, no he did not.

Modern vectors arguably evolved from Hamilton's quaternions[15], the set of numbers of the form

$$a + bi + cj + dk$$

with $a, b, c, d \in \mathbb{R}$ and

$$i^2 = j^2 = k^2 = ijk = -1.$$

As we will see at the end of this article (keep reading), not only vectors but also the inner product and the cross product of vectors are hiding inside quaternions. Better still, they gave us the concept of non-commutative number systems. I write this to convince you, the reader, that it is worth your time to learn how quaternions were discovered. The thing is, we are blessed with an unusually detailed account of the birth of quaternions[16] from the correspondence of Hamilton himself, so we can follow his thought process step by step.

But first, a few words about the ancestors of the quaternions themselves, the complex numbers. The square root of -1 was introduced by Cardano to solve cubic equations. Complex numbers share many properties with the reals, such as

$$\begin{aligned} z + w &= w + z \\ wz &= zw \\ w(z + x) &= wz + wx \\ \forall z \ \ \exists z' \ \text{ s.t. } \ zz' &= z'z = 1 \\ |wz| &= |w||z|. \end{aligned}$$

The last property was known to Hamilton as the law of the moduli, and will play a crucial role in what follows.

Mathematicians were very queasy about complex numbers at first – back then, even negative numbers were suspicious – but eventually a whole host of people developed a geometric interpretation of complex numbers as lines in a plane (the Argand plane, for those of you who went to high school). Each number $a + bi$ corresponds to a line stretching from the origin to (a, b), and multiplication corresponds to addition of angles and multiplication of lengths in the plane.

Well, that's complex numbers for you. Now fast forward to the 1800s, and enter William Rowan Hamilton. Given the interpretation of complex numbers as directed lines in a 2D plane, it's quite natural to desire a similar system for 3D space. Hamilton suspected that physical concepts like velocity and force could find natural expressions in this hypothetical system; the existing notation was extremely cumbersome.

His goal was to forge a system of "triplets" $a + bi + cj$ that shared most of the properties listed above, among them the law of the moduli, the existence of inverses and distributivity. The catch is that there's no such system: Hurwitz would prove half a century later that any finite dimensional normed division algebra[17] over the reals has dimension either 1, 2, 4 or 8. Real and complex numbers are dimension 1 and 2 normed division algebras – Hamilton was going for 3 dimensions, but his efforts would yield the 4 dimensional quaternions instead.

Let's see how he went about looking for his triplets. In order to define triplets completely, he only needed to choose (or find, if you will) the values of the products ij, ji, i^2 and j^2 – ev-

[15] For brevity's sake we will neglect the other great ancestor of vectors, Grassman's exterior calculus.

[16] Much of this information is summarized by a nice article [1] from B.L. van der Waerden, which I encourage you to seek out on the internet.

[17] An **algebra** over the reals is roughly a vector space V over the reals equipped with a bilinear operation $* : V \times V \to V$ which you can think of as a multiplication of vectors. The complex numbers can be viewed as a vector space over the reals with 1 and i as basis vectors. When we define $1 * 1$, $1 * i$, $i * 1$ and $i^2 = i * i$ and use distributivity to extend the product to the whole vector space, it becomes an algebra. A **division algebra** is an algebra where very element except 0 has a multiplicative inverse. A **normed** algebra satisfies the law of the moduli.

[18] The "purely real" triplets $a + 0i + 0j$ were presumed to commute with all other triplets.

erything else would follow by distributivity and the properties of real numbers.[18] The challenge was to choose values that would satisfy the basic properties listed above.

First he went after i^2 and j^2. With the norm squared defined as the sum of the squares of the coefficients (this comes to us from Pythagoras' theorem), the law of the moduli for Hamilton's hypothetical triplets reads

$$|(a+bi+cj)(\alpha+\beta i+\gamma j)| = |a+bi+cj||\alpha+\beta i+\gamma j|$$

with

$$|a+bi+cj| = \sqrt{a^2+b^2+c^2}$$

Considering the subset of triplets of the form $a+bi$ or $a+bj$, Hamilton used distributivity to get

$$(a+bi)(\alpha+\beta i) = a\alpha + (a\beta + b\alpha)i + b\beta i^2.$$

Setting $i^2 = A + Bi$, with A and B as yet undetermined, he then used the law of the moduli to get

$$\begin{aligned} |(a+bi)(\alpha+\beta i)| = &\quad (a\alpha + b\beta A)^2 + (a\beta + \\ &\qquad b\alpha + b\beta B)^2 \\ = &\quad (a^2+b^2)(\alpha^2+\beta^2). \end{aligned}$$

To make this equality hold, Hamilton had to set $i^2 = -1$. The exact same procedure also gave $j^2 = -1$. Now, he had both i^2 and j^2, so he was only missing ij and ji. If we assume commutativity, then $(ij)^2$ must be 1, so ij should be either 1 or -1. Alas, both of these options violate the law of moduli (check for yourself).

So he forgot commutativity for a second and considered the square of the triplet $a+bi+cj$. He noticed that the very demanding law of moduli

$$|(a+bi+cj)^2| = |a+bi+cj|^2$$

would be satisfied with $ij = 0$, and that furthermore it gave the product a geometric interpretation, just like the complex number product. Just as the square of a complex number z lies at twice the angle with respect to the real axis as z itself, the square of a triplet (with $ij = 0$) lies at twice the angle with respect to the "real axis", i.e. the axis of $(a, 0, 0)$ triplets. But happiness is fleeting, as this quote reveals:

> *Behold me therefore tempted for a moment to fancy that $ij = 0$. But this seemed odd and uncomfortable, and I perceived that the same suppression of the term which was* de trop *might be attained by assuming what seemed to me less harsh, namely that $ji = -ij$. I made therefore $ij = k, ji = -k$, reserving to myself to inquire whether k was 0 or not.*

The reader will note that Hamilton's ambitions first leaned toward poetry, but his friend Wordsworth wisely advised him to study mathematics instead. Anyway, Hamilton next considered the product

$$\begin{aligned} (a+bi+cj)(x+bi+cj) = ax - b^2 - c^2 + \\ b(a+x)i + c(a+x)j + (bc-bc)k \end{aligned}$$

He wrote:

> *The coefficient of k still vanishes; and $ax - b^2 - c^2$, $(a+x)b$, $(a+x)c$ are easily found to be the correct coordinates of the product-point in the sense that the rotation from the unit line to the radius vector of a, b, c being added in its own plane to the rotation from the same unit-line to the radius vector of the other factor-point x, b, c conducts to the radius vector of the lately mentioned product-point; and that this latter radius vector is in length the product of the two former. Confirmation of $ij = -ji$; but no information yet of the value of k.*

Mysterious k. Well, the next thing Hamilton did was to bravely consider the general product of two triplets:

$$\begin{aligned} (a+bi+cj)(x+yi+zj) = ax - by - cz + \\ (ay+bx)i + (az+cx)j + (bz-cy)k \end{aligned}$$

He then checked whether the law of the moduli was happy with k being null, and found that with $k = 0$, the left-hand side's norm was smaller than the right-hand side's by $(bz-cy)^2$. Curious – exactly the square of the coefficient of k. As Hamilton and his wife strolled along Dublin's Royal Canal on the 16th of October 1843, inspiration struck: if he set k on equal footing with i and j, and hence added a fourth dimension, then the norms would match. This was the key result, the crucial step. With k given the respect it deserved, everything flowed smoothly:

> *I saw that we had probably $ik = -j$, because $ik = iij$, and $i^2 = -1$; and that in like manner we might expect to find $kj = ijj = -i$;*

Hamilton writes "probably" because he was not sure whether or not his quaternions – as they must now be called – were associative. He was a wise, cautious man and you would do well to emulate him.[19] The same reasoning that led to $i^2 = j^2 = -1$ now led him to $k^2 = -1$, and he happily carved the defining equations

$$i^2 = j^2 = k^2 = ijk = -1$$

in the stone flank of Broom Bridge.

And what about vectors? Well, Hamilton called the real component of his quaternions the "scalar part" whereas the i, j, k component was the "vector part". Now, consider the product of two "pure vector" quaternions:

$$(bi + cj + dk)(\beta i + \gamma j + \delta k) = -b\beta - c\gamma - d\delta + (c\delta - d\gamma)i + (d\beta - b\delta)j + (b\gamma - \beta c)k$$

Note that the scalar part of the product is simply minus the scalar product of the two quaternions viewed as vectors with components (b, c, d) and (β, γ, δ), while the vector part is their vector product (cross product). The geometric interpretation of these operations was known, and when the modern vector was born, they were carried over. Of course, this is another story, given with much scholarly detail by M. J. Crowe in [2]. But let us end on a high note with this poem, by Hamilton himself, about his creations:

Or high Mathesis, with her charm severe,
Of line and number, was our theme; and we
Sought to behold her unborn progeny,
And thrones reserved in Truth's celestial sphere:
While views, before attained, became more clear;
And how the One of Time, of Space the Three,
Might, in the Chain of Symbol, girdled be:
And when my eager and reverted ear
Caught some faint echoes of an ancient strain,
Some shadowy outlines of old thoughts sublime,
Gently he smiled to see, revived again,
In later age, and occidental clime,
A dimly traced Pythagorean lore,
A westward floating, mystic dream of FOUR.

References

[1] B. L. van der Waerden, *Hamilton's discovery of quaternions*, Mathematics Magazine, 49 (1976), pp. 227–234.

[2] M. J. Crowe, *A History of Vector Analysis*, Dover Publications, Reprint edition (1994)

Jokes

This is $|BS|$! □

e^x and x^2 are walking down the road when they suddenly see a differential operator. x^2 says to e^x, "Run, or we'll be differentiated!" e^x calmly replies that he cannot be differentiated. So e^x walks up to the differential operator and says, "Hi, I'm e^x." To which the differential operator replies, "Hi, I'm $\frac{d}{dy}$." □

Mathematics is made of 50 percent formulas, 50 percent proofs, and 50 percent imagination. □

An engineer, a physicist and a mathematician find themselves in an anecdote, indeed an anecdote quite similar to many that you have no doubt already heard. After some observations and rough calculations the engineer realizes the situation and starts laughing. A few minutes later the physicist understands too and chuckles to himself happily as he now has enough experimental evidence to publish a paper. This leaves the mathematician somewhat perplexed, as he had observed right away that he was the subject of an anecdote, and deduced quite rapidly the presence of humor from similar anecdotes, but considers this anecdote to be too trivial a corollary to be significant, let alone funny. □

[19] It turns out that they really are associative.

Getting Acquainted with Φ

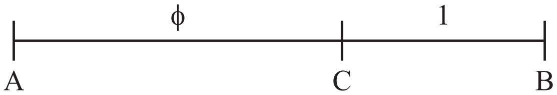
Juan Manuel Martinez

One can say so many cheesy things about the number $\Phi = \frac{1+\sqrt{5}}{2}$, that there's no way of picking which one to start with. Read on to find out what makes it so special.

At the beginning of the sixteenth century, the Italian mathematician Luca Pacioli published *Divina Proportione* in which he called Φ the "Divine Proportion". Since the nineteenth century, Φ has been also called the "Golden Section, Ratio or Number".

How to derive Φ

It was Euclid of Alexandria who first gave a proper definition of the number Φ in his famous book *The Elements.* He derived it using the following geometric construction. Consider a straight line.



If you divide the line so that the ratio of the entire line AB to that of the larger line segment AC is the same as the ratio of the larger line segment AC to that of the smaller line segment CB, you will get Φ. The lines are then said to have been cut in extreme and mean ratio.

Φ can also be derived from the Fibonacci sequence $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$ Each term (except the first two) is the sum of the two preceding ones. We're going to prove here that the ratio of successive terms in the Fibonacci sequence converges to Φ. Consider the set V of real-valued sequences, $\{a_n\},\ n \geq 0$ satisfying $a_{n+1} = a_n + a_{n-1}$ for all $n \geq 1$. It can easily be shown that this set, equipped with the usual addition and scalar multiplication of sequences, is a vector space of dimension 2. A geometric progression is a sequence satisfying $a_n = \alpha^n$. We will now build a basis for V consisting of two geometric progressions. By definition, $\alpha^{n+1} = \alpha^n + \alpha^{n-1}$. Dividing by α^{n-1} and solving for α, we get $\alpha = \Phi$ and its algebraic conjugate, denoted $\overline{\Phi}$. To show that the sequences generated by Φ^n and $\overline{\Phi}^n$ form a basis for V, we need only show that they are linearly independent (since the space is 2-dimensional), i.e. show that if $a(1+\sqrt{5})/2)^n + b((1-\sqrt{5})/2)^n = 0$ then $a = 0$ and $b = 0$. We can show this easily by successively setting n=0 and n=1. Since the Fibonacci sequence is an element of V, we can express it as a linear combination of the two basis vectors. Hence there exist $u, v \in \mathbb{R}$ such that $a_n = u(\Phi^n) + v(\overline{\Phi}^n)$. Setting $n = 0$ and $n = 1$, we have $u + v = 0$ and $(\Phi)u + (\overline{\Phi})v = 1$. Solving for u and v, we get a closed expression for the n-th term of the Fibonacci sequence given by $a_n = (\Phi^n - \overline{\Phi}^n)/\sqrt{5}$.

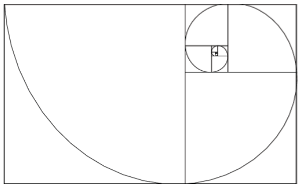
We can write the ratio of consecutive terms of the Fibonacci sequence as $(\Phi/(1-(\overline{\Phi}/\Phi)^n)) - \overline{\Phi}/((\Phi/\overline{\Phi})^n - 1)$. Since $\|\overline{\Phi}\| < \Phi$, the second term vanishes as n goes to infinity and the first term yields Φ. This proves the initial claim.

Φ can also be expressed in terms of various limits. Consider the following expression: $\sqrt{1+\sqrt{1+\sqrt{1+...}}}$ To find the value of this expression, let $x = \sqrt{1+\sqrt{1+\sqrt{1+....}}}$ Then $x^2 = 1+\sqrt{1+\sqrt{1+...}}$ and hence $x^2 = x+1$ or $\Phi = \sqrt{1+\sqrt{1+\sqrt{1+....}}}$

A similar expression for Φ is the sequence $1+(1/(1+(1/(1+...)$. Letting x be this expression, we have that $x = 1+(1/x)$, or $x^2 = x+1$. Hence, the Golden ratio can also be expressed as a continued fraction.

The golden rectangle

A rectangle whose sides are in the ratio of the golden number is known as a golden rectangle. Here is an unusual procedure. Consider a golden rectangle.



Cut a square out of this rectangle and you'll get a golden rectangle whose dimensions are smaller by a factor of Φ. Moreover, by joining the successive points of division we obtain a logarithmic spiral.

Relationships between Φ and trigonometric functions

Here are some exact trigonometric formulas involving Φ.

$$\begin{aligned}\Phi &= 2\cos(\pi/5) \\ &= (1/2)\sec(2\pi/5) \\ &= (1/2)\csc(\pi/10)\end{aligned}$$

Here are some surprisingly simple and interesting expressions involving Φ and complex numbers. Recall that $\sin(x) = (e^{ix} - e^{-ix})/(2i)$ and note that $(1/\Phi) = \Phi - 1$.

$$\begin{aligned}\sin(i\log\Phi) &= (e^{-\log\Phi} - e^{\log(\Phi)/(2i)}) \\ &= ((1/\Phi) - \Phi)/(2i) \\ &= (-1)/(2i) = i/2\end{aligned}$$

An even more interesting expression obtained by using the Euler formula $e^{ix} = \cos x + i\sin x$:

$$\begin{aligned}\sin\left(\frac{\pi}{2} - i\log(\Phi)\right) = \frac{1}{2i}(e^{\pi i/2}e^{\log(\Phi)} \\ -e^{-\pi i/2}e^{\log(\Phi^{-1})})\end{aligned}$$

Using the Euler formula again, we obtain

$$\begin{aligned}(i\Phi + i/\Phi)/(2i) &= (\Phi + (1/\Phi))/2 \\ &= (2\Phi - 1)/2 \\ &= (1 + \sqrt(5) - 1)/2 \\ &= \sqrt(5)/2\end{aligned}$$

Occurrences of the Golden Ratio

The Golden Number makes multiple appearances in art, historical monuments and life in general. Totally unrelated phenomena, such as the petal arrangement in a red rose and the breeding of rabbits, have this proportion in common. The Great Pyramid at Giza in Egypt has dimensions based on the Golden ratio. In fact, the area of a triangular lateral side is equal to the square of the height, as stated by the Greek historian Herodotus. With simple geometric manipulations, one can show that the height of a lateral side divided by half the side of the base gives Φ. This ratio is accurate to less than 0.1 percent. This suggests that the Egyptians knew about the Golden Section. The spiral growth of sea shells is also based on the Golden Number.

It is widely believed that the Golden ratio appears in Leonardo Da Vinci's "Mona Lisa" and in Salvador Dali's "Sacrament of the Last Supper". There are other cases where the Golden ratio's role is still uncertain. A very good example is the Parthenon in Greece. Some sources, such as the Random House Encyclopedia, state that the Golden Section appears in the building of the Parthenon. George Markowsky's "Misconceptions about the Golden Ratio" states that this is not true.

References

[1] Livio, Mario. The Golden Ratio: The Story of Phi, The World's Most Astonishing Number. New York: Broadway Books, 2002.

[2] Borowski, E.J. and Borwein, J.M. Collins internet-linked dictionary of Mathematics. Collins, 2nd ed. HarperCollins Publishers: Glasgow, 2002. 239-240.

[3] Weisstein, Eric W. "Golden Ratio." From Mathworld–A Wolfram Web Resource. [mathworld.wolfram.com/GoldenRatio.html]

[4] Weisstein, Eric W. "Golden Rectangle." From Mathworld–A Wolfram Web Resource. [mathworld.wolfram.com/GoldenRectangle.html]

Jokes

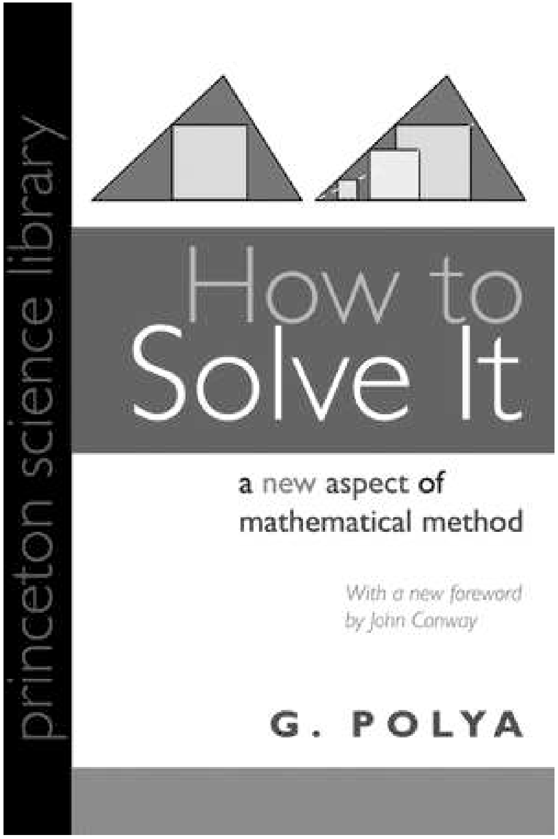
A mathematician is someone who thinks "A", writes "B", says "C" when it should be "D". □

A mathematician is asked to design a table. He first designs a table with no legs. Then he designs a table with infinitely many legs. He then spends the rest of his life generalizing results for the table with N legs (where N is not necessarily a natural number). □

Review of George Pólya's *How to Solve it?*

Nan Yang

Heuristics is not often taught systematically in mathematics, or indeed in general today, at least not directly; while we are mostly taught how to solve a problem or even a series of problems, 'how to solve problems' is a question that is usually left to the students, as if it is an instinct that is best left to develop on its own. Evidently, George pólya disagrees, and *How to Solve It* is his answer to this unspoken question.


princeton science library
How to Solve It
a new aspect of mathematical method
With a new foreword by John Conway
G. POLYA

How to Solve It is literally a book of dialogues; it contains dialogues between a fictional teacher and a fictional student through which Pólya illustrates the process of how the student is gradually nudged into the right direction by the teacher; that is, his ideal process of learning. The dialogues need not be strictly between the fictional characters; when the teacher says, 'What is the unknown? Can you think of a similar problem?' it is obviously directed at the reader.

A recurring theme throughout the book is that if you can not solve a problem, then you should find an easier but similar one. 'Do you know a related problem?' pólya would ask. For example, suppose the student has just learned the Pythagorean theorem, and is now asked to find the length of the spatial diagonal of a parallelepiped. A small amount of ingenuity is required to make the jump from the plane to the space, and the student is naturally stumped. 'Do you know a problem with a similar unknown?' the teacher asks. The student gets a brilliant insight – a previously solved problem. 'Good! Here's a problem related to yours and solved before. Can you use it?' the teacher presses on. Eventually the solution is found, and all is well. Of course, the point of that passage is not to introduce to the readers the Pythagorean theorem in higher dimensions, but to show the readers the process with which one can use to find an 'auxiliary problem' that has been solved and use it to solve the harder problem. pólya himself is often accredited the quote, 'For every problem you can't solve there exists an easier problem that you can: find it.'

Unfortunately, the effectiveness of the book is debatable. As Feynman said, 'You can't learn to solve problems by reading about it.' There is no other way of gaining problem-solving experience save for actually solving problems. Hence, ironically, it is hard to appreciate the book unless you no longer need it.

References

[1] G. Pólya, How to Solve it, 2nd ed., Princeton University Press, 1957.

The Adventures of A and B

Joël Perras and Nan Yang

> AUTHORS' NOTE: A and B are *fictitious* characters. Any resemblance to any real-world persons is purely coincidental.

On a bright and sunny afternoon in early May, two students sat in the corner of a tiny restaurant, far back where the deadly rays of sun could not touch them – deadly because years of isolation in the bleak dungeons of Burnside basement had left them extremely vulnerable to natural light. Over a meal consisting of poutine and hot-dogs, they celebrated the end of the final examination period. With that dreaded ordeal behind them, they were now able to concentrate on the independent projects they had planned for the summer. Somehow, the conversation drifted onto lottery numbers. Both having studied probability and statistics, they lamented over the superstitions that the general public usually attached to the choice of numbers.

'What possible difference would it make what numbers you choose?' A said. 'For instance, I make my point to buy the numbers *one* through *six*. Uniform distribution... you know...' He trailed off as he dug up another fork-load of poutine.

'I agree,' said B. 'But I always choose *nine eight seven six five*. That way, if I win, I will have a better chance of getting the whole prize.'

A nodded, but after a second hesitated and looked at B quizzically, 'Better chance? What do you mean? The distribution is *uniform*. Surely you're not one of those *superstitious* types...' He looked at B as if he were Goldstein himself and was considering whether or not to denounce him to Big Brother.

'You don't know about *Benford's law*?' B asked with traces of genuine surprise in his voice. A shaked his head, still suspicious of the true motives of B.

'Why, Benford's law states that:

THEOREM: (Benford's law) In listings, tables of statistics, etc., the digit 1 tends to occur with $\approx 30\%$ probability , much greater than the expected 11.1% (i.e., one digit out of 9).

Surprising, is it not? Hence, since most people see numbers that begin with a 1 more often than any other single number, they are more likely to choose a number that begins with 1 when given the choice.'

'Impossible!'

'It's true. Why, I might even publish an article in that upcoming *delta-epsilon* magazine about it.' B said as he shifted his attention back to the poutine.

A period of silence lingered between them while an idea formed in A's head. Suddenly he looked up and said, 'Gee, I wonder of this law applies to the prime numbers. Let's see... $2, 3, 5, 7, 11, 13, 17, 19, 23$. Why, of the first ten primes, four of them begin with 1!'

'Interesting... but we better not jump to conclusions here. We need more data.'

'Yes. We're going to need quite a list, maybe fifty million or so,' A replied.

On the very next day, they went to Rosenthall library and asked the librarian whether they kept a list of at least fifty million primes.

'Fifty million primes?' the librarian said, shocked. 'I don't think anyone keeps anything like that on hand. It's much faster just to generate them yourselves unless your a disproportionally fast connection.'

'Oh.'

So they went to the computer lab and began working on a prime list generator. Fortunately, B had some programming experience; they began working and half an hour later they were already generating primes. Eight hours later, the first ten million primes were ready. An analysis of the first ten million primes showed that over 40 percent of them began with a 1. A and B were stunned, but were too caught up in the moment of excitement to think properly. All they knew was that they needed a bigger sample size. But their program would take at least another 80 hours to cough up the next forty million primes. *Surely a better algorithm exists*, they thought. A bit of googling revealed such a program which, after only 5 minutes, was able to generate fifty million primes.

'Well, that would have saved us some time,' B said.

After analyzing the new set of data, they were shocked to discover that almost fifty percent of the first fifty million primes began with 1. A and B had some past experience with number theory, and never had they heard of this startling result. Surely something this trivial could not have gone unnoticed since the time of Euclid!

Before they had time to contemplate further, they realized that they had to attend a Montreal Symphony Orchestra concert. A professional mathematician would never have given up new math for a concert but, as it were, they put down the problem temporarily and went. Their speculations and visions of grandeur had to wait; they could not forfeit their monthly allowance of culture and appreciation of the arts.

Nonetheless, while the bodies of A and B were in the most expensive section they can afford (the balcony), their minds were decidedly elsewhere. After an hour of almost-listening and near-appreciation, A gripped the arms of his seat, his eyes bulging wide with apprehension.

'B, when we generated that list of primes, did you specify that we wanted the first fifty million primes, or all the primes *up to* fifty million?'

'The first fifty million primes. Why do you – we made a sampling mistake, didn't we?' B caught on.

'I think so. After this concert, we need to go verify this,' he answered.

After the curtain fell and the applause still ringing in their ears, both rushed to the most important place they needed to be on a Saturday night: a computer lab at McGill. B ran the prime generator program once again, but this time instructed it to halt after producing all the primes *up to* one billion. The results showed what they both feared: a near uniform distribution of the first digits.

'There goes our theory,' said A.

'Oh well, it was fun and interesting while it lasted,' commented B. 'Let's go have some more poutine.'

Several weeks later A mentioned their null result to a mathematician, who readily replied, 'The logarithmic density of the primes actually does follow Benford's law. It's an old result. You can check it up in the February 1972 issue of American Mathematical Monthly, page 150. I believe it's in a paper by R. E. Whitney.'

Unbelieving, A browsed the shelves of Rosenthall Library, quite sure that the mathematician he had spoken to had been mistaken. He scanned the volumes of journals until he found the one with the article in question and opened it to the correct page. It read:

'It is well known that the logarithmic density of D in the sequence of positive integers is $\log_{10}(1 + 1/a)$. The purpose of this note is to show that the relative logarithmic density of D in P is also $\log_{10}(1+1/a)$. This is an unusual result because of the irregular distribution of the primes. As a consequence of this result, one might say that 1 is the preferred initial digit for the sequence of primes.'

A closed the journal and swore in frustration, '$\epsilon < 0$!'

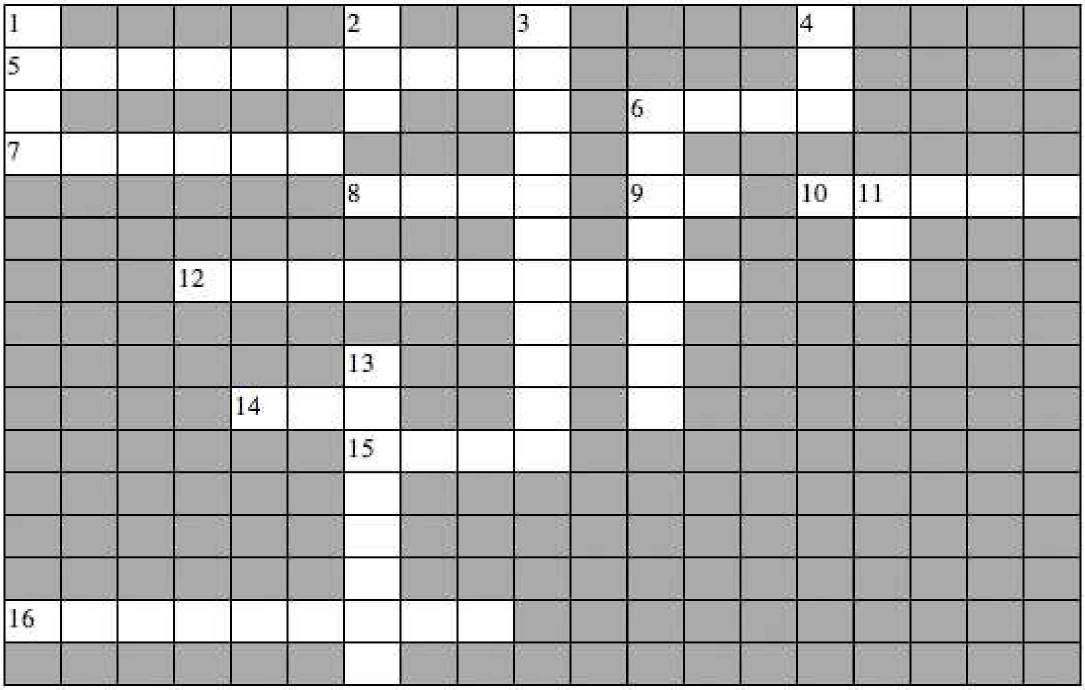
The article, more than thirty years old, disproved in two pages the fallacy that had tied them in ropes for days. In the end, A and B learned a valuable lesson:

COROLLARY: Never abuse statistics.□

Jokes

Several scientists were all posed the following question: "What is 2*2?" The engineer whips out his slide ruler and shuffles it back and forth, and finally announces, "4.01." The physicist consults his technical references, sets up the problem on his computer, and announces, "it lies between 3.98 and 4.02." The mathematician cogitates for a while, then says, "I don't know what the answer is, but I can tell you an answer exists!" The philosopher smiles and replies, "But what do you mean by 2*2?" The logician replies, "Please define 2*2 more precisely." The sociologist says, "I don't know, but it was nice talking about it." The medical student replies, "4!" All the others are amazed. "How did you know?" they asked. The medical student replies, "I memorized it." □

Puzzle by Nan Yang



ACROSS

5. First 10 digit prime in the consecutive digits of e.
6. Odds of successfully navigating an asteroid field, to 1.
7. Feynman point.
8. Index of largest known repunit prime.
9. The answer to life, the universe, and everything.
10. Largest known Fermat prime.
12. Google's initial public offering, in dollars.
14. Number of kilobytes Bill Gates claimed 'ought to be enough for anybody'.
15. Largest known Wieferich prime.
16. Largest number with increasing digits.

DOWN

1. Smallest integer expressible as the sum of 2 cubes in 2 different ways.
2. Sum of the squares of the first 7 primes.
3. 11th repunit.
4. Page number in the 1st edition of Principia Mathematica where 1+1=2 is proven.
6. Exponent of the largest known Mersenne prime (43rd).
11. Number of people with Erdos number one.
13. Position of digit of pi where '31415926' first occurs.

Credits

THE DELTA-EPSILON EDITING TEAM
In alphabetical order

Agnès F. Beaudry

Juan Manual Martinez

Michael McBreen

Alexandra Ortan

Joël Perras

Vincent Quenneville-Bélair

Nan Yang

COVER ART & DESIGN
Michael McBreen

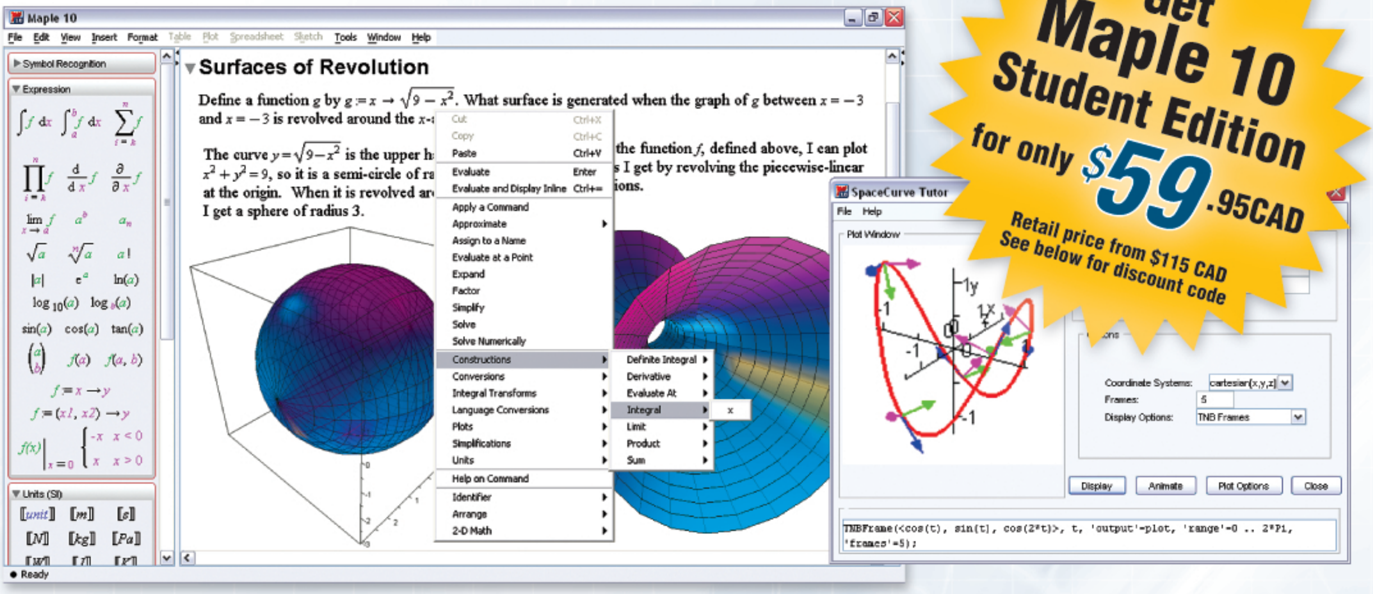
Mathieu Ménard

Acknowledgements

It's a little exaggerated to claim that some accomplishment would not have been possible without the help of person x and y. If an "event" isn't "independent" of the participation of some individual, it might remain possible, but become less probable. Therefore, I would like to thank the following people for their help in increasing the probability of the publication of this journal, but above all, for surely increasing its quality. Thank you prof. Henri Darmon, prof. Niky Kamran and prof. Nilima Nigam for taking the time to speak with us as well as for smiling for the picture. Thank you SUMS and especially Marc-André Rousseau for supporting the project from beginning to end. We are grateful to the Science Undergraduate Society (SUS), the Faculty of Science and the McGill Mathematics Department for providing the funding. Graduate students of McGill, in particular David Campbell, Daniel Simeone, Dennis The and Michael Lennox Wong, many mistakes have been avoided thanks to your keen eyes (and those that remain are left as exercises.) Samuel Altarac-Hofmann, your assistance for printing the magazine is much appreciated. Finally, reader, thank Adèle Flannery *pour la beauté de la rhétorique*, or at least for polishing our broken, mathematical English: she tried to wean us from our ways and sometimes she succeeded, so be thankful to her that you can actually read this!

Professional Math Software at a Student Price

Maplesoft is offering McGill Students a special promotion on all Maple 10 purchases.



Surfaces of Revolution

Define a function g by $g := x \rightarrow \sqrt{9 - x^2}$. What surface is generated when the graph of g between $x = -3$ and $x = -3$ is revolved around the x-

The curve $y = \sqrt{9 - x^2}$ is the upper h $x^2 + y^2 = 9$, so it is a semi-circle of ra at the origin. When it is revolved ar I get a sphere of radius 3.

the function f, defined above, I can plot s I get by revolving the piecewise-linear ions.

Coordinate Systems: cartesian[x,y,z]
Frames: 5
Display Options: TNB Frames

```
TNBFrame(<cos(t), sin(t), cos(2*t)>, t, 'output'=plot, 'range'=0 .. 2*Pi,
'frames'=5);
```

Get Maple 10 Student Edition for only $59.95CAD

Retail price from $115 CAD
See below for discount code

"Maple is the first piece of software I've ever seen that displays mathematical graphs, functions and symbols the same way I'd write them. Never has it been so easy to figure out where I made my mistakes."

— Nathan Sokalski, Student,
Kutztown University of Pennsylvania

Maple 10
Student Edition...

can be purchased through the Maplesoft Webstore at *webstore.maplesoft.com*

To get the special McGill Student price enter promo code: MCIS6794

Maple™ 10

With thousands of functions covering all technical majors and courses - calculus, differential equations, linear algebra, statistics - Maple 10 brings the problem solving power of expert mathematicians to your computer.

- Improve your comprehension levels and scholastic performance with the ability to illustrate mathematical concepts, visualize problems and illuminate theory.
- See the steps and thinking behind the math problem in addition to providing the answer.
- Finish assignments and projects faster while improving your understanding of even the most difficult subjects.
- BONUS - Maple Graphing Calculator* offers high-end calculation and graphing power through a familiar graphing calculator interface. Perfect for quick calculations and plotting.

Maplesoft
command the brilliance

www.maplesoft.com | info@maplesoft.com